



Elias Rodrigues da Silva Júnior

Técnicas de extração de características para cenários com multimodalidade intraclasse

Tese de Doutorado

Recife

2016

Elias Rodrigues da Silva Júnior

**Técnicas de extração de características para cenários com multimodalidade
intraclasse**

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Doutor em Ciência da Computação.

Orientador: George Darmiton da Cunha Cavalcanti

Coorientador: Tsang Ing-Ren

Recife

2016

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S586t Silva Júnior, Elias Rodrigues da
Técnicas de extração de características para cenários com multimodalidade
intraclasse / Elias Rodrigues da Silva Júnior. – 2016.
121 f.: il., fig., tab.

Orientador: George Darmiton da Cunha Cavalcanti.
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da
Computação, Recife, 2016.

Inclui referências.

1. Ciência da computação. I. Cavalcanti, George Darmiton da Cunha
(orientador). II. Título.

004

CDD (23. ed.)

UFPE- MEI 2016-096

Elias Rodrigues da Silva Júnior

Metodologia de extração de características para cenários com multimodalidade intraclasse

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de **Doutor** em Ciência da Computação.

Aprovado em: 23/02/2016.

Prof. Dr. George Darmiton da Cunha Cavalcanti

Orientador do Trabalho de Tese

BANCA EXAMINADORA

Prof. Dr. Adriano Lorena Inácio de Oliveira
Centro de Informática / UFPE

Prof. Dr. Cleber Zanchettin
Centro de Informática / UFPE

Prof. Dr. Carlos Henrique Costa Ribeiro
Divisão de Ciência da Computação / ITA

Prof. Dr. Juliano Bandeira Lima
Departamento de Eletrônica e Sistemas / UFPE

Prof. Dr. Getúlio Jose Amorim do Amaral
Departamento de Estatística / UFPE

*Dedico este trabalho a minha família, por todo apoio dispensado para a conclusão deste.
E, em especial a minha mãe, cujo apoio e investimento em mim se acumulam por toda a
minha vida.*

Agradecimentos

Com alegria concluo esse trabalho, estando ciente de que nessa jornada fui muito auxiliado. Assim, não posso deixar de reconhecer o apoio que recebi de:

(i) Minha família, que sempre participou de minha formação e me supriu de toda motivação e condições necessárias para a dedicação de todos esses anos de estudo.

(ii) Cristiano de Santana e Nara Portela, amigos que partilharam das mesmas lutas e com os quais discutimos inúmeras possibilidades de solucionar os desafios que surgiam.

(iii) George Darmiton, pela sua orientação próxima, sábia e paciente. Muito obrigado por me guiar e encorajar desde o mestrado.

(iv) Rayza Keonária e Amanda Ferreira, obrigado pela dedicada revisão de meu texto e das sugestões que, seguramente, o tornaram mais legível.

(v) Ao CNPq e Facepe, por respectivamente financiar meus estudos através de bolsa e custear parte da publicação de um artigo.

Resumo

Técnicas de Extração de Características (TECs) clássicas não consideram a possibilidade dos padrões se dispersarem segundo distribuição multimodal. A falta de percepção quanto a esse tipo de comportamento frequentemente conduz a reduções de dimensionalidade para espaços que não preservam a multimodalidade. A não preservação da estrutura multimodal intraclasse implica na sobreposição entre suas modas, o que pode degradar o desempenho de classificação sobre essa classe. Encontra-se na literatura TECs que buscam por reduções dimensionais que minimizam as alterações nas relações métricas entre os padrões nesse novo espaço. Contudo, a realização desse objetivo não é suficiente para garantir a preservação da multimodalidade. Além disso, não foi encontrada na literatura TEC que avalie a dispersão dos padrões da classe e, em caso de multimodalidade, obtenha redução que preserve ao máximo esse comportamento. Uma possível razão para essa ausência é a complexidade associada a determinação da ocorrência de multimodalidade, principalmente em espaços multidimensionais. Estimar a multimodalidade em uma dispersão de padrões envolve questões como determinar: o número de modas na dispersão, a função de distribuição que cada moda segue, se a mistura dessas funções apresenta comportamento multimodal e quão evidente é esse comportamento. Na literatura, vários autores estimam multimodalidade através de reduções no escopo dessa tarefa, como restrições a: cenários bimodais e distribuições Normais. Adicionalmente, a restrição a espaços unidimensionais é comum a todas as técnicas de estimativa de multimodalidade encontradas. Desse modo, este trabalho apresenta uma metodologia para o desenvolvimento de TECs que buscam preservar o comportamento multimodal da dispersão dos padrões intraclasse. Para isso, são encontradas as direções ortogonais com maiores evidências de preservar o comportamento multimodal original. Essas direções são utilizadas como os vetores de base de espaços de redução dimensional que preservam ao máximo esse comportamento. Através dessa metodologia são propostas três TECs: *Class-dependent Multimodal Decomposition* (CdMD), *Class-dependent Conditional Multimodal Decomposition* (CdCMD) e *Class-dependent Conditional Detachment of Modes* (CdCDM). Essas TECs estendem *Multimodal Decomposition* (MD), que é uma TEC proposta para realizar a decomposição espacial orientada pela multimodalidade. Essas decomposições estão associadas a valores de estimativa de multimodalidade, os quais são obtidos com a métrica proposta, denominada *Index of Area-based Normal Intersection* (IANI). Os resultados experimentais mostram que IANI obteve sucesso em estimar a multimodalidade mesmo em cenários com acentuado desbalanceamento entre as modas. Isso contribuiu para MD obter desempenho superior ao PCA e LPP. Além disso, os maiores valores médios de desempenho em cenários de classificação foram alcançados com CdMD, CdCMD e CdCDM.

Palavras-chave: Técnicas de Extração de Características. Multimodalidade Intraclasse. Estimativa de Multimodalidade. Preservação do Comportamento Multimodal.

Abstract

Classical Feature Extraction Techniques (FETs) do not consider the possibility of patterns scattered following multimodal distribution. No perception about this kind of behavior often leads to dimensionality reduction spaces that do not preserve multimodality. Do not preserve within-class multimodal structure implies overlap between classes modes, which can degrade performance classification over the class. There is in literature FETs that seek for dimensionality reductions which minimize the changes in metric relations between the patterns inside the new space. However, perform this goal is not enough to ensure preservation of multimodality. Besides that, there was not found in literature FET that evaluates scattering of class patterns and, in case of multimodality, it reaches reduction that preserve this behavior at most. A possible reason for this absence is related with the complexity to determine the occurrence of multimodality, mainly in multidimensional spaces. Estimate multimodality in a patterns scattering involves questions like determine: the number of modes in the scattering, the distribution function that each mode follows, if the mixture of these functions present multimodal behavior and how evident is this behavior. In literature, many authors estimate multimodality through scope reductions of this task, like restrictions to: bimodal scenarios and Normal distributions. Additionally, the restriction to unidimensional spaces is common to all multimodal estimation techniques that were found. Thereby, this work presents a methodology to develop FETs that aim to preserve the multimodal behavior of within-class scattering of patterns. For this, the orthogonal directions with more evidence to preserve the original multimodal behavior are found. These directions are used as basis vectors of dimensional reduction spaces that preserve this behavior at most. Using this methodology, three FETs are proposed: TECs: *Class-dependent Multimodal Decomposition*, *Class-dependent Conditional Multimodal Decomposition* and *Class-dependent Conditional Detachment of Modes*. These FETs extend *Multimodal Decomposition*, an FET proposed to perform the spatial decomposition driven by multimodality. These decompositions are associated with multimodal estimation values, which are achieved with a proposed metric, called *Index of Area-based Normal Intersection*. Experimental results show IANI was well succeed in estimating multimodality even at scenarios with large unbalance between modes. This contributed to MD reach superior performances than PCA and LPP. Furthermore, the largest average values of performance for classification scenarios were achieved with CdMD, CdCMD e CdCDM.

Key-words: Feature Extraction Techniques. Within-class Multimodality. Estimation of Multimodality. Preservation of Multimodal behavior.

Lista de ilustrações

Figura 1 – Exemplo de conjunto de padrões multimodal.	19
Figura 2 – Falha de técnicas de extração em preservar a multimodalidade.	21
Figura 3 – Distinção entre padrões de diferentes modas para a melhoria da classificação intraclasse.	22
Figura 4 – Técnicas de extração que não garantem preservação da estrutura multimodal.	26
Figura 5 – Classificador <i>classwise</i>	32
Figura 6 – Extração de características do CPCA.	34
Figura 7 – Extração de características do CdLPP.	35
Figura 8 – Características das técnicas <i>classwise</i> com investigação de multimodalidade.	39
Figura 9 – Decomposição em espaços unidimensionais.	42
Figura 10 – Configurações para o raio da vizinhança do PSO.	47
Figura 11 – Configurações específicas à aplicação do PSO.	48
Figura 12 – Evidência de multimodalidade em espaços unidimensionais.	51
Figura 13 – Variação da probabilidade <i>a priori</i>	55
Figura 14 – Comportamentos da área de interseção quando $\mu_1 = \mu_2$ e $\sigma_1 = \sigma_2$	56
Figura 15 – Comportamentos da área de interseção quando $\mu_1 \neq \mu_2$ e $\sigma_1 = \sigma_2$	57
Figura 16 – Comportamentos da área de interseção quando $\sigma_1 \neq \sigma_2$	57
Figura 17 – Alteração no aspecto de distribuições Normais.	58
Figura 18 – Influência dos parâmetros da mistura sobre a área de interseção entre as Normais.	59
Figura 19 – Diferenças entre as abordagens de treinamento do classificador.	61
Figura 20 – Agrupamentos das técnicas de extração de características.	61
Figura 21 – Influência da técnica de agrupamento sobre a FDP.	63
Figura 22 – Procedimento de extração de características do CdMD.	64
Figura 23 – Consequência da diminuição da evidência de multimodalidade.	66
Figura 24 – Procedimento de extração de características do CdCMD.	66
Figura 25 – Extração de características condicional do CdCMD.	67
Figura 26 – Elevação da separação entre modas.	68
Figura 27 – Extração de características condicional do CdCDM.	68
Figura 28 – Representação da dispersão dos padrões gerados para as bases dos Grupos 1-3. Essas dispersões seguem distribuição $\mathcal{N}(0, 0, 25^2)$ em y e diferem entre si quanto a dispersão sobre x , que é modelada segundo duas modas Normais, ambas com desvio padrão $\sigma = 1$. As médias das Normais em x são: ± 2 (28a), ± 1 (28b) e $\pm \frac{1}{2}$ (28c).	73

Figura 29 – Representação da dispersão dos padrões gerados para as bases dos Grupos 4-6. Essas dispersões seguem distribuição $\mathcal{N}(0, 1^2)$ em y e diferem entre si quanto a dispersão sobre x , que é modelada segundo duas modas Normais, ambas com desvio padrão $\sigma = 1$. As médias das Normais em x são: ± 2 (29a), ± 1 (29b) e $\pm \frac{1}{2}$ (29c).	73
Figura 30 – Representação da dispersão dos padrões gerados para as bases dos Grupos 7-9. Essas dispersões seguem distribuição $\mathcal{N}(0, 2^2)$ em y e diferem entre si quanto a dispersão sobre x , que é modelada segundo duas modas Normais, ambas com desvio padrão $\sigma = 1$. As médias das Normais em x são: ± 2 (30a), ± 1 (30b) e $\pm \frac{1}{2}$ (30c).	74
Figura 31 – Ângulos com o eixo x das primeiras direções retornadas pela decomposição espacial utilizando cada uma das métricas de estimativa de multimodalidade.	75
Figura 32 – Comparação entre as técnicas de estimativa de multimodalidade através do teste de Friedman, com confiança de 95%.	76
Figura 33 – Dispersão das bases <i>toy</i> para as quais IANI não obteve a maior precisão média (visão não-supervisionada).	80
Figura 34 – Comparação entre as técnicas de estimativa de multimodalidade através do teste de Friedman, com confiança de 95%.	81
Figura 35 – Exemplos de cenários com comportamento multimodal utilizado no experimento.	82
Figura 36 – Exemplos das bases de dados bidimensionais construídas.	83
Figura 37 – Valores de TNR obtidos para as técnicas de extração com diferentes valores de θ	86
Figura 38 – Comparação entre as técnicas de extração de características através do teste de Friedman, com confiança de 95%, utilizando $\theta = 0, 5$	87
Figura 39 – Protocolo metodológico para determinação dos desempenhos das técnicas.	89
Figura 40 – Representação bidimensional da dispersão dos padrões por classes das bases <i>toy</i>	90
Figura 41 – Comparação do desempenho das variações do CPCA com o teste de Friedman.	94
Figura 42 – Comparação do desempenho das configurações do CdLPP r_n com o teste de Friedman.	96
Figura 43 – Comparação do desempenho das configurações do CdLPP com o teste de Friedman.	97
Figura 44 – Comparação do desempenho das configurações do CdMD r_n com o teste de Friedman.	98
Figura 45 – Comparação do desempenho das configurações do CdMD com o teste de Friedman.	99

Figura 46 – Comparação do desempenho das configurações do CdCMD com o teste de Friedman.	102
Figura 47 – Comparação do desempenho das configurações do CdCMD com o teste de Friedman.	105
Figura 48 – Comparação do desempenho entre as técnicas de extração de características com o teste de Friedman.	109
Figura 49 – Complexidade na estimativa de multimodalidade com mais de duas modas.	114
Figura 50 – Confusão na categorização de dispersões devido a limiarização sobre o valor de IANI.	115
Figura 51 – Controle do número de interseções quando uma das modas Normais está inteiramente contida sob a outra, sendo suas médias iguais.	116
Figura 52 – Transformação do espaço de características realizada pelo SVM.	117

Lista de tabelas

Tabela 1 – Tomada de decisão por classificador <i>classwise</i>	33
Tabela 2 – Restrições ao escopo de <i>Multimodal Decomposition</i> devido ao uso das métricas para estimativa de multimodalidade.	52
Tabela 3 – Configuração dos grupos sobre as bases geradas de acordo com os eixos x e y	73
Tabela 4 – Ordenação das técnicas segundo a corretude da estimativa de multimodalidade.	76
Tabela 5 – <i>Rank</i> médio das bases agrupadas pelo valor de ρ	77
Tabela 6 – Configurações para geração dos padrões das bases <i>Toy</i>	78
Tabela 7 – Valores médios de precisão de classificação (%) obtidos pelas estimativas de multimodalidade para as bases <i>Toy</i> . Valores entre parênteses representam os desvios padrão obtidos.	79
Tabela 8 – Configurações dos parâmetros de geração dos conjuntos de treinamento.	83
Tabela 9 – Quantidades de TNR nulas (TNR = 0,000%) obtidas pelas técnicas de extração de características.	87
Tabela 10 – Parâmetros de configuração das modas Normais que compõem a função de distribuição dos padrões da classe 1.	91
Tabela 11 – Parâmetros de configuração das modas Normais que compõem a função de distribuição dos padrões da classe 2.	92
Tabela 12 – Rótulos das variações das técnicas <i>classwise</i> segundo suas configurações.	93
Tabela 13 – CPCA vs. CPCA r_b	94
Tabela 14 – CdLPP $r_n(1)$ vs. CdLPP $r_n(2)$	95
Tabela 15 – Comparações entre CdLPP, CdLPP r_b e CdLPP $r_n(2)$	96
Tabela 16 – CdMD $r_n(1)$ vs. CdMD $r_n(2)$	98
Tabela 17 – Comparações entre CdMD, CdMD r_b e CdMD $r_n(1)$	99
Tabela 18 – Sumário dos desempenhos do CdCMD nos diferentes limiares de decisão.	100
Tabela 19 – Sumário dos desempenhos do CdCMD r_b nos diferentes limiares de decisão.	100
Tabela 20 – Sumário dos desempenhos do CdCMD $r_n(1)$ nos diferentes limiares de decisão.	100
Tabela 21 – Sumário dos desempenhos do CdCMD $r_n(2)$ nos diferentes limiares de decisão.	101
Tabela 22 – Comparação entre os desempenhos das configurações do CdCMD r_n	101
Tabela 23 – Comparações entre CdCMD ^{50%} , CdCMD ^{30%} r_b e CdCMD ^{10%} $r_n(1)$	102
Tabela 24 – Sumário dos desempenhos do CdCMD nos diferentes limiares de decisão.	103

Tabela 25 – Sumário dos desempenhos do CdCDM r_b nos diferentes limiares de decisão.	103
Tabela 26 – Sumário dos desempenhos do CdCDM $r_n(1)$ nos diferentes limiares de decisão.	103
Tabela 27 – Sumário dos desempenhos do CdCDM $r_n(2)$ nos diferentes limiares de decisão.	103
Tabela 28 – Comparação entre os desempenhos das configurações do CdCDM r_n	104
Tabela 29 – Comparações entre CdCDM ^{30%} , CdCDM ^{60%} r_b e CdCDM ^{60%} $r_n(1)$	104
Tabela 30 – Configurações melhores adaptadas para cada uma das técnicas <i>classwise</i>	105
Tabela 31 – Frequência das melhores reduções obtidas pelas técnicas de extração de características comparadas.	106
Tabela 32 – Taxas de acerto (%) obtidas com as técnicas de extração de características.	107
Tabela 33 – <i>Ranks</i> obtidos com as técnicas de extração de características.	108
Tabela 34 – Sumário das técnicas propostas.	110
Tabela 35 – Sumário das avaliações de desempenho das técnicas propostas.	111

Lista de abreviaturas e siglas

CdCDM	<i>Class-dependent Conditional Detachment of Modes</i>
CdCMD	<i>Class-dependent Conditional Multimodal Decomposition</i>
CdLPP	<i>Class-dependent Locality Preserving Projections</i>
CdMD	<i>Class-dependent Multimodal Decomposition</i>
CLPP	<i>Classwise Locality Preserving Projections</i>
CPCA	<i>Classwise Principal Component Analysis</i>
EM	<i>Expectation Maximization</i>
FDP	Função Densidade de Probabilidade
GA	<i>Genetic Algorithm</i>
GMM	<i>Gaussian Mixture Model</i>
IANI	<i>Index of Area-based Normal Intersection</i>
LDA	<i>Linear Discriminant Analysis</i>
LI	Linearmente Independente
LPP	<i>Locality Preserving Projections</i>
LFDA	<i>Local Fisher Discriminant Analysis</i>
MDS	<i>Multi Dimensional Scale</i>
MD	<i>Multimodal Decomposition</i>
PSO	<i>Particle Swarm Optimization</i>
PCA	<i>Principal Component Analysis</i>
SSS	<i>Small Sample Size</i>
SVM	<i>Support Vector Machine</i>
TEC	Técnica de Extração de Características
TNR	<i>True Negative Rate</i>

Lista de símbolos

\mathbf{x}	Padrão de treinamento
\mathbf{z}	Padrão de consulta
\mathbf{TR}	Conjunto de treinamento
\mathcal{X}	Conjunto de dados
N	Número de padrões
N_l	Número de padrões da classe l
d	Dimensão do espaço original
r	Dimensão do espaço resultante da transformação
r^*	Melhor dimensão do espaço resultante da transformação
C	Número de classes do problema
\mathbf{A}	Matriz de afinidade
\mathbf{A}^s	Matriz de afinidade supervisionada
\mathbf{D}	Matriz de auxiliar para construção do grafo Laplaciano dos padrões
\mathbf{L}	Matriz que representa o grafo Laplaciano dos padrões
\mathbf{T}	Matriz de transformação
\mathbf{T}_{PCA}	Matriz de transformação do PCA
\mathbf{T}_{LDA}	Matriz de transformação LDA
\mathbf{T}_{MD}	Matriz de transformação MD
μ	Média populacional unidimensional
σ	Desvio padrão populacional
ρ	Probabilidades <i>a priori</i>
$\tilde{\boldsymbol{\mu}}$	Vetor de média amostral
$\tilde{\boldsymbol{\mu}}_l$	Vetor de média amostral da classe l
\mathbf{S}	Matriz de dispersão

\mathbf{S}_t	Matriz de dispersão total
\mathbf{S}^n	Matriz de dispersão negativa
\mathbf{S}^b	Matriz de dispersão entreclasses
\mathbf{S}^w	Matriz de dispersão intraclasse
$\tilde{\mathbf{S}}^b$	Matriz de dispersão local entreclasses
$\tilde{\mathbf{S}}^w$	Matriz de dispersão local intraclasse
$\widetilde{\mathbf{W}}^b$	Matriz de pesos locais entreclasses
$\widetilde{\mathbf{W}}^w$	Matriz de pesos locais intraclasse
$\boldsymbol{\rho}$	Vetor de probabilidades <i>a priori</i>
λ	Autovalor
$\mathbf{\Lambda}$	Matriz de autovalores
\mathbf{e}	Autovetor
\mathbf{E}	Matriz de autovetores
A_{\cap}	Área de interseção entre Normais
A_{\cap}^*	Área máxima de interseção entre Normais
ρ_q	Medida de equilíbrio entre distribuições Normais
A_q	Medida padronizada da área de interseção entre distribuições Normais
\mathbf{v}	Vetor indicando direção em espaço multidimensional
\mathbf{v}_*	Vetor indicando direção com maior valor de estimativa de multimodalidade
\mathbf{V}	Matriz que armazena direções ortogonais de um espaço multidimensional
\mathbb{V}	Espaço euclidiano
ξ	Valor de estimativa de multimodalidade
ξ_*	Maior valor de estimativa de multimodalidade
$\boldsymbol{\xi}$	Vetor de valores de estimativa de multimodalidade
\mathcal{V}	Função de custo ou valor de métrica de estimativa de multimodalidade

\mathbf{v}_i^l	[PSO] Posição atual da partícula i
\mathbf{p}_i	[PSO] Melhor posição visitada pela partícula i
\mathbf{g}_i	[PSO] Melhor posição visitada pelos vizinhos da partícula i
Δ_i	[PSO] Deslocamento atual da partícula i
t_i	[PSO] Idade da partícula i sem mutação
ω	[PSO] Controla influência do deslocamento anterior
ϕ_1	[PSO] Controla tendência do deslocamento na direção de \mathbf{p}_i
ϕ_2	[PSO] Controla tendência do deslocamento na direção de \mathbf{g}_i
\mathbf{R}	[PSO] Matriz diagonal aleatória

Sumário

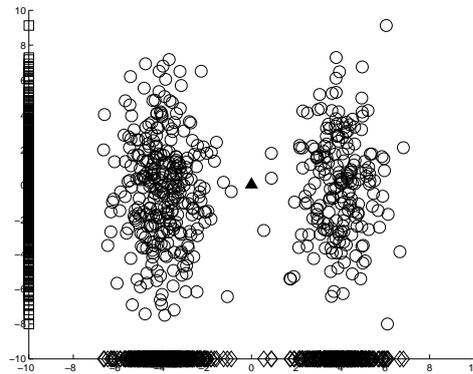
1	INTRODUÇÃO	19
1.1	Objetivos	21
1.2	Estrutura do documento	23
2	ESTADO DA ARTE	24
2.1	Investigação da multimodalidade	24
2.1.1	Técnicas de identificação da multimodalidade	24
2.1.2	Técnicas de verificação da multimodalidade	25
2.2	Técnicas de extração de características	26
2.2.1	Técnicas não- <i>classwise</i>	28
2.2.1.1	<i>Principal Component Analysis</i>	28
2.2.1.2	<i>Linear Discriminant Analysis</i>	29
2.2.1.3	<i>Locality Preserving Projections</i>	30
2.2.1.4	<i>Classwise Locality Preserving Projections</i>	30
2.2.1.5	<i>Local Fisher Discriminant Analysis</i>	31
2.2.2	Técnicas <i>classwise</i>	32
2.2.2.1	<i>Classwise Principal Component Analysis</i>	33
2.2.2.2	<i>Class-dependent Locality Preserving Projections</i>	35
2.2.2.3	Classificador das técnicas <i>classwise</i>	36
2.3	Considerações finais	37
3	TÉCNICA CLASSWISE COM INVESTIGAÇÃO DE MULTIMODALIDADE	39
3.1	Operar classes individualmente	39
3.2	Preservar estrutura multimodal das classes	40
3.2.1	Decomposição espacial	41
3.2.2	Busca das direções com maior multimodalidade	44
3.2.2.1	Direção com maior multimodalidade	45
3.2.3	Métricas de multimodalidade	51
3.2.3.1	Eisenberger value	52
3.2.3.2	Behboodian value	53
3.2.3.3	Silverman test	53
3.2.3.4	dip test	54
3.2.3.5	<i>Index of Area-based Normal Intersection</i>	54
3.2.3.5.1	Área de interseção entre as duas Normais	55
3.2.3.5.2	Área máxima de interseção entre as duas Normais	58

3.3	Utilizar informação de multimodalidade na classificação	60
3.4	TECs <i>classwise</i> com investigação de multimodalidade	64
3.4.1	Class-dependent Multimodal Decomposition	64
3.4.2	Class-dependent Conditional Multimodal Decomposition	65
3.4.3	Class-dependent Conditional Detachment of Modes	67
3.4.4	Classificador das TECs <i>classwise</i> com investigação de multimodalidade	69
4	EXPERIMENTOS	71
4.1	Comparação entre as métricas de estimativa de multimodalidade	71
4.1.1	Influência da ponderação entre as modas	72
4.1.1.1	Definição da base de dados	72
4.1.1.2	Resultados	74
4.1.2	Influência sobre a discriminação entre classes	77
4.1.2.1	Definição da base de dados	77
4.1.2.2	Resultados	78
4.1.3	Conclusões do experimento	81
4.2	Preservação do comportamento multimodal	82
4.2.1	Definição da base de dados	83
4.2.2	Classificador	84
4.2.3	Resultados	85
4.2.4	Conclusões do experimento	87
4.3	Desempenhos de classificação das extrações de características	88
4.3.1	Protocolo metodológico	88
4.3.2	Definição da base de dados	90
4.3.3	Resultados	92
4.3.3.1	Variações das técnicas <i>classwise</i>	93
4.3.3.2	Comparação entre as técnicas	105
4.3.4	Conclusões do experimento	109
5	CONSIDERAÇÕES FINAIS	110
5.1	Conclusões experimentais	111
5.2	Trabalhos futuros	112
5.2.1	Emprego de bases reais	113
5.2.2	Classificador de multimodalidade	114
5.2.3	Métrica de estimativa de multimodalidade multidimensional	117
	REFERÊNCIAS	119

1 Introdução

Multimodalidade em um conjunto de padrões é caracterizada pela presença de mais de um máximo em sua função de densidade de probabilidade. Em outras palavras, mais de uma moda na dispersão dos padrões. A Figura 1 ilustra um conjunto de padrões 2D (\circ) agrupados em duas modas.

Figura 1 – Exemplo de conjunto de padrões multimodal.



Fonte: Autoria própria.

A incapacidade de lidar com dados multimodais pode conduzir a erros indesejados quando se desempenha redução de dimensionalidade. Estes erros são relativos a não preservação da estrutura multimodal dos dados após sua projeção no espaço de menor dimensionalidade. A Figura 1 mostra a influência da projeção sobre a preservação da estrutura multimodal. Os padrões \square representam as projeções dos padrões \circ em algum espaço vertical 1D. Nessa projeção, as modas originais se sobrepõem, perdendo a estrutura multimodal original. Os padrões \diamond representam as projeções dos padrões \circ em algum espaço horizontal 1D. Essa projeção não gera sobreposição entre as modas, mantendo a estrutura multimodal.

A importância de preservar a estrutura multimodal dos dados originais é entendida supondo um ruído no ponto $(0, 0)$ da Figura 1, representado por \blacktriangle . Qualquer projeção na vertical conduzirá o ruído a uma região com alta densidade de padrões, o que aumenta a chance do ponto ser erroneamente classificado. Por outro lado, qualquer projeção na horizontal conduzirá o ruído a uma região de baixa densidade de padrões, diminuindo a chance de erro em sua classificação.

Para desempenhar reduções de dimensionalidade que mantenham a estrutura multimodal pode-se, de algum modo, obter algum conhecimento prévio sobre o comportamento (unimodal ou multimodal) da distribuição dos padrões. As técnicas que possibilitam a obtenção desse conhecimento prévio podem ser categorizadas como: (i) técnicas de

identificação ou (ii) técnicas de verificação da multimodalidade.

Definição 1.1 (Identificação da multimodalidade). *São as técnicas que informam o número de modas e/ou os agrupamentos dos padrões relativos à cada moda encontrada.*

Definição 1.2 (Verificação da multimodalidade). *São as técnicas que avaliam se a dispersão de padrões de modas distintas segue distribuição multimodal ou não, podendo expressar a evidência da estrutura multimodal por um valor.*

Porém, desenvolver técnicas para ambas categorias é uma tarefa complexa devido a fatores que dificultam a proposição de uma solução geral para o problema. Alguns desses fatores são: (i) como determinar a quantidade de modas existente na dispersão dos padrões? (ii) como determinar a distribuição estatística que modela cada uma das modas? (iii) como avaliar se a mistura das modas apresenta comportamento multimodal? (iv) espaços de maiores dimensionalidades requerem maior cardinal de padrões a fim de possibilitar inferências com determinado grau de corretude; como a variabilidade do cardinal dos padrões por moda influencia no modelo multimodal proposto?

Muitos autores decidem pela redução do escopo do problema para a proposição de suas técnicas. Reduções de escopo frequentemente encontradas na literatura incluem a modelagem das técnicas para: (i) espaços unidimensionais (HALL; YORK, 2001; CHAN; TONG, 2004), (ii) modas com distribuição estatística Normal (GNANADESIKAN; GUPTA, 1970; RAY; LINDSAY, 2005) e (iii) cenários bimodais (BEHBOODIAN, 1970; SARMA; RAO; RAO, 1990).

Não foi encontrada na literatura Técnica de Extração de Características (TEC) específica para dados com estrutura multimodal. Ou seja, não foi encontrada TEC que, utilizando conhecimento prévio da estrutura multimodal dos dados, empregasse este conhecimento para obter redução de dimensionalidade que preserve ao máximo a estrutura multimodal original.

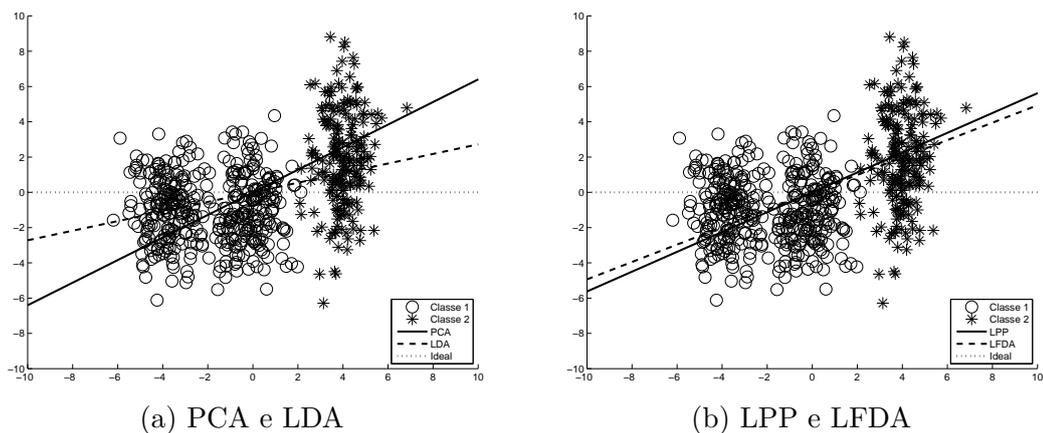
Encontra-se na literatura TECs como *Principal Component Analysis* (PCA) (JOLLIFFE, 2002; LEITE et al., 2007) e *Linear Discriminant Analysis* (LDA) (FISHER, 1936; BALAKRISHNAMA; GANAPATHIRAJU, 1998; NAZ; FAROOQ; NAZ, 2006), as quais modelam suas soluções sem considerar o tipo de distribuição (unimodal ou multimodal) que os dados seguem. Consequentemente, essas técnicas não objetivam a preservação da estrutura multimodal dos dados.

Outro grupo de TECs visa minimizar as alterações na estrutura local da vizinhança dos padrões após a redução de dimensionalidade, por exemplo, *Locality Preserving Projections* (LPP) (HE; NIYOGLI, 2003; SILVA JR; CAVALCANTI; REN, 2010) e *Local Fisher Discriminant Analysis* (LFDA) (SUGIYAMA, 2007). Contudo, a preservação das

relações métricas entre os padrões não é suficiente para garantir a preservação da estrutura multimodal dos dados.

A Figura 2 ilustra a falha do PCA, LDA, LPP e LFDA na preservação da estrutura multimodal dos dados. Essa figura apresenta três modas distintas, as quais encontram em qualquer projeção na horizontal solução ideal quanto a maximização da preservação da estrutura multimodal original. PCA busca maximização da variância dos dados, enquanto que o LDA busca maximizar a dispersão entreclasses e minimizar a dispersão intraclasse. Como os objetivos dessas técnicas não envolvem percepção da estrutura multimodal dos dados, suas soluções não alcançaram a solução ideal quanto a preservação da estrutura multimodal (Figura 2a). Tanto LPP quanto LFDA buscam minimizar as alterações na estrutura local da vizinhança dos padrões, porém, este objetivo não foi suficiente para conduzir a solução obtida por cada técnica à solução ideal (Figura 2b).

Figura 2 – Falha de técnicas de extração em preservar a multimodalidade nos dados. 2a ilustra a falha do PCA e LDA enquanto que 2b ilustra a falha do LPP e LFDA.



Fonte: Autoria própria.

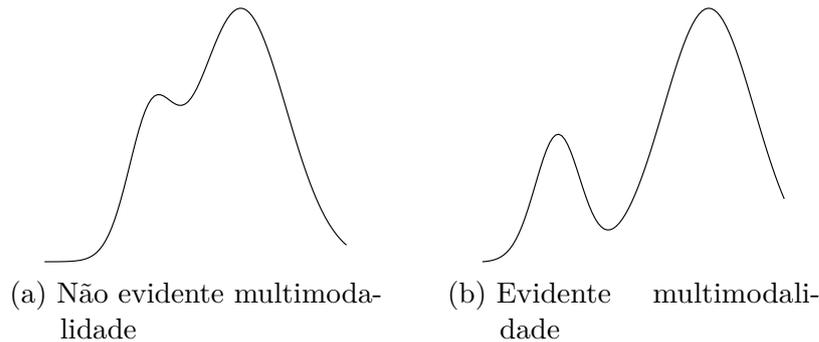
1.1 Objetivos

O objetivo primário desse trabalho é a proposição de Técnicas de Extração de Características (TECs) que preservem ao máximo a estrutura multimodal dos dados quando aplicadas a cenários com multimodalidade em suas classes (multimodalidade intraclasse).

Para o desenvolvimento dessa metodologia, um dos problemas primeiramente enfrentados é a identificação da multimodalidade intraclasse em conjuntos de padrões, uma vez que, para se avaliar o desempenho das TECs derivadas dessa metodologia, é necessário conhecer a estrutura dos dados quanto ao comportamento de sua dispersão (unimodal ou multimodal).

Identificadas as modas, o próximo problema é relativo à avaliação da multimodalidade expressa por elas. A Figura 3 mostra que a dispersão das modas podem ocorrer produzindo menor (Figura 3a) ou maior (Figura 3b) evidência da multimodalidade entre elas.

Figura 3 – Distinção entre padrões de diferentes modas para a melhoria da classificação intraclasses. 3a ilustra a dificuldade na realização desse objetivo quando a multimodalidade for pouco expressiva e 3b mostra que esse objetivo é facilitado quando há evidente multimodalidade.



Fonte: Autoria própria.

Se as modas identificadas tem multimodalidade pouco evidente (Figura 3a), implica que as modas possuem grande interferência ou interseção entre si. Conseqüentemente, distinguir os padrões da classe segundo suas modas oferece pouca informação. Desse modo, para fins de classificação sobre a classe, é mais vantajoso entender sua dispersão como unimodal.

Contudo, se as modas identificadas guardam evidente multimodalidade (Figura 3b), há também evidente distinção entre os padrões das diferentes modas. Nesse caso, se a classe fosse entendida como unimodal, ruídos entre as modas seriam frequentemente classificados por padrões da classe. Assim, distinguir os padrões da classe segundo suas modas oferece informação necessária para melhorar o acerto da classificação sobre a classe.

Assim, as TECs derivadas da metodologia proposta identificam e avaliam as modas encontradas, ou seja, verificam a multimodalidade dos dados. O resultado da verificação determinará o modo como a técnica tratará esses dados, como unimodais ou como multimodais. Desse modo, a realização de verificação de multimodalidade é um objetivo secundário e necessário à obtenção do objetivo primário.

Síntese dos objetivos

- Desenvolver TECs adaptadas a cenários com multimodalidade intraclasses.
 - Realizar verificação de multimodalidade sobre dados.

1.2 Estrutura do documento

Este documento apresenta a pesquisa realizada para o cumprimento dos objetivos relacionados na Seção 1.1. Para realizar essa apresentação, esse documento foi estruturado do seguinte modo:

Seção 2 - *Estado da arte* - contém a pesquisa bibliográfica que foi conduzida com o objetivo de encontrar na literatura soluções que contribuam para a realização dos objetivos dessa pesquisa. Os resultados dessa pesquisa bibliográfica foram agrupados segundo suas contribuições. Nesse caso, a Seção 2.1 apresenta técnicas que inferem sobre o comportamento da dispersão dos dados (unimodal ou multimodal); enquanto que a Seção 2.2 apresenta as TECs que possuem características de interesse a essa pesquisa.

Seção 3 - *Técnica classwise com investigação de multimodalidade* - Essa seção apresenta a arquitetura através da qual são desenvolvidas as TECs que cumprem o objetivo primário desse trabalho. Esse grupo de TECs é uma especialização das técnicas *classwise*. Sendo a característica da especialização a investigação da multimodalidade que é realizada sobre os padrões de cada classe. Desse modo, esse novo grupo de TECs é denominado: *classwise* com investigação de multimodalidade.

Seção 4 - *Experimentos* - Essa seção apresenta os experimentos realizados a fim de comparar e validar as técnicas propostas para cenários com multimodalidade intraclasses. As Seções 4.1-4.3 respectivamente apresentam os experimentos para avaliação de *Index of Area-based Normal Intersection*, *Multimodal Decomposition* e das três instâncias propostas das técnicas *classwise* com investigação de multimodalidade.

Seção 5 - *Considerações finais* - Essa seção salienta as contribuições que foram obtidas com esse trabalho, além de sumarizar as conclusões que foram observadas experimentalmente e apresentar rumo futuro para a continuação dessa pesquisa.

2 Estado da arte

Não se encontrou na literatura Técnica de Extração de Características (TEC) que atenda ao objetivo desse trabalho, ou seja, preservar ao máximo a estrutura multimodal dos dados em cenários com multimodalidade intraclasse. Contudo, foi encontrada na literatura uma variedade de técnicas de base estatística que possibilitam alguma inferência sobre a dispersão dos padrões quanto multimodal ou não. Adicionalmente, algumas TECs foram investigadas, dado que suas modelagens possuem características de interesse a esse trabalho.

A Seção 2.1 trata das técnicas encontradas que inferem sobre o comportamento multimodal ou não dos dados. Enquanto que a Seção 2.2 apresenta as TECs de interesse a esse trabalho. Por fim, a Seção 2.3 apresenta uma síntese do que foi encontrado no estado da arte.

2.1 Investigação da multimodalidade

Foi encontrada na literatura um conjunto de técnicas que visam estimar a ocorrência de multimodalidade nos dados. Como as técnicas que investigam multimodalidade podem ser agrupadas em quanto técnicas de identificação da multimodalidade (Definição 1.1 na página 20) e técnicas de verificação da multimodalidade (Definição 1.2 na página 20). Assim, as técnicas encontradas são apresentadas segundo sua pertinência a cada um desses grupos, sendo a Seção 2.1.1 referente identificação da multimodalidade, enquanto que a Seção 2.1.2 trata das técnicas de verificação da multimodalidade encontradas.

2.1.1 Técnicas de identificação da multimodalidade

Técnicas desse grupo estimam o número de modas na dispersão dos padrões.

Silverman (1981) propõe uma técnica que investiga a quantidade de modas existente em distribuição univariável, através de estimativas da densidade, analisadas via teste de hipótese, com hipótese nula de que a distribuição possui k modas, contra a alternativa de ter mais do que k modas. Considerando que modas se caracterizam por serem regiões de máximo local na função de densidade de uma distribuição, Sawitzki (1996) modela técnica para determinação da quantidade de modas da distribuição baseado na contagem das regiões de máximo local, dado um limiar para determinação dos máximos locais de interesse.

Técnicas de agrupamento também podem ser utilizadas na identificação de modas, com a vantagem de serem empregadas em cenários multidimensionais. Uma técnica

de agrupamento largamente utilizada na literatura é o K-médias (HARTIGAN, 1975; HARTIGAN; WONG, 1979), cujo objetivo é dividir os padrões que lhe são apresentados em K grupos, de modo que seja mínima a soma das distâncias dos padrões do grupo em relação a média do grupo. Na implementação clássica do K-médias o valor de K é dado ao algoritmo, o qual retorna K grupos mesmo que não hajam K máximos na dispersão dos padrões. *Gaussian Mixture Model* (GMM) (BILMES, 1998) é uma técnica de agrupamento com base estatística que encontra a mistura de M distribuições Normais que aproximam a dispersão dos padrões. Ela desempenha algoritmo *Expectation Maximization* (EM) (BIERNACKI; CELEUX; GOVAERT, 2003) a fim de ajustar os parâmetros das M distribuições Normais da mistura o mais verossímil possível em relação a dispersão dos padrões. Porém, assim como o K-médias, GMM necessita ser informada do número de modas que deve determinar.

2.1.2 Técnicas de verificação da multimodalidade

Técnicas desse grupo avaliam se modas previamente encontradas expressam entre si comportamento multimodal ou não.

Hartigan e Hartigan (1985) propõem um teste estatístico, denominado *dip*, que mede o desvio de distribuição univariável dos padrões quanto à unimodalidade. Isto é, o *dip* verifica se a distribuição é unimodal. Um índice como medida da multimodalidade de uma distribuição univariável é dado em Nason e Sibson (1992), que é calculado via *projection pursuit* (Friedman e Tukey (1974) e Jones e Sibson (1987)), que envolve encontrar a projeção dos padrões de maior interesse ao problema. Porém, o índice de Nason e Sibson (1992) não tem limite superior, desse modo, um valor maior do índice indica maior tendência à multimodalidade, o que torna dependente do problema em análise o entendimento quanto ao valor que indique uma distribuição de fato multimodal ou não. Minnotte (1997) propõe teste não-paramétrico na investigação da ocorrência de bimodalidade nos dados.

Outros trabalhos também investigam testes de multimodalidade (Fischer, Mammen e Marron (1994), Hall e York (2001), Chan e Tong (2004)), porém todas as técnicas encontradas na literatura foram modeladas para cenários unidimensionais. Até então não se encontrou na literatura técnica que possibilite um índice de multimodalidade sobre dados multidimensionais. Alguns fatores podem justificar tal ausência: (i) maior complexidade matemática na construção da solução multidimensional e (ii) a necessidade de modas suficientemente densas para serem percebidas é um problema em espaços multidimensionais devido a maldição da dimensionalidade (DUDA; HART; STORK, 2001; BELLMAN, 1961).

2.2 Técnicas de extração de características

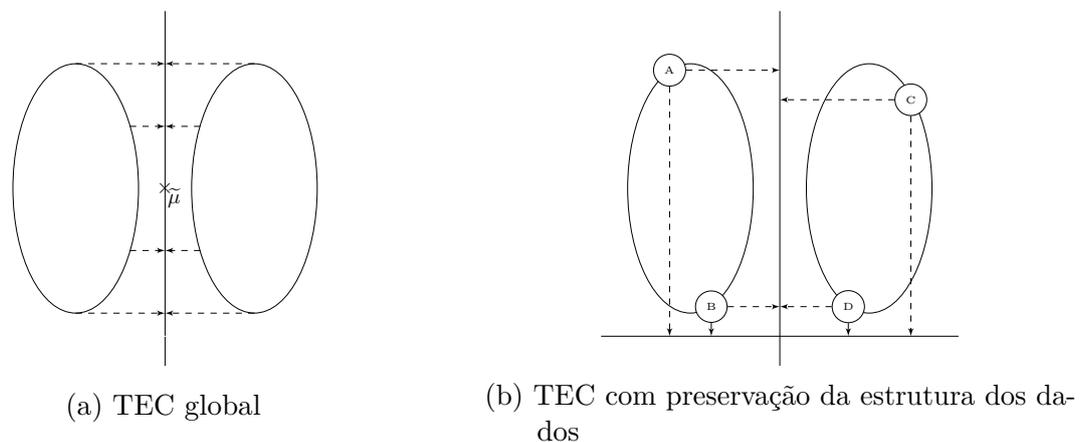
Não se encontrou na literatura TEC que preserve a estrutura multimodal de dados intraclasses. Contudo, foram encontradas TECs: (i) que objetivam preservar a estrutura da dispersão dos padrões e (ii) que operam individualmente sobre cada classe do problema.

(i) As TECs que preservam a estrutura da dispersão dos padrões são de interesse porque podem preservar a multimodalidade na dispersão dos padrões, caso exista. Assim, as TECs podem ser agrupadas como: TECs globais e TECs com preservação da estrutura dos dados.

São exemplos de TECs globais: *Principal Component Analysis* (LEITE et al., 2007) e *Linear Discriminant Analysis* (NAZ; FAROOQ; NAZ, 2006). Enquanto que *Locality Preserving Projections* (SILVA JR; CAVALCANTI; REN, 2010) e *Classwise Locality Preserving Projections* (LI; PAN; CHU, 2008) são exemplos de TECs com preservação da estrutura dos dados.

Entretanto, as TECs desses dois grupos não garantem preservação da estrutura multimodal dos dados, como ilustra a Figura 4.

Figura 4 – Técnicas de extração que não garantem preservação da estrutura multimodal. 4a apresenta esquema de funcionamento de TEC global, as quais calculam suas soluções através de parâmetros globais da dispersão dos padrões. 4b apresenta esquema de funcionamento de TEC com preservação da estrutura dos dados.



Fonte: Autoria própria.

TECs globais calculam a matriz de transformação através de parâmetros globais da dispersão dos padrões (como, média amostral global ou média amostral da classe), não havendo preocupação em explorar a estrutura dos dados. Exemplo de TEC global é *Principal Component Analysis* (PCA) (LEITE et al., 2007), que busca redução de dimensionalidade que preserve ao máximo a variância dos padrões entorno de sua média amostral global. A Figura 4a generaliza o funcionamento das TECs globais pelo funcionamento do PCA. Nessa figura, a maior variância entorno do parâmetro global ($\bar{\mu}$, média amostral global) está no eixo vertical, o que causa perda da estrutura multimodal quando os padrões

são projetados neste eixo. Objetivos similares a esse não são adequados a problemas com multimodalidade intraclasse por não perceberem e não garantirem sua preservação.

TECs com preservação da estrutura dos dados buscam por reduções que minimizem as alterações nas relações métricas entre os padrões. Ou seja, padrões que no espaço original são próximos devem permanecer próximos no espaço de menor dimensionalidade obtido pela TEC; enquanto que padrões que no espaço original são distantes devem permanecer distantes no novo espaço. Esse objetivo geral parece promissor, pois, se a estrutura dos dados é preservada, então a estrutura multimodalidade original também deve ser preservada. Entretanto, essa afirmação é falsa. Como mostra a Figura 4b, a depender do eixo escolhido para a projeção dos padrões, variam as relações métricas entre os padrões (círculos: A, B, C e D). Tomando os eixos vertical e horizontal como as únicas reduções de dimensionalidade possíveis à TEC com preservação da estrutura dos dados, então essa TEC apresentará como sua solução a matriz de transformação que projeta os padrões sobre o eixo que menos altera as relações métricas originais. Assim, o eixo vertical pode ser escolhido como solução mesmo que ele não preserve a estrutura multimodal original; basta que o eixo vertical seja o espaço com menor alteração das relações métricas originais.

A preservação do comportamento multimodal através de TEC com preservação da estrutura dos dados é obtida colateralmente, sem que haja busca explícita por redução de dimensionalidade que preserve esse comportamento. A prioridade pela minimização nas alterações das relações métricas descarta a manutenção do comportamento multimodal, caso esta manutenção prejudique a minimização. Assim, esse grupo de TECs também não garante a preservação da estrutura multimodal dos dados.

(ii) As TECs que operam individualmente sobre cada classe do problema são ditas *classwise* e elas motivam outro agrupamento das TECs, as quais podem ser agrupadas enquanto: *não-classwise* e *classwise*.

Definição 2.1 (*Classwise*). *Nomenclatura dada a abordagem que avalia cada classe de um problema separadamente.*

TECs *não-classwise* são técnicas supervisionadas ou não que calculam suas soluções observando todos os padrões do conjunto de treinamento. São exemplos desse grupo: *Principal Component Analysis* (LEITE et al., 2007), *Linear Discriminant Analysis* (NAZ; FAROOQ; NAZ, 2006), *Locality Preserving Projections* (SILVA JR; CAVALCANTI; REN, 2010), *Classwise Locality Preserving Projections* (LI; PAN; CHU, 2008) e *Local Fisher Discriminant Analysis* (SUGIYAMA, 2007).

TECs *classwise* são técnicas supervisionadas que particionam os padrões segundo suas classes. Cada classe é individualmente avaliada por uma TEC não-supervisionada, sendo sua solução definida como a solução da TEC *classwise* para a classe corrente.

Classwise Principal Component Analysis (DAS; NENADIC, 2009) e *Class-dependent Locality Preserving Projections* (SILVA JR; CAVALCANTI; REN, 2012) são exemplos de TECs *classwise*. Adicionalmente, as TECs *classwise* possuem procedimento de classificação próprio, o qual avalia um padrão de consulta segundo a solução específica obtida para cada classe, determinando em qual dessas soluções o padrão é melhor classificado.

Nas seções seguintes são definidas algumas TECs que são utilizadas nesse trabalho, as quais são agrupadas em não-*classwise* (Seção 2.2.1) e *classwise* (Seção 2.2.2).

2.2.1 Técnicas não-*classwise*

Uma TEC não-*classwise* pode ser supervisionada ou não. Sua característica principal é a obtenção de sua matriz de transformação única a partir da computação de todos os padrões que lhe são apresentados.

As seções seguintes definem algumas TECs não-*classwise* de interesse para esse trabalho, são elas: *Principal Component Analysis* (Seção 2.2.1.1), *Linear Discriminant Analysis* (Seção 2.2.1.2), *Locality Preserving Projections* (Seção 2.2.1.3), *Classwise Locality Preserving Projections* (Seção 2.2.1.4) e *Local Fisher Discriminant Analysis* (Seção 2.2.1.5).

2.2.1.1 *Principal Component Analysis*

Principal Component Analysis (PCA) (LEITE et al., 2007) é uma técnica não-supervisionada que visa encontrar um conjunto de direções que maximizem a variância de um dado conjunto de padrões. A razão para buscar essas direções se deve a pouca informação sobre os padrões que é fornecida pelas direções com baixa variância. As direções com baixa variância podem ser descartadas através da projeção dos padrões sobre as direções mais informativas (chamadas de componentes principais).

A função objetivo do PCA é dada pela Equação 2.1. \mathbf{S}_t é a matriz de dispersão total entre os padrões, definida segundo a Equação 2.2 e $\tilde{\boldsymbol{\mu}}$ é o vetor da média amostral dos padrões, Equação 2.3. Sendo o número de padrões \mathbf{x}_i dado por N .

$$\mathbf{T}_{PCA} = \arg \max_{T \in \mathbb{R}^{d \times r}} |\mathbf{T}' \mathbf{S}_t \mathbf{T}| \quad (2.1)$$

$$\mathbf{S}_t = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \tilde{\boldsymbol{\mu}}) (\mathbf{x}_i - \tilde{\boldsymbol{\mu}})' \quad (2.2)$$

$$\tilde{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (2.3)$$

Os componentes principais em \mathbf{T}_{PCA} são os autovetores correspondentes aos r maiores autovalores da matriz de dispersão total (\mathbf{S}_t); $1 \leq r \leq d$ sendo que d é a dimensionalidade original do problema.

2.2.1.2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) (NAZ; FAROOQ; NAZ, 2006) é uma técnica supervisionada que busca por uma transformação que: (i) maximize a dispersão entreclasses e (ii) minimize a dispersão intraclasse. Ou seja, LDA busca por redução de dimensionalidade que torne os agrupamentos de classes distintas o mais afastados possível e os padrões de uma mesma classe o mais próximos possível. Para isto, LDA desempenha a decomposição em autovalores e autovetores da Equação 2.4.

$$\frac{\mathbf{S}^b}{\mathbf{S}^w} \mathbf{e}_i = \lambda_i \mathbf{e}_i \quad (2.4)$$

\mathbf{T}_{LDA} é obtida como o conjunto dos autovetores $\{\mathbf{e}\}_{k=1}^r$ correspondentes aos r maiores autovalores do conjunto $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq \dots \geq \lambda_d\}$ encontrado na resolução da Equação 2.4.

$$\mathbf{S}^w = \sum_{l=1}^C \sum_{\forall \mathbf{x} \in \text{classe } l} (\mathbf{x} - \tilde{\boldsymbol{\mu}}_l) (\mathbf{x} - \tilde{\boldsymbol{\mu}}_l)' \quad (2.5)$$

$$\mathbf{S}^b = \sum_{l=1}^C N_l (\tilde{\boldsymbol{\mu}}_l - \tilde{\boldsymbol{\mu}}) (\tilde{\boldsymbol{\mu}}_l - \tilde{\boldsymbol{\mu}})' \quad (2.6)$$

$$\tilde{\boldsymbol{\mu}}_l = \frac{1}{N_l} \sum_{\forall \mathbf{x} \in \text{classe } l} \mathbf{x} \quad (2.7)$$

As matrizes \mathbf{S}^w e \mathbf{S}^b são, respectivamente, as matrizes de dispersão intraclasse (Equação 2.5) e entreclasses (Equação 2.6). \mathbf{e}_i e λ_i são, respectivamente, autovetores e autovalores da decomposição de $\mathbf{S}^b/\mathbf{S}^w$. N_l é o número de padrões de treinamento pertencentes à classe l e $\tilde{\boldsymbol{\mu}}_l$ representa o vetor amostral da média dessa classe (Equação 2.7). $\tilde{\boldsymbol{\mu}}$ é o vetor da média entre todos os padrões como definido na Equação 2.3.

Para a obtenção de \mathbf{T}_{LDA} é necessário que $\mathbf{T}' \mathbf{S}^w \mathbf{T}$ seja inversível. Para tanto, \mathbf{S}^w deve ser uma matriz não-singular. Caso \mathbf{S}^w seja singular, isto pode ser corrigido fazendo $\mathbf{S}^w = \mathbf{S}^w + \alpha \mathbf{I}_d$ (YE et al., 2004), sendo α é um número positivo pequeno e \mathbf{I}_d é a matriz identidade de ordem d .

A matriz \mathbf{S}^b tem no máximo $C-1$ vetores linearmente independentes, o que implica que LDA pode no máximo obter $C-1$ características significativas, $1 \leq r < C$.

2.2.1.3 Locality Preserving Projections

Locality Preserving Projections (LPP) (SILVA JR; CAVALCANTI; REN, 2010) é uma técnica não-supervisionada de redução de dimensionalidade que visa preservar a estrutura local da vizinhança dos padrões após a transformação. O cálculo de sua matriz de transformação é baseado na teoria espectral dos grafos (CHUNG, 1997), o qual interpreta cada padrão como um nó do grafo do problema. A percepção da estrutura local da vizinhança dos padrões é dada pela matriz de afinidade $\mathbf{A} \in \mathbb{R}^{N \times N}$ (N número de padrões). $\mathbf{A}_{i,j} \in [0, 1]$ contém a afinidade entre os padrões \mathbf{x}_i e \mathbf{x}_j . Se $\mathbf{A}_{i,j} \rightarrow 1$ então \mathbf{x}_i e \mathbf{x}_j são próximos, se $\mathbf{A}_{i,j} \rightarrow 0$ então \mathbf{x}_i e \mathbf{x}_j são afastados. Silva Jr, Cavalcanti e Ren (2010) apresentam métricas para o cálculo da afinidade entre padrões.

Na matriz $\mathbf{X}\mathbf{L}\mathbf{X}'/\mathbf{X}\mathbf{D}\mathbf{X}'$ estão contidas as relações métricas entre os padrões, as quais definem a estrutura da dispersão deles. Os autovetores associados aos menores autovalores da decomposição de $\mathbf{X}\mathbf{L}\mathbf{X}'/\mathbf{X}\mathbf{D}\mathbf{X}'$ indicam as direções de menores alterações nas relações métricas dos padrões. Assim, a matriz de transformação do LPP é formada dos autovetores correspondentes aos r menores autovalores obtidos da Equação 2.8.

$$\frac{\mathbf{X}\mathbf{L}\mathbf{X}'}{\mathbf{X}\mathbf{D}\mathbf{X}'}\mathbf{e}_i = \lambda_i\mathbf{e}_i \quad (2.8)$$

$$\mathbf{D}_{i,i} = \sum_{j=1}^N \mathbf{A}_{i,j} \quad (2.9)$$

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \quad (2.10)$$

\mathbf{D} é uma matriz diagonal de ordem N (Equação 2.9), a qual é utilizada na construção do grafo laplaciano \mathbf{L} (Equação 2.10). \mathbf{e}_i e λ_i são respectivamente autovetores e autovalores da decomposição de $\mathbf{X}\mathbf{L}\mathbf{X}'/\mathbf{X}\mathbf{D}\mathbf{X}'$.

2.2.1.4 Classwise Locality Preserving Projections

Classwise Locality Preserving Projections (CLPP) (LI; PAN; CHU, 2008) é uma extensão do LPP para problemas supervisionados. Esta extensão é dada pela alteração do cálculo da matriz de afinidade, que passa a observar o rótulo de classe dos padrões na determinação de seu grau de afinidade. A nova matriz de afinidade é dita matriz de afinidade supervisionada (\mathbf{A}^s) e, como visto em Silva Jr, Cavalcanti e Ren (2012), pode ser generalizada da matriz de afinidade para o caso não-supervisionado (\mathbf{A}) pela Equação 2.11. Ou seja, apenas se calcula afinidade para padrões de mesma classe (y_i é o rótulo da classe

do padrão \mathbf{x}_i).

$$\mathbf{A}_{i,j}^s = \begin{cases} \mathbf{A}_{i,j} & \text{se } y_i = y_j \\ 0 & \text{se } y_i \neq y_j \end{cases} \quad (2.11)$$

Uma vez definida a matriz de afinidade supervisionada, a obtenção da matriz de transformação do CLPP é análoga a obtenção da matriz de transformação do LPP (Equações 2.8-2.10).

Mesmo tendo em seu título o termo *classwise*, CLPP não está de acordo com a definição para técnicas *classwise* adotada nesse trabalho, dado que ela não observa cada classe separadamente para o cálculo de sua solução.

2.2.1.5 Local Fisher Discriminant Analysis

Local Fisher Discriminant Analysis (LFDA) (SUGIYAMA, 2007) propõe sua matriz de transformação através de objetivos herdados do LDA e LPP. Ou seja, o LFDA visa, localmente, maximizar a dispersão entre classes e minimizar a dispersão intraclasse.

Para a determinação de sua matriz de transformação são definidas as matrizes de dispersão local intraclasses $\tilde{\mathbf{S}}^w$ (Equação 2.12) e de dispersão local entreclasses $\tilde{\mathbf{S}}^b$ (Equação 2.13).

$$\tilde{\mathbf{S}}^w = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \tilde{\mathbf{W}}_{i,j}^w (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)' \quad (2.12)$$

$$\tilde{\mathbf{S}}^b = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \tilde{\mathbf{W}}_{i,j}^b (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)' \quad (2.13)$$

As matrizes $\tilde{\mathbf{W}}^w$ (Equação 2.14) e $\tilde{\mathbf{W}}^b$ (Equação 2.15) são, respectivamente, as matrizes de pesos locais intraclasse e entreclasses, que ponderam a variância entre pares de padrões. O termo $\mathbf{A}_{i,j}$ é a afinidade entre os padrões \mathbf{x}_i e \mathbf{x}_j (SILVA JR; CAVALCANTI; REN, 2010).

$$\tilde{\mathbf{W}}_{i,j}^w = \begin{cases} \mathbf{A}_{i,j}/N_l & \text{se } y_i = y_j = l \\ 0 & \text{se } y_i \neq y_j \end{cases} \quad (2.14)$$

$$\tilde{\mathbf{W}}_{i,j}^b = \begin{cases} \mathbf{A}_{i,j} (1/N - 1/N_l) & \text{se } y_i = y_j = l \\ 1/N & \text{se } y_i \neq y_j \end{cases} \quad (2.15)$$

A matriz de transformação do LFDA é composta dos autovetores associados aos r maiores autovalores da decomposição de $\tilde{\mathbf{S}}^b/\tilde{\mathbf{S}}^w$.

Sugiyama (2007) obtém experimentalmente que o LFDA alcança $1 \leq r \leq d$ reduções da dimensionalidade.

2.2.2 Técnicas *classwise*

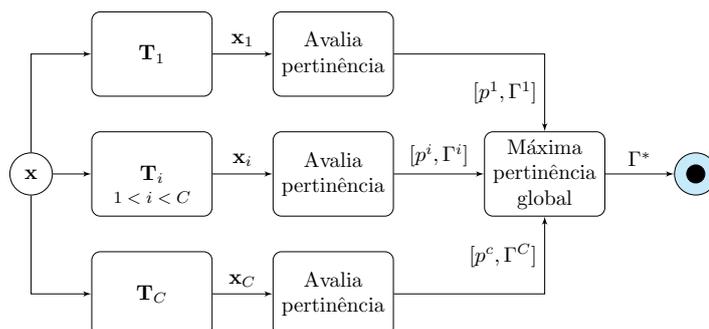
TECs não-*classwise* determinam suas matrizes de transformações avaliando as classes em conjunto, o que incorre em perda de informação específica das classes. A fim de manter a informação específica de cada classe, as TECs *classwise* avaliam cada classe separadamente, determinando uma matriz de transformação específica para cada uma dessas classes e mantendo essas estruturas na memória. Como as classes são avaliadas individualmente, implica que os padrões avaliados possuem o mesmo rótulo, o que torna essa informação não significativa para a avaliação dos padrões da classe. Assim, TECs *classwise* agregam em sua arquitetura uma TEC não-*classwise* e não-supervisionada para realizar essas avaliações.

Em resumo, TECs *classwise* são técnicas supervisionadas que determinam C matrizes de transformações (C é o número de classes do problema) por meio de TECs não-supervisionadas.

Como a matriz de transformação da classe exprime melhor suas informações específicas, se espera que no subespaço de projeção alcançado por essa matriz, os padrões relativos a essa mesma classe sejam melhor discriminados. Desse modo, um padrão com rótulo de classe desconhecido é projetado no subespaço de cada classe a fim de avaliar sua pertinência a todas as classes. Esse comportamento conduz a abordagem *classwise* a adotar procedimento de classificação particular. Detalhes sobre o procedimento de classificação adotado por TECs *classwise* serão dados na Seção 2.2.2.3.

A Figura 5 apresenta a visão geral do funcionamento do classificador para abordagem *classwise*.

Figura 5 – Classificador *classwise*. \mathbf{x} padrão de consulta a ser classificado. \mathbf{T}_i matriz de transformação obtida sobre os padrões da $i^{\text{ésima}}$ classe do problema ($i \in [1, C]$). $\mathbf{x}_i = \mathbf{T}_i' \mathbf{x}$. Γ^i é o rótulo de classe obtido para \mathbf{x}_i . p^i valor da pertinência de \mathbf{x}_i à classe Γ^i . Γ^* é o rótulo de classe obtido para \mathbf{x} . $\Gamma^* = \Gamma^i$ se $\max(p^1, p^2, \dots, p^C) = p^i$.



Fonte: Autoria própria.

Para determinar o rótulo de um padrão de consulta, o classificador *classwise* utiliza as C matrizes de transformações obtidas (de \mathbf{T}_1 até \mathbf{T}_C) para projetar o padrão em cada um dos subespaços definidos por essas matrizes. Assim, \mathbf{x}_j é a projeção do padrão de consulta \mathbf{x} no subespaço da classe j , com $1 \leq j \leq C$. No subespaço relativo à cada classe, a pertinência do padrão a todas as classes é mensurada, sendo p^j o maior valor de pertinência encontrado no subespaço relativo a classe j e Γ^j o rótulo da classe que obteve tal valor ($\Gamma^j \in [1, C]$). Dentre as classes que obtiveram os maiores valores de pertinência, aquela com o maior valor de pertinência global rotulará o padrão (Γ^* , rótulo da classe com maior valor de pertinência global).

A Tabela 1 exemplifica a classificação feita por classificador *classwise* para um problema de três classes.

		Pertinência		
		Classe 1	Classe 2	Classe 3
Subespaço	\mathbf{T}_1	0,09	0,26	0,65
	\mathbf{T}_2	0,10	0,87	0,03
	\mathbf{T}_3	0,37	0,57	0,06

Tabela 1 – Tomada de decisão por classificador *classwise*.

Nesse exemplo, a linha que contém \mathbf{T}_i é relativa a classificação do padrão de consulta quando projetado no subespaço da classe i . Ou seja, a linha que contém \mathbf{T}_i informa os valores de pertinência de \mathbf{x}_i a cada classe do problema. Assim, quando projetado por \mathbf{T}_1 , os valores de pertinência de \mathbf{x}_1 às classes 1, 2 e 3 são 0,09, 0,26 e 0,65 respectivamente. Isso implica que, no subespaço de projeção dado por \mathbf{T}_1 , o padrão de consulta \mathbf{x} é classificado como $\Gamma^1 = 3$, sendo $p^1 = 0,65$. Analogamente, para o subespaço definido por \mathbf{T}_2 , $\Gamma^2 = 2$ e $p^2 = 0,87$. Enquanto que no subespaço definido por \mathbf{T}_3 , $\Gamma^3 = 2$ e $p^3 = 0,57$. Como em cada subespaço o padrão de consulta pode ser rotulado de modo distinto, o rótulo final que representa a tomada de decisão por classificador *classwise* (Γ^*) é o rótulo obtido no subespaço que alcançou maior valor de pertinência. No caso, $\Gamma^* = \Gamma^2$ uma vez que $\max(p^1, p^2, p^3) = p^2$.

As seções seguintes definem os procedimentos de extração de características das TECs *classwise*: *Classwise Principal Component Analysis* (Seção 2.2.2.1) e *Class-dependent Locality Preserving Projections* (Seção 2.2.2.2).

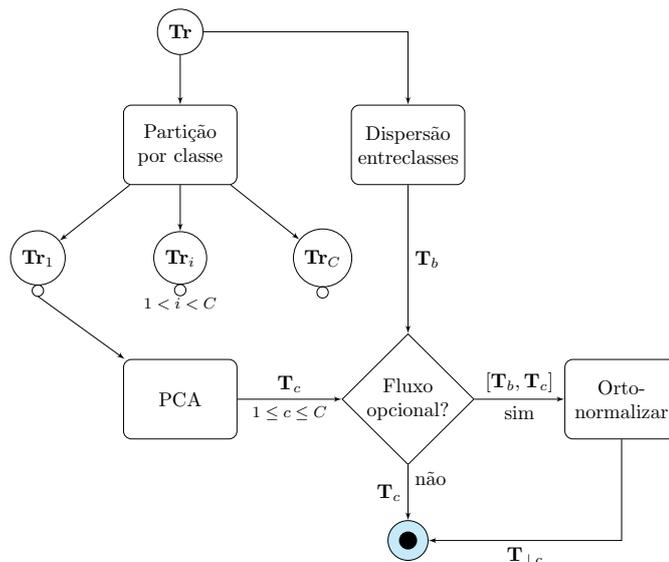
2.2.2.1 *Classwise Principal Component Analysis*

Classwise Principal Component Analysis (CPCA) (DAS; NENADIC, 2009) foi desenvolvida para tratar problemas *Small Sample Size* (SSS), que são problemas cujo número dos padrões de treinamento são insuficientes para possibilitar inferências estatísticas a determinado grau de significância.

Essa proposição se deve ao seguinte entendimento: aprender a generalizar uma classe requer menos padrões do que aprender a generalizar várias classes. Assim, a ideia principal do CPCA é particionar os padrões segundo suas classes e verificar os eixos de maiores variâncias em cada uma das partições.

A Figura 6 apresenta o funcionamento da extração de características do CPCA. Essa técnica possui dois fluxos distintos, sendo um obrigatório e o outro opcional.

Figura 6 – Extração de características do CPCA.



Fonte: Autoria própria.

O objetivo do fluxo obrigatório é obter as informações específicas à discriminação de cada classe. Para isso, no fluxo obrigatório os padrões do conjunto de treinamento (\mathbf{TR}) são particionados segundo suas classes, sendo \mathbf{Tr}_j o conjunto dos padrões de treinamento que pertencem à classe j ($1 \leq j \leq C$). Sobre as partições de cada classe é aplicado o PCA, o qual obtém a matriz de transformação que guarda os eixos de maiores variância em relação à dispersão da classe corrente (\mathbf{T}_c é a matriz de transformação do PCA quando aplicado sobre \mathbf{Tr}_c).

O fluxo opcional visa contribuir com discriminação entre as classes, extraíndo do conjunto de treinamento a matriz de transformação \mathbf{T}_b . Essa matriz guarda em suas colunas os autovetores relativos aos maiores autovalores da decomposição da matriz de dispersão entre classes, a qual é calculada segundo a Equação 2.6.

Caso o fluxo opcional não seja empregado, \mathbf{T}_c é o resultado da extração de características do CPCA para a classe c ($1 \leq c \leq C$). Caso contrário, $\mathbf{T}_b \in \mathbb{R}^{d \times r_b}$ e $\mathbf{T}_c \in \mathbb{R}^{d \times r_c}$ são concatenadas de modo que a matriz resultante pertence a $\mathbb{R}^{d \times (r_b + r_c)}$ ($1 < r_b + r_c \leq d$). A matriz resultante da concatenação é aplicada a processo de ortogonalização, o qual pode ser realizado por Gram-Schmidt (BAPAT, 2000). Sendo a matriz ortogonalizada ($\mathbf{T}_{\perp c}$) é o resultado da extração de características do CPCA para a classe c .

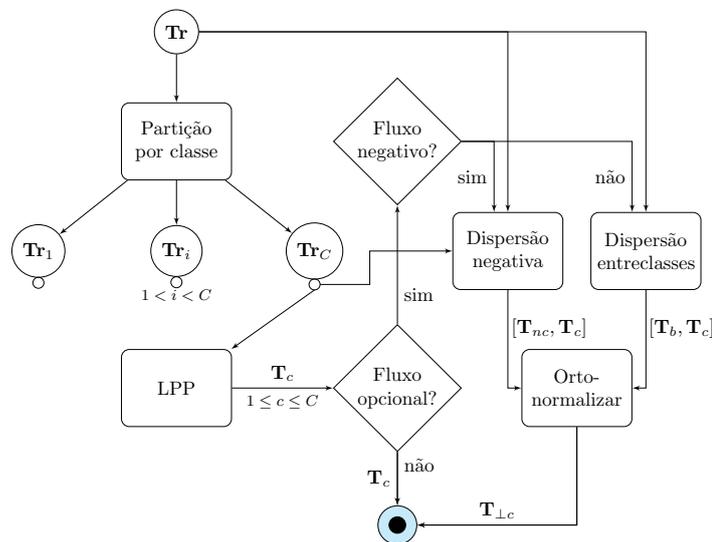
2.2.2.2 Class-dependent Locality Preserving Projections

Class-dependent Locality Preserving Projections (CdLPP) (SILVA JR; CAVALCANTI; REN, 2012) tem como objetivo preservar a estrutura local da vizinhança dos padrões de cada classe do problema. Para isso, após a partição das classes, cada uma delas é submetida individualmente ao LPP.

Assim como o CPCA, CdLPP também realiza um fluxo obrigatório, o qual desempenha abordagem *classwise* com o objetivo de manter informações específicas de cada classe. No caso, os eixos ortogonais que melhor preservam a estrutura local da vizinhança de cada classe. CdLPP também realiza um fluxo opcional, de caráter global (utiliza todo conjunto de treinamento), com o objetivo de contribuir para a discriminação entre as classes. Assim como no CPCA, a contribuição na discriminação visa maximizar a dispersão entreclasses. Contudo, um segundo fluxo opcional é proposto com o CdLPP (SILVA JR; CAVALCANTI; REN, 2012). Esse segundo fluxo opcional segue abordagem *classwise*, ou seja, calcula contribuição específica para cada classe do problema. Seu objetivo é elevar a separação da classe corrente em relação às demais classes.

A Figura 7 apresenta o funcionamento da extração de características do CdLPP. Essa técnica possui três fluxos distintos, sendo um obrigatório e os outros dois opcionais.

Figura 7 – Extração de características do CdLPP.



Fonte: Autoria própria.

No fluxo obrigatório, os padrões do conjunto de treinamento (\mathbf{Tr}) são particionados segundo suas classes, sendo \mathbf{Tr}_j o conjunto dos padrões de treinamento que pertencem a classe j ($1 \leq j \leq C$). Sobre as partições de cada classe é aplicado o LPP, o qual obtém a matriz de transformação que guarda os eixos que melhor preservam a estrutura local da vizinhança da classe corrente (\mathbf{T}_c é a matriz de transformação do LPP quando aplicado sobre \mathbf{Tr}_c).

Caso se deseje elevar a dispersão entre as classes de modo global, indica-se a utilização do fluxo opcional que calcula sobre todo conjunto de treinamento a matriz de transformação \mathbf{T}_b , a qual guarda em suas colunas os eixos que maximizam a dispersão entreclasses. Tais eixos são obtidos pela decomposição da matriz de dispersão entreclasses (Equação 2.6) em autovetores e autovalores.

Porém, caso se deseje elevar a dispersão entre as classes de modo específico para cada classe, Silva Jr, Cavalcanti e Ren (2012) propõem a utilização da matriz de dispersão negativa, a qual é calculada individualmente para cada classe do problema. Sendo \mathbf{T}_c a matriz de transformação obtida pelo LPP para a classe corrente, então \mathbf{T}_{nc} é a matriz negativa da classe corrente. \mathbf{T}_{nc} é composta pelos autovetores relativos aos maiores autovalores da decomposição da matriz negativa da classe c (\mathbf{S}^n_c), a qual é calculada pela Equação 2.16 para uma classe c qualquer ($1 \leq c \leq C$).

$$\mathbf{S}^n_c = \sum_{\forall \mathbf{x} \notin \text{classe } c} (\mathbf{x} - \tilde{\boldsymbol{\mu}}_c)(\mathbf{x} - \tilde{\boldsymbol{\mu}}_c)' \quad (2.16)$$

$\tilde{\boldsymbol{\mu}}_c$ é calculada pela Equação 2.7.

Caso nenhum dos fluxos opcionais sejam empregados, \mathbf{T}_c é o resultado da extração de características do CdLPP para a classe c ($1 \leq c \leq C$). Caso um dos fluxos opcionais seja adotado, a matriz de transformação por ele é concatenada com \mathbf{T}_c . Na abordagem global a concatenação realizada é $[\mathbf{T}_b, \mathbf{T}_c]$. Enquanto que na abordagem *classwise* a concatenação usada é $[\mathbf{T}_{nc}, \mathbf{T}_c]$.

A matriz resultante da concatenação é aplicada a processo de ortonormalização (Gram-Schmidt (BAPAT, 2000)). Sendo a matriz ortonormalizada ($\mathbf{T}_{\perp c}$) é o resultado da extração de características do CdLPP para a classe c .

2.2.2.3 Classificador das técnicas *classwise*

As técnicas *classwise* são modeladas através de dois procedimentos: extração de características e classificação. Essa seção apresenta o algoritmo de classificação utilizado pelas técnicas *classwise* convencionais.

O algoritmo pelo qual esse procedimento de classificação é realizado é apresentado no Algoritmo 1. Para a realização desse procedimento, o classificador das TECs *classwise* convencionais deve receber, além do padrão \mathbf{z} a ser classificado, as estruturas que exprimem o conhecimento obtido sobre a dispersão de cada uma das C classes durante o procedimento de extração de características. Essas estruturas, obtidas para cada classe c ($1 \leq c \leq C$) são \mathbf{T}_c , $\tilde{\boldsymbol{\mu}}_c$, \mathbf{S}_c e $\boldsymbol{\rho}_c$ respectivamente a matriz de transformação, o vetor de média, a matriz de covariância e a probabilidade *a priori* estimados para a classe c . Com essas informações o Algoritmo 1 avalia a pertinência de \mathbf{z} para cada classe. Essa avaliação é

repetida dentro do espaço resultante da projeção por cada matriz de transformação \mathbf{T}_c . Assim, esse algoritmo avalia C^2 possibilidades de pertinência a fim de determinar a classe da matriz de transformação i^* e a classe j^* que resultaram no maior valor de pertinência dado \mathbf{z} .

Algoritmo 1: Procedimento de classificação para TECs *classwise* convencionais.

```

Input:  $\mathbf{z}$  // Padrão de consulta
Input:  $\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_C,$  // Matrizes de transformações para cada classe
Input:  $\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\mu}}_2, \dots, \tilde{\boldsymbol{\mu}}_C$  // Vetores das médias de cada classe
Input:  $\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_C$  // Matrizes de covariância de cada classe
Input:  $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots, \boldsymbol{\rho}_C$  // Valores das prob. a priori de cada classe
Output:  $i^*$ ; // Rótulo da transformação que melhor representa  $\mathbf{z}$ 
Output:  $j^*$ ; // Rótulo de classe obtido para  $\mathbf{z}$ 

1  $\mathbf{P} \leftarrow \text{zeros}(C, C);$ 
2 for  $i \leftarrow 1$  to  $C$  do // Para cada transformação  $\mathbf{T}_i$  */
3    $\mathbf{T} \leftarrow \mathbf{T}_i;$ 
4    $\mathbf{z}_i \leftarrow \mathbf{T}' \times \mathbf{z};$  // Projeta  $\mathbf{z}$  com  $\mathbf{T}$ 
5   for  $j \leftarrow 1$  to  $C$  do // Para cada classe  $j$  */
6      $\tilde{\boldsymbol{\mu}}_{ij} \leftarrow \mathbf{T}' \times \tilde{\boldsymbol{\mu}}_j;$  // Projeta  $\tilde{\boldsymbol{\mu}}_j$  com  $\mathbf{T}$ 
7      $\mathbf{S}_{ij} \leftarrow \mathbf{T}' \times \mathbf{S}_j \times \mathbf{T};$  // Projeta  $\mathbf{S}_j$  com  $\mathbf{T}$ 
8      $\mathbf{P}_{ij} \leftarrow \mathcal{N}(\mathbf{z}_i, \tilde{\boldsymbol{\mu}}_{ij}, \mathbf{S}_{ij}) \boldsymbol{\rho}_j;$ 
9   end
10   $\mathbf{P}_i \leftarrow \mathbf{P}_i / \text{sum}(\mathbf{P}_i);$  // Padroniza valores da linha  $i$  de  $\mathbf{P}$  entre  $[0, 1]$ 
11 end
12  $[i^*, j^*] \leftarrow \text{max}(\mathbf{P});$  // Índices de linha e coluna do elemento máximo em  $\mathbf{P}$ 

```

A primeira operação desse algoritmo é iniciar a matriz \mathbf{P} que armazenará os C^2 valores de pertinência de \mathbf{z} . Nas linhas 2 à 11, são calculadas as pertinências de \mathbf{z} para todas as classes, porém dentro do espaço resultante da projeção com a matriz de transformação da classe corrente. Assim, na linha 4, \mathbf{z}_i representa a projeção de \mathbf{z} com a matriz \mathbf{T}_i ($1 \leq i \leq C$). De modo análogo, para cada uma das classes j ($1 \leq j \leq C$), $\tilde{\boldsymbol{\mu}}_{ij}$ (linha 6) representa a projeção do vetor $\tilde{\boldsymbol{\mu}}_j$ com a matriz \mathbf{T}_i e, de mesmo modo, \mathbf{S}_{ij} (linha 7) representa a projeção da matriz de covariância em \mathbf{S}_j com a matriz \mathbf{T}_i . Realizadas essas projeções, \mathbf{P}_{ij} (linha 8) é o valor da pertinência de \mathbf{z} à classe j no espaço de projeção da classe i . Sendo esse valor calculado com base na regra de Bayes. Os índices i^* e j^* respectivamente da linha e coluna do maior valor em \mathbf{P} são retornados por esse algoritmo. i^* indica a classe cujo espaço de projeção obteve o maior valor de pertinência para \mathbf{z} . Enquanto que j^* indica a classe que obteve o maior valor de pertinência para \mathbf{z} dentro do espaço de projeção de i^* . Assim, \mathbf{z} é rotulado como pertencente a classe j^* .

2.3 Considerações finais

Não se encontrou na literatura TEC que realize o objetivo de preservar a estrutura multimodal dos dados em cenários com multimodalidade intraclasses. Contudo, a realização desse objetivo depende do sucesso na execução de três etapas: (i) agrupar os padrões

segundo suas classes; (ii) identificar a estrutura multimodal na dispersão de cada classe (caso exista) e (iii) verificar se as modas, previamente identificadas na dispersão da classe, guardam entre si multimodalidade tal que possibilite melhoria na classificação sobre a classe observada.

A etapa (i) é realizada por abordagem *classwise* como definida para as TECs da Seção 2.2.2.

A Seção 2.1.1 apresenta técnicas que realizam a etapa (ii). Algumas dessas técnicas são modeladas para cenários unidimensionais. Porém, técnicas de agrupamento podem ser alternativas viáveis para identificação de multimodalidade em cenários multidimensionais. Contudo, o emprego de técnicas de agrupamento na realização da etapa (ii) não é direto, fazendo-se necessário a resolução de alguns problemas, como: adoção de heurística para determinação autônoma da quantidade de modas encontradas na dispersão dos padrões e decidir quanto a utilização de procedimentos paramétricos ou não-paramétricos na busca pelas modas.

Na Seção 2.1.2 encontram-se testes estatísticos que inferem se uma dada dispersão univariável é unimodal ou multimodal. Das técnicas encontradas, a técnica proposta por Nason e Sibson (1992) é que mais se aproxima da realização da etapa (iii), uma vez que essa técnica gera um índice que exprime a evidência da multimodalidade entre duas modas. Contudo, o índice de Nason e Sibson (1992) só se aplica a cenários unidimensionais e, por não possuir um limite superior, a interpretação de seu valor não possibilita a definição de limiares nos quais se possa afirmar que, abaixo de determinado valor, o par de modas segue distribuição unimodal ou, acima de tal valor, o par de modas segue distribuição multimodal.

3 Técnica *classwise* com investigação de multimodalidade

Como foi visto na Seção 2, não se encontrou na literatura Técnica de Extração de Características (TEC) que preserve a estrutura multimodal intraclasse para um conjunto de padrões. Assim, essa Seção apresenta a arquitetura geral através da qual são desenvolvidas TECs que cumpram o objetivo primário desse trabalho. Essas TECs, aqui denominadas TECs *classwise* com investigação de multimodalidade, possuem as seguintes características: (i) operam individualmente sobre cada classe do problema, (ii) preservam a estrutura multimodal de cada classe (caso exista) e (iii) utilizam a informação da multimodalidade para melhorar a classificação. A Figura 8 apresenta a divisão em subseções desta seção, as quais explicam o modo como essas características são alcançadas.

Figura 8 – Características das técnicas *classwise* com investigação de multimodalidade.



Fonte: Autoria própria.

A Seção 3.1 apresenta a abordagem utilizada para o cumprimento da característica (i). Na Seção 3.2 é definida a TEC não-supervisionada que atende a característica (ii), referente a preservação da estrutura multimodal dos dados. A utilização da informação de multimodalidade para melhorar a classificação é explicada na Seção 3.3. As abordagens encontradas para o cumprimento das características (i)-(iii) das técnicas *classwise* com investigação de multimodalidade são aplicadas na Seção 3.4 que apresenta a proposição de três TECs desse grupo.

3.1 Operar classes individualmente

Encontra-se na literatura TECs que seguem abordagem *classwise* (Definição 2.1 na página 27). Essa abordagem possibilita a investigação intraclasse do problema, sendo

utilizada para o cumprimento da característica (i) das técnicas *classwise* com investigação de multimodalidade.

As TECs podem ser agrupadas em não-*classwise* e *classwise*. TECs não-*classwise* são aquelas que obtêm única matriz de transformação calculada sobre todos os padrões do conjunto de treinamento (Seção 2.2.1). As TECs que seguem abordagem *classwise* (Seção 2.2.2) particionam o conjunto de treinamento segundo suas classes. Em seguida, cada classe é submetida a alguma outra TEC, não-supervisionada, a qual gera matriz de transformação específica para a classe corrente. Cabe à TEC não-supervisionada que foi escolhida, o cumprimento dos objetivos que se desejam alcançar por classe.

O objetivo que se deseja alcançar em cada classe do problema é a preservação de sua estrutura multimodal (caso exista). Como não foi encontrada na literatura TEC (supervisionada ou não) que realize a preservação da estrutura multimodal dos dados (Seção 2), um objetivo específico deste trabalho é a construção de uma TEC não-supervisionada que investigue a multimodalidade intraclasse e que preserve ao máximo essa estrutura após a redução de dimensionalidade (Seção 3.2). Essa TEC não-supervisionada, proposta na Seção 3.2, deve ser empregada junto a abordagem *classwise* a fim de cumprir as duas primeiras características (i e ii).

3.2 Preservar estrutura multimodal das classes

Nesta seção é proposta TEC denominada *Multimodal Decomposition* (MD). Essa técnica, não supervisionada e não linear, busca preservar ao máximo a estrutura multimodal da dispersão dos padrões.

Para desempenhar reduções de dimensionalidade que mantenham a estrutura multimodal deve-se, de algum modo, estimar o comportamento (unimodal ou multimodal) da distribuição dos padrões. E, a partir dessa estimativa, propor reduções que privilegiem a preservação multimodalidade original. Porém, é complexa a tarefa de estimar o comportamento multimodal da dispersão de padrões. Essa complexidade se deve a dificuldade na resolução das seguintes questões: (i) como determinar a quantidade de modas existente na dispersão dos padrões? (ii) como determinar a distribuição estatística que modela cada uma das modas? (iii) como avaliar se a mistura das modas apresentam comportamento multimodal? (iv) espaços de maiores dimensionalidades requerem maior cardinal de padrões a fim de possibilitar inferências com determinado grau de corretude; como a variabilidade do cardinal dos padrões por moda influencia no modelo multimodal proposto?

Para reduzir a complexidade da tarefa de estimar a multimodalidade na dispersão de padrões, muitos autores decidem pela redução do escopo do problema para a proposição de suas técnicas. Reduções de escopo frequentemente encontradas na literatura incluem a modelagem das técnicas para: (i) espaços unidimensionais (HALL; YORK, 2001; CHAN;

TONG, 2004), (ii) modas com distribuição estatística Normal (GNANADESIKAN; GUPTA, 1970; RAY; LINDSAY, 2005) e (iii) cenários bimodais (BEHBOODIAN, 1970; SARMA; RAO; RAO, 1990).

Encontrou-se na literatura uma variedade de técnicas que possibilitam alguma inferência sobre a dispersão dos padrões quanto multimodal ou não. Por exemplo, Silverman (1981) propõe uma técnica que investiga a quantidade de modas existente em distribuição unidimensional, através de estimativas da densidade, analisadas via teste de hipótese, com hipótese nula de que a distribuição possui k modas, contra a alternativa de ter mais do que k modas. Considerando que modas se caracterizam por serem regiões de máximo local na função de densidade de uma distribuição, Sawitzki (1996) modela técnica para determinação da quantidade de modas da distribuição baseado na contagem das regiões de máximo local, dado um limiar para determinação dos máximos locais de interesse. Hartigan e Hartigan (1985) propõem um teste estatístico, denominado *dip*, que mede o desvio de distribuição univariável dos padrões quanto à unimodalidade. Isto é, o *dip* verifica se a distribuição é unimodal. Um índice como medida da multimodalidade de uma distribuição univariável é dado em Nason e Sibson (1992), que é calculado via *projection pursuit* (Friedman e Tukey (1974) e Jones e Sibson (1987)), que envolve encontrar a projeção dos padrões de maior interesse ao problema. Porém, o índice de Nason e Sibson (1992) não tem limite superior, desse modo, um valor maior do índice indica maior tendência à multimodalidade, o que torna dependente do problema em análise o entendimento quanto ao valor que indique uma distribuição de fato multimodal ou não. Minnotte (1997) propõe teste não-paramétrico na investigação da ocorrência de bimodalidade nos dados. Outros trabalhos também investigam testes de multimodalidade ((FISCHER; MAMMEN; MARRON, 1994), (HALL; YORK, 2001), (CHAN; TONG, 2004)), porém todas as técnicas encontradas na literatura foram modeladas para cenários unidimensionais.

Devido a isso, MD realiza procedimento de busca pelas direções (espaços unidimensionais) que contêm maior comportamento multimodal referente a dispersão multidimensional dos padrões originais.

As seções seguintes tratam da proposição do MD. A Seção 3.2.1 apresenta a herística utilizada na modelagem do MD, enquanto que a Seção 3.2.2 aborda o algoritmo de otimização utilizado para busca das direções com maiores evidências de apresentar comportamento multimodal. A evidência de comportamento multimodal é traduzida segundo uma medida a qual é utilizada como critério de escolha das direções. A Seção 3.2.3 apresenta formas pelas quais a multimodalidade pode ser medida.

3.2.1 Decomposição espacial

Devido a limitação a cenários unidimensionais das técnicas que se destinam a estimar o comportamento multimodal da dispersão dos padrões, a técnica *Multimodal*

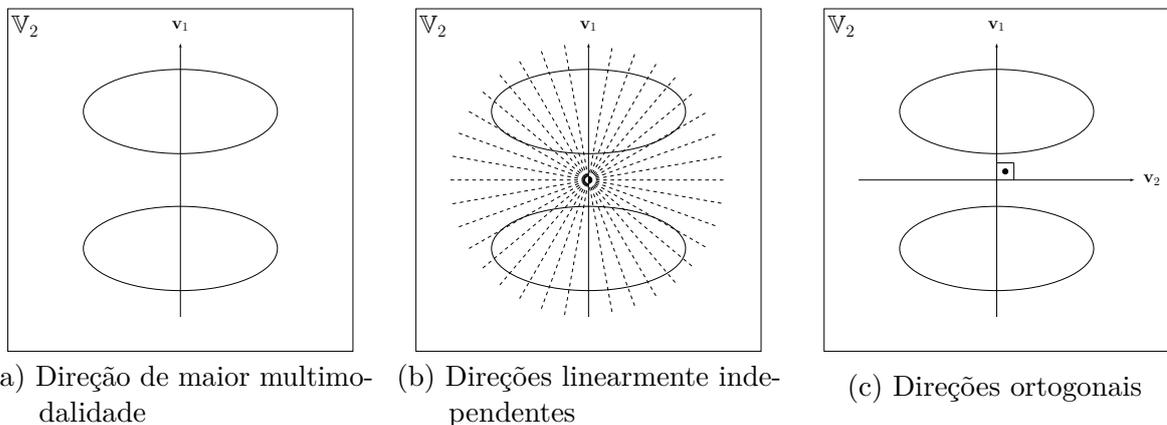
Decomposition (MD) foi modelada para buscar por direções unidimensionais com maiores valores de estimativa de multimodalidade.

Semelhante ao *Principal Component Analysis*, a primeira direção retornada pelo MD (\mathbf{v}_1) possui o maior valor de estimativa de multimodalidade. A segunda direção (\mathbf{v}_2) possui o maior valor de estimativa de multimodalidade dada a restrição de ser ortogonal a primeira direção ($\mathbf{v}_2 \perp \mathbf{v}_1$). Seguindo esse procedimento, a r^{th} direção retornada pelo MD possui o maior valor de estimativa de multimodalidade dada a restrição de ser ortogonal a todas direções anteriormente obtidas ($\mathbf{v}_r \perp \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{r-1}\}$).

As direções retornadas pelo MD são obtidas através de algoritmo de busca e otimização (Seção 3.2.2) utilizando uma função de custo (Seção 3.2.3) que realiza a estimativa da multimodalidade da projeção dos padrões sobre a direção em análise.

A Figura 9 ilustra o procedimento de busca realizado no espaço bidimensional \mathbb{V}_2 , que contém dispersão de padrões agrupados em duas modas, as quais são representadas pelas elipses.

Figura 9 – Decomposição em espaços unidimensionais do espaço \mathbb{V}_2 (bidimensional). 9a apresenta \mathbf{v}_1 , a decomposição unidimensional de \mathbb{V}_2 na direção da maior multimodalidade presente na dispersão de padrões representada pelas elipses. 9b salienta as infinitas direções sobre as quais um vetor linearmente independentes à \mathbf{v}_1 pode ser encontrado. Dentre essas infinitas possibilidades, 9c mostra que apenas a direção que contém o vetor \mathbf{v}_2 é tal que $\mathbf{v}_1 \perp \mathbf{v}_2$.



Fonte: Autoria própria.

A aplicação do MD sobre essa dispersão multimodal primeiramente obtém \mathbf{v}_1 (Figura 9a), a direção da maior multimodalidade presente na dispersão. Na busca pela segunda direção (Figura 9b), há infinitas direções candidatas cujas estimativas de multimodalidade seriam próximas à obtida para \mathbf{v}_1 . Mais precisamente, uma rotação infinitesimal de \mathbf{v}_1 resulta em um vetor linearmente independente a ele e que possui estimativa de multimodalidade aproximadamente igual à máxima. Nessa situação $\mathbf{v}_2 \approx \mathbf{v}_1$ e não representaria ganho significativo de informação. Assim, para garantir unicidade e maior distinção da solução, MD busca $\mathbf{v}_2 \perp \mathbf{v}_1$ (Figura 9c).

O Algoritmo 2 (`MultimodalDecomposition`) apresenta o procedimento de decomposição espacial com base na multimodalidade. Esse algoritmo recebe um conjunto de padrões \mathcal{X} em um espaço d -dimensional e retorna (i) a matriz $\mathbf{V} \in \mathbb{R}^{d \times d}$ cujas colunas são as direções ortogonais com maiores estimativas de multimodalidade e (ii) o vetor $\boldsymbol{\xi} \in \mathbb{R}^{1 \times d}$ que contém os valores de estimativa de multimodalidade obtidos para cada coluna de \mathbf{V} .

Algoritmo 2: MultimodalDecomposition - Encontra direções com maior comportamento multimodal

```

Input:  $\mathcal{X}$  // Conjunto de padrões em espaço  $d$ -dimensional
Output:  $\mathbf{V}$  // Matriz direções com maiores multimodalidade
Output:  $\boldsymbol{\xi}$  // Vetor das estimativas de multimodalidade das colunas de  $\mathbf{V}$ 
1  $d \leftarrow \text{NumDimensions}(\mathcal{X});$  // Dimensionalidade dos padrões em  $\mathcal{X}$ 
2  $\mathbf{V} \leftarrow \emptyset;$  // Inicializa  $\mathbf{V}$ 
3  $\mathbf{T} \leftarrow \mathbf{I}_d;$  // Inicializa  $\mathbf{T}$ 
4 for  $i = 1$  to  $d$  do
5   if  $\mathbf{V} \neq \emptyset$  then
6      $\mathbf{Y} \leftarrow \mathbf{V} \times \mathbf{V}';$  // Matriz auxiliar
7      $[\mathbf{E}, \boldsymbol{\Lambda}] \leftarrow \text{Decompose}(\mathbf{Y});$  // Autovetores ( $\mathbf{E}$ ) e autovalores ( $\boldsymbol{\Lambda}$ )
8      $\text{nulls} \leftarrow \text{Find}(\text{MainDiag}(\boldsymbol{\Lambda}) == 0);$  // Encontra autovalores nulos
9      $\mathbf{T} \leftarrow \mathbf{E}(:, \text{nulls});$  // Obtém autovetores dos autovalores nulos
10  end
11   $[\mathbf{v}_*, \xi_*] \leftarrow \text{SearchMostMultimodalDirection}(\mathbf{T}^t \times \mathcal{X});$ 
12   $\boldsymbol{\xi} \leftarrow \text{Concatenate}(\boldsymbol{\xi}, \xi_*);$ 
13   $\mathbf{V} \leftarrow \text{Concatenate}(\mathbf{V}, \mathbf{T} \times \mathbf{v}_*);$ 
14 end

```

A partir de \mathcal{X} , `MultimodalDecomposition` obtém a dimensionalidade do problema (linha 1). Inicialmente, $\mathbf{V} \leftarrow \emptyset$, dado que nenhuma direção foi até então pesquisada. A matriz \mathbf{T} é utilizada para projetar o espaço de busca pela direção com maior estimativa de multimodalidade para o espaço ortogonal às colunas de \mathbf{V} . Assim, como inicialmente \mathbf{V} está vazio, o espaço d -dimensional original é o próprio espaço ortogonal à \mathbf{V} . Desse modo, na linha 3, \mathbf{T} recebe os vetores de base do espaço original (colunas de \mathbf{I}_d). Nas linhas 4-14, um laço com d iterações é realizado para a busca das direções ortogonais com maior estimativa de multimodalidade. Em sua primeira iteração \mathbf{V} está vazio, implicando que a busca pela primeira direção é realizada no espaço original d -dimensional pela função `SearchMostMultimodalDirection` (linha 11), a qual terá seu funcionamento explicado na Seção 3.2.2.

A função `SearchMostMultimodalDirection` recebe o conjunto de padrões projetados sobre o espaço ortogonal às colunas de \mathbf{V} ($\mathbf{T}^t \times \mathcal{X}$), retornando \mathbf{v}_* e ξ_* respectivamente a direção com maior estimativa de multimodalidade e o valor da estimativa de multimodalidade obtido para a direção \mathbf{v}_* .

Nas linhas 12 e 13, as estruturas $\boldsymbol{\xi}$ e \mathbf{V} são atualizadas com os valores de ξ_* e $\mathbf{T} \times \mathbf{v}_*$ recém obtidos. A multiplicação $\mathbf{T} \times \mathbf{v}_*$ projeta \mathbf{v}_* do espaço ortogonal às colunas de \mathbf{V} para o espaço d -dimensional original. Assim, a concatenação realizada na linha 13 incrementa o

conjunto das colunas de \mathbf{V} com mais um vetor ortogonal às colunas anteriormente contidas em \mathbf{V} e que indica a direção com maior estimativa de multimodalidade. Desse modo, ao final da primeira iteração de `MultimodalDecomposition`, $\mathbf{V} = [\mathbf{v}_1]$.

Nas iterações seguintes de `MultimodalDecomposition`, \mathbf{V} não está mais vazia, sendo executadas as operações entre as linhas 5 e 10. A linha 6 apresenta a obtenção da matriz $\mathbf{Y} \in \mathbb{R}^{d \times d}$. O modo como essa matriz é calculada indica que suas linhas ou colunas são combinações lineares das colunas de \mathbf{V} . Assim, se $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$ com $1 \leq r < d$, obrigatoriamente, a decomposição de \mathbf{Y} (linha 7) obtém r autovalores não nulos (linha 8). Com isso, dado que $\mathbf{Y}' = (\mathbf{V} \times \mathbf{V}')' = (\mathbf{V}')' \times \mathbf{V}' = \mathbf{V} \times \mathbf{V}' = \mathbf{Y}$ (\mathbf{Y} é simétrica), a sua decomposição espectral pode ser reduzida à Equação 3.1, sendo λ_i um autovalor não nulo pertencente à $\mathbf{\Lambda}$ e \mathbf{e}_i o autovetor pertencente à matriz \mathbf{E} ; sendo \mathbf{e}_i associado à λ_i .

$$\mathbf{Y} = \sum_{i=1}^r \lambda_i \mathbf{e}_i \mathbf{e}_i' \quad (3.1)$$

Isso mostra que as linha ou colunas de \mathbf{Y} também são combinações lineares dos autovetores associados aos r autovalores não nulos de \mathbf{Y} (considere $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_r\}$). Desse modo, as colunas de \mathbf{V} também são combinações lineares de $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_r\}$. A consequência disso é que qualquer vetor resultante da combinação linear de $\{\mathbf{e}_{r+1}, \mathbf{e}_{r+2}, \dots, \mathbf{e}_d\}$ é ortogonal as colunas de \mathbf{V} . Portanto, o espaço que tem como bases os autovetores associados aos $d - r$ autovalores nulos da decomposição de \mathbf{Y} definem o espaço no qual a direção \mathbf{v}_{r+1} deve ser procurada. Assim, $\mathbf{T} = [\mathbf{e}_{r+1}, \mathbf{e}_{r+2}, \dots, \mathbf{e}_d]$ (linha 9).

3.2.2 Busca das direções com maior multimodalidade

Essa seção explica o procedimento de busca pela direção com maior estimativa de multimodalidade que é ortogonal as colunas da matriz \mathbf{V} (Seção 3.2.1). Essa direção é encontrada no espaço nulo de \mathbf{V} , o qual tem vetores de base armazenados nas colunas da matriz $\mathbf{T} \in \mathbb{R}^{d \times (d-r)}$, com $1 \leq r < d$.

A direção \mathbf{v}_* tem maior estimativa de multimodalidade no espaço nulo de \mathbf{V} . Sendo a obtenção dessa direção realizada através de algoritmo de busca e otimização. Duas estratégias de otimização frequentemente utilizadas, com sucesso, para uma variedade de problemas são: *Genetic Algorithm* (GA) (CHAMBERS, 2001) e *Particle Swarm Optimization* (PSO) (KENNEDY; EBERHART, 1995).

GA é inspirado nos princípios da genética e evolução das espécies. Ele simula o comportamento reprodutivo observado em populações biológicas para realizar a busca pela solução ótima no espaço de soluções. Essa estratégia emprega o princípio da sobrevivência do mais apto em seu processo de busca. Isso permite que, a cada geração (iteração), os indivíduos (soluções) melhor adaptados (melhor desempenho medido por uma função

de custo) tenham maiores chances de reprodução, passando para os novos indivíduos as características desejáveis à otimização da função de custo.

PSO é inspirado pela capacidade de adaptação ao meio ambiente observada em bandos de pássaros, cardumes de peixes ou enxame de insetos. O comportamento colaborativo dos indivíduos (partículas) do enxame possibilita encontrar fontes ricas de alimentos (locais do espaço de soluções com melhores desempenhos medidos por uma função de custo). Essa colaboração é implementada através da troca de informação entre os indivíduos do enxame, que exploram o espaço de soluções em busca dos locais (soluções) onde uma dada função de custo é otimizada. Com o passar das iterações, as partículas entorno de um mínimo ou máximo local da função de custo tem seus deslocamentos reduzidos, a fim de possibilitar a convergência por parte dessa estratégia.

GA e PSO são similares no sentido de que elas são duas estratégias de busca baseadas em população e ambas compartilham informações entre os membros da população; o que possibilita melhorar a busca que realizam. Devido a essas semelhanças, Hassan et al. (2005) comparam o GA e o PSO usando uma abordagem formal de teste de hipóteses. Seus objetivos são verificar a qualidade da solução obtida e a eficiência computacional dessas estratégias. Em seus resultados, Hassan et al. (2005) mostram que GA e PSO obtêm soluções de alta qualidade (nível de confiança de 99%). Porém, o esforço computacional requeridos pelo PSO para chegar às soluções de elevada qualidade é menor do que o esforço necessário ao GA para chegar às mesmas soluções. Hassan et al. (2005) relacionam maior diferença entre a superioridade da eficiência computacional do PSO em relação ao GA quando utilizados para resolver problemas não-lineares (sem restrições) com variáveis contínuas; sendo menos significativa essa diferença quando as estratégias são aplicadas à problemas não-lineares (com restrições) com variáveis contínuas ou discretas.

Assim, com base nos resultados de Hassan et al. (2005), a estratégia de otimização implementada na função `SearchMostMultimodalDirection`, que é empregada no Algoritmo 2, é baseada no PSO (KAMEYAMA, 2009). A Seção 3.2.2.1 apresenta a configuração do PSO para a busca da direção \mathbf{v}_* obteve maior estimativa de multimodalidade no espaço nulo de \mathbf{V} .

3.2.2.1 Direção com maior multimodalidade

Na modelagem da função `SearchMostMultimodalDirection`, a aplicação do PSO possibilita encontrar, no espaço ortogonal às colunas de \mathbf{V} , o vetor $\mathbf{v}_* \in \mathbb{R}^{(d-r) \times 1}$ que maximiza a estimativa de multimodalidade obtida nesse espaço. A busca por \mathbf{v}_* é realizada por N_p partículas. Cada partícula i ($1 \leq i \leq N_p$) do PSO guarda as seguintes informações:

Partícula i :	\mathbf{v}_i^\perp	\mathbf{p}_i	\mathbf{g}_i	Δ_i
	$\mathcal{V}(\mathbf{v}_i^\perp)$	$\mathcal{V}(\mathbf{p}_i)$	$\mathcal{V}(\mathbf{g}_i)$	t_i

\mathbf{v}_i^\perp é a posição em que a partícula i se encontra no espaço de soluções; \mathbf{p}_i a posição com maior valor na função de custo dentre as posições visitadas pela partícula i ; \mathbf{g}_i a posição com maior valor na função de custo dentre as posições visitadas por todas as partículas que compõem a vizinhança da partícula i ; Δ_i é o deslocamento realizado pela partícula i para alcançar a posição atual; t_i escalar que informa por quantas iterações do PSO a partícula i se deslocou sem sofrer mutação. Enquanto que $\mathcal{V}(\mathbf{v}_i^\perp)$, $\mathcal{V}(\mathbf{p}_i)$ e $\mathcal{V}(\mathbf{g}_i)$ são os respectivos valores de estimativa de multimodalidade (função de custo \mathcal{V}) obtidos sobre \mathbf{v}_i^\perp , \mathbf{p}_i e \mathbf{g}_i .

A cada iteração do PSO, \mathbf{v}_i^\perp e Δ_i de cada partícula é alterado respectivamente segundo as Equações 3.2 e 3.3.

$$\mathbf{v}_i^\perp(it+1) = \mathbf{v}_i^\perp(it) + \Delta_i(it+1) \quad (3.2)$$

$$\Delta_i(it+1) = \omega^{t_i} \Delta_i(it) + \phi_1 \mathbf{R}_1 (\mathbf{p}_i - \mathbf{v}_i^\perp(it)) + \phi_2 \mathbf{R}_2 (\mathbf{g}_i - \mathbf{v}_i^\perp(it)) \quad (3.3)$$

O deslocamento $\Delta_i(it+1)$, responsável pela alteração da posição da partícula i , varia com a configuração dos parâmetros ω^{t_i} , \mathbf{p}_i , \mathbf{g}_i , ϕ_1 , ϕ_2 , \mathbf{R}_1 e \mathbf{R}_2 . $\omega \in (0, 1)$ pondera a influência do deslocamento anterior ($\Delta_i(it)$) no deslocamento atual. Kameyama (2009) sugere que o valor de ω^{t_i} (inicialmente $\omega^{t_i} = 1$) sofra um decaimento exponencial com o passar das iterações ($\omega^{t_i} \rightarrow 0$), a fim de conferir ao PSO estabilidade e convergência. Como na primeira iteração do PSO $t_i = 0$ e nas iterações seguintes, caso a partícula não sofra mutação, t_i é incrementado ($t_i = t_i + 1$), então o decaimento exponencial de ω^{t_i} é realizado. \mathbf{p}_i é a posição com maior valor de estimativa de multimodalidade dentre as posições visitadas pela partícula i . Assim, $(\mathbf{p}_i - \mathbf{v}_i^\perp(it))$ indica que o deslocamento da partícula i é realizado entorno de \mathbf{p}_i . \mathbf{g}_i é a posição com maior valor de estimativa de multimodalidade dentre as posições visitadas pelas partículas que pertencem a vizinhança da partícula i . Desse modo, $(\mathbf{g}_i - \mathbf{v}_i^\perp(it))$ indica que o deslocamento da partícula i também é realizado entorno de \mathbf{g}_i . A tendência do deslocamento da partícula i na direção de \mathbf{p}_i ou \mathbf{g}_i é controlada, respectivamente, por ϕ_1 e ϕ_2 , ambos não-negativos. Kennedy e Eberhart (1995) sugerem igualdade na tendência dos deslocamentos, eles utilizaram em seus experimentos $\phi_1 = \phi_2 = 2$. \mathbf{R}_1 e \mathbf{R}_2 são matrizes diagonais aleatoriamente geradas, contendo valores $[0, 1]$. Essas matrizes dão aleatoriedade ao deslocamento da partícula, simulando o deslocamento do enxame.

Vizinhança entre partículas

O conceito de vizinhança confere ao PSO o comportamento colaborativo necessário

a generalização da solução a ser encontrada. Sem a influência da melhor solução da vizinhança (\mathbf{g}_i), a partícula i teria seu deslocamento atualizado por $\Delta_i(it+1) = \omega^{t_i} \Delta_i(it) + \phi_1 \mathbf{R}_1(\mathbf{p}_i - \mathbf{v}_i^\perp(it))$. Isso implica que, com o passar das iterações, i convergiria para \mathbf{p}_i . Ou seja, o PSO tenderia à máximos locais da função de custo.

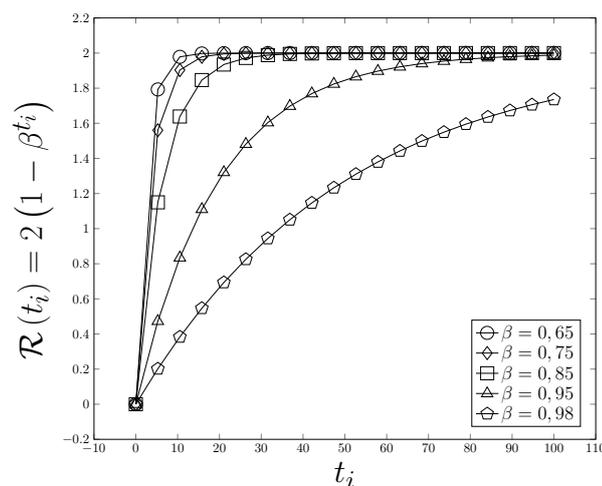
Kameyama (2009) apresenta abordagens estáticas e dinâmicas para a determinação da vizinhança entre partículas. As abordagens estáticas são de simples implementação e, nas configurações ideais ao problema empregado, favorecem a obtenção de elevados desempenhos. Contudo, a configuração ideal varia com cada problema empregado e sua determinação depende da experiência do projetista da solução. Entretanto, a adoção de abordagem dinâmica na determinação da vizinhança das partículas é uma alternativa eficiente, que permite ao PSO ser aplicado a uma variedade de problemas sem requerer configurações específicas quanto a vizinhança.

A abordagem dinâmica apresentada em Kameyama (2009) redefine \mathcal{R} , o raio de vizinhança, a cada nova iteração. Tendo a partícula i como centro, todas partículas dentro da esfera cujo raio é calculado pela Equação 3.4 compõem a vizinhança de i .

$$\mathcal{R}(t_i) = 2(1 - \beta^{t_i}) \quad (3.4)$$

Nessa equação, $\beta \in (0, 1)$ controla o crescimento de \mathcal{R} a cada iteração. A Figura 10 salienta a influência do valor de β na velocidade com que $\mathcal{R} \rightarrow 2$.

Figura 10 – Configurações para o raio da vizinhança do PSO. Essa figura apresenta a variação do raio da vizinhança do PSO ao longo de 100 iterações com diferentes valores de β . A medida que $\beta \rightarrow 1$, \mathcal{R} assintota 2 suavemente.



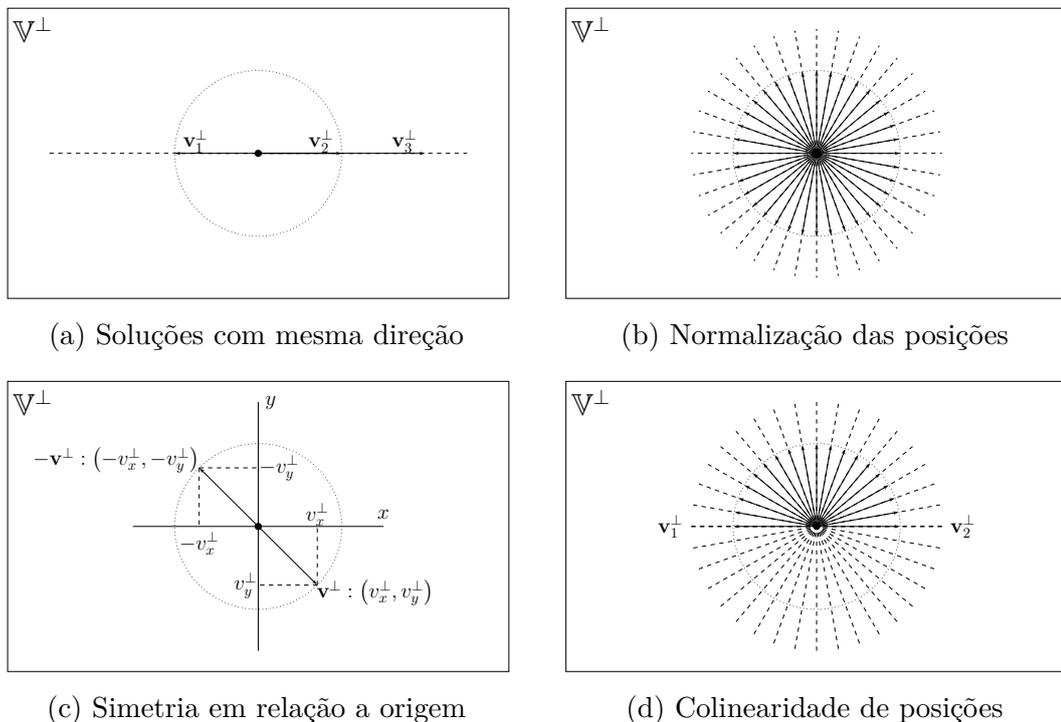
Fonte: Autoria própria.

Vê-se que a medida que $\beta \rightarrow 1$, \mathcal{R} assintota 2 mais suavemente. A modelagem da Equação 3.4 de modo a \mathcal{R} assintotar 2, se deve a esse valor ser o limiar máximo da distância entre duas posições alcançadas por essa modelagem do PSO.

Posição da partícula

Configurações adicionais ao PSO são necessárias devido às características particulares da aplicação na qual o PSO é submetido, no caso, determinação da direção com maior evidência de multimodalidade na dispersão dos padrões multidimensionais apresentados. Como mostra a Figura 11, o espaço de soluções \mathbb{V}^\perp admite vetores \mathbf{v}^\perp que representam as posições que indicam as possíveis soluções do PSO.

Figura 11 – Configurações específicas à aplicação do PSO. O espaço \mathbb{V}^\perp admite vetores \mathbf{v}^\perp que indicam direções como possíveis soluções do PSO. Em 11a, os vetores \mathbf{v}_1^\perp , \mathbf{v}_2^\perp e \mathbf{v}_3^\perp indicam possíveis posições que partículas do PSO podem assumir no espaço \mathbb{V}^\perp . Essas posições estão sobre a mesma direção, portanto, \mathbf{v}_1^\perp , \mathbf{v}_2^\perp e \mathbf{v}_3^\perp representam uma mesma solução do PSO. \mathbf{v}_3^\perp é um múltiplo de \mathbf{v}_2^\perp e não representa uma solução diferente do acpspo. Por isso, as posições das partículas são normalizadas a cada iteração, como mostra 11b. Assim, as posições \mathbf{v}^\perp possíveis são tomadas numa casca esférica de raio unitário. Porém, vê-se em 11c que ainda ocorre duplicidade nas soluções devido à simetria em relação a origem. Isso é minimizado reduzindo a busca à uma semiesfera de raio unitário, que pode ser obtida adotando $-\mathbf{v}^\perp$ sempre que $\mathbf{v}^\perp : (v_x^\perp, v_y^\perp)$ tiver $v_y^\perp < 0$. Contudo, esse procedimento não é suficiente para eliminar todos os problemas de simetria. Como vê-se em 11d, se $v_y^\perp = 0$ ainda pode ocorrer esse problema.



Fonte: Autoria própria.

A execução do PSO segundo as Equações 3.2 e 3.3, admite os vetores \mathbf{v}_1^\perp , \mathbf{v}_2^\perp e \mathbf{v}_3^\perp (Figura 11a) como possíveis posições que partículas do PSO podem assumir no espaço \mathbb{V}^\perp . Porém, como essas posições são colineares em relação a origem, indica que \mathbf{v}_1^\perp , \mathbf{v}_2^\perp e \mathbf{v}_3^\perp estão sobre a mesma direção e, devido à isso, não representam soluções distintas.

Nessa figura, \mathbf{v}_1^\perp e \mathbf{v}_2^\perp são vetores unitários (opostos), enquanto que \mathbf{v}_3^\perp é um múltiplo de \mathbf{v}_2^\perp . Sendo a reta tracejada a direção (solução do PSO) indicada pelos três vetores. Para contornar esse problema, as posições das partículas são normalizadas a cada iteração, como mostra 11b. O procedimento de normalização é posterior a alteração da posição da partícula na iteração corrente do PSO e ocorre segundo a Equação 3.5.

$$\mathbf{v}_i^\perp(it+1) = \frac{\mathbf{v}_i^\perp(it+1)}{\|\mathbf{v}_i^\perp(it+1)\|} \quad (3.5)$$

Com isso, as posições \mathbf{v}_i^\perp alcançáveis pelo PSO são tomadas numa casca esférica de raio unitário. E com isso se evita tomar posições que simplesmente são múltiplos de outras. Porém, permanece o problema de simetria em relação a origem, como é salientado na Figura 11c. Esse problema é minimizado reduzindo a busca pelas posições à uma semiesfera de raio unitário. A Equação 3.6 exemplifica como isso pode ser obtido. No caso, adota-se $-\mathbf{v}^\perp$ sempre que $\mathbf{v}^\perp : (v_x^\perp, v_y^\perp)$ tiver $v_y^\perp < 0$ (v_y^\perp representa a última posição nos vetores $\mathbf{v}^\perp \in \mathbb{V}^\perp$).

$$\mathbf{v}^\perp = \begin{cases} (v_x^\perp, v_y^\perp) & \text{se } v_y^\perp \geq 0 \\ (-v_x^\perp, -v_y^\perp) & \text{se } v_y^\perp < 0 \end{cases} \quad (3.6)$$

Contudo, esse procedimento não é suficiente para eliminar todos os problemas de simetria em relação a origem, dado que essa situação ainda pode ocorrer quando $v_y^\perp = 0$, como mostra a Figura 11d. Nessa figura, $\langle \mathbf{v}_1^\perp, \mathbf{v}_2^\perp \rangle = -1$, indicando a simetria em relação a origem desses vetores, ou seja, \mathbf{v}_1^\perp e \mathbf{v}_2^\perp indicam a mesma direção (solução do PSO). Situação equivalente ocorre quando $\mathbf{v}_1^\perp = \mathbf{v}_2^\perp$. Nesse caso, $\langle \mathbf{v}_1^\perp, \mathbf{v}_2^\perp \rangle = 1$ e os vetores também indicam a mesma direção. Assim, caso $\langle \mathbf{v}_1^\perp, \mathbf{v}_2^\perp \rangle = \pm 1$, um operador de mutação é aplicado sobre \mathbf{v}_2^\perp .

Operadores evolucionários

Kameyama (2009) comenta a importação de operadores de outras estratégias evolucionárias de otimização. A importação desses operadores tem como objetivo aumentar a diversidade das soluções visitadas e diminuir a chance do PSO ficar preso à um máximo local da função de custo \mathcal{V} . Dentre os operadores evolucionários comentados por Kameyama (2009) estão os operadores de (i) seleção e (ii) mutação.

(i) o operador de seleção ordena as partículas segundo seus valores na função de custo. Com base nessa ordenação, o conjunto das partículas é dividido ao meio (maiores e menores valores na função de custo). As partículas pertencentes ao subconjunto com os menores valores na função de custo têm suas posições alteradas para as posições das

partículas pertencentes ao subconjunto com os maiores valores na função de custo (essa operação não altera o \mathbf{p}_i e \mathbf{g}_i de cada partícula i alterada). Essa modelagem do operador de seleção favorece a convergência do PSO a um máximo local, porém não favorece a diversidade nas posições visitadas pelas partículas. Alternativamente à essa modelagem, pode-se aplicar o operador de mutação sobre as partículas pertencentes ao subconjunto com os menores valores na função de custo. Essa segunda modelagem do operador de seleção favorece a diversidade quanto as posições visitadas e, por isso, é empregada no PSO utilizado nesse trabalho.

(ii) o operador de mutação aleatoriamente muda a posição da partícula e faz seu parâmetro $t_i = 0$. A alteração na posição \mathbf{v}_i^\perp pelo operador de mutação é feita segundo a Equação 3.7. Nessa equação, \mathbf{R}_3 é uma matriz diagonal aleatoriamente gerada, cujos valores em sua diagonal principal variam entre $[-1, 1]$. Porém, se alguma célula de \mathbf{v}_i^\perp for nula, essa célula não mudará devido a multiplicação $\mathbf{R}_3 \mathbf{v}_i^\perp$. Por isso, o operador de mutação utiliza \mathbf{R}_4 , um vetor aleatoriamente gerado cujas células variam entre $[-1, 1]$.

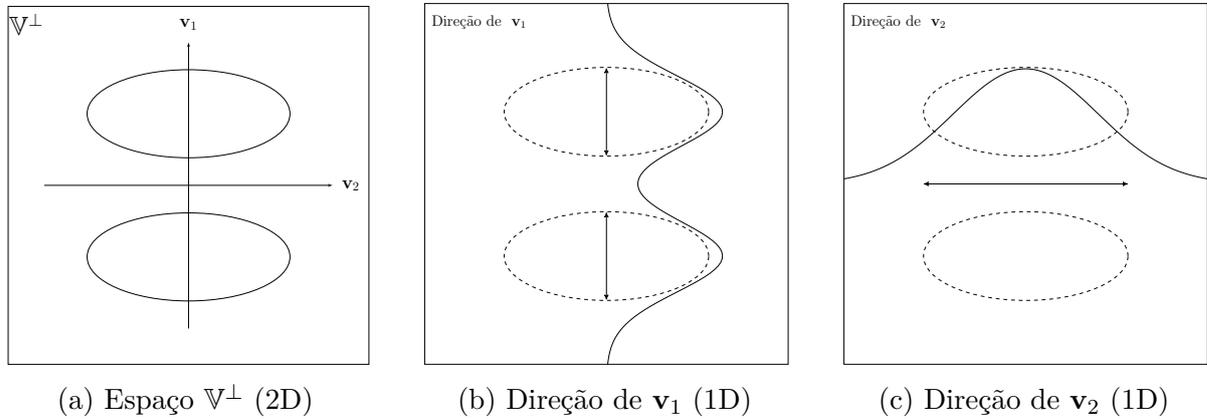
$$\mathbf{v}_i^\perp = \mathbf{R}_3 \mathbf{v}_i^\perp + \mathbf{R}_4 \quad (3.7)$$

Fazer $t_i = 0$ na partícula i mutada, possibilita à partícula realizar sua investigação do espaço de soluções, a princípio, localmente ($\omega^{t_i} = 1$ e $\mathcal{R} = 0$) e, com o passar das iterações, elevar a influência da melhor solução global (\mathbf{g}_i) sobre seu deslocamento ($\omega^{t_i} \rightarrow 0$ e $\mathcal{R} \rightarrow 2$).

Função de custo

A função de custo \mathcal{V} realiza a estimativa de multimodalidade no espaço unidimensional indicado pela posição corrente da partícula avaliada. Em outras palavras, a função de custo \mathcal{V} estima a multimodalidade da projeção dos padrões em \mathcal{X} no espaço unidimensional indicado por $\mathbf{v}^\perp \in \mathbb{R}^{(d-r) \times 1}$. A projeção de \mathcal{X} sobre diferentes direções gera representações unidimensionais dos padrões, o que possibilita diferentes percepções do comportamento multimodal da dispersão projetada, como ilustra a Figura 12. Na Figura 12a o espaço bidimensional \mathbb{V}^\perp apresenta dispersão de padrões segundo as modas representadas pelas elipses. \mathbf{v}_1 e \mathbf{v}_2 são vetores quaisquer em \mathbb{V}^\perp e a projeção dos padrões na direção de \mathbf{v}_1 e na direção de \mathbf{v}_2 causam diferentes percepções da multimodalidade. No espaço unidimensional da direção de \mathbf{v}_1 (Figura 12b), a Função Densidade de Probabilidade (FDP) que modela a mistura das distribuições Normais apresenta evidente multimodalidade. Enquanto que, no espaço unidimensional da direção de \mathbf{v}_2 (Figura 12c), a FDP que modela a mistura das distribuições Normais apresenta comportamento unimodal. Como o objetivo da função de custo é expressar a evidência do comportamento multimodal dos dados em espaços unidimensionais, sua aplicação nesse exemplo deve indicar a direção de \mathbf{v}_1 como mais favorável à multimodalidade do que a direção de \mathbf{v}_2 ($\mathcal{V}(\mathbf{v}_1) > \mathcal{V}(\mathbf{v}_2)$).

Figura 12 – Evidência de multimodalidade em espaços unidimensionais. 12a apresenta um espaço bidimensional com padrões dispersos segundo modas representadas pelas elipses. 12b e 12c apresentam a projeção dos padrões, respectivamente, nas direções dos vetores \mathbf{v}_1 e \mathbf{v}_2 .



Fonte: Autoria própria.

Assim, a cada iteração do Algoritmo 2, a direção \mathbf{v}_* computada no espaço ortogonal à \mathbf{V} é obtida satisfazendo a Equação 3.8.

$$\mathbf{v}_* = \arg \max_{\mathbf{v}^\perp \in \mathbb{R}^{(d-r) \times 1}} \mathcal{V}(\mathbf{v}^\perp, \mathbf{T}' \times \mathcal{X}) \quad (3.8)$$

As métricas investigadas como função de custo do PSO com o fim de estimar a multimodalidade em dispersões unidimensionais são apresentadas na Seção 3.2.3.

3.2.3 Métricas de multimodalidade

Essa seção apresenta as métricas utilizadas para estimar a multimodalidade em dispersões de padrões unidimensionais. As métricas empregadas são denominadas: *Eisenberger*, *Behboodian*, *Silverman*, *dip* e *IANI*. Algumas das cinco métricas empregadas trazem restrições ao escopo de MD. Essas restrições estão sumarizadas na Tabela 2.

Assim, as proposições das métricas *Eisenberger*, *Behboodian* e *IANI* consideram a dispersão dos padrões em \mathcal{X} seguindo distribuição \mathcal{F} bimodal (Equação 3.9) composta por modas Normais (Equação 3.10).

Tabela 2 – Restrições ao escopo de *Multimodal Decomposition* devido ao uso das métricas para estimativa de multimodalidade.

Métrica	Restrições
Eisenberger	Distribuições Normais
	Dispersão com máximo de duas modas
Behboodian	Distribuições Normais
	Dispersão com máximo de duas modas
Silverman	
dip	
IANI	Distribuições Normais
	Dispersão com máximo de duas modas

$$\mathcal{F}(x, \rho) = \rho \mathcal{N}(x, \mu_1, \sigma_1^2) + (1 - \rho) \mathcal{N}(x, \mu_2, \sigma_2^2) \quad (3.9)$$

$$\mathcal{N}(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (3.10)$$

As Seções 3.2.3.1-3.2.3.5 apresentam as métricas de estimativa de multimodalidade anteriormente citadas.

3.2.3.1 Eisenberger value

Eisenberger (1964) estabelece condições para rotular duas distribuições Normais como unimodais ou bimodais. Em seu trabalho encontra-se as condições seguintes:

- Uma condição suficiente para que $\mathcal{F}(x, \rho)$ seja unimodal para todo valor de $\rho \in (0, 1)$ é

$$(\mu_2 - \mu_1)^2 < \frac{27}{4} \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

- Uma condição suficiente para que exista valores de ρ para os quais $\mathcal{F}(x, \rho)$ seja bimodal é

$$(\mu_2 - \mu_1)^2 > 8 \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

Com base nessas condições de Eisenberger (1964), entende-se que o crescimento do valor $\mathcal{V}_{Eisenberger}$ (Equação 3.11) indica maior evidência de comportamento multimodal. Podendo $\mathcal{V}_{Eisenberger}$ ser utilizada como métrica de estimativa de multimodalidade.

$$\mathcal{V}_{Eisenberger} = \frac{(\mu_2 - \mu_1)^2 (\sigma_1^2 + \sigma_2^2)}{\sigma_1^2 \sigma_2^2} \quad (3.11)$$

Contudo, $\mathcal{V}_{Eisenberger}$ não considera a ponderação que ρ causa sobre as modas de \mathcal{F} e, conseqüentemente, seu efeito sobre a evidência da multimodalidade nessa distribuição.

3.2.3.2 Behboodian value

Behboodian (1970) também estabelece condição sobre distribuições bimodais Normais. Segundo Behboodian (1970), a unimodalidade de \mathcal{F} é suficientemente garantida se a seguinte condição for obedecida:

$$|\mu_2 - \mu_1| \leq 2 \min(\sigma_1, \sigma_2)$$

Assim, através da negação da condição de Behboodian (1970), pode-se encontrar $\mathcal{V}_{Behboodian}$ (Equação 3.12), cujos maiores valores indicam maior evidência de comportamento multimodal.

$$\mathcal{V}_{Behboodian} = \frac{|\mu_2 - \mu_1|}{\min(\sigma_1, \sigma_2)} \quad (3.12)$$

Essa métrica também não considera a ponderação que ρ causa sobre as modas de \mathcal{F} e, conseqüentemente, seu efeito sobre a evidência da multimodalidade nessa distribuição.

3.2.3.3 Silverman test

Silverman (1981) propôs um teste de multimodalidade, não paramétrico, que é baseado na estimativa da função de densidade de uma população unidimensional através da adoção de uma função de *kernel* \mathcal{K} , uma largura de banda b e um conjunto de amostras $\{x_1, x_2, \dots, x_n\}$, como mostra a Equação 3.13.

$$\hat{f}(t, b) = \frac{1}{nb} \sum_{i=1}^n \mathcal{K}\left(\frac{t - x_i}{b}\right) \quad (3.13)$$

Usualmente \mathcal{K} é modelada segundo uma função gaussiana, uma vez que esse tipo de função confere a estimativa \hat{f} a propriedade de que seu número de modas é uma função não crescente de b . Ou seja, maiores valores de b conduzem a estimativas de densidade mais suave e com menos modas.

Assim, o teste de Silverman busca pelo menor valor de largura de banda b_{min} capaz de produzir em \hat{f} comportamento unimodal (um único máximo). Com isso, se um alto valor de b é necessário para suavizar \hat{f} e tornar seu comportamento unimodal, então, a função de densidade populacional (f) tem duas ou mais modas.

Com base nessa ideia, Silverman (1981) sugeriu um teste estatístico com hipótese nula que f é unimodal. Através de *bootstrap*, o *valor-p* (p_{val}) é calculado de modo que seus menores valores indicam maior evidência de multimodalidade. Desse modo, a estimativa de multimodalidade baseada no teste de Silverman ($\mathcal{V}_{Silverman}$) é expressa pela Equação 3.14.

$$\mathcal{V}_{Silverman} = 1 - p_{val} \quad (3.14)$$

Implementation of Silverman test can be found at www.uni-marburg.de/fb12/stoch/research/rpackage.

3.2.3.4 dip test

Hall e York (2001) mostram que, no teste de Silverman, o intervalo no qual a função de densidade populacional é estimada deve ser restringida de modo a reduzir a influência de valores extremos (dados isolados) que produzem modas espúrias. Para contornar esse problema, Hartigan e Hartigan (1985) desenvolveram um método não paramétrico, baseado em histograma e que não requer estimativa da função de densidade populacional. Esse método é denominado teste *dip*, o qual verifica, para um conjunto de amostras, se sua dispersão é unimodal.

A ideia por trás do *dip* é que se uma distribuição é unimodal com máximo em m , então essa distribuição é convexa em $(-\infty, m]$ e é côncava em $[m, +\infty)$. Com base nessa ideia, o teste *dip* estima a unimodalidade calculando a máxima diferença entre a função de distribuição empírica (sobre todos os pontos da amostra) e a função de distribuição unimodal que minimiza essa máxima diferença.

Utilizando *bootstrap*, um *valor-p* (p_{val}) do *dip* teste é calculado de modo que seus menores valores indicam maior evidência de multimodalidade. Desse modo, a estimativa de multimodalidade baseada no teste *dip* (\mathcal{V}_{dip}) é expressa pela Equação 3.15.

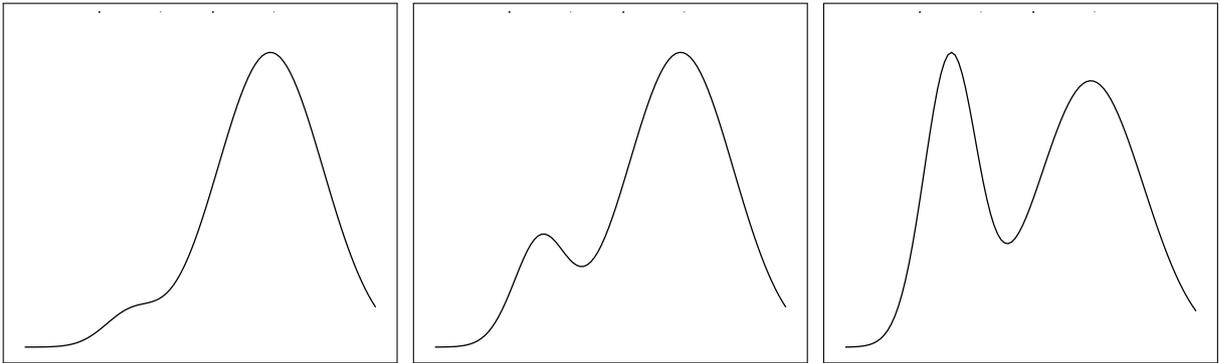
$$\mathcal{V}_{dip} = 1 - p_{val} \quad (3.15)$$

Implementation of *dip* test can be found at www.nicprice.net/diptest.

3.2.3.5 Index of Area-based Normal Intersection

Eisenberger (1964) e Behboodian (1970) propõem condições claras para a ocorrência de unimodalidade. Contudo, a condição para a ocorrência de bimodalidade em Eisenberger (1964) é dependente da probabilidade ρ que pondera as modas Normais na mistura \mathcal{F} (Equação 3.9). Porém nem Eisenberger (1964) ou Behboodian (1970) tratam da influência de ρ na percepção do comportamento multimodal. Essa influência é salientada na Figura 13. Nessa figura, três pares de modas Normais idênticas são ponderadas por diferentes valores de ρ , resultando em diferentes percepções da multimodalidade.

Para contornar esse problema, é proposto um índice que mede a interferência entre duas modas Normais com base na área de interseção entre elas. Essa proposta é denominada *Index of Area-based Normal Intersection* (IANI) e é expressa pela Equação 3.16, que assume

Figura 13 – Variação da probabilidade *a priori*

Fonte: Autoria própria.

valores limitados entre $[0, 1)$.

$$\mathcal{V}_{IANI} = \rho_q \times A_q \quad (3.16)$$

Nessa equação, o fator ρ_q expressa o equilíbrio na ponderação entre as Normais (Equação 3.17). Se $\rho_q \rightarrow 0$ há severo desequilíbrio na ponderação entre as Normais, o que favorece o comportamento unimodal. Caso $\rho_q \rightarrow 1$, a ponderação entre as Normais se aproxima do equilíbrio ($\rho \approx 0,5$), o que favorece a percepção do comportamento multimodal. O fator A_q (Equação 3.18) exprime a interferência de uma Normal sobre a outra através da área de interseção entre suas Normais (A_\cap). Assim, se $A_q \rightarrow 0$, a sobreposição entre as Normais é expressiva indicando unimodalidade. caso $A_q \rightarrow 1$, a interferência entre as Normais é mínima, indicando evidência de comportamento multimodal.

$$\rho_q = \frac{\min[\rho, (1 - \rho)]}{\max[\rho, (1 - \rho)]} \quad (3.17)$$

$$A_q = 1 - \frac{A_\cap}{A_\cap^*} \quad (3.18)$$

Na Equação 3.18, A_\cap é a área de interseção entre as duas Normais. Enquanto que A_\cap^* é área máxima de interseção obtida para essas Normais com a variação dos valores de suas médias. Assim, A_\cap^* padroniza os valores de $A_q \in [0, 1)$, dado que $A_\cap \leq A_\cap^*$.

Nos Parágrafos 3.2.3.5.1 e 3.2.3.5.2 são explicados os cálculos de A_\cap e A_\cap^* respectivamente.

3.2.3.5.1 Área de interseção entre as duas Normais

Sendo a função \mathcal{F} composta pela mistura de duas funções de distribuição Normal como mostra a Equação 3.9, então, a área de interseção entre as funções $\mathcal{N}_1(x) = \rho \mathcal{N}(x, \mu_1, \sigma_1^2)$ e $\mathcal{N}_2(x) = (1 - \rho) \mathcal{N}(x, \mu_2, \sigma_2^2)$ é dada pelo valor $A_\cap(\mu_1, \sigma_1, \rho, \mu_2, \sigma_2)$.

Assim, o cálculo desse valor é obtido descobrindo-se os pontos de interseção entre \mathcal{N}_1 e \mathcal{N}_2 e integrando-se essas funções em seus trechos que se sobrepõe.

Os pontos de interseção são descobertos através da igualdade $\mathcal{N}_1 = \mathcal{N}_2$, qual pode ser reduzida na equação $ax^2 + bx + c = 0$, com:

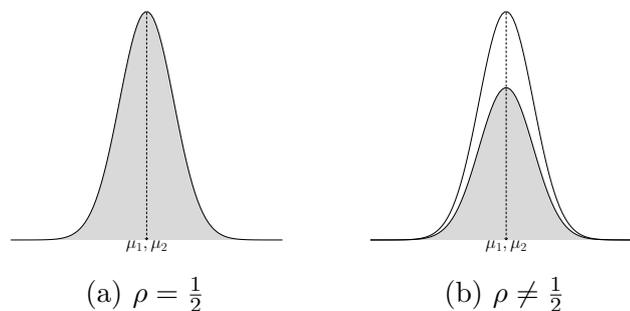
$$\begin{aligned} a &= \sigma_1^2 - \sigma_2^2 \\ b &= 2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) \\ c &= -\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 - 2\sigma_1^2\sigma_2^2 \ln\left(\frac{\sigma_1}{\sigma_2} \frac{1-\rho}{\rho}\right) \end{aligned}$$

Se $\mu_1 = \mu_2 = \mu$ e $\sigma_1 = \sigma_2 = \sigma$, obrigatoriamente $a = 0$ e $b = 0$. Com isso, para que $c = 0$ seja verdade, obrigatoriamente:

$$c = \ln\left(\frac{1-\rho}{\rho}\right) = 0 \Rightarrow \frac{1-\rho}{\rho} = 1 \Rightarrow \rho = \frac{1}{2}$$

Ou seja, se $\mu_1 = \mu_2$ e $\sigma_1 = \sigma_2$, só há interseção entre as Normais se $\rho = \frac{1}{2}$. Nesse caso (Figura 14a), \mathcal{N}_1 e \mathcal{N}_2 são idênticas, fazendo $A_\cap(\mu_1, \sigma_1, \rho, \mu_2, \sigma_2) = \frac{1}{2}$. Porém, se $\rho \neq \frac{1}{2}$ (Figura 14b), não há ponto de interseção entre \mathcal{N}_1 e \mathcal{N}_2 , dado que uma das Normais está contida sob a outra. Nesse cenário, $A_\cap(\mu_1, \sigma_1, \rho, \mu_2, \sigma_2) = \min[\rho, (1-\rho)]$.

Figura 14 – Comportamentos da área de interseção quando $\mu_1 = \mu_2$ e $\sigma_1 = \sigma_2$.



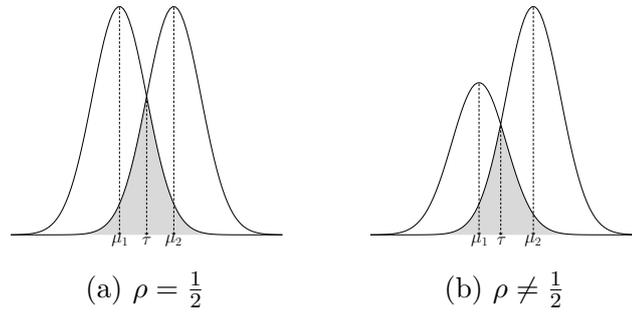
Fonte: Autoria própria.

Se $\mu_1 \neq \mu_2$ e $\sigma_1 = \sigma_2 = \sigma$, obrigatoriamente $a = 0$ e $b \neq 0$. Assim, existe um ponto de interseção τ entre \mathcal{N}_1 e \mathcal{N}_2 (Figura 15), o qual é dado por $\tau = -\frac{c}{b}$. Assim, considerando $\mu_1 < \mu_2$, tem-se que $A_\cap(\mu_1, \sigma_1, \rho, \mu_2, \sigma_2) = \int_{-\infty}^{\tau} \mathcal{N}_2(x) dx + \int_{\tau}^{+\infty} \mathcal{N}_1(x) dx$.

Se $\sigma_1 \neq \sigma_2$, obrigatoriamente $a \neq 0$. Isso implica que existem no máximo dois pontos de interseção reais entre \mathcal{N}_1 e \mathcal{N}_2 (τ_1 e τ_2). O comportamento da área de interseção quando $\sigma_1 \neq \sigma_2$ é apresentado na Figura 16. Essa figura salienta que a existência dos pontos de interseção é determinada pelo sinal de $\Delta = b^2 - 4ac$.

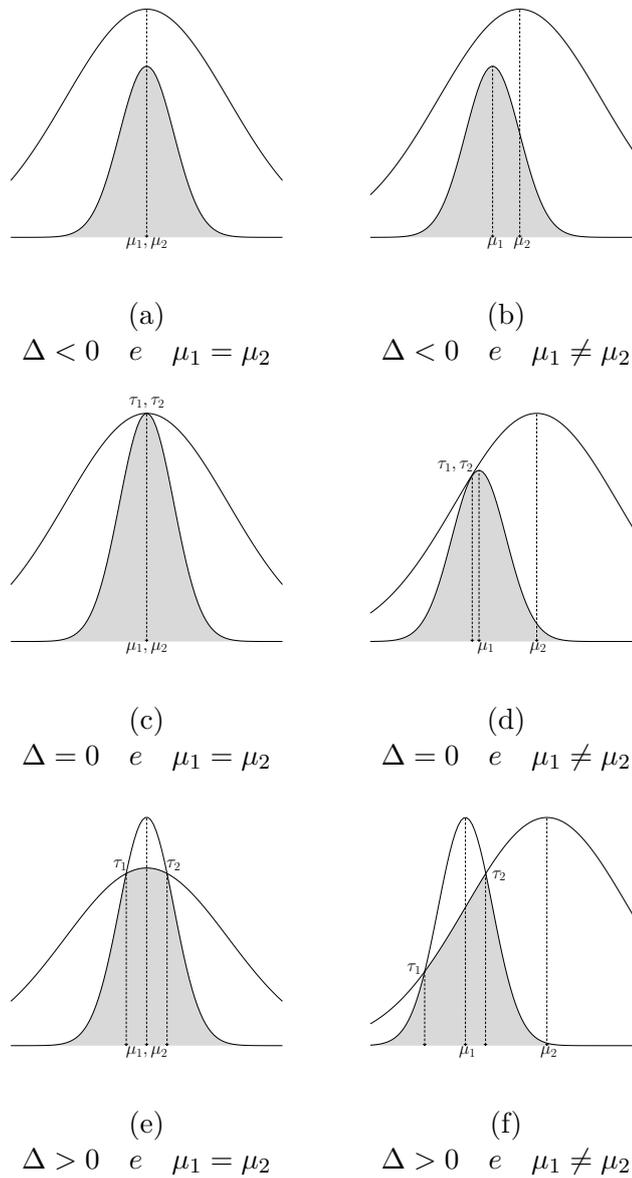
Assim, se $\Delta < 0$ obrigatoriamente $\nexists \tau_1, \tau_2 \in \mathbb{R}$, como mostram as Figuras 16a e 16b. Nesse caso, uma das Normais está contida sob a outra, de modo que $A_\cap(\mu_1, \sigma_1, \rho, \mu_2, \sigma_2) =$

Figura 15 – Comportamentos da área de interseção quando $\mu_1 \neq \mu_2$ e $\sigma_1 = \sigma_2$.



Fonte: Autoria própria.

Figura 16 – Comportamentos da área de interseção quando $\sigma_1 \neq \sigma_2$.



Fonte: Autoria própria.

$\min[\rho, (1 - \rho)]$. Sendo $\Delta = 0$ obrigatoriamente $\exists \tau_1, \tau_2 \in \mathbb{R}$ tal que $\tau_1 = \tau_2$, como mostram as Figuras 16c e 16d. Isso também implica na inscrição de uma das Normais sob a outra, de modo que:

$$A_{\cap}(\mu_1, \sigma_1, \rho, \mu_2, \sigma_2) = \min[\rho, (1 - \rho)]$$

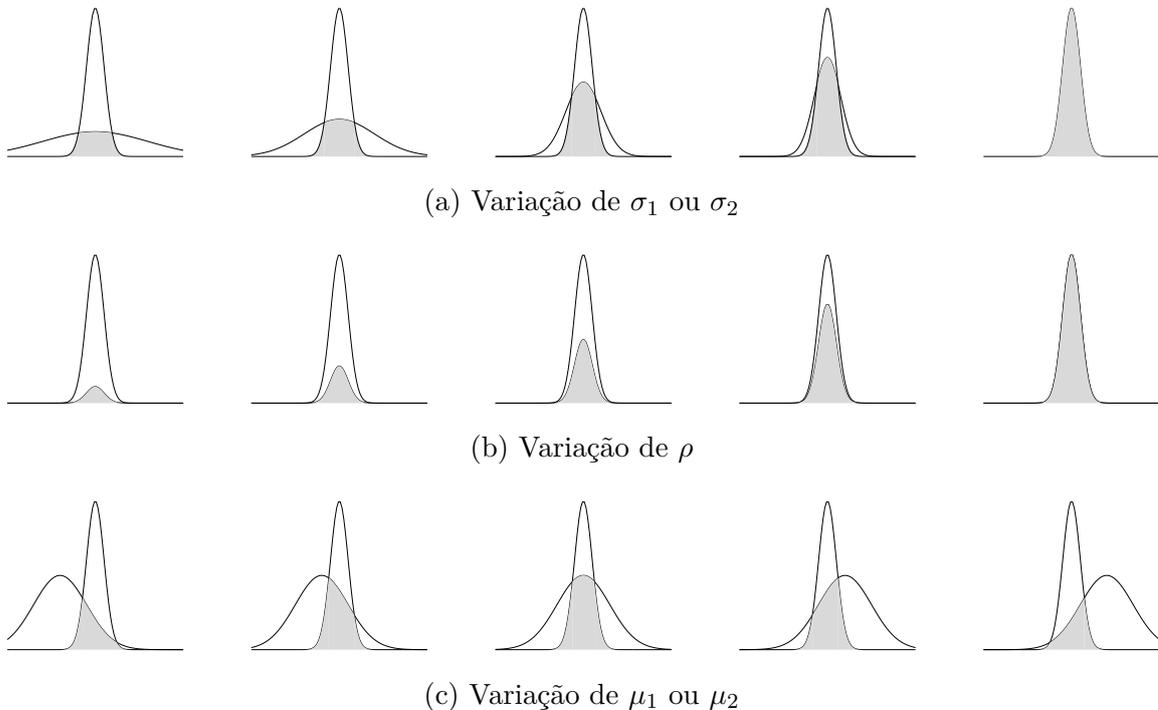
Porém, quando $\Delta > 0$ obrigatoriamente $\exists \tau_1, \tau_2 \in \mathbb{R}$ tal que $\tau_1 \neq \tau_2$, como mostram as Figuras 16e e 16f. Assim, considerando $\tau_1 < \tau_2$ e $\sigma_1 > \sigma_2$, tem-se que:

$$A_{\cap}(\mu_1, \sigma_1, \rho, \mu_2, \sigma_2) = \int_{-\infty}^{\tau_1} \mathcal{N}_2(x) dx + \int_{\tau_1}^{\tau_2} \mathcal{N}_1(x) dx + \int_{\tau_2}^{+\infty} \mathcal{N}_2(x) dx$$

3.2.3.5.2 Área máxima de interseção entre as duas Normais

O cálculo de A_{\cap}^* é realizado a fim de padronizar dos valores de A_q entre $[0, 1]$. Para isso, A_{\cap}^* deve ser a maior área de interseção obtida com a variação dos parâmetros das distribuições Normais envolvidas na mistura. Contudo, essa variação não deve alterar o aspecto (forma) de cada distribuição individualmente. A Figura 17 mostra que a variação de σ_1, σ_2 ou ρ (Figuras 17a e 17b) altera o aspecto da distribuição Normal cujo parâmetro foi variado.

Figura 17 – A variação dos parâmetros σ_1, σ_2 (17a) ou ρ (17b) alteram o aspecto das distribuições Normais. Enquanto que a variação de μ_1 ou μ_2 (17c) simplesmente desloca as distribuições sem alterar seus aspectos (forma), individualmente.



Fonte: Autoria própria.

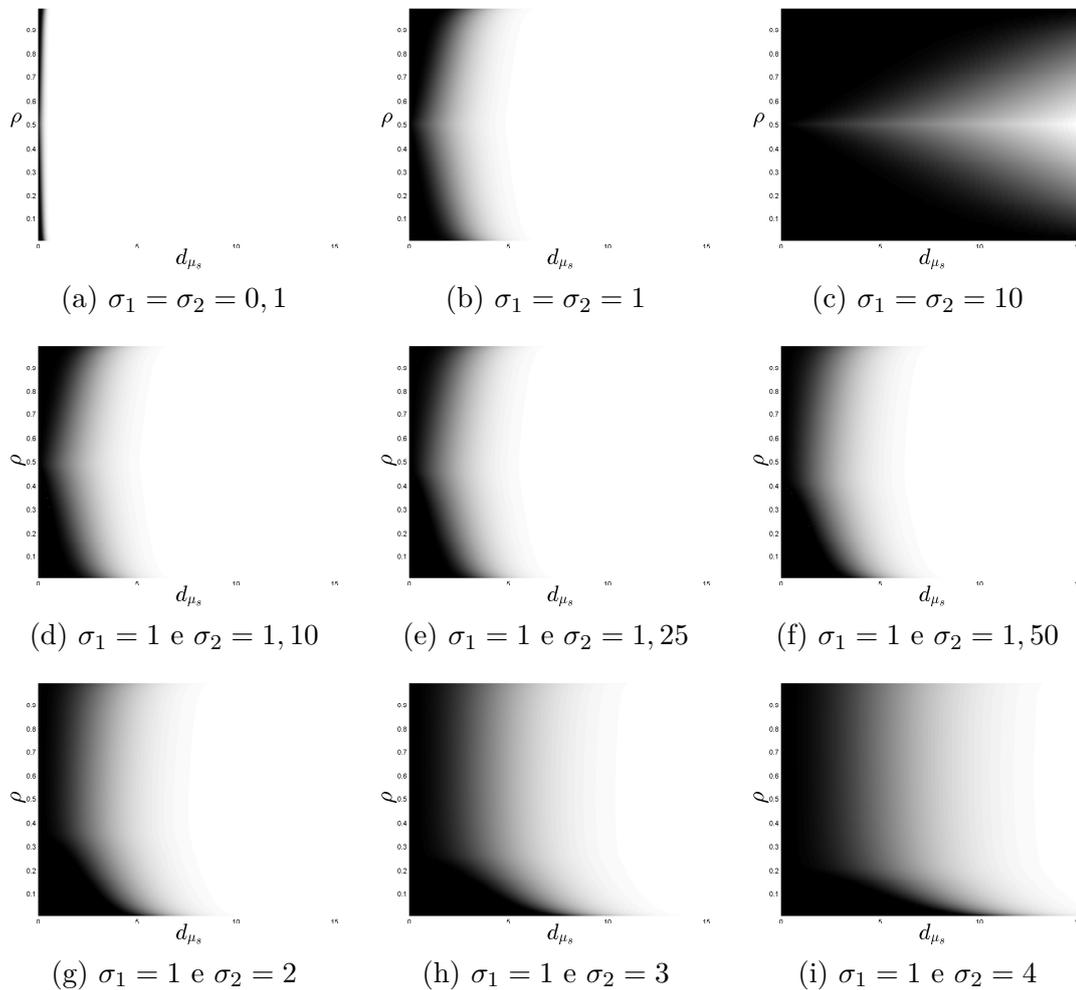
Entretanto, a variação dos parâmetros μ_1 e μ_2 (Figura 17c) simplesmente desloca as distribuições Normais sem alterar seus aspectos (formas). Desse modo, a busca por A_{\cap}^*

traduz-se na busca pela configuração dos parâmetros μ_1 e μ_2 que conduzem ao maior valor de A_\cap dado que são constantes os valores de σ_1 , σ_2 e ρ .

$$A_\cap^* = \max_{\mu_1, \mu_2 \in \mathbb{R}} A_\cap(\mu_1, \mu_2) \quad \text{com } \sigma_1, \sigma_2 \text{ e } \rho \text{ constantes}$$

A configuração dos parâmetros μ_1 e μ_2 de modo a maximizar A_\cap pode ser determinada empiricamente, como mostra a Figura 18.

Figura 18 – Influência dos parâmetros da mistura sobre a área de interseção entre as Normais. Em cada linhas horizontal desses gráficos, o valor de A_\cap é calculado variando-se apenas os parâmetros μ_1 e μ_2 (expressos em d_{μ_s}) da misturas entre distribuições Normais. Cada valor de A_\cap foi dividido por A_\cap^* (maior valor de A_\cap obtido por linha horizontal).



Fonte: Autoria própria.

As Figuras 18a-18i relacionam o valor de $\rho \in (0, 1)$ com $d_{\mu_s} \in [0, 15]$ (a distância entre as médias μ_1 e μ_2), para valores fixos de σ_1 e σ_2 (par de valores apresentados na legenda de cada uma das figuras). Para cada ponto no gráfico dessas figuras foi calculado o valor de $A_\cap(\mu_1, \sigma_1, \rho, \mu_2, \sigma_2)$. A fim de salientar a ocorrência do maior valor de A_\cap , dados valores fixos de σ_1 , σ_2 e ρ , realizou-se padronização dos valores de A_\cap obtidos para o mesmo valor

de ρ . Ou seja, foi encontrado o maior valor de A_\cap (A_\cap^*) para cada linha horizontal do gráfico (σ_1 , σ_2 e ρ constantes). Em seguida, para essa mesma linha horizontal, todos os valores obtidos de A_\cap foram divididos por A_\cap^* . Assim, os valores em cada linha variam entre $(0, 1]$, de modo que $0 \Rightarrow$ *branco* e $1 \Rightarrow$ *preto*. Com isso, observa-se nas Figuras 18a-18i que A_\cap^* (*pixels* pretos) ocorre sempre que $d_{\mu_s} = 0$ ($\mu_1 = \mu_2$). A depender da configuração de σ_1 , σ_2 e ρ , A_\cap^* pode ser alcançado também para valores de $d_{\mu_s} \neq 0$. Esse comportamento é mais evidente nas Figuras 18c, 18g-18i, ocorrendo para desvios padrão distintos sendo a distribuição Normal com menor desvio padrão ponderada por probabilidade *a priori* significativamente baixa. Nessa configuração, alguns deslocamentos das Normais são tais que uma delas está inteiramente contida sobre a outra. Para todos esses deslocamentos, $A_\cap = A_\cap^*$.

Assim, a evidência empírica indica que, para valores constantes de σ_1 , σ_2 e ρ ; A_\cap^* ocorre quando $\mu_1 = \mu_2$. Assim, sendo \mathcal{F} definida pelos parâmetros $\mu_1, \mu_2 \in \mathbb{R}$; $\sigma_1, \sigma_2 \in \mathbb{R}_+^*$ e $\rho \in (0, 1)$ tem-se que:

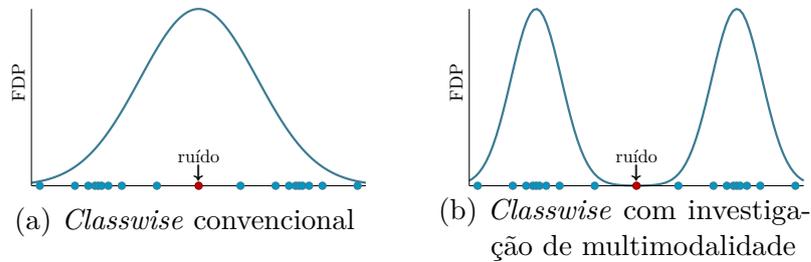
$$A_\cap^* = A_\cap(0, \sigma_1, \rho, 0, \sigma_2) \geq A_\cap(\mu_1, \sigma_1, \rho, \mu_2, \sigma_2)$$

3.3 Utilizar informação de multimodalidade na classificação

O emprego da abordagem *classwise* utilizando TEC não-supervisionada que investiga e preserva a multimodalidade nos dados não é suficiente para cumprir a característica (iii) definida para as técnicas *classwise* com investigação de multimodalidade (página 39). Isso acontece porque a abordagem *classwise*, em sua modelagem convencional (Seção 2.2.2), apenas informa ao classificador as matrizes de transformações que foram obtidas. A determinação das regiões de decisão é feita pelo classificador sobre todo o conjunto de treinamento, sem que ele obtenha qualquer informação da estrutura da dispersão de cada classe. Desse modo, mesmo que a classe possua dispersão multimodal, o classificador de TEC *classwise* convencional calcula a FDP da classe considerando sua dispersão como unimodal, como mostra a Figura 19a. Essa figura apresenta padrões unidimensionais com dispersão multimodal. Como o classificador *classwise* entende qualquer dispersão de padrões como unimodal, conseqüentemente, um ruído entre as modas (região de baixa densidade de padrões) recebe alta valoração na FDP calculada para a classe, o que eleva a chance de classificar o ruído como padrão da classe (eleva o erro de classificação).

Alternativamente, o emprego da abordagem *classwise* utilizando TEC não supervisionada que investiga e preserva a multimodalidade nos dados realizará a característica (iii) se, sua modelagem informar ao classificador as matrizes de transformações que foram obtidas para cada classe e, em caso de multimodalidade intraclasses, informar a estrutura multimodal identificada. Com a informação da estrutura multimodal da classe, a nova abordagem de classificação será capaz de calcular FDP que condiz com a dispersão

Figura 19 – Diferenças entre as abordagens de treinamento do classificador para as TECs *classwise* e *classwise* com investigação de multimodalidade. As técnicas *classwise* com investigação de multimodalidade (19b) percebem a multimodalidade intraclasse possibilitando treinar o classificador de modo a não confundir ruído que se encontra entre as modas. O mesmo não se observa no treinamento *classwise* convencional (19a).

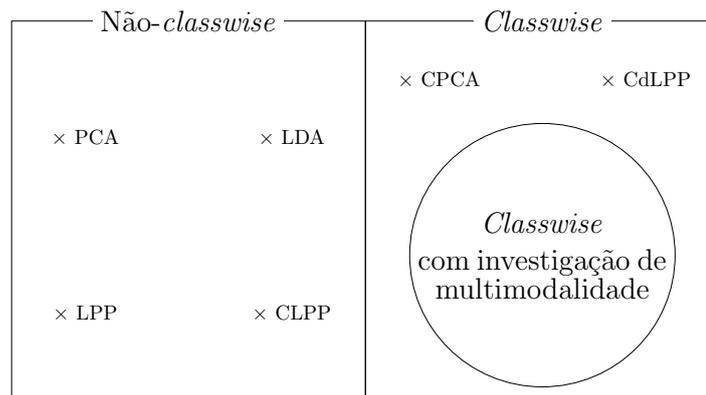


Fonte: Autoria própria.

multimodal identificada, como mostra a Figura 19b. Essa figura apresenta os mesmos padrões unidimensionais com dispersão multimodal vistos na Figura 19a. Porém, diferente do classificador *classwise* convencional, a multimodalidade da dispersão dos padrões da classe é expressa em sua FDP. Isso faz com que um ruído entre as modas (região de baixa densidade de padrões) receba baixa valoração na FDP da classe, o que reduz a chance de classificar o ruído como padrão da classe (reduz o erro de classificação).

A Figura 20 ilustra o agrupamento das TECs segundo seus tipos. Essa figura salienta a especialização que as TECs *classwise* com investigação de multimodalidade representam em relação as TECs *classwise* convencionais.

Figura 20 – Agrupamentos das técnicas de extração de características.



Fonte: Autoria própria.

O classificador associado à TEC *classwise* com investigação de multimodalidade requer a prévia identificação das modas na dispersão dos padrões. Essa investigação visa encontrar o modelo de mistura que otimiza a FDP para os padrões apresentados. O modelo ótimo deve ser selecionado de uma variedade de modelos que variam o número de modas que compõem a dispersão dos padrões e os valores dos parâmetros dessas modas. Contudo, as restrições apresentadas na Tabela 2 para as métricas *Eisenberger*, *Behboodian* e *IANI*

limitam os modelos avaliados à duas modas. Sendo essa limitação uma restrição adicional às TECs *classwise* com investigação de multimodalidade que são propostas nesse trabalho (Seção 3.4).

Como a quantidade de modas nos modelos investigados está fixa (duas modas), pode-se utilizar técnica de agrupamento de padrões para a determinação dessas modas.

K-médias (HARTIGAN, 1975; HARTIGAN; WONG, 1979) é uma técnica de agrupamento largamente utilizada na literatura. Seu objetivo é dividir os N padrões em K grupos. Esses agrupamentos são obtidos de modo que seja mínima a soma das distâncias dos padrões do grupo corrente em relação a média desses padrões.

Um dos fatores que justificam a larga utilização do K-médias é sua implementação não paramétrica. Ou seja, o K-médias é desempenhado sem a necessidade de conhecer o tipo de distribuição que cada moda segue. Por esse mesmo motivo, os agrupamentos obtidos pelo K-médias podem seguir distribuições variadas, sem que haja um prévio conhecimento do tipo de distribuição a ser obtida.

Por outro lado, os classificadores utilizados pelas TECs *classwise* derivam de classificador Bayesiano, o qual representa o aprendizado sobre uma classe do problema através da FDP que rege a dispersão dos padrões dessa classe. Para a simplificação de toda solução e baseado no teorema do limite central, assumiu-se que as FDPs de todas as classes seguem distribuição Normal, ou mistura de distribuições Normais, no caso do classificador para TECs *classwise* com investigação de multimodalidade.

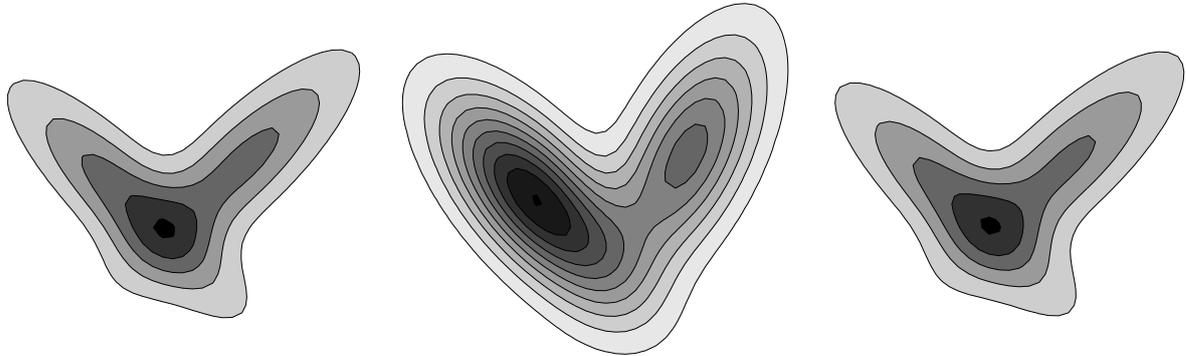
Assim, a falta de controle do K-médias quanto à distribuição que os agrupamentos obtidos devem seguir, pode conduzir a ineficiente modelagem dos classificadores, como ilustra a Figura 21.

Para construir a Figura 21, padrões foram aleatoriamente gerados em um espaço bidimensional, de modo a formar duas modas distintas, ambas seguindo distribuição Normal. Na Figura 21a é apresentada a FDP obtida da mistura das distribuições Normais utilizadas para a geração aleatória dos padrões.

Ao K-médias são apresentados os padrões que foram gerados, sem qualquer outra informação adicional, exceto $K = 2$, que é o número de grupos (modas) que essa técnica deve encontrar. Após a aplicação do K-médias dois grupos de padrões são obtidos e, para cada um deles, é calculada a média e matriz de dispersão de seus padrões. A partir dos parâmetros obtidos para cada grupo, distribuições Normais são configuradas e da mistura delas obtém-se a FDP da Figura 21b. A diferença entre as FDPs das Figuras 21a e 21b é expressa nos valores do erro de classificação sobre os padrões gerados. O erro de classificação utilizando a FDP da Figura 21a é de 10,75%, enquanto que o erro de classificação utilizando a FDP da Figura 21b é de 85,50%.

Uma alternativa ao K-médias é a *Gaussian Mixture Model* (GMM) (BILMES,

Figura 21 – Influência da técnica de agrupamento sobre a FDP. Essa figura apresenta as FDPs obtidas sobre padrões sintéticos 2D, gerados segundo modas com distribuição Normal. [21a](#) apresenta a FDP obtida com os parâmetros utilizados para a geração dos padrões. A FDP [21b](#) é obtida a partir de parâmetros calculados pelo K-médias com $K = 2$. Em [21c](#) a FDP é obtida por GMM configurada para procurar por duas modas Normais.



(a) FDP parâmetros originais (b) FDP parâmetros K-médias (c) FDP parâmetros GMM

Fonte: Autoria própria.

1998). Essa técnica é utilizada para encontrar mistura de M distribuições Normais que aproxime uma dada dispersão. GMM utiliza o algoritmo *Expectation Maximization* (EM) (BIERNACKI; CELEUX; GOVAERT, 2003) a fim de ajustar os parâmetros das M distribuições Normais por ele encontradas. EM tem como objetivo maximizar a verossimilhança desses parâmetros, o que torna a mistura das modas encontradas pela GMM o mais verossímil possível em relação à dispersão dos padrões. Assim como o K-médias, GMM possui como parâmetro externo o número de modas que deve determinar. Contudo, sua modelagem específica para distribuições Normais está de acordo com a redução de escopo assumida para o projeto dos classificadores desse trabalho (FDPs Normais ou mistura de distribuições Normais).

A adequação da GMM à modelagem dos classificadores desse trabalho é exemplificada na Figura [21c](#). Essa figura representa a FDP obtida da mistura das $M = 2$ distribuições Normais encontradas pela GMM sobre os padrões que foram aleatoriamente gerados. Sendo o número de modas a ser encontradas a única informação adicional dada a GMM. As FDPs das Figuras [21a](#) e [21c](#) são semelhantes. Essa semelhança é expressa no erro de classificação utilizando a FDP da Figura [21c](#), o qual foi 11,00%. Esse valor é próximo do obtido pela FDP da Figura [21a](#) e é significativamente menor do que o valor de erro obtido com o K-médias. Portanto, nesse trabalho, as TECs *classwise* com investigação de multimodalidade obtêm a informação de multimodalidade através da aplicação do GMM para otimização do modelo de mistura com duas modas Normais.

3.4 TECs *classwise* com investigação de multimodalidade

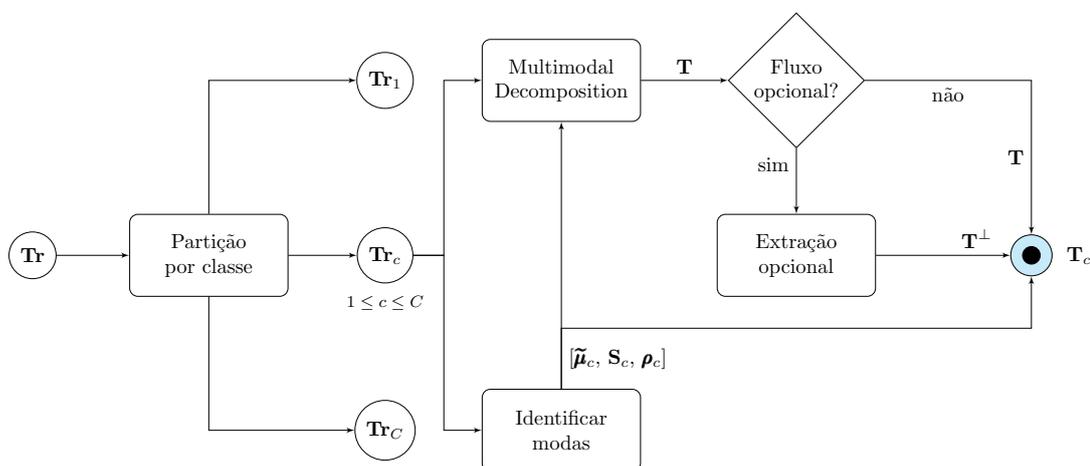
Nessa seção são propostas três TECs para a categoria das técnicas *classwise* com investigação de multimodalidade. Como seguem abordagem *classwise*, as TECs aqui apresentadas possuem procedimentos próprios de extração de características e classificação. Assim, as Seções 3.4.1-3.4.3 apresentam o procedimento de extração de características particular a cada uma das TECs propostas; enquanto que a Seção 3.4.4 apresenta o procedimento de classificação que é comum a elas.

3.4.1 Class-dependent Multimodal Decomposition

Nessa seção é apresentada a primeira TEC *classwise* com investigação de multimodalidade proposta nesse trabalho, a qual é denominada por *Class-dependent Multimodal Decomposition* (CdMD). Essa técnica tem como objetivo encontrar reduções de dimensionalidade que, para cada classe do problema, preservam ao máximo a estrutura multimodal original da dispersão. A realização desse objetivo, associado com a utilização de classificador que produz FDP melhor ajustada sobre a dispersão dos padrões, visa diminuir o erro na classificação de ruídos que se encontrem na região entre as modas, como foi representado na Figura 19.

O diagrama da Figura 22 representa o procedimento de extração de características do CdMD. Inicialmente, o conjunto de treinamento \mathbf{Tr} é particionado segundo suas classes. Em seguida, essa técnica estima os parâmetros do modelo de mistura Normal que melhor se ajustam a dispersão dos padrões de cada classe. Assim, para a classe corrente c ($1 \leq c \leq C$), $\tilde{\boldsymbol{\mu}}_c$, \mathbf{S}_c e $\boldsymbol{\rho}_c$ representam respectivamente os vetores de média, as matrizes de covariância e as probabilidades *a priori* estimadas no bloco *Identificar modas*. Esse bloco utiliza algoritmo GMM para realizar a estimativa dos parâmetros Normais que modelam a dispersão da classe corrente, como discutido na Seção 3.3.

Figura 22 – Procedimento de extração de características do CdMD.



Fonte: Autoria própria.

Os parâmetros estimados para o modelo de mistura da classe corrente são utilizados por MD, que determina as direções ortogonais com maiores evidências de multimodalidade. As direções encontradas compõem as colunas de \mathbf{T} .

Assim como ocorre para as demais técnicas *classwise* (CPCA e CdLPP), é opcional o uso de um procedimento de extração de características adicional na composição da matriz de transformação para a classe corrente. Exemplos do uso dessa opção são encontrados na Seção 2.2.2 (página 32) e resultam na matriz \mathbf{T}^\perp que é resultado de procedimento de ortonormalização sobre a concatenação entre colunas da matriz de transformação retornada pela extração de características adicional e colunas da matriz \mathbf{T} . Caso o procedimento de extração opcional seja utilizado, \mathbf{T}^\perp é apresentada como resposta do CdMD aos padrões da classe c ($\mathbf{T}_c = \mathbf{T}^\perp$). Caso contrário, CdMD retorna \mathbf{T} para a classe c ($\mathbf{T}_c = \mathbf{T}$).

Independente da configuração utilizada para o CdMD, essa técnica sempre retorna, para cada classe, os parâmetros do modelo de mistura que foram estimados, os quais serão utilizados no procedimento de classificação da técnica.

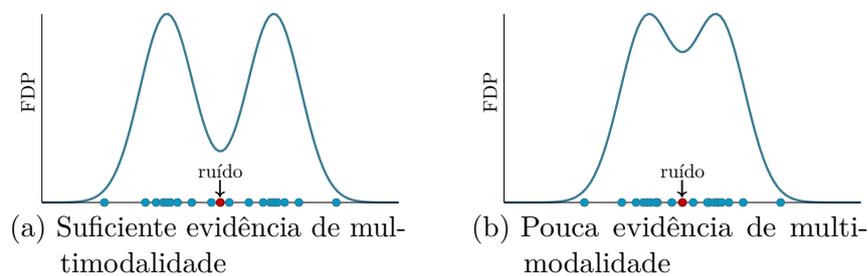
3.4.2 Class-dependent Conditional Multimodal Decomposition

A segunda TEC *classwise* com investigação de multimodalidade proposta nesse trabalho é denominada *Class-dependent Conditional Multimodal Decomposition* (CdCMD). Essa TEC difere da CdMD por adotar diferentes métodos não-supervisionados de extração de características a depender da evidência da multimodalidade encontrada para a dispersão da classe corrente. Assim, sendo a multimodalidade encontrada suficientemente evidente, CdCMD realiza procedimento de extração de características análogo ao do CdMD, que implica na busca pela preservação da multimodalidade na dispersão dos padrões da classe. Caso contrário, CdCMD utiliza em seu procedimento de extração de características o LPP, de modo a buscar pela preservação da estrutura local da vizinhança dos padrões da classe corrente.

A motivação para adoção desse processo de decisão é representado na Figura 23. A melhoria que se deseja alcançar ao buscar preservar a multimodalidade depende da preservação da estrutura multimodal associada à obtenção de FDP melhor ajustada às características dessa dispersão. Assim, como mostra a Figura 23a, havendo suficiente evidência de multimodalidade, é maior a chance de corretamente classificar padrões que não pertencem à classe corrente (ruídos), mas que se encontram entre as modas dessa classe. Contudo, como mostra a Figura 23b, a diminuição da evidência de multimodalidade também diminui a chance de corretamente classificar os ruídos entre as modas dessa classe. Desse modo, caso a multimodalidade na dispersão da classe seja pouco evidente, a melhoria que se deseja alcançar com a preservação desse comportamento pode ser inexpressiva. Alternativamente, percebendo-se pouca evidência de comportamento multimodal na dispersão da classe, uma TEC que realize outro objetivo pode ser empregada a fim de não

degradar o desempenho de classificação. Silva Jr, Cavalcanti e Ren (2012) mostraram com o CdLPP que a preservação da estrutura local da vizinhança dos padrões da classe conduz à elevado desempenho de classificação para diversos problemas e sem restrições a cenários multimodais. Desse modo, a utilização do LPP (ao invés do MD) quando há pouca evidência de multimodalidade, tem como objetivo conferir ao CdCMD maior robustez quando aplicada à esse tipo de cenário.

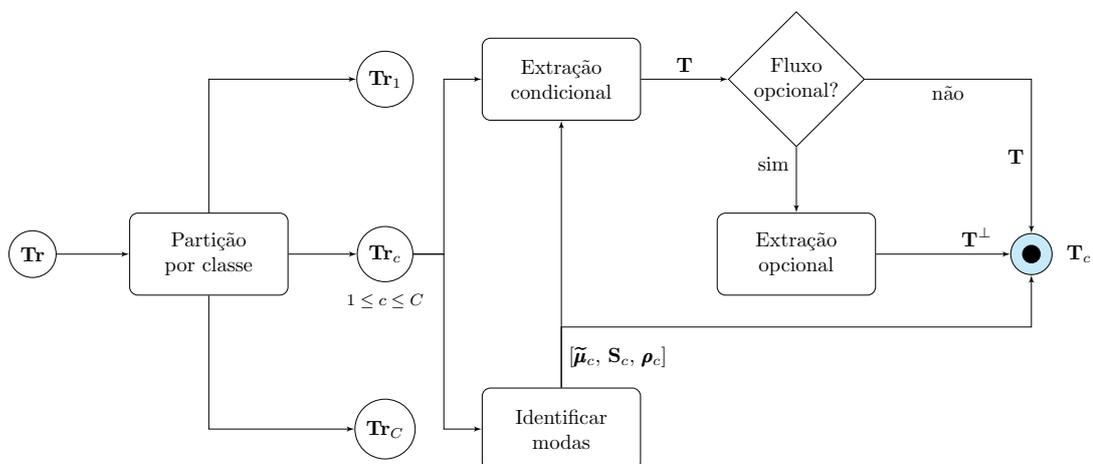
Figura 23 – Consequência da diminuição da evidência de multimodalidade sobre a classificação do CdMD. As Figuras 23a e 23b apresentam dispersões de padrões com diferentes evidências de comportamento multimodal, salientando que a diminuição na percepção da multimodalidade dificulta a obtenção da melhoria na classificação que se deseja obter com o CdMD.



Fonte: Autoria própria.

O procedimento de extração de características do CdCMD é apresentado na Figura 24. Essa TEC difere da CdMD por adotar diferentes métodos não-supervisionados de extração de características a depender da evidência da multimodalidade calculada sobre a dispersão da classe c . A seleção do método não-supervisionado apropriado à classe corrente é realizada no bloco *Extração condicional*, que tem sua arquitetura representada na Figura 25.

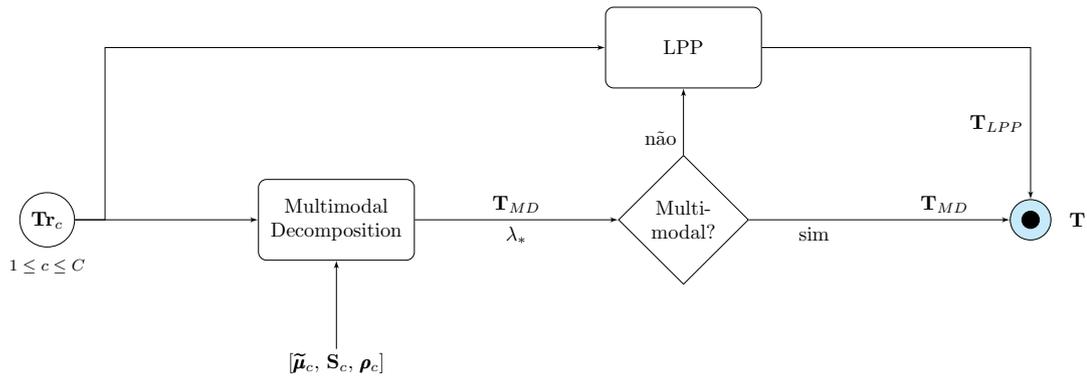
Figura 24 – Procedimento de extração de características do CdCMD.



Fonte: Autoria própria.

A extração de características condicional do CdCMD recebe como entradas as estruturas que contêm os padrões de treinamento pertencentes a classe corrente (\mathbf{Tr}_c) e

Figura 25 – Extração de características condicional do CdCMD.



Fonte: Autoria própria.

os parâmetros do modelo de mistura estimados no bloco *Identificar modas* ($\tilde{\mu}_c$, S_c e ρ_c). Sobre essas estruturas, MD é executada retornando a matriz de transformação \mathbf{T}_{MD} que contém em suas colunas as direções ortogonais com maiores evidências de multimodalidade para a classe corrente. A aplicação de MD também retorna λ_* , que é o valor da função de custo do PSO obtido para a primeira coluna de \mathbf{T}_{MD} (direção com maior evidência de multimodalidade). Assumindo um limiar sobre os valores da função de custo, é possível determinar se λ_* representa evidência de multimodalidade suficiente ou não. Em caso afirmativo, \mathbf{T}_{MD} é a matriz de transformação retornada pela extração de características condicional do CdCMD ($\mathbf{T} = \mathbf{T}_{MD}$). Caso contrário, CdCMD abandona o objetivo de preservar ao máximo a multimodalidade intraclasse, pelo objetivo de preservar ao máximo a estrutura local da vizinhança dos padrões dessa classe. Nesse caso, LPP é aplicado sobre \mathbf{Tr}_c e sua matriz resultante é retornada pela extração de características condicional do CdCMD ($\mathbf{T} = \mathbf{T}_{LPP}$).

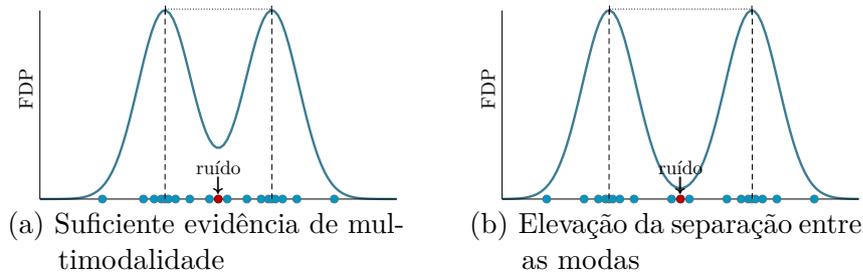
A escolha do valor do limiar de decisão, que classifica os valores λ_* da função de custo do MD, depende do conjunto de treinamento apresentado ao CdCMD. Desse modo, esse valor deve ser ajustado experimentalmente para cada problema a ser tratado com essa TEC.

3.4.3 Class-dependent Conditional Detachment of Modes

A TEC apresentada nessa seção é denominada *Class-dependent Conditional Detachment of Modes* (CdCDM). Essa TEC é uma variação do CdCMD que tem como objetivo adicional elevar a separação entre as modas quando a multimodalidade é classificada como suficientemente evidente. A Figura 26 ilustra a consequência da realização desse objetivo adicional sobre a classificação. Sendo a dispersão dos padrões da classe suficientemente multimodal (Figura 26a), indicando que um ruído entre suas modas tem pouca chance de ser classificado erroneamente, a elevação da separação entre essas modas (Figura 26b) aumenta a evidência do comportamento multimodal, consequentemente, diminui a chance

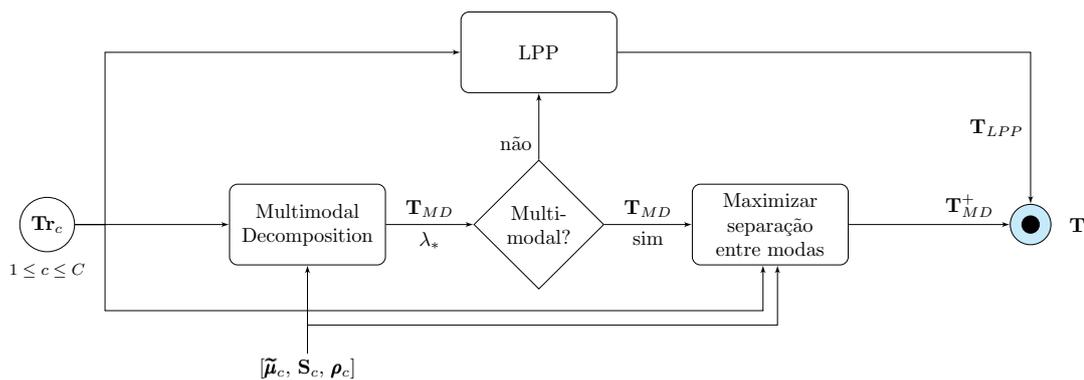
de erro na classificação desse ruído. Desse modo, CdCDM difere do CdCMD apenas na arquitetura do bloco *Extração condicional*, a qual é apresentada na Figura 27.

Figura 26 – Elevação da separação entre modas com comportamento multimodal suficientemente evidente. A Figura 26a representa dispersão de padrões com comportamento multimodal suficientemente evidente. A Figura 26b mostra que a maximização da separação entre essas modas eleva a evidência do comportamento multimodal.



Fonte: Autoria própria.

Figura 27 – Extração de características condicional do CdCDM.



Fonte: Autoria própria.

A extração de características condicional do CdCDM define a matriz de transformação \mathbf{T} apropriada para o cumprimento de seus objetivos sobre a dispersão dos padrões da classe c (\mathbf{Tr}_c). Isso é feito, inicialmente, determinando-se as \mathbf{T}_{MD} , matriz que contém as direções ortogonais com maiores evidências de multimodalidade. Com a determinação dessas direções obtém-se também λ_* , o valor que expressa a evidência de multimodalidade da primeira coluna de \mathbf{T}_{MD} (direção com maior evidência de multimodalidade). Com a adoção de um valor de limiar, λ_* é classificado quanto suficiente multimodal ou não. Caso a multimodalidade da dispersão dos padrões da classe não seja considerada suficientemente evidente, esse bloco retorna a matriz de transformação obtida com a aplicação do LPP sobre esses padrões ($\mathbf{T} = \mathbf{T}_{LPP}$). Caso contrário, \mathbf{Tr}_c , \mathbf{T}_{MD} , $\tilde{\boldsymbol{\mu}}_c$, \mathbf{S}_c e $\boldsymbol{\rho}_c$ são aplicados ao bloco *Maximizar separação entre modas* que tem como objetivo produzir a matriz de transformação \mathbf{T}_{MD}^+ que eleva a separação entre as modas. Nesse caso, a extração de características condicional do CdCDM retorna $\mathbf{T} = \mathbf{T}_{MD}^+$.

Inicialmente, para a determinação da matriz \mathbf{T}_{MD}^+ , os padrões de \mathbf{Tr}_c são rotulados segundo sua pertinência às modas estimadas. Isso é feito atribuindo o rótulo m ($m \in \{1, 2\}$)

da moda para qual o padrão \mathbf{x} produziu maior valor de probabilidade *a posteriori*. Esse procedimento é sintetizado na Equação 3.19, sendo $\tilde{\boldsymbol{\mu}}_c^m$, \mathbf{S}_c^m e ρ_c^m respectivamente o vetor média, a matriz de covariância e a probabilidade *a priori* da moda m encontrada na classe c .

$$m = \begin{cases} 1, & \text{se } \mathcal{N}(\mathbf{x}, \tilde{\boldsymbol{\mu}}_c^1, \mathbf{S}_c^1) \rho_c^1 \geq \mathcal{N}(\mathbf{x}, \tilde{\boldsymbol{\mu}}_c^2, \mathbf{S}_c^2) \rho_c^2 \\ 2, & \text{se caso contrário} \end{cases} \quad (3.19)$$

Em seguida, calcula-se a matriz de dispersão entre modas \mathbf{S}_c^b , como expressa na Equação 3.20. Nessa equação, N_m é o número de padrões que foram rotulados pela moda m . $\tilde{\boldsymbol{\mu}}_c^g$ é o vetor da média global dos padrões da classe c , calculado segundo a Equação 3.21. $\tilde{\boldsymbol{\mu}}_c^m$, a média dos padrões pertencentes à classe m , é calculada pela Equação 3.22.

$$\mathbf{S}_c^b = \sum_{m=1}^2 N_m (\tilde{\boldsymbol{\mu}}_c^m - \tilde{\boldsymbol{\mu}}_c^g) (\tilde{\boldsymbol{\mu}}_c^m - \tilde{\boldsymbol{\mu}}_c^g)' \quad (3.20)$$

$$\tilde{\boldsymbol{\mu}}_c^g = \rho_c^1 \tilde{\boldsymbol{\mu}}_c^1 + \rho_c^2 \tilde{\boldsymbol{\mu}}_c^2 \quad (3.21)$$

$$\tilde{\boldsymbol{\mu}}_c^m = N_m^{-1} \sum_{\forall \mathbf{x} \in \text{moda } m} \mathbf{x} \quad (3.22)$$

A matriz de autovetores da decomposição de \mathbf{S}_c^b possui uma única coluna linearmente independente das demais. Desse modo, apenas \mathbf{e}_* , o autovetor associado ao maior autovalor dessa decomposição, contribui para a maximização da separação entre as modas. Assim, o bloco *Maximizar separação entre modas* retorna \mathbf{T}_{MD}^+ como resultado da ortonormalização da concatenação entre \mathbf{e}_* e as colunas de \mathbf{T}_{MD} com exceção da última. Ou seja, dado que a última coluna de \mathbf{T}_{MD} indica a direção com menor evidência de multimodalidade, ela é desprezada em favor da componente (\mathbf{e}_*) que auxilia na maximização da separação entre as modas.

3.4.4 Classificador das TECs *classwise* com investigação de multimodalidade

TECs *classwise* com investigação de multimodalidade possuem dois procedimentos associados, um de extração de características e outro de classificação. Esse seção apresenta algoritmo pelo qual esse procedimento de classificação é realizado, sendo sua descrição resumizada no Algoritmo 3. Para a realização desse procedimento, o classificador das TECs *classwise* com investigação de multimodalidade deve receber, além do padrão \mathbf{z} a ser classificado, as estruturas que exprimem o conhecimento obtido sobre a dispersão de cada uma das C classes durante o procedimento de extração de características. Essas estruturas, obtidas para cada classe c ($1 \leq c \leq C$) são \mathbf{T}_c , $\tilde{\boldsymbol{\mu}}_c$, \mathbf{S}_c , ρ_c e $\boldsymbol{\eta}_c$ respectivamente

a matriz de transformação, os vetores de média das modas estimadas, as matrizes de covariância associadas à esses vetores, as probabilidades *a priori* de cada moda e o valor da probabilidade *a priori* da classe c . Com essas informações o Algoritmo 3 avalia a pertinência de \mathbf{z} para cada classe. Essa avaliação é repetida dentro do espaço resultante da projeção por cada matriz de transformação \mathbf{T}_c . Assim, esse algoritmo avalia C^2 possibilidades de pertinência a fim de determinar a classe da matriz de transformação i^* e a classe j^* que resultaram no maior valor de pertinência dado \mathbf{z} .

Algoritmo 3: Procedimento de classificação para TEC *classwise* com investigação de multimodalidade

```

Input:  $\mathbf{z}$  // Padrão de consulta
Input:  $\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_C$ , // Matrizes de transformações para cada classe
Input:  $\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\mu}}_2, \dots, \tilde{\boldsymbol{\mu}}_C$  // Vetores das médias das modas de cada classe
Input:  $\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_C$  // Matrizes de covariância das modas de cada classe
Input:  $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots, \boldsymbol{\rho}_C$  // Vetores das prob. a priori das modas de cada classe
Input:  $\eta_1, \eta_2, \dots, \eta_C$  // Valores das prob. a priori das classes
Output:  $i^*$ ; // Rótulo da transformação que melhor representa  $\mathbf{z}$ 
Output:  $j^*$ ; // Rótulo de classe obtido para  $\mathbf{z}$ 

1  $\mathbf{P} \leftarrow \text{zeros}(C, C)$ ;
2 for  $i \leftarrow 1$  to  $C$  do // Para cada transformação  $\mathbf{T}_i$  */
3    $\mathbf{T} \leftarrow \mathbf{T}_i$ ;
4    $\mathbf{z}_i \leftarrow \mathbf{T}' \times \mathbf{z}$ ; // Projeta  $\mathbf{z}$  com  $\mathbf{T}$ 
5   for  $j \leftarrow 1$  to  $C$  do // Para cada classe  $j$  */
6      $\tilde{\boldsymbol{\mu}}_{ij} \leftarrow \mathbf{T}' \times \tilde{\boldsymbol{\mu}}_j$ ; // Projeta  $\tilde{\boldsymbol{\mu}}_j$  com  $\mathbf{T}$ 
7      $\mathbf{S}_{ij} \leftarrow \mathbf{T}' \times \mathbf{S}_j \times \mathbf{T}$ ; // Projeta  $\mathbf{S}_j$  com  $\mathbf{T}$ 
8      $\mathbf{P}_{ij} \leftarrow [\mathcal{N}(\mathbf{z}_i, \tilde{\boldsymbol{\mu}}_{ij}^1, \mathbf{S}_{ij}^1) \boldsymbol{\rho}_j^1 + \mathcal{N}(\mathbf{z}_i, \tilde{\boldsymbol{\mu}}_{ij}^2, \mathbf{S}_{ij}^2) \boldsymbol{\rho}_j^2] \eta_j$ ;
9   end
10   $\mathbf{P}_i \leftarrow \mathbf{P}_i / \text{sum}(\mathbf{P}_i)$ ; // Padroniza valores da linha  $i$  de  $\mathbf{P}$  entre  $[0, 1]$ 
11 end
12  $[i^*, j^*] \leftarrow \text{max}(\mathbf{P})$ ; // Índices de linha e coluna do elemento máximo em  $\mathbf{P}$ 

```

A primeira operação desse algoritmo é iniciar a matriz \mathbf{P} que armazenará os C^2 valores de pertinência de \mathbf{z} . Nas linhas 2 à 11, a pertinência de \mathbf{z} para todas as classes, porém dentro do espaço resultante da projeção com a matriz de transformação da classe corrente. Assim, na linha 4, \mathbf{z}_i representa a projeção de \mathbf{z} com a matriz \mathbf{T}_i ($1 \leq i \leq C$). De modo análogo, para cada uma das classes j ($1 \leq j \leq C$), $\tilde{\boldsymbol{\mu}}_{ij}$ (linha 6) representa a projeção dos vetores em $\tilde{\boldsymbol{\mu}}_j$ com a matriz \mathbf{T}_i e, de mesmo modo, \mathbf{S}_{ij} (linha 7) representa a projeção das matrizes de covariância em \mathbf{S}_j com a matriz \mathbf{T}_i . Realizadas essas projeções, \mathbf{P}_{ij} (linha 8) é o valor da pertinência de \mathbf{z} à classe j no espaço de projeção da classe i . Sendo esse valor calculado com base na regra de Bayes e considerado a dispersão segundo um modelo de mistura de duas modas Normais. Os índices i^* e j^* respectivamente da linha e coluna do maior valor em \mathbf{P} são retornados por esse algoritmo. i^* indica a classe cujo espaço de projeção obteve o maior valor de pertinência para \mathbf{z} . Enquanto que j^* indica a classe que obteve o maior valor de pertinência para \mathbf{z} dentro do espaço de projeção de i^* . Assim, \mathbf{z} é rotulado como pertencente a classe j^* .

4 Experimentos

Na Seção 3 foi definido o grupo das Técnicas de Extração de Características (TECs) *classwise* com investigação de multimodalidade, que visa encontrar espaços de redução dimensional adaptados a presença de multimodalidade intraclasse. Para esse grupo foram propostas três TECs: *Class-dependent Multimodal Decomposition* (CdMD - página 64), *Class-dependent Conditional Multimodal Decomposition* (CdCMD - página 65) e *Class-dependent Conditional Detachment of Modes* (CdCDM - página 67). Todas essas técnicas utilizam TEC não-supervisionada para realizar a investigação e preservação da multimodalidade presente na dispersão dos padrões de cada classe. *Multimodal Decomposition* (MD - página 40) é a TEC não-supervisionada proposta para o cumprimento desses objetivos. MD busca por direções ortogonais que melhor evidenciam a multimodalidade existente na dispersão original dos padrões. Para que isso seja possível, é necessário medir e comparar as direções investigadas quanto a sua aptidão em evidenciar/ preservar a multimodalidade original. Na Seção 3.2.3 (página 51) são apresentadas cinco métricas para estimar a multimodalidade em dispersões unidimensionais. Dessas cinco, *Index of Area-based Normal Intersection* (IANI - página 54) é proposta neste trabalho.

Nessa seção são apresentados os experimentos desenvolvidos para comparação das técnicas propostas neste trabalho com técnicas de referência na literatura. Inicialmente, na Seção 4.1, são comparadas as métricas utilizadas para estimativa da multimodalidade. Em seguida, na Seção 4.2, são comparados os desempenhos de TECs não-supervisionadas na tarefa de preservar a multimodalidade em dispersões de padrões. Enquanto que, na Seção 4.3, compara-se o desempenho de classificação das técnicas *classwise* com investigação de multimodalidade com outras TECs, *classwise* ou não.

4.1 Comparação entre as métricas de estimativa de multimodalidade

Nessa seção são avaliadas as métricas utilizadas para estimar multimodalidade em dispersões de padrões unidimensionais aplicadas a tarefa de decomposição de espaços multidimensionais.

A utilização das métricas *Eisenberger*, *Behboodian* e IANI requer a estimativa da Função Densidade de Probabilidade (FDP) que melhor aproxima distribuição bimodal sobre os dados. Assim, dado que em um cenário real não se conhece os parâmetros que modelam a mistura entre as duas modas que compõem a dispersão dos padrões, deve-se inferir os valores desses parâmetros com a utilização dessas métricas. Como discutido na

Seção 3.3, foi utilizado GMM (BILMES, 1998) para estimar valores dos parâmetros que melhor modelam a dispersão dos padrões segundo distribuição composta pela mistura entre duas Normais (Equação 3.9).

A avaliação dessas métricas possibilita a comparação entre seus desempenhos, sendo essa comparação realizada através de dois experimentos. No primeiro experimento (Seções 4.1.1), verifica-se a influência do valor do parâmetro ρ sobre as respostas das métricas. Enquanto que no experimento seguinte (Seções 4.1.2), as métricas são aplicadas a cenário de classificação a fim de relacionar suas influências sobre a discriminação entre classes. A Seção 4.1.3 apresenta as conclusões obtidas com a realização desses experimentos.

4.1.1 Influência da ponderação entre as modas

Esse experimento visa avaliar o desempenho das métricas de estimativa da multimodalidade com a variação dos valores do parâmetro ρ , que é responsável pela ponderação das Normais na mistura expressa na Equação 3.9.

A relação entre o valor de ρ e a categoria do comportamento da dispersão (unimodal ou multimodal) é de tal modo complexa que, a exemplo de Eisenberger (1964) e Behboodian (1970), as condições estabelecidas para a ocorrência dessas categorias não observam a de variação desse parâmetro. Contudo, como evidencia a Figura 13 (página 55), a variação de ρ influencia na percepção da multimodalidade. Assim, esse experimento é projetado de modo a salientar o comportamento das diferentes métricas de multimodalidade com a variação de ρ e avaliar quais delas são mais robustas a essa variação.

As Seções 4.1.1.1 e 4.1.1.2 apresentam respectivamente as bases de dados utilizadas e os resultados obtidos nesse experimento.

4.1.1.1 Definição da base de dados

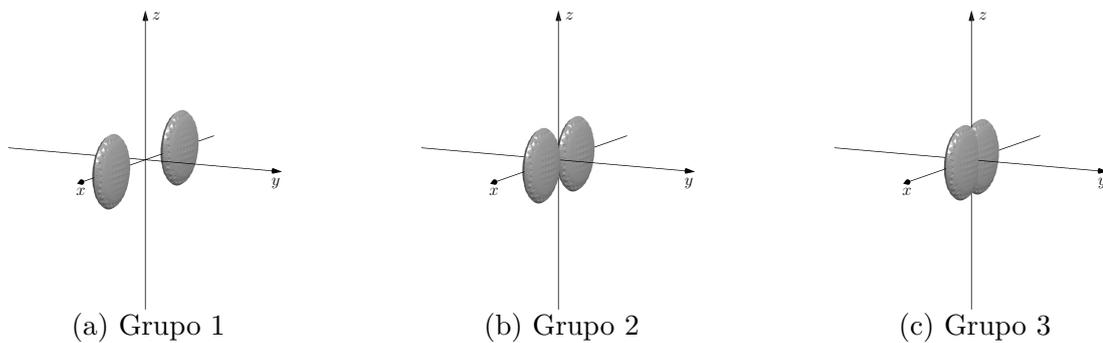
Para a realização desse experimento foram geradas 90 bases de dados tridimensionais. O valor x da primeira dimensão dos padrões dessas bases segue distribuição bimodal formada pela mistura de modas Normais. Enquanto que as demais dimensões (y e z) são geradas de função de distribuição Normal e unimodal. Essa modelagem faz com que a multimodalidade ocorra na direção do eixo x . Assim, o ângulo \hat{d}_1^x entre x e d_1 , a primeira direção retornada pela decomposição espacial, é um parâmetro da corretude dessa direção quanto a preservação da multimodalidade (a corretude de d_1 aumenta a medida que $\hat{d}_1^x \rightarrow 0^\circ$ e diminui quando $\hat{d}_1^x \rightarrow 90^\circ$). Para todas as bases, a dispersão em z segue $\mathcal{N}(0, 1^2)$. Entretanto, as dispersões em x e y causam partição das bases em 9 grupos, como mostra a Tabela 3.

Adicionalmente, as Figuras 28-30 ilustram a dispersão das bases em cada grupo segundo as configurações especificadas para cada eixo.

Tabela 3 – Configuração dos grupos sobre as bases geradas de acordo com os eixos x e y .

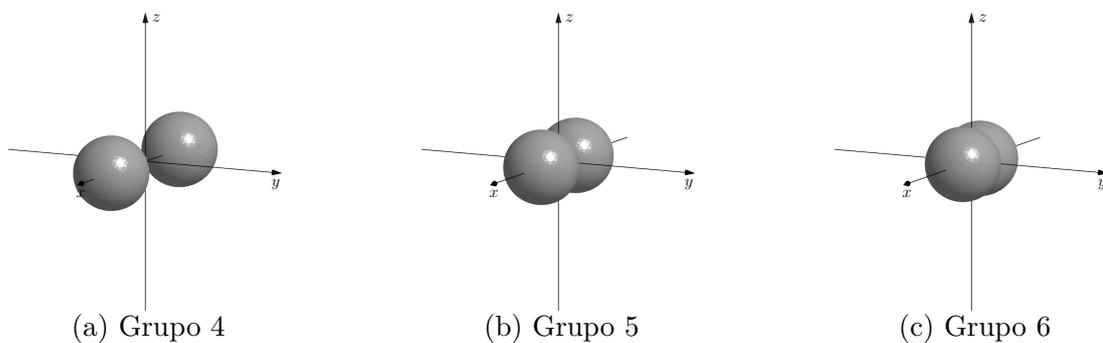
		Dispersão em x		
		Configuração	$\mathcal{N}(\pm 2, 1^2)$	$\mathcal{N}(\pm 1, 1^2)$
Dispersão em y	$\mathcal{N}(0, 0, 25^2)$	Grupo 1 Figura 28a	Grupo 2 Figura 28b	Grupo 3 Figura 28c
	$\mathcal{N}(0, 1^2)$	Grupo 4 Figura 29a	Grupo 5 Figura 29b	Grupo 6 Figura 29c
	$\mathcal{N}(0, 2^2)$	Grupo 7 Figura 30a	Grupo 8 Figura 30b	Grupo 9 Figura 30c

Figura 28 – Representação da dispersão dos padrões gerados para as bases dos Grupos 1-3. Essas dispersões seguem distribuição $\mathcal{N}(0, 0, 25^2)$ em y e diferem entre si quanto a dispersão sobre x , que é modelada segundo duas modas Normais, ambas com desvio padrão $\sigma = 1$. As médias das Normais em x são: ± 2 (28a), ± 1 (28b) e $\pm \frac{1}{2}$ (28c).



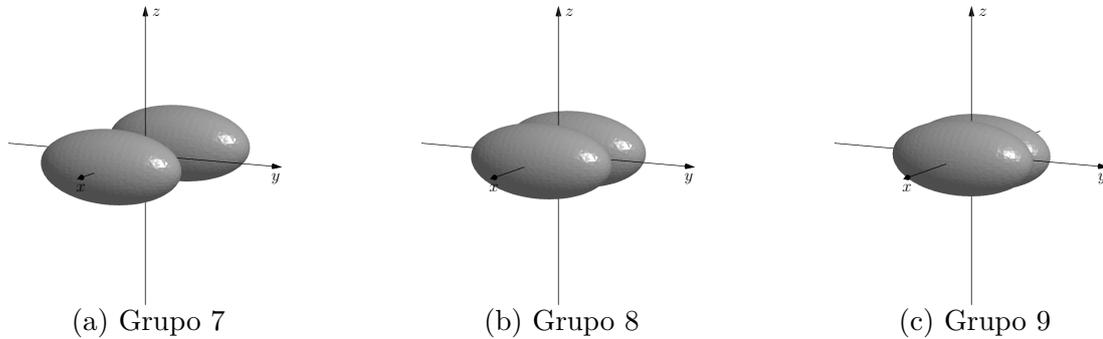
Fonte: Autoria própria.

Figura 29 – Representação da dispersão dos padrões gerados para as bases dos Grupos 4-6. Essas dispersões seguem distribuição $\mathcal{N}(0, 1^2)$ em y e diferem entre si quanto a dispersão sobre x , que é modelada segundo duas modas Normais, ambas com desvio padrão $\sigma = 1$. As médias das Normais em x são: ± 2 (29a), ± 1 (29b) e $\pm \frac{1}{2}$ (29c).



Fonte: Autoria própria.

Figura 30 – Representação da dispersão dos padrões gerados para as bases dos Grupos 7-9. Essas dispersões seguem distribuição $\mathcal{N}(0, 2^2)$ em y e diferem entre si quanto a dispersão sobre x , que é modelada segundo duas modas Normais, ambas com desvio padrão $\sigma = 1$. As médias das Normais em x são: ± 2 (30a), ± 1 (30b) e $\pm \frac{1}{2}$ (30c).



Fonte: Autoria própria.

Portanto, cada um dos grupos contém dez bases de dados que diferem entre si quanto o valor de ρ utilizado na geração de seus padrões. Sendo o valor de ρ variado com $\{ 5\%, 10\%, 15\%, 20\%, 25\%, 30\%, 35\%, 40\%, 45\%, 50\% \}$.

4.1.1.2 Resultados

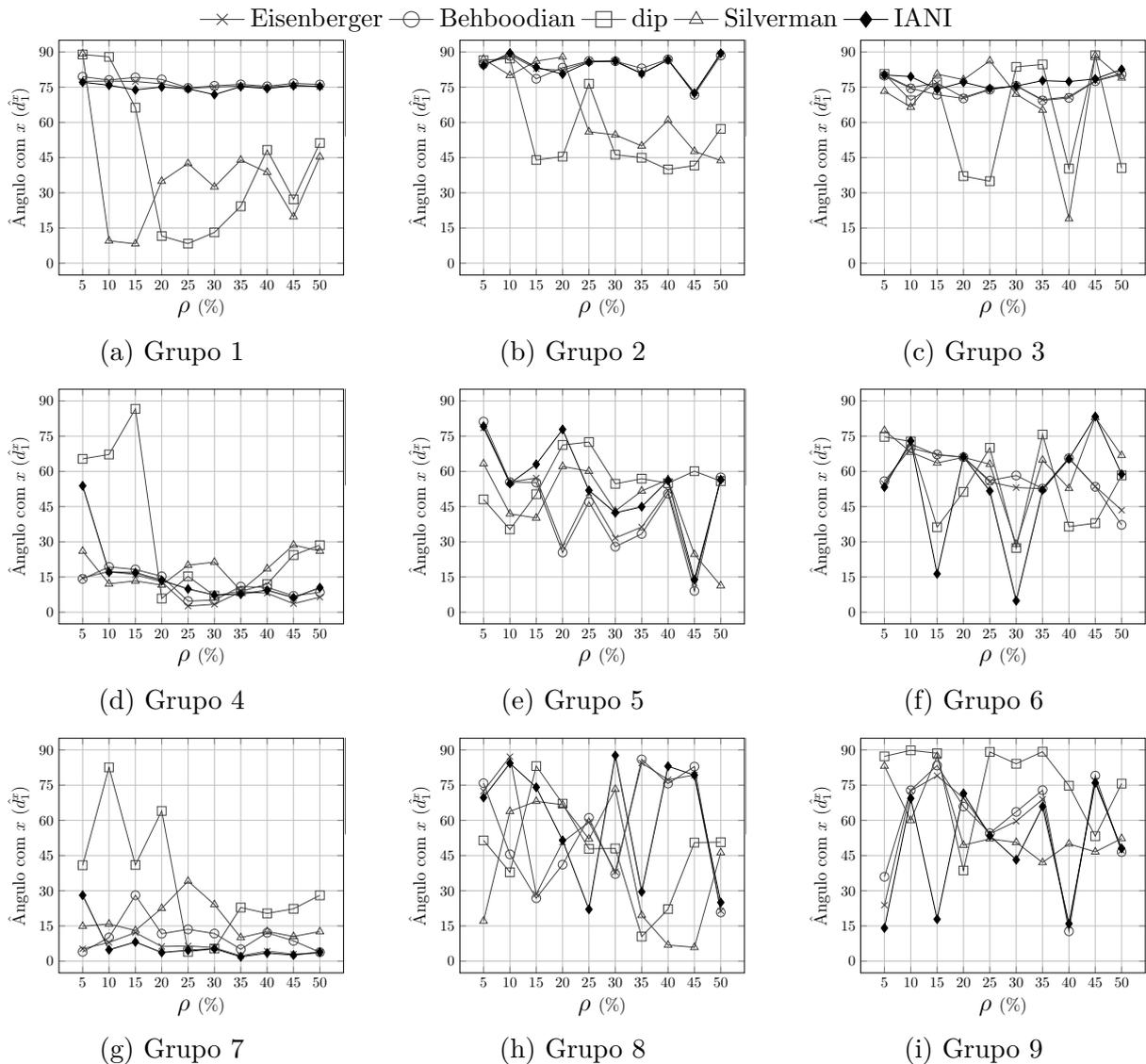
Essa seção apresenta os resultados experimentais obtidos para as 90 bases de dados tridimensionais descritas na Seção 4.1.1.1. Dado que, para todas as bases, a dispersão multimodal é possível apenas no eixo x , utiliza-se \hat{d}_1^x , o ângulo entre x e d_1 , como parâmetro da corretude de d_1 quanto a preservação da multimodalidade. Assim, a Figura 31 apresenta, para todos grupos de bases de dados, os valores de \hat{d}_1^x relacionados com a variação dos valores de ρ .

Os resultados obtidos para os grupos $\{1, 2, 3\}$ são apresentados nas Figuras 31a-31c respectivamente. Vê-se que, para esses grupos, as estimativas de multimodalidade por dip e Silverman obtiveram os melhores valores de corretude. E, como indicam os resultados para o grupo 3, a correta estimativa é dificultada quando a multimodalidade em x é menos evidente.

As Figuras 31d-31f apresentam os resultados obtidos para os grupos $\{4, 5, 6\}$ respectivamente. Nos grupos 4 e 5, Eisenberger, Behboodian e IANI apresentam os melhores valores de corretude com o aumento dos valores de ρ . Contudo, para o grupo 6, nenhuma das técnicas apresenta um comportamento bem definido quanto a estimativa da multimodalidade, o que é devido à menor evidência da multimodalidade contida nas bases desse grupo.

Os resultados obtidos para os grupos $\{7, 8, 9\}$ são apresentados nas Figuras 31g-31i respectivamente. IANI apresentou os melhores valores de corretude para a maioria das bases do grupo 7, sendo seguida por Eisenberger. No grupo 8, dip e Silverman alcançam

Figura 31 – Ângulos com o eixo x das primeiras direções retornadas pela decomposição espacial utilizando cada uma das métricas de estimativa de multimodalidade.



Fonte: Autoria própria.

os maiores valores de corretude para os maiores valores de ρ , exceto quando $\rho = 50\%$, cujos melhores valores são alcançados por Eisenberger, Behboodian e IANI. Para todas as técnicas, a menor evidência de multimodalidade que ocorre nas bases do grupo 9 conduz a valores de \hat{d}_1^x elevados (menor corretude na estimativa da multimodalidade), destacando-se dip devido a obtenção dos piores valores de corretude.

Além dessas observações, vê-se que o aumento da dispersão em y eleva a corretude de Eisenberger, Behboodian e IANI quando a multimodalidade em x é a mais evidente (Figuras 31a, 31d e 31g).

A fim de comparar o desempenho das técnicas de estimativa de multimodalidade, por base de dados, seus valores de corretude receberam posições que expressam a ordem das técnicas quanto ao sucesso na estimativa da multimodalidade (DEMŠAR, 2006). Esse

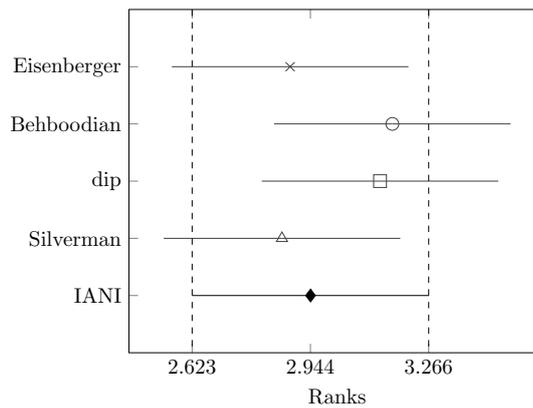
procedimento é exemplificado na Tabela 4, a qual apresenta valores de corretude (\hat{d}_1^x) para uma base de dados hipotética. As técnicas com as maiores corretudes (menores valores de \hat{d}_1^x) recebem os menores valores de posição (*Rank*).

Tabela 4 – Ordenação das técnicas segundo a corretude da estimativa de multimodalidade.

Statistics	Eisenberger	Behboodian	dip	Silverman	IANI
\hat{d}_1^x	42,44	82,88	83,89	23,22	20,56
<i>Rank</i>	3,0	4,0	5,0	2,0	1,0

Realizado o posicionamento das técnicas para todas as 90 bases de dados, verifica-se através de teste de Friedman (DEMŠAR, 2006) que todas as técnicas de estimativa de multimodalidade comparadas possuem desempenhos que são estatisticamente equivalentes entre si (nível de confiança = 95%), como mostra a Figura 32.

Figura 32 – Comparação entre as técnicas de estimativa de multimodalidade através do teste de Friedman, com confiança de 95%.



Fonte: Autoria própria.

Porém, a comparação dessas técnicas limitada pelos valores de ρ salienta a relação entre seus desempenhos com a variação desse parâmetro (ρ). Para isso, valores médios dos posicionamentos das técnicas foram calculados para cada valor de ρ , sendo esses valores apresentados na Tabela 5.

Ou seja, para um valor fixo de ρ e para uma técnica qualquer, existem nove valores de posicionamento (um posicionamento para cada grupo de base de dados). Assim, a referida técnica, para o valor fixado de ρ , possui posicionamento médio como a média desses nove valores obtidos de cada grupo. Desse modo, a Tabela 5 apresenta os valores de posicionamento médio dos desempenhos de cada técnica de estimativa de multimodalidade para cada valor de ρ . Essa tabela mostra que, para os menores desbalanceamentos da ponderação entre as modas ($\rho \geq 30\%$), Silverman e IANI alcançam cada uma 2/5 das melhores posições de desempenho. Porém, IANI alcança sozinha a maioria das melhores posições de desempenho (2/5) para os menores valores de ρ ($\rho \leq 25\%$), para os quais a estimativa de multimodalidade tem corretude degradada devido ao maior desbalanceamento entre as modas.

Tabela 5 – *Rank* médio das bases agrupadas pelo valor de ρ .

ρ	Eisenberger	Behboodian	dip	Silverman	IANI
5%	2,444	2,889	3,889	3,222	2,556
10%	3,556	3,222	3,333	1,667	3,222
15%	2,889	3,111	3,444	2,889	2,667
20%	2,889	3,111	2,222	3,222	3,667
25%	2,889	3,444	2,889	3,333	2,444
30%	2,889	3,222	3,000	3,111	2,778
35%	2,889	3,667	3,444	2,667	2,333
40%	2,889	3,222	2,778	2,778	3,333
45%	3,000	3,111	3,000	2,889	3,000
50%	2,556	2,667	3,333	2,889	3,556

4.1.2 Influência sobre a discriminação entre classes

Esse experimento é realizado para comparação das métricas de estimativa de multimodalidade aplicadas a cenários de classificação de padrões.

Considerando que diferentes classes possuem dispersão de seus padrões segundo diferentes modas, é trivial a percepção de que uma maior sobreposição entre essas modas eleva a confusão na classificação dos padrões que estão contidos na região de conflito (região da sobreposição entre as modas). Desse modo, quanto menor a região de conflito: (i) há maior separação entre as modas, o que implica em (ii) maior multimodalidade em sua dispersão conjunta e, conseqüentemente, (iii) maior precisão na discriminação dos padrões entre as classes. Porém, a projeção desses padrões para espaço de menor dimensionalidade pode acentuar a sobreposição entre as modas das classes, degradando a precisão de classificação. Alternativamente, as técnicas de estimativa de multimodalidade podem ser utilizadas para determinar espaços de redução dimensional que preservem ao máximo a multimodalidade original contida entre a dispersão das classes e com isso, minimizar a degradação da precisão de classificação. Portanto, esse experimento é realizado para comparar o desempenho das técnicas de estimativa de multimodalidade na tarefa de encontrar espaços de projeção que reduzam a dimensionalidade dos padrões. Sendo o desempenho de cada técnica mensurado pela precisão de classificação.

As Seções 4.1.2.1 e 4.1.2.2 apresentam respectivamente as bases de dados utilizadas nesse experimento e os resultados que foram obtidos.

4.1.2.1 Definição da base de dados

Para mensurar o desempenho de classificação com o emprego das técnicas de estimativa de multimodalidade, 24 bases de dados bidimensionais foram aleatoriamente geradas. A Tabela 6 apresenta os valores dos parâmetros envolvidos na construção dessas bases de dados.

Cada uma dessas bases possuem 500 padrões igualmente divididos em duas classes, ambas

Tabela 6 – Configurações para geração dos padrões das bases *Toy*.

<i>Toy</i>	Classe 1					Classe 2				
	$\tilde{\mu}_x$	$\tilde{\mu}_y$	\tilde{s}_x	\tilde{s}_y	α	$\tilde{\mu}_x$	$\tilde{\mu}_y$	\tilde{s}_x	\tilde{s}_y	α
1	7,911	8,778	0,902	4,815	280,653	-4,539	-6,442	6,438	7,742	221,753
2	6,361	-8,517	7,517	0,283	260,172	-2,230	4,502	8,354	8,028	266,146
3	-2,790	-9,612	1,206	4,337	336,145	5,646	5,737	3,590	5,637	12,419
4	2,397	-2,707	0,427	0,102	151,581	-2,053	-9,414	4,828	9,635	6,005
5	0,139	1,868	2,257	6,356	241,532	-6,089	8,059	2,667	3,682	95,550
6	8,923	-4,957	5,374	8,816	96,231	-0,521	5,283	6,882	0,891	337,962
7	-1,674	-9,123	6,909	9,890	256,775	-2,751	9,198	1,043	1,405	281,770
8	-6,952	-5,373	4,913	9,040	346,944	8,002	2,648	0,123	8,708	26,193
9	1,446	0,985	2,080	0,564	41,874	-2,134	-6,952	5,529	9,955	37,556
10	-2,528	-0,418	2,973	1,094	190,837	-2,230	3,979	5,129	5,848	28,709
11	5,815	9,733	8,157	7,771	64,249	5,891	8,339	0,615	7,840	91,351
12	6,922	1,079	0,140	2,233	228,727	-1,478	-0,271	2,414	0,348	203,078
13	0,931	-2,398	6,585	7,829	72,824	1,553	-2,629	1,679	7,524	248,255
14	4,345	-7,249	5,019	5,546	247,668	-8,301	-2,114	8,564	6,261	27,281
15	7,261	-7,319	5,764	4,413	278,052	-3,916	1,231	0,707	6,431	215,321
16	-0,799	-6,878	9,604	0,176	114,166	8,516	8,046	5,741	1,954	69,309
17	-2,524	-9,483	8,861	3,246	69,224	8,092	7,331	0,448	3,956	317,073
18	7,805	9,761	4,430	8,292	252,340	-6,864	-1,094	7,566	4,589	272,055
19	-8,393	-3,009	5,775	5,367	40,468	8,667	8,257	3,802	8,815	155,979
20	6,015	5,621	1,686	8,251	216,433	0,949	-1,839	1,914	8,042	351,962
21	-1,459	6,650	7,997	8,718	332,258	6,159	7,227	5,781	2,474	235,367
22	-8,876	-1,640	3,305	6,614	20,073	-9,815	-7,929	0,429	3,550	270,323
23	-1,170	-3,624	5,651	2,448	3,198	-1,851	9,857	9,750	7,585	113,164
24	4,368	-9,304	7,112	5,266	266,068	8,274	3,165	7,181	0,129	307,211

com dispersão unimodal. Os 250 padrões de cada classe foram obtidos com base na variação de cinco parâmetros responsáveis pela determinação da função de distribuição Normal da classe: $\tilde{\mu}_x$ e $\tilde{\mu}_y$, projeções da média $\tilde{\mu}$ respectivamente sobre os eixos das abscissas e ordenadas, variados no intervalo $[-10, 10]$; \tilde{s}_x e \tilde{s}_y , desvios padrão da projeção dos padrões respectivamente sobre os eixos das abscissas e ordenadas, variados no intervalo $(0, 10]$; $\alpha \in [0^\circ, 360^\circ]$, ângulo utilizado para rotação dos padrões da classe.

$$\mathcal{N}(\tilde{\mu}, \tilde{s}) \quad \tilde{\mu} = \begin{bmatrix} \tilde{\mu}_x \\ \tilde{\mu}_y \end{bmatrix} \quad \tilde{s} = \begin{bmatrix} \tilde{s}_x & 0 \\ 0 & \tilde{s}_y \end{bmatrix} \quad rot_\alpha = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}$$

4.1.2.2 Resultados

Utilizando validação cruzada em cinco partes ($P = \{p_1, p_2, p_3, p_4, p_5\}$) sobre essas bases *toy*, as técnicas de estimativa de multimodalidade foram aplicadas para cada partição dos dados (p_i com $1 \leq i \leq 5$), sendo calculadas cinco direções de maior evidência de multimodalidade, por técnica de estimativa de multimodalidade. Cada direção obtida foi utilizada para projetar os padrões das partições restantes ($P - \{p_i\}$) no espaço unidimensional com maior evidência de multimodalidade segundo a técnica corrente. Em cada novo espaço unidimensional, os padrões $P - \{p_i\}$ foram classificados, sendo calculado o valor de precisão de classificação. Assim, como mostra a Tabela 7, para cada técnica de estimativa de multimodalidade são calculados o valor médio e desvio padrão dos cinco valores de

precisão obtidos para cada base *toy*.

Tabela 7 – Valores médios de precisão de classificação (%) obtidos pelas estimativas de multimodalidade para as bases *Toy*. Valores entre parênteses representam os desvios padrão obtidos.

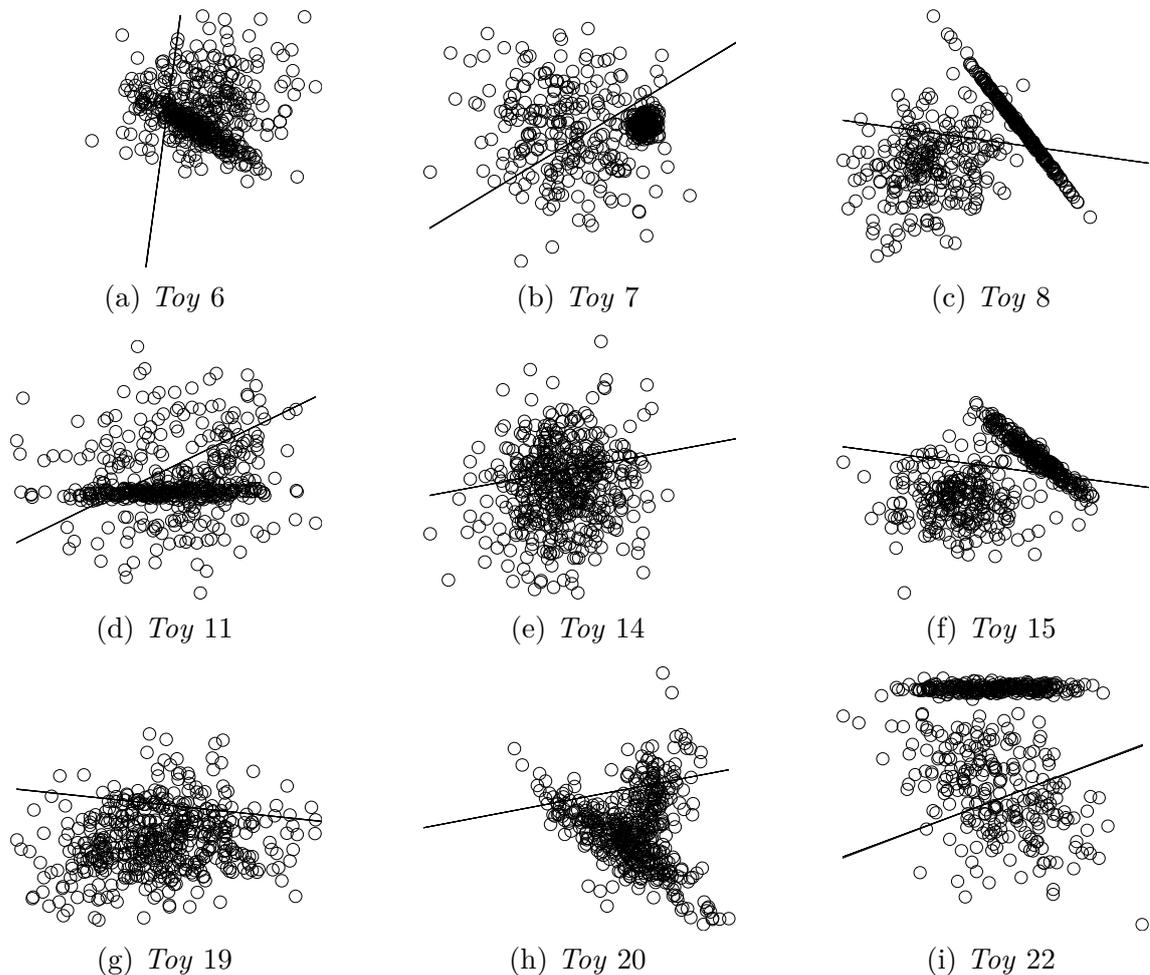
<i>Toy</i>	Eisenberger	Behboodian	dip	Silverman	IANI
1	97,68 (0,168)	97,68 (0,168)	96,66 (2,062)	88,09 (5,528)	97,92 (0,104)
2	12,48 (0,220)	12,48 (0,220)	4,86 (2,480)	12,97 (8,177)	50,00 (0,000)
3	87,40 (0,366)	87,12 (0,327)	97,72 (0,220)	97,70 (0,203)	98,68 (0,076)
4	66,12 (0,295)	66,12 (0,295)	73,26 (5,961)	72,82 (7,845)	76,31 (0,531)
5	5,80 (0,190)	5,80 (0,190)	11,05 (2,537)	11,18 (0,979)	11,80 (0,302)
6	10,72 (0,358)	10,72 (0,358)	49,96 (0,055)	30,11 (15,039)	39,24 (0,185)
7	94,52 (0,192)	94,52 (0,192)	94,00 (0,838)	94,72 (0,214)	94,16 (0,288)
8	67,08 (0,323)	67,08 (0,323)	96,76 (0,263)	96,77 (0,309)	90,16 (0,370)
9	5,60 (0,117)	5,60 (0,117)	15,07 (11,091)	20,85 (17,174)	32,84 (0,363)
10	20,28 (0,214)	20,24 (0,216)	34,25 (13,066)	30,72 (4,869)	43,40 (0,359)
11	92,44 (0,171)	92,44 (0,171)	50,00 (0,000)	64,74 (20,411)	50,00 (0,000)
12	69,68 (0,259)	69,68 (0,259)	78,87 (0,811)	78,66 (0,647)	99,84 (0,042)
13	35,92 (0,697)	35,92 (0,697)	48,27 (3,868)	45,70 (6,158)	50,00 (0,000)
14	50,00 (0,000)	60,86 (0,580)	50,93 (2,080)	52,52 (3,631)	50,00 (0,000)
15	97,88 (0,157)	97,88 (0,157)	99,28 (0,067)	99,29 (0,065)	81,82 (0,271)
16	72,68 (0,202)	72,68 (0,202)	89,55 (3,979)	84,26 (2,803)	94,61 (0,248)
17	13,60 (0,405)	13,61 (0,420)	42,39 (11,479)	36,47 (13,909)	49,36 (0,082)
18	52,60 (0,146)	52,55 (0,127)	56,64 (10,929)	62,46 (13,096)	90,65 (0,209)
19	74,96 (0,668)	72,77 (0,347)	53,00 (6,708)	51,70 (3,801)	74,32 (0,571)
20	72,72 (0,346)	72,72 (0,346)	54,33 (5,987)	50,45 (0,825)	50,37 (0,067)
21	34,00 (0,361)	33,92 (0,377)	47,15 (5,575)	44,03 (6,129)	49,12 (0,115)
22	92,44 (0,175)	92,44 (0,175)	92,43 (0,199)	92,48 (0,214)	79,35 (0,433)
23	47,64 (0,074)	47,54 (0,096)	48,15 (4,025)	49,96 (0,055)	50,00 (0,000)
24	69,60 (0,391)	69,60 (0,391)	94,99 (0,297)	94,43 (0,673)	95,60 (0,215)
%Wins	3	3	1	4	15
%Global	55,993 (0,271)	56,332 (0,281)	61,649 (3,941)	60,962 (5,532)	66,648 (0,201)

Em cada linha dessa tabela, as células destacadas indicam as técnicas de estimativa de multimodalidade que obtiveram maior precisão de classificação média. Vê-se que 15 dentre as 24 bases *toy* obtiveram maior valor de precisão média através de IANI (linha %Wins). Além disso, a linha %Global apresenta os valores médios e desvios padrão calculados sobre os 24 valores de precisão obtidos por cada técnica de estimativa de multimodalidade. Nessa linha, IANI obtém o maior valor de precisão média com menor valor de desvio padrão, indicando maior estabilidade nas precisões calculadas com a variação das partições dos dados.

Na Figura 33 são apresentadas as dispersões das bases *toy* para as quais IANI não obteve o maior valor médio de precisão. Dado que as técnicas de estimativa de multimodalidade não consideram a informação de supervisão dos padrões, essas figuras não salientam a divisão dos padrões segundo suas classes. Nessa figura, as retas indicam as direções de maior evidência de multimodalidade estimadas por IANI para cada partição das bases.

As bases das Figuras 33a, 33c, 33d, 33f e 33i (respectivamente: *toys* 6, 8, 11, 15 e 22) apresentam como característica comum que, em uma das dimensões, uma das modas possui baixo valor de desvio padrão, $s_{x\text{ou}y} = \{0, 891; 0, 123; 0, 615; 0, 707; 0, 429\}$ respectivamente. Nessas bases, IANI estimou direções que não preservaram ao máximo

Figura 33 – Dispersão das bases *toy* para as quais IANI não obteve a maior precisão média (visão não-supervisionada).



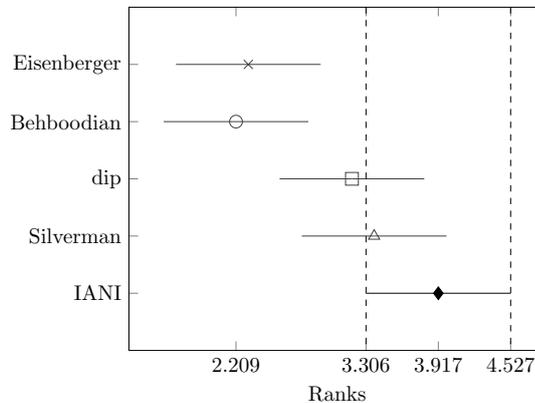
Fonte: Autoria própria.

a multimodalidade original entre as classes. Contudo, em outras bases com essa mesma característica, IANI obtém elevados valores médio de precisão, como, por exemplo, para as bases *toy 12* (99,84%), *16* (94,61%) e *24* (95,60%). Vê-se que o baixo desempenho de IANI para esse tipo de cenário foi também identificado no experimento anterior, o que é salientado nos resultados obtidos para as bases do grupos 1, 2 e 3 respectivamente Figuras 31a, 31b e 31c. Esses grupos de bases possuem desvio padrão de 0,25 no eixo y .

Não se percebe uma relação comum entre as características das demais bases *toy* (Figuras 33b, 33e, 33g e 33h) de modo a justificar os baixos valores médio de precisão que foram obtidos. Pontualmente: a base *toy 7* (Figura 33b) apresenta valor de precisão equivalente à maior precisão obtida; as *toys 14* e *19* (Figuras 33e e 33g) não apresentam evidente dispersão multimodal, de modo que não se esperam elevados desempenhos de classificação por nenhuma das técnicas de estimativa de multimodalidade; *toy 20* (Figura 33h) apresenta dispersão multimodal das classes, porém o arranjo entre as dispersões de suas classes é de tal modo complexo que dificultou a estimativa realizada por IANI.

A Figura 34 apresenta o resultado da comparação de desempenho das técnicas de estimativa de multimodalidade através do teste de Friedman, para as 24 bases *toy*.

Figura 34 – Comparação entre as técnicas de estimativa de multimodalidade através do teste de Friedman, com confiança de 95%.



Fonte: Autoria própria.

Nessa figura, os maiores valores de *rank* indicam os melhores desempenhos. Assim, a ordenação das técnicas segundo os valores decrescentes do *rank* é: IANI, Silverman, dip, Eisenberger e Behboodian. Contudo, Silverman, dip, Eisenberger e Behboodian são estatisticamente equivalentes entre si, com confiança de 95%. Sendo IANI estatisticamente equivalente à Silverman e dip.

4.1.3 Conclusões do experimento

Os resultados experimentais apresentados nas Seções 4.1.1 e 4.1.2 salientam a utilização de IANI como métrica de estimativa de multimodalidade mais adequada para uso geral.

Na Seção 4.1.1 viu-se que a comparação das métricas sobre as 90 bases geradas não possibilita a distinção entre seus desempenhos quanto a corretude da direção de maior multimodalidade que são encontradas através delas. Ou seja, de modo geral, as cinco métricas de estimativa de multimodalidade são estatisticamente equivalentes (confiança de 95%). Contudo, comparando-se as métricas para cenários mais restritos (diferentes valores de ρ na ponderação das modas), vê-se que IANI obtém a maioria dos melhores desempenhos médios dentre os dez cenários avaliados ($\rho = \{ 15\%, 25\%, 30\%, 35\% \}$).

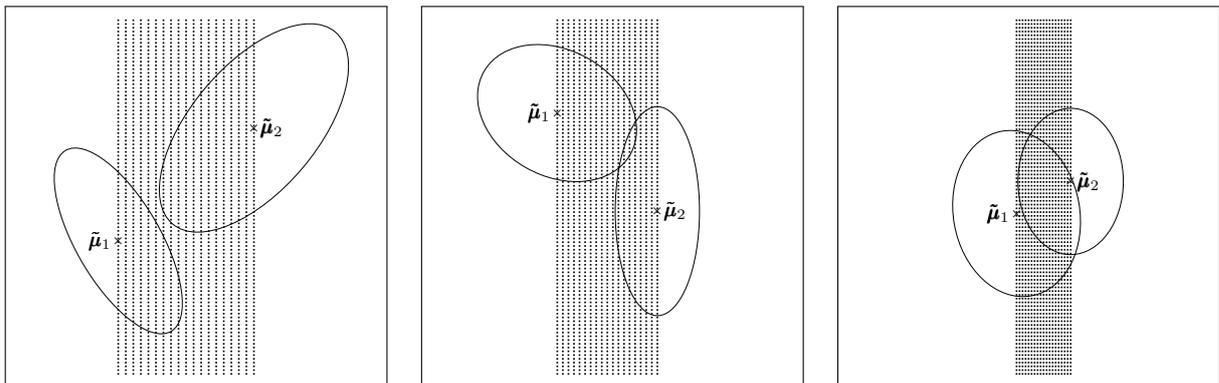
Na Seção 4.1.2 foram comparados os desempenhos de classificação obtidos com o uso das métricas de estimativa de multimodalidade. Sobre as 24 bases geradas e com confiança de 95%, dip, Silverman e IANI são estatisticamente equivalentes. Porém, IANI alcançou os maiores desempenhos médios de classificação.

4.2 Preservação do comportamento multimodal

Nessa seção, MD tem seu desempenho de classificação comparado ao de TECs largamente difundidas na literatura: PCA e LPP. Essas três TECs são não supervisionadas, porém modeladas para satisfazerem objetivos distintos: PCA busca preservar ao máximo a variância da dispersão dos padrões; LPP visa preservar a estrutura local da vizinhança dos padrões e MD tem como objetivo preservar ao máximo a estrutura multimodal na dispersão dos padrões.

Considerando o objetivo para qual o MD é proposto, o experimento aqui apresentado compara o desempenho das TECs em cenários cuja dispersão dos padrões possuem diferentes evidências de comportamento multimodal. A Figura 35 exemplifica os cenários construídos para a comparação entre as TECs.

Figura 35 – Exemplos de cenários com comportamento multimodal utilizado no experimento.



Fonte: Autoria própria.

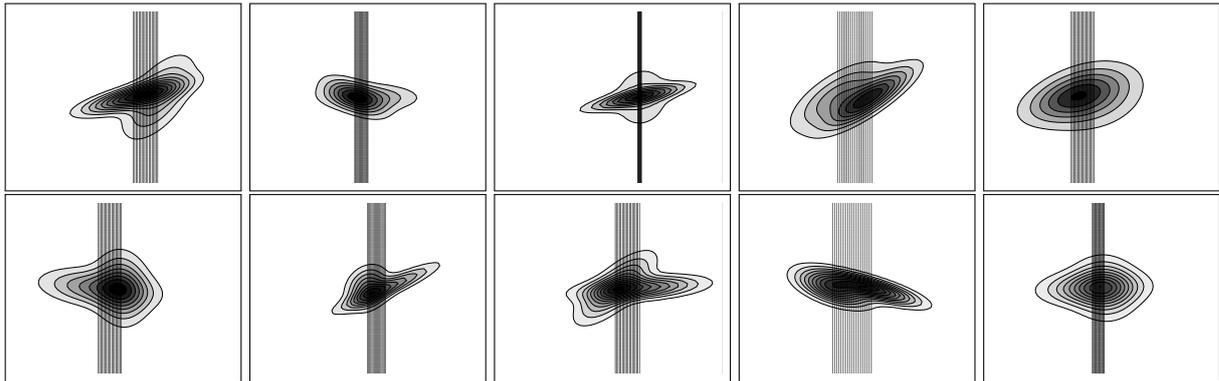
Nessa figura, cada elipse representa uma moda existente na dispersão dos padrões. Sobre esses padrões as TECs empregadas calculam suas respectivas reduções dimensionais para espaço unidimensional. A eficiência com que esse novo espaço preserva o comportamento multimodal original é medido através do desempenho de classificação em distinguir ruídos dispersos entre as médias de cada moda. Na Figura 35, $\tilde{\mu}_1$ e $\tilde{\mu}_2$ representam as médias das modas (elipses), sendo os ruídos representados por \bullet . Assim, o espaço unidimensional que preserva a multimodalidade original de modo mais eficiente tende a obter melhor taxa de correta classificação sobre os ruídos projetados nesse espaço.

Portanto, para a realização desse experimento, são construídas na Seção 4.2.1 bases de dados bidimensionais formadas por duas modas Normais e que apresentam diferentes evidências de comportamento multimodal. O classificador utilizado para mensurar o desempenho das TECs em preservar a multimodalidade da dispersão dos padrões é apresentado na Seção 4.2.2. Os resultados obtidos com esse experimento são apresentados na Seção 4.2.3. Enquanto que, na Seção 4.2.4, estão sumarizadas as conclusões obtidas com esse experimento.

4.2.1 Definição da base de dados

Para a realização desse experimento foram criadas 36 bases de dados bidimensionais (*toys*). Cada uma das bases possui conjuntos de treinamento e teste. A Figura 36 apresenta a dispersão dos padrões de cada um desses conjuntos para algumas das bases de dados construídas.

Figura 36 – Exemplos das bases de dados bidimensionais construídas.



Fonte: Autoria própria.

O conjunto de treinamento é gerado por função de distribuição que mistura duas modas Normais (Equação 3.9), sendo apresentados na Tabela 8 valores dos parâmetros que configuram essa função para a geração dos padrões de cada conjunto de treinamento.

Tabela 8 – Configurações dos parâmetros de geração dos conjuntos de treinamento.

Mode 1																	
<i>Toy</i>	μ_x	μ_y	σ_x	σ_y	α_z	<i>Toy</i>	μ_x	μ_y	σ_x	σ_y	α_z	<i>Toy</i>	μ_x	μ_y	σ_x	σ_y	α_z
1	9,35	0,00	9,60	4,30	118,31	13	3,26	0,00	4,81	9,67	63,44	25	-9,42	0,00	4,54	1,64	75,41
2	0,18	0,00	2,88	8,27	90,98	14	-9,39	0,00	6,04	2,31	54,73	26	-5,73	0,00	8,02	2,02	195,60
3	6,71	0,00	8,37	1,52	344,02	15	0,80	0,00	2,32	7,79	64,49	27	-3,64	0,00	7,05	5,04	208,73
4	-5,17	0,00	9,75	4,24	136,97	16	9,24	0,00	9,17	8,60	351,27	28	-1,30	0,00	2,03	8,62	110,83
5	-8,16	0,00	4,46	8,81	244,30	17	3,57	0,00	2,06	6,81	315,96	29	-8,84	0,00	2,91	4,79	218,90
6	-6,69	0,00	3,82	9,28	278,42	18	3,25	0,00	6,26	7,82	305,16	30	6,36	0,00	9,26	3,45	329,50
7	-0,11	0,00	4,48	3,19	288,05	19	9,97	0,00	2,33	4,86	297,65	31	5,49	0,00	2,76	6,66	83,97
8	7,34	0,00	9,88	2,31	354,42	20	-6,52	0,00	3,08	6,73	30,23	32	0,23	0,00	7,26	3,30	230,06
9	-6,58	0,00	2,83	6,23	115,68	21	-2,44	0,00	2,24	6,83	136,60	33	-1,79	0,00	8,12	4,14	163,56
10	-2,49	0,00	3,55	8,27	86,55	22	-9,67	0,00	3,18	5,89	79,44	34	3,32	0,00	5,27	7,93	66,06
11	3,75	0,00	8,32	4,85	223,45	23	1,42	0,00	3,81	2,83	17,62	35	-9,61	0,00	1,94	6,07	246,09
12	4,10	0,00	3,58	7,75	284,73	24	4,05	0,00	9,06	1,66	348,52	36	2,13	0,00	3,74	3,14	292,55

Mode 2																	
<i>Toy</i>	μ_x	μ_y	σ_x	σ_y	α_z	<i>Toy</i>	μ_x	μ_y	σ_x	σ_y	α_z	<i>Toy</i>	μ_x	μ_y	σ_x	σ_y	α_z
1	2,65	0,00	2,14	8,54	70,98	13	-3,18	0,00	3,95	2,95	293,63	25	1,79	0,00	8,08	8,68	158,53
2	-3,50	0,00	2,80	5,88	125,92	14	0,40	0,00	7,66	5,09	13,31	26	-3,61	0,00	1,68	6,98	78,97
3	7,73	0,00	7,20	5,17	98,48	15	-9,17	0,00	3,58	5,21	251,60	27	9,46	0,00	9,97	1,17	64,29
4	3,93	0,00	9,04	2,56	317,88	16	-1,48	0,00	2,31	8,37	355,56	28	7,33	0,00	7,40	3,16	41,23
5	-2,01	0,00	8,89	6,93	150,99	17	-3,78	0,00	7,37	4,42	38,85	29	5,71	0,00	1,64	6,51	264,94
6	-0,42	0,00	4,28	7,35	356,10	18	8,97	0,00	4,83	7,89	280,65	30	3,80	0,00	8,28	2,37	141,72
7	4,86	0,00	9,09	1,67	329,84	19	-4,47	0,00	5,55	8,42	121,45	31	-6,03	0,00	3,26	2,01	151,90
8	0,74	0,00	3,23	8,51	39,69	20	-0,78	0,00	2,76	9,22	56,26	32	-3,57	0,00	9,38	1,28	182,70
9	3,96	0,00	7,96	1,98	21,68	21	-8,20	0,00	7,60	5,62	183,15	33	8,40	0,00	3,42	8,14	78,16
10	0,77	0,00	6,29	5,13	82,60	22	9,50	0,00	2,91	1,95	341,40	34	0,87	0,00	3,96	4,83	289,07
11	5,93	0,00	4,15	5,23	38,43	23	-2,11	0,00	9,27	3,53	181,47	35	8,57	0,00	2,78	3,92	39,77
12	9,85	0,00	9,94	2,46	258,02	24	-5,30	0,00	4,39	8,78	233,55	36	1,68	0,00	6,15	9,92	270,84

μ_x e μ_y correspondem aos valores das coordenadas da média de cada moda respectivamente sobre os eixos X e Y . Enquanto que σ_x e σ_y correspondem aos valores de desvio padrão obtidos com a projeção dos padrões de cada moda respectivamente sobre os eixos X e Y . Utilizando os valores de configuração de cada moda Normal, 500 padrões são aleatoriamente gerados ($\rho = 0,5$). Os padrões de cada moda são rotacionados em torno do eixo Z por ângulo de α_z graus.

Para a geração dos padrões de cada moda os valores de média foram variados entre $[-10, 10]$, os desvios padrão foram variados entre $[0, 10]$ e os ângulos de rotação entre $[0, 360]$.

Os padrões dos conjuntos de teste simulam os ruídos que se encontram na faixa entre as médias das modas do respectivo conjunto de treinamento; sendo sua variação em Y igual a $[-30, 30]$.

4.2.2 Classificador

O classificador utilizado nesse experimento é configurado apenas com as informações do conjunto de treinamento e a matriz de transformação obtida pela corrente TEC.

Dado que os padrões desse conjunto não possuem rótulos de classe, o classificador desenvolvido categoriza os padrões de consulta entre padrões (como os de treinamento) ou ruídos. A realização dessa classificação, para cada base *toy*, é baseada na função de distribuição dos padrões de treinamento. Isso possibilita à classificação realizada alcançar o limite superior do desempenho de classificação sobre o respectivo conjunto de padrões. Desse modo, para cada base *toy*, $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$ são respectivamente os vetores de média das modas 1 e 2. Essas modas possuem respectivamente $\boldsymbol{\sigma}_1$ e $\boldsymbol{\sigma}_2$, que são suas matrizes de covariância. Como em todas as bases *toy* a quantidade de padrões gerados por moda são iguais, obrigatoriamente, $\rho = 0,5$.

$$\boldsymbol{\mu}_{\{1,2\}} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \quad \boldsymbol{\sigma}_{\{1,2\}} = \boldsymbol{\alpha}'_{\{1,2\}} \times \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \times \boldsymbol{\alpha}_{\{1,2\}}$$

$$\boldsymbol{\alpha}_{\{1,2\}} = \begin{bmatrix} \cos(\alpha_z) & -\sin(\alpha_z) \\ \sin(\alpha_z) & \cos(\alpha_z) \end{bmatrix}$$

As TECs aplicadas sobre os padrões do conjunto de treinamento retornam matrizes de transformação que mapeiam o espaço original em um espaço de menor dimensionalidade. Para o caso desse experimento, a matriz de transformação $\mathbf{T} \in \mathbb{R}^{2 \times 1}$ de qualquer das técnicas empregadas mapeia o espaço bidimensional original em um espaço unidimensional.

Com a obtenção desses parâmetros, a Equação 4.1 é configurada de modo a definir a função de distribuição dos padrões do conjunto de treinamento quando projetados no

espaço mapeado por \mathbf{T} .

$$g(\mathbf{x}|\mathbf{T}) = \frac{1}{2} \left[\mathcal{N}(\mathbf{T}'\mathbf{x}, \mathbf{T}'\boldsymbol{\mu}_1, \mathbf{T}'\boldsymbol{\sigma}_1\mathbf{T}) + \mathcal{N}(\mathbf{T}'\mathbf{x}, \mathbf{T}'\boldsymbol{\mu}_2, \mathbf{T}'\boldsymbol{\sigma}_2\mathbf{T}) \right] \quad (4.1)$$

Contudo, a classificação do padrão de consulta \mathbf{x} com base exclusivamente no valor $g(\mathbf{x}|\mathbf{T})$ é dificultada devido a não padronização dos valores que a função g pode alcançar (diferentes valores de $\boldsymbol{\mu}_{\{1,2\}}$, $\boldsymbol{\sigma}_{\{1,2\}}$ e \mathbf{T} conduzem a diferentes valores máximos de g). Devido a isso, a classificação é realizada utilizando a função h , definida na Equação 4.2.

$$h(\mathbf{x}|\mathbf{T}) = \frac{g(\mathbf{x}|\mathbf{T})}{g_{max}} \quad (4.2)$$

Nessa equação, g_{max} é computado a partir da diferenciação da função g em relação a \mathbf{x} , de modo a ser o maior valor que essa função pode alcançar. Desse modo, qualquer configuração da função h é padronizada entre $[0, 1]$. Assim, como mostra a Equação 4.3, o padrão de consulta \mathbf{x} (no espaço mapeado por \mathbf{T}) é categorizado como: (i) padrões, se seu valor em h for maior ou igual ao limiar $\theta \in [0, 1]$ ou (ii) ruído, caso contrário.

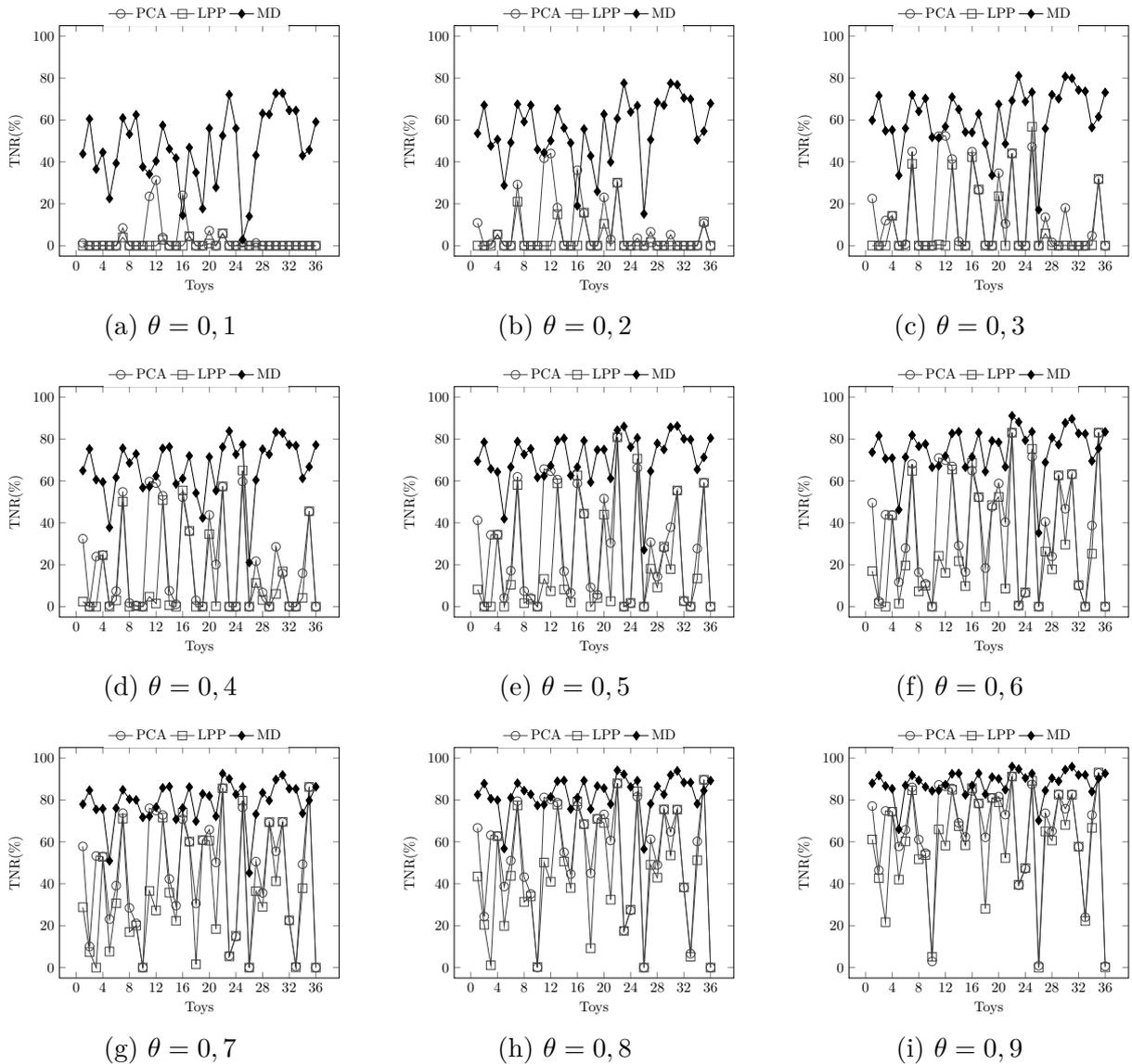
$$categoria(\mathbf{x}|\mathbf{T}) = \begin{cases} \text{padrão,} & \text{se } h(\mathbf{x}|\mathbf{T}) \geq \theta \\ \text{ruído,} & \text{se } h(\mathbf{x}|\mathbf{T}) < \theta \end{cases} \quad (4.3)$$

4.2.3 Resultados

Através do experimento realizado na Seção 4.1, verificou-se que, dentre as métricas de estimativa de multimodalidade avaliadas, IANI alcançou os maiores desempenhos médios de classificação. Devido a isso, nesse experimento, MD é configurado utilizando IANI como função de custo.

Nessa seção são apresentados os desempenhos de classificação obtidos para o PCA, LPP e MD sobre os padrões do conjunto de teste das bases *toy*. Para cada uma dessas bases são realizadas 50 repetições do procedimento de treinamento e classificação do conjunto de teste. Em cada repetição: (i) 80% dos padrões do conjunto de treinamento são aleatoriamente selecionados e utilizados para o cálculo das matrizes de transformação das TECs empregadas; (ii) os parâmetros $\boldsymbol{\mu}_{\{1,2\}}$ e $\boldsymbol{\sigma}_{\{1,2\}}$ da base *toy* corrente são utilizados para configuração do classificador; (iii) todos os padrões do conjunto de teste são categorizados com uso desse classificador no espaço de redução dimensional de cada TEC; (iv) sendo a métrica de desempenho utilizada, *True Negative Rate* (TNR), computada como a taxa com que os padrões de teste foram classificados como ruído, para um dado valor de θ .

A Figura 37 apresenta os valores médios de TNR obtidos pelas TECs empregadas, sendo as Figuras 37a-37i obtidas respectivamente com valores de $\theta = \{0, 1; 0, 2; 0, 3; 0, 4; 0, 5; 0, 6; 0, 7; 0, 8; 0, 9\}$.

Figura 37 – Valores de TNR obtidos para as técnicas de extração com diferentes valores de θ .

Fonte: Autoria própria.

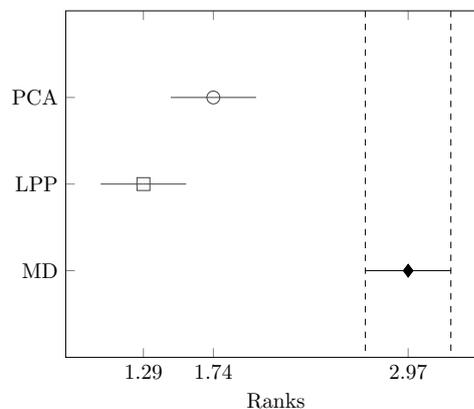
Os resultados nessas figuras mostram que, independente do valor de θ , MD obteve os mais elevados valores de TNR, para a maioria das bases *toy*, indicando seu melhor desempenho em relação ao PCA e LPP para lidar com cenários que apresentam multimodalidade intraclasses. A melhor adaptação de MD a cenários com multimodalidade intraclasses é mais evidente com o emprego dos mais baixos valores de θ (valores inferiores à 0,5). Nessa condição, é mais difícil classificar um padrão de consulta como ruído. Consequentemente, como mostra a Tabela 9, PCA e LPP obtiveram quantidades de TNR = 0,000% elevadas. Por outro lado, mesmo nos menores valores de θ , MD conseguiu realizar correta classificação dos padrões de teste quanto ruídos, de modo a não apresentar TNR = 0,000% em qualquer das configurações de θ .

Tabela 9 – Quantidades de TNR nulas (TNR = 0,000%) obtidas pelas técnicas de extração de características.

	θ								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
PCA	24	18	15	8	5	4	3	3	0
LPP	30	27	21	14	9	6	4	2	1
MD	0	0	0	0	0	0	0	0	0

A fim de elevar o rigor estatístico na comparação entre as TECs, seus valores médios de TNR obtidos com $\theta = 0,5$ foram submetidos ao teste de Friedman (DEMŠAR, 2006). Esse teste posiciona as TECs, para cada base de dados, de acordo com a ordem de seus desempenhos. Ou seja, a TEC com menor valor de TNR médio recebe menor valor de posicionamento (*rank*), enquanto que a TEC com maior valor de TNR médio recebe o maior valor de posicionamento. Realizado o posicionamento das TECs para todas as bases de dados empregadas, o teste de Friedman analisa a frequência dos posicionamentos alcançados por cada TEC, determinando seu posicionamento médio e intervalo de confiança. Assim, através desse teste, verifica-se que, com respeito as 36 bases *toy*, PCA e LPP são estatisticamente equivalentes entre si (nível de confiança = 95%), como mostra a Figura 38.

Figura 38 – Comparação entre as técnicas de extração de características através do teste de Friedman, com confiança de 95%, utilizando $\theta = 0,5$.



Fonte: Autoria própria.

Nessa figura, os maiores valores de *rank* indicam os melhores desempenhos. Assim, além de MD atingir o posicionamento mais elevado, há evidência estatística suficiente para afirmar que o desempenho de MD é superior aos do PCA e LPP.

4.2.4 Conclusões do experimento

O desempenho de classificação de MD foi comparado ao do PCA e LPP para 36 cenários com diferentes comportamentos multimodais. Sendo a métrica de desempenho a taxa com que cada técnica corretamente classifica padrões de ruído.

Os resultados dessas comparações, utilizando diferentes limiares de decisão durante a classificação, mostram que MD obteve desempenho médio superior aos do PCA e LPP para a maioria das bases, o que evidencia MD como técnica de extração de características bem sucedida para aplicação em cenários com multimodalidade. Além disso, a comparação do desempenho entre essas técnicas sobre todas as bases de dados (utilizando teste de Friedman à 95% de confiança) atesta o melhor desempenho de MD em relação ao PCA e LPP (utilizando limiar de classificação em 0,5).

4.3 Desempenhos de classificação das extrações de características

Nesse seção são comparados os desempenhos de classificação obtidos com o emprego das TECs *classwise* com investigação de multimodalidade em relação a outras TECs *classwise* ou não.

Para a realização desse experimento, a Seção 4.3.1 apresenta o protocolo metodológico empregado sobre as TECs com o objetivo de determinar: (i) a redução dimensional na qual cada TEC obtém melhor desempenho e (ii) o desempenho de classificação da TEC nessa melhor redução. A Seção 4.3.2 apresenta as bases de dados utilizadas nesse experimento. Sendo os resultados e conclusões respectivamente apresentados nas Seções 4.3.3 e 4.3.4.

4.3.1 Protocolo metodológico

Dado que as TECs não-*classwise* calculam uma única matriz de transformação, a classificação de padrões utilizando esse tipo de TEC projeta os padrões em outro espaço, através da matriz de transformação obtida. Nesse outro espaço, determina-se a pertinência da projeção a cada classe do problema.

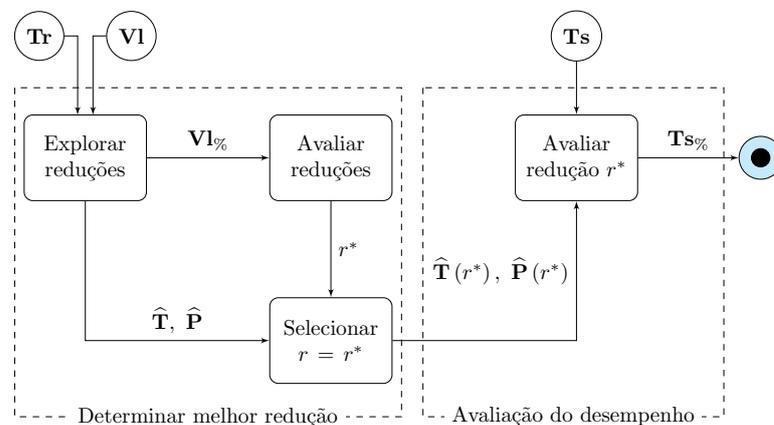
As TECs *classwise* calculam uma matriz de transformação para cada classe do problema. Assim, a classificação de padrões requer que eles sejam projetados no espaço específico de cada classe. Em cada um desses espaços, determina-se a pertinência da projeção à cada classe do problema. Sendo C o número de classes do problema, as TECs *classwise* obtêm C classificações, cada uma obtida em um espaço específico das classes. A classificação final é escolhida dentre uma das C classificações, sendo a escolha feita pela classificação com maior valor de pertinência.

Devido as diferenças entre as abordagens não-*classwise* e *classwise*, a metodologia desenvolvida nessa seção é proposta com a finalidade de possibilitar comparação de desempenho adequada entre as TECs das diferentes categorias (não-*classwise*, *classwise* e *classwise* com investigação de multimodalidade).

A Figura 39 apresenta a visão geral do protocolo metodológico utilizado para deter-

minação do desempenho de cada TEC. Esse protocolo realiza duas tarefas: (i) determinar a redução de dimensionalidade para a qual a TEC é melhor adaptada e (ii) avaliar o desempenho de classificação dos padrões de teste utilizando a TEC configurada para a melhor redução de dimensionalidade encontrada.

Figura 39 – Protocolo metodológico para determinação dos desempenhos das TECs. Sobre o conjunto de treinamento (\mathbf{Tr}) a TEC é realizada para cada redução de dimensionalidade possível de ser alcançada. As matrizes de transformações obtidas para essas reduções são armazenadas em $\hat{\mathbf{T}}$. $\hat{\mathbf{P}}$ é a coleção dos parâmetros de configuração do classificador em cada redução. $\mathbf{VI}_{\%}$ contém o desempenho de classificação sobre os padrões do conjunto de validação (\mathbf{VI}) para cada redução empregada. Através de testes estatísticos sobre $\mathbf{VI}_{\%}$ é possível determinar a redução r^* com melhor desempenho na classificação sobre os padrões de \mathbf{VI} . O desempenho da TEC é calculado na redução r^* sobre os padrões do conjunto de teste (\mathbf{Ts}), sendo esse valor expresso em $\mathbf{Ts}_{\%}$.



Fonte: Autoria própria.

Utilizando metodologia de validação cruzada e estratificada com cinco partições, são determinados conjuntos de treinamento (\mathbf{Tr}), validação (\mathbf{VI}) e teste (\mathbf{Ts}).

Em cada ciclo da validação cruzada, o correspondente conjunto de treinamento é utilizado para determinação da matriz de transformação e parâmetros de configuração do classificador, os quais são obtidos para cada redução de dimensionalidade r que é alcançável pela TEC. Os conjuntos $\hat{\mathbf{T}}$, $\hat{\mathbf{P}}$ colecionam respectivamente essas matrizes de transformações e parâmetros do classificador que foram determinados em cada redução r .

Em cada redução r , o par de elementos $\{\hat{\mathbf{T}}(r), \hat{\mathbf{P}}(r)\}$ é utilizado para classificação dos padrões do correspondente conjunto de validação. O desempenho da TEC sobre a classificação dos padrões de \mathbf{VI} na redução r é armazenado na matriz $\mathbf{VI}_{\%}$. As colunas dessa matriz contém os desempenhos obtidos para uma mesma redução de dimensionalidade nos diferentes ciclos da validação cruzada.

O completo preenchimento das células de $\mathbf{VI}_{\%}$ finaliza a exploração das possíveis reduções de dimensionalidade para a TEC corrente. Após isso, é possível avaliar os desempenhos das diferentes reduções a fim de determinar a redução r^* que obteve melhor

desempenho de classificação sobre os padrões de validação.

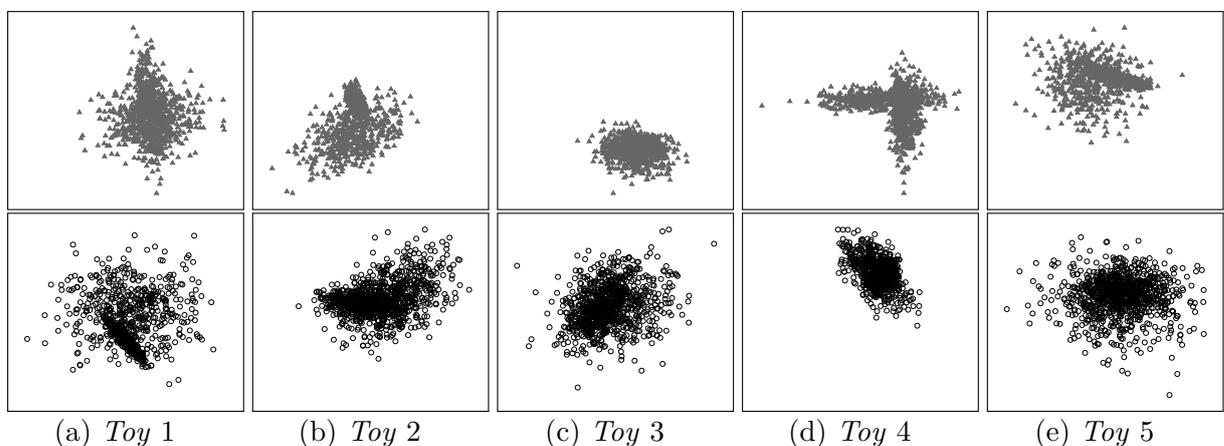
A determinação de r^* é feita através de teste estatístico (teste- t pareado) que compara as colunas de $\mathbf{VI}_\%$. Inicialmente, encontra-se a redução de dimensionalidade r^{ref} com maior valor médio de desempenho. Desse modo, cada uma das demais reduções são comparadas com r^{ref} tendo como hipótese nula: $r^{ref} > r$. Caso essa hipótese não seja rejeitada para nenhuma das reduções, implica que há evidência estatística para afirmar que, na redução r^{ref} , a TEC obteve melhor desempenho de classificação dos padrões de validação. Nesse caso, $r^* = r^{ref}$. Contudo, havendo reduções que rejeitem a hipótese nula, implica que essas reduções alcançaram desempenhos médios equivalentes ao desempenho médio em r^{ref} ; não sendo possível afirmar que r^{ref} é a redução com melhor desempenho. Nessa situação, r^* é escolhida como a redução equivalente à r^{ref} (isso inclui r^{ref}) que possui menor desvio padrão nos valores de desempenho obtidos com a validação cruzada.

Uma vez determinada a redução de dimensionalidade r^* , para a qual a TEC é melhor adaptada, o par de elementos $\{\hat{\mathbf{T}}(r^*), \hat{\mathbf{P}}(r^*)\}$ é utilizado para classificação dos padrões do correspondente conjunto de teste. O desempenho da TEC sobre a classificação dos padrões de $\mathbf{T}s$ na redução r^* é armazenado no vetor $\mathbf{T}s_\%$. As células desse vetor contém os desempenhos obtidos para r^* nos diferentes ciclos da validação cruzada.

4.3.2 Definição da base de dados

Para esse experimento foram aleatoriamente geradas 50 bases de dados tridimensionais. Cada base possui duas classes, cada uma com 1000 padrões. Os padrões de cada classe foram gerados por função que mistura duas modas Normais como na Equação 3.9 (página 52), fazendo $\rho = 0,5$. A Figura 40 apresenta representações bidimensionais das 5 primeiras bases *toy*.

Figura 40 – Representação bidimensional da dispersão dos padrões por classes das bases *toy*.



Fonte: Autoria própria.

Essas representações foram geradas utilizando algoritmo *Multi Dimensional Scale* (MDS)

(SEBER, 2004), o qual possibilita representar um conjunto de padrões em menor dimensionalidade preservando ao máximo as relações métricas entre eles. Assim, como pode ser visto na Figura 40, a aleatoriedade na geração dessas bases conduz a classes cujas dispersões de seus padrões atingem diferentes evidências de comportamento multimodal.

As Tabelas 10 e 11 apresentam respectivamente os valores dos parâmetros utilizados para determinação das modas das classe 1 e 2 de cada base *toy*.

Tabela 10 – Parâmetros de configuração das modas Normais que compõem a função de distribuição dos padrões da classe 1.

Toy	Classe 1																	
	Moda 1									Moda 2								
	$\tilde{\mu}_x$	$\tilde{\mu}_y$	$\tilde{\mu}_z$	\tilde{s}_x	\tilde{s}_y	\tilde{s}_z	α_x	α_y	α_z	$\tilde{\mu}_x$	$\tilde{\mu}_y$	$\tilde{\mu}_z$	\tilde{s}_x	\tilde{s}_y	\tilde{s}_z	α_x	α_y	α_z
1	-1,21	1,52	3,66	7,81	9,24	0,13	199,91	285,55	147,53	-5,94	-3,43	5,75	7,92	6,14	0,20	192,53	17,21	179,78
2	6,06	8,09	9,56	6,20	3,35	8,45	111,18	57,19	334,42	7,18	3,86	-2,64	6,15	1,19	1,15	262,96	290,47	253,85
3	1,90	9,98	3,17	2,58	1,04	5,28	8,04	160,13	167,03	-9,11	-5,08	-4,86	2,82	4,56	4,28	151,18	128,37	240,14
4	1,46	-4,23	6,62	3,26	7,69	2,41	60,14	271,99	192,71	-3,42	-9,38	-3,23	0,80	7,01	9,85	358,65	65,78	274,95
5	0,19	-2,53	2,71	3,48	1,03	0,89	336,47	330,90	168,57	1,99	-1,07	4,72	1,99	8,98	5,32	67,18	82,80	211,04
6	5,26	2,69	-2,22	0,23	8,05	7,38	61,61	232,08	299,53	2,94	9,59	2,96	2,24	3,73	6,06	21,57	99,59	204,06
7	0,28	0,70	-2,57	6,16	9,20	0,45	27,70	5,20	116,40	4,17	-2,44	0,76	9,96	6,12	5,89	301,22	80,73	128,82
8	-0,96	-8,64	0,81	7,94	3,34	3,37	122,29	159,03	277,65	-7,86	-7,77	2,93	9,52	8,68	6,11	93,31	158,33	344,50
9	-5,40	9,15	4,79	4,75	9,09	0,19	342,84	35,23	85,77	-5,44	0,69	-5,46	8,20	4,89	5,03	344,81	186,31	325,04
10	5,39	5,88	-3,65	7,04	8,54	6,96	344,56	3,65	153,10	-3,96	-3,28	-5,43	7,17	1,72	1,69	270,86	3,09	187,36
11	0,66	8,12	3,23	0,41	7,36	6,30	191,63	15,56	233,74	-8,07	4,12	-3,25	9,95	8,05	9,69	246,29	92,10	271,03
12	-8,24	1,34	2,28	5,11	2,89	0,29	83,49	29,03	125,41	0,40	4,09	-2,29	5,77	6,62	9,07	81,59	37,42	305,71
13	6,37	-0,56	9,88	1,74	8,24	8,22	246,64	342,02	137,56	-1,53	9,12	1,27	1,13	4,83	2,08	6,77	85,64	36,85
14	6,83	9,71	8,94	2,18	2,55	3,26	84,56	208,59	84,13	7,71	-0,01	-8,68	7,75	6,53	1,46	289,19	62,19	350,02
15	5,21	1,63	-2,18	1,77	2,59	2,55	46,73	249,23	237,11	-6,47	7,93	-0,60	6,08	2,39	5,86	253,10	176,11	295,08
16	7,44	-3,34	4,80	9,55	2,77	3,96	273,14	325,69	242,70	-8,22	-1,83	-6,87	7,29	0,28	6,99	340,17	15,88	218,83
17	7,17	2,25	-6,66	9,06	3,51	0,52	11,35	216,63	134,12	-3,66	-9,22	-1,39	8,10	7,28	8,51	95,49	98,33	204,45
18	7,41	-0,49	-4,79	0,63	8,49	7,40	47,57	211,23	45,73	-3,30	6,43	-0,22	1,93	7,32	0,31	32,19	169,44	111,60
19	-5,19	4,06	-4,01	9,60	7,30	9,15	271,92	275,29	310,94	8,36	7,99	2,67	6,63	7,23	9,69	203,71	212,77	33,12
20	6,67	-1,02	-5,67	5,54	5,70	9,52	263,05	69,26	330,37	7,32	-4,02	-5,56	6,67	4,55	5,02	322,20	26,66	17,59
21	-5,96	-9,91	-4,33	4,45	2,77	6,71	54,70	91,41	204,07	-6,34	-3,55	-6,64	4,80	7,18	9,29	91,17	172,09	295,59
22	-2,28	-7,54	-9,74	5,01	8,80	4,14	92,09	181,85	9,98	0,65	5,43	-3,71	2,91	9,67	8,64	255,40	237,35	329,25
23	-8,94	7,04	1,16	4,28	3,50	7,68	164,77	343,35	44,42	-5,60	7,02	3,44	8,96	2,30	6,22	59,90	203,32	235,28
24	0,38	-2,33	-7,48	7,40	5,62	8,44	244,75	63,79	75,40	0,47	0,46	-5,18	3,33	1,24	7,63	287,65	302,99	186,31
25	-6,59	0,98	5,98	1,95	4,17	1,19	145,79	328,25	117,13	7,41	-8,78	-3,18	2,92	5,80	4,97	232,28	342,57	136,26
26	1,61	2,41	-2,73	6,02	4,05	4,89	289,60	72,13	163,51	2,11	0,37	-8,54	6,43	2,46	7,65	253,47	152,24	158,60
27	3,88	-4,10	0,30	6,79	6,55	9,04	342,18	16,18	348,49	4,89	1,98	-0,91	2,98	4,91	0,14	290,57	62,73	291,77
28	-7,51	6,61	-7,83	2,14	4,59	6,11	53,69	175,75	327,25	-9,91	-2,52	-9,74	1,34	6,74	0,91	145,11	47,73	339,90
29	-8,44	9,60	0,47	1,71	1,03	0,02	13,20	348,85	113,73	-1,71	5,81	9,97	9,62	0,88	7,70	5,70	97,93	125,78
30	1,43	5,35	4,23	1,37	3,27	6,95	344,91	61,90	257,01	-3,12	-8,69	-3,25	6,42	9,06	3,18	58,69	340,60	324,26
31	-5,41	6,90	5,75	4,26	1,39	3,59	252,70	61,71	308,24	8,47	8,94	-9,05	1,61	8,29	8,18	299,01	294,66	64,04
32	-5,95	8,06	6,57	0,19	3,03	1,37	190,56	279,23	239,61	2,28	6,48	-3,86	4,53	5,81	3,46	118,80	80,25	303,01
33	-6,56	1,37	4,86	5,07	3,59	4,75	355,66	90,98	129,14	-0,59	4,80	4,47	6,43	0,65	8,46	85,77	357,34	55,95
34	8,14	-2,08	-8,83	1,98	2,07	8,59	254,60	81,62	252,56	-9,62	9,36	-9,83	9,69	5,95	5,15	231,55	98,14	16,51
35	-0,56	-8,10	-6,17	5,95	3,62	9,64	0,34	205,88	96,87	0,72	1,10	-2,64	0,74	2,74	7,73	347,75	331,43	310,47
36	-3,30	1,69	7,99	5,80	0,35	7,52	345,70	81,81	28,30	-9,47	-6,31	-1,42	5,45	5,25	8,53	98,56	141,53	211,93
37	5,35	6,11	-6,51	0,14	4,50	7,12	125,84	247,22	22,37	4,16	8,30	-0,34	8,95	7,40	9,83	152,29	175,91	251,07
38	-2,07	9,55	8,61	2,10	6,96	3,25	181,88	139,65	267,33	-0,05	-2,06	1,22	9,56	0,71	5,68	270,24	298,18	208,20
39	-5,19	-4,88	-6,99	2,24	8,06	0,79	18,00	255,23	168,47	7,15	5,46	-1,86	2,98	9,32	7,70	169,64	88,76	299,50
40	-0,32	4,74	-2,44	8,69	9,71	5,06	146,30	89,15	168,71	4,77	4,42	9,30	2,86	1,66	8,25	92,01	65,14	228,97
41	-2,05	8,24	-8,79	7,53	3,08	5,16	331,44	154,74	103,31	8,63	2,40	1,54	1,42	6,71	9,26	121,06	155,08	120,46
42	5,63	-7,03	-0,16	3,15	8,17	8,22	4,84	312,65	336,95	6,17	-6,42	-4,07	4,24	9,81	5,04	282,87	314,46	189,76
43	5,17	3,83	-6,82	9,91	7,07	9,59	355,11	258,78	73,02	7,92	9,26	-2,23	0,61	6,59	7,12	336,26	348,48	112,31
44	8,44	9,11	6,87	3,91	8,40	9,83	70,01	238,93	312,12	3,20	-7,31	-0,95	4,12	6,92	9,29	40,28	83,94	326,04
45	1,24	-1,91	-5,87	5,62	1,73	0,52	83,00	119,13	51,25	-8,69	8,51	-2,66	4,79	8,76	4,90	291,15	184,57	160,37
46	-6,97	7,39	-9,74	9,19	4,75	3,26	349,53	343,00	133,50	-5,78	1,18	8,69	3,71	2,22	3,73	138,47	9,23	122,21
47	-1,65	6,18	-9,11	0,84	3,64	7,94	344,88	309,75	332,61	-9,38	7,05	2,54	9,24	8,46	8,54	93,04	58,07	103,99
48	1,61	3,02	-2,80	7,07	4,49	9,87	126,01	176,55	202,82	4,38	-7,86	4,17	8,12	3,89	4,49	240,23	229,95	289,79
49	-6,79	3,04	-7,69	8,56	9,54	7,48	173,72	320,61	164,23	5,43	-2,90	-8,92	5,11	0,42	7,04	178,70	311,54	253,30
50	6,15	-7,21	-4,28	5,76	2,08	6,03	334,10	322,87	345,72	7,87	-0,51	1,19	2,97	7,15	9,12	15,50	263,01	311,14

Nessas tabelas, para cada moda das classes, os valores $\tilde{\mu}_i$ e \tilde{s}_i correspondem respectivamente o valor da média e desvio padrão dos padrões da moda quando projetados sobre o eixo i

Tabela 11 – Parâmetros de configuração das modas Normais que compõem a função de distribuição dos padrões da classe 2.

Toy	Classe 2																	
	Moda 1									Moda 2								
	$\tilde{\mu}_x$	$\tilde{\mu}_y$	$\tilde{\mu}_z$	\tilde{s}_x	\tilde{s}_y	\tilde{s}_z	α_x	α_y	α_z	$\tilde{\mu}_x$	$\tilde{\mu}_y$	$\tilde{\mu}_z$	\tilde{s}_x	\tilde{s}_y	\tilde{s}_z	α_x	α_y	α_z
1	-7,28	5,29	-3,24	0,58	4,41	1,39	177,64	327,14	319,65	-2,83	-3,91	3,13	8,88	8,50	9,66	37,01	102,79	332,52
2	-3,48	-0,92	-2,04	0,95	6,37	5,65	210,03	183,48	263,40	-8,12	2,30	-9,27	5,93	3,83	9,60	282,83	133,31	357,24
3	-1,87	-8,66	6,12	5,89	4,72	9,03	145,70	47,91	40,45	5,66	8,55	1,19	7,05	3,00	2,88	114,00	151,76	244,19
4	5,84	5,15	5,65	5,75	2,39	7,32	5,39	172,73	26,10	-3,44	-2,97	-5,64	1,27	2,55	1,09	112,54	315,56	116,73
5	-3,19	-0,20	-7,68	8,07	8,97	7,49	224,18	349,75	153,43	-7,03	4,11	-1,80	5,72	3,10	4,16	82,23	168,69	315,87
6	-1,18	4,94	2,30	9,86	3,51	8,30	350,21	287,76	269,04	5,65	4,49	3,07	2,78	9,48	1,45	268,74	232,56	329,82
7	5,68	3,52	-3,25	4,64	5,73	7,78	1,50	156,65	314,02	-3,15	8,53	9,00	2,10	8,73	9,21	95,22	163,98	236,25
8	6,68	7,08	2,13	5,32	8,17	7,17	182,97	334,29	240,72	9,13	9,25	-6,63	3,85	0,68	0,61	124,24	26,60	149,20
9	1,95	1,76	-7,81	0,84	3,34	5,05	232,12	28,49	277,74	-6,41	-1,30	6,13	5,90	2,90	0,62	189,56	164,34	17,83
10	-1,07	6,57	0,33	1,75	5,23	5,04	94,03	16,57	308,58	2,73	-2,18	-3,86	7,37	8,05	6,19	77,09	126,74	331,48
11	8,76	9,95	-3,92	1,66	9,03	7,91	163,10	260,02	252,33	-7,09	-3,20	4,90	2,92	1,67	2,42	160,93	161,55	132,24
12	3,73	-2,79	2,20	6,34	8,66	0,98	116,52	333,63	34,43	-2,32	2,38	-2,14	3,55	8,78	9,25	224,85	111,26	62,55
13	4,72	-1,80	-2,98	2,45	4,94	0,33	49,59	327,69	103,45	-0,03	-2,89	-2,64	4,63	7,76	7,22	242,03	265,18	325,19
14	-2,69	0,27	-1,37	1,14	1,84	5,73	94,67	28,23	243,20	-8,60	-7,90	2,86	6,73	9,16	1,89	315,25	239,62	107,63
15	-4,29	-4,68	-5,13	9,84	9,97	2,68	88,36	185,86	109,16	-4,78	5,90	4,23	1,68	3,73	8,82	317,13	50,23	10,43
16	8,95	7,00	-5,96	5,42	0,63	0,36	80,86	128,38	198,44	1,58	-7,51	-1,91	6,92	4,13	5,36	264,03	353,77	98,50
17	4,41	3,53	3,82	7,57	6,45	4,98	47,15	172,14	302,70	1,46	3,54	3,28	1,22	1,79	6,70	178,85	292,41	325,54
18	0,26	1,23	5,25	9,45	0,19	1,18	188,77	7,45	249,62	6,25	5,40	0,37	5,14	0,81	5,15	85,64	84,77	203,44
19	-6,30	-5,36	1,73	1,72	9,09	5,31	358,45	78,54	19,84	-6,51	5,24	2,47	8,61	6,20	1,59	214,94	9,07	268,93
20	-2,79	3,82	5,69	8,11	9,00	7,14	115,15	171,46	335,53	5,41	1,81	1,36	6,33	7,69	5,60	48,05	63,61	147,77
21	3,35	-5,23	-9,15	5,73	8,89	6,08	216,24	347,50	281,71	8,93	-8,04	-1,12	5,77	9,57	1,10	24,21	155,54	60,21
22	9,49	-5,91	2,67	7,06	5,22	8,53	314,50	22,94	101,76	2,45	-3,92	-0,84	8,36	4,66	7,10	113,83	98,03	270,69
23	-4,92	4,72	9,15	5,42	4,81	0,88	333,36	148,13	183,26	-7,79	1,75	0,99	6,55	1,71	7,37	87,65	325,83	59,16
24	-1,57	-5,85	-4,49	2,85	0,07	7,68	128,08	82,14	47,97	-4,31	-0,46	-0,07	0,67	5,14	9,87	276,06	201,26	37,79
25	8,82	8,47	2,40	0,55	8,00	5,28	57,31	198,66	3,36	7,58	1,08	2,83	1,10	4,09	2,11	166,98	201,91	262,77
26	8,42	-2,44	0,52	5,10	8,37	5,32	22,72	300,04	25,81	-6,79	-8,34	-5,09	6,34	3,05	9,65	238,71	87,61	211,90
27	9,91	1,63	-3,21	9,12	7,23	0,84	107,57	133,32	324,46	-9,10	2,57	-2,99	1,54	6,76	5,88	304,17	52,84	349,55
28	5,88	8,59	5,80	6,22	7,03	8,44	309,47	304,98	155,14	1,73	-2,64	-5,90	3,51	1,97	5,50	340,54	156,93	149,02
29	-1,56	4,52	5,99	3,33	6,15	5,31	301,89	301,58	90,04	-5,18	-5,81	7,77	2,71	7,66	6,03	129,90	239,31	119,60
30	0,43	7,81	-2,75	3,17	8,55	1,33	311,56	135,03	306,62	1,64	-3,75	6,92	0,87	7,77	6,56	189,41	351,87	265,67
31	3,52	5,99	-4,94	1,65	4,81	7,19	307,43	253,54	136,56	4,06	0,99	5,77	6,82	3,00	4,57	134,62	118,35	220,97
32	4,57	1,71	0,07	2,45	8,89	9,77	299,23	217,73	40,90	5,92	9,31	1,93	3,36	3,43	6,15	75,85	196,24	243,80
33	-1,89	3,74	-4,22	6,91	4,80	8,48	147,53	113,43	72,75	0,89	-1,18	9,63	4,19	2,56	9,48	255,45	256,41	187,51
34	-6,06	-3,99	-9,55	3,78	3,24	6,48	83,21	251,65	178,82	8,95	2,19	5,44	4,82	8,26	0,40	240,33	63,63	6,28
35	-3,61	-7,52	0,27	0,10	2,54	6,34	7,11	21,02	122,99	-3,50	-9,63	-2,31	6,81	8,06	2,84	90,89	151,66	46,81
36	-1,20	-1,90	9,66	0,68	6,90	3,24	209,87	293,53	57,97	8,64	2,26	9,15	2,92	1,34	8,20	309,62	346,45	60,00
37	-2,05	-4,51	-4,82	7,73	5,76	3,88	336,86	54,41	302,56	-7,14	7,37	4,44	0,58	5,42	6,79	30,06	215,52	122,65
38	2,68	5,22	6,22	0,79	8,52	2,09	112,36	40,70	45,39	0,31	6,25	6,68	9,76	4,84	4,75	329,71	318,38	31,23
39	-4,70	-8,26	7,86	3,31	8,41	2,28	350,79	13,08	106,61	2,31	-9,01	0,92	6,99	2,30	3,94	50,19	291,76	249,22
40	1,07	0,01	4,46	8,94	2,79	6,92	218,11	212,26	57,88	0,79	7,99	-3,26	7,89	7,95	8,20	72,15	115,01	311,74
41	-1,59	4,25	9,12	3,13	9,90	5,32	33,28	67,78	276,95	-8,70	1,76	-0,24	8,88	8,63	4,10	328,28	205,62	15,81
42	9,29	6,54	9,51	6,99	0,26	8,39	156,15	219,38	103,66	-0,11	-9,18	8,01	5,18	0,58	5,01	330,80	270,22	38,70
43	9,44	-4,49	-3,55	8,25	2,33	8,75	46,44	345,32	293,85	-9,38	-3,00	-6,46	6,57	8,32	5,04	64,86	298,64	350,58
44	3,14	-4,28	-6,04	3,25	9,00	3,00	200,64	34,34	161,56	-4,96	3,23	0,92	2,94	9,67	5,43	1,24	103,57	44,59
45	5,59	-6,90	1,58	7,37	0,58	2,02	81,32	221,12	74,79	3,64	6,90	1,31	9,00	0,77	8,54	193,53	219,43	246,50
46	4,42	-7,62	-0,28	0,54	8,09	3,43	340,63	28,60	60,81	-6,82	-1,85	3,69	8,06	5,28	4,94	245,66	230,80	324,78
47	-0,75	-7,42	-1,76	5,18	2,58	3,25	357,72	324,80	70,28	5,06	1,94	0,99	0,14	8,99	8,00	149,38	208,87	201,45
48	2,38	-1,84	3,57	2,15	5,87	9,60	105,11	218,32	173,47	-8,44	3,94	-2,51	2,27	6,50	2,17	14,30	85,73	130,76
49	8,02	5,41	7,16	7,67	0,03	4,31	62,96	100,42	124,89	-7,11	-1,50	-5,32	8,81	9,50	3,09	101,18	347,76	121,13
50	-5,34	-9,55	0,83	2,24	4,71	8,56	129,89	4,28	106,84	-7,74	10,00	5,32	9,74	8,77	5,56	158,97	125,41	122,10

($i = \{X, Y, Z\}$); enquanto que α_i é o valor do ângulo com que os padrões dessa moda são rotacionados em torno do eixo i .

4.3.3 Resultados

Nessa seção são apresentados os resultados experimentais obtidos com a aplicação de TECs segundo o protocolo metodológico definido na Seção 4.3.1 e sobre as bases de dados apresentadas na Seção 4.3.2. A apresentação dos resultados obtidos está particionada nas Seções 4.3.3.1 e 4.3.3.2.

Devido as TECs *classwise* admitirem configurações opcionais que competem com suas proposições mais simples (Seção 2.2.2, página 32), na Seção 4.3.3.1 são comparados os desempenhos entre as variações das TECs *classwise*, a fim de verificar quais dessas variações melhor representam a TEC para o cenário de multimodalidade intraclasse representado nas bases de dados. Em seguida, na Seção 4.3.3.2, as variações das TECs *classwise* melhor adaptadas aos cenários de multimodalidade intraclasse são comparadas entre si e com as demais TECs de referência que são utilizadas nesse trabalho.

4.3.3.1 Variações das técnicas *classwise*

Nessa seção são comparados os desempenhos das diferentes variações de cada TEC *classwise*, o que inclui as TECs *classwise* com investigação de multimodalidade. A Tabela 12 apresenta, para cada TEC *classwise*, suas variações segundo o uso ou não das configurações opcionais (dispersões entreclasses ou negativa).

Tabela 12 – Rótulos das variações das técnicas *classwise* segundo suas configurações obrigatórias ou opcionais.

TEC	Variações		
	Configuração obrigatória	Dispersão entreclasses	Dispersão negativa
CPCA (Seção 2.2.2.1)	CPCA	CPCA r_b	
CdLPP (Seção 2.2.2.2)	CdLPP	CdLPP r_b	CdLPP r_n
CdMD (Seção 3.4.1)	CdMD	CdMD r_b	CdMD r_n
CdCMD (Seção 3.4.2)	CdCMD	CdCMD r_b	CdCMD r_n
CdCDM (Seção 3.4.3)	CdCDM	CdCDM r_b	CdCDM r_n

O objetivo dessas comparações é verificar quais das variações são melhor adaptadas ao cenário de interesse a esse trabalho, ou seja, cenários com multimodalidade intraclasse.

Comparação entre as variações do CPCA

A Tabela 13 apresenta os resultados obtidos com a classificação dos padrões das 50 bases de dados *toy* utilizando CPCA ou CPCA r_b . Vê-se que, para a maioria das bases, foram iguais os resultados obtidos pelas variações do CPCA. Além disso, a maior diferença absoluta entre essas técnicas não ultrapassou 1, 50% de acerto. A equivalência entre essas variações do CPCA é reforçada pela proximidade dos valores médios de seus acertos e *ranks*.

Para cada base de dados, o valor do *rank* (DEMŠAR, 2006) indica a posição de uma técnica em relação as demais, referente aos desempenhos que elas atingem. Assim, as técnicas com: menores valores de *rank* indicam menores desempenhos, maiores valores de

Tabela 13 – Comparação entre os desempenhos do CPCA e CPCA r_b .

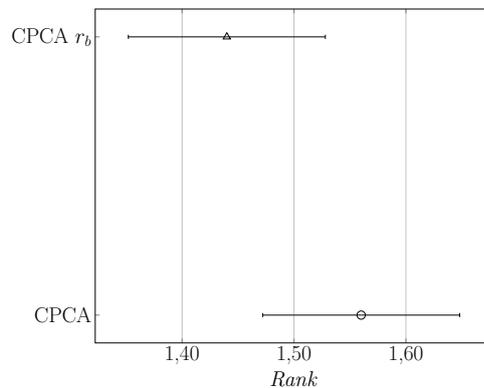
CPCA						CPCA r_b					
<i>Toy</i>	Acertos (%)	<i>Rank</i>									
01	79,35	1,50	26	79,05	2,00	01	79,35	1,50	26	78,30	1,00
02	89,25	1,50	27	88,30	2,00	02	89,25	1,50	27	87,90	1,00
03	97,25	1,50	28	84,85	1,00	03	97,25	1,50	28	86,15	2,00
04	92,50	1,50	29	98,20	2,00	04	92,50	1,50	29	97,70	1,00
05	92,20	1,50	30	83,50	1,00	05	92,20	1,50	30	83,95	2,00
06	74,65	1,50	31	97,25	1,50	06	74,65	1,50	31	97,25	1,50
07	81,45	1,50	32	92,45	1,50	07	81,45	1,50	32	92,45	1,50
08	79,45	1,00	33	87,80	2,00	08	80,45	2,00	33	87,55	1,00
09	80,70	2,00	34	97,05	1,50	09	80,50	1,00	34	97,05	1,50
10	74,80	2,00	35	70,80	1,00	10	74,05	1,00	35	72,15	2,00
11	89,25	2,00	36	96,00	1,50	11	87,95	1,00	36	96,00	1,50
12	72,00	1,50	37	90,00	1,50	12	72,00	1,50	37	90,00	1,50
13	89,50	1,50	38	76,20	1,50	13	89,50	1,50	38	76,20	1,50
14	96,55	1,50	39	99,00	1,50	14	96,55	1,50	39	99,00	1,50
15	72,95	1,50	40	78,95	1,50	15	72,95	1,50	40	78,95	1,50
16	87,30	1,50	41	78,20	1,50	16	87,30	1,50	41	78,20	1,50
17	86,35	2,00	42	85,05	1,50	17	85,55	1,00	42	85,05	1,50
18	88,50	1,50	43	80,75	2,00	18	88,50	1,50	43	79,30	1,00
19	72,70	1,50	44	76,05	1,50	19	72,70	1,50	44	76,05	1,50
20	78,10	1,00	45	86,10	2,00	20	79,40	2,00	45	84,70	1,00
21	80,80	1,50	46	91,30	1,50	21	80,80	1,50	46	91,30	1,50
22	86,15	2,00	47	87,25	1,50	22	85,40	1,00	47	87,25	1,50
23	77,20	1,50	48	76,90	2,00	23	77,20	1,50	48	76,35	1,00
24	68,50	1,50	49	81,25	1,00	24	68,50	1,50	49	82,50	2,00
25	95,75	2,00	50	84,50	1,00	25	95,60	1,00	50	84,95	2,00

Médias		
	CPCA	CPCA r_b
Acertos	84,60	84,56
<i>Ranks</i>	1,56	1,44
+ / ~ / -	12/30/8	8/30/12

rank indicam maiores desempenhos e iguais valores de *rank* indicam iguais desempenhos. Desse modo, os valores médios de *rank* auxiliam na ordenação das técnicas quanto a frequência com que atingem maiores desempenhos que as demais, considerando todas as bases de dados.

A Figura 41 mostra, segundo o teste de Friedman (DEMŠAR, 2006), os intervalos de confiança para os *ranks* de cada técnica a $\tilde{5}\%$ de significância. Assim, devido a existência de interseção entre esses intervalos, os desempenhos do CPCA e CPCA r_b são considerados estatisticamente equivalentes.

Figura 41 – Comparação do desempenho das variações do CPCA com o teste de Friedman.



Fonte: Autoria própria.

Todavia, o CPCA possui maior *rank* médio do que o CPCA r_b , o que indica que essa variação atinge melhores valores de desempenho com frequência maior do que a técnica rival. Isso é verificado através da simples contagem da quantidade de vezes com que uma das variações obteve desempenho superior (+), similar (~) ou inferior (-) as demais.

Portanto, para as comparações de desempenho realizadas na Seção 4.3.3.2, a variação obrigatória do CPCA será utilizada para representar essa TEC.

Comparação entre as variações do CdLPP

As variações do CdLPP são: (i) sua configuração obrigatória (CdLPP), (ii) sua configuração opcional utilizando a dispersão entreclasses (CdLPP r_b) e (iii) sua configuração opcional utilizando a dispersão negativa (CdLPP r_n).

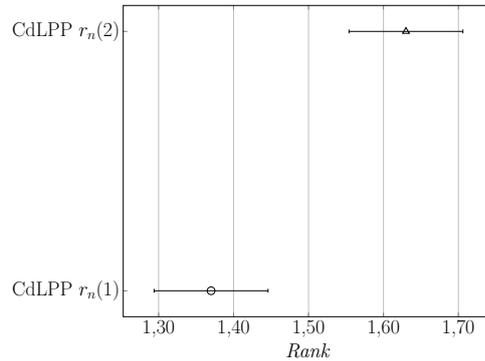
Enquanto que CdLPP r_b admite apenas uma configuração, a qual é a participação de apenas um vetor da matriz de dispersão entreclasses na sua matriz de transformação ($r_b = 1$); o CdLPP r_n admite duas configurações ($r_n = \{1, 2\}$). Desse modo, antes de comparar as variações do CdLPP, é preciso definir a configuração ($r_n = 1$ ou $r_n = 2$) que representará a variação CdLPP r_n . Para isso, na Tabela 14 são apresentados os desempenhos de classificação alcançados com CdLPP $r_n(1)$ e CdLPP $r_n(2)$ que respectivamente utilizam $r_n = 1$ e $r_n = 2$.

Tabela 14 – Comparação entre os desempenhos das configurações do CdLPP r_n .

CdLPP $r_n(1)$						CdLPP $r_n(2)$					
<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>
01	79,35	1,50	26	79,05	1,50	01	79,35	1,50	26	79,05	1,50
02	89,25	1,50	27	88,25	1,00	02	89,25	1,50	27	88,30	2,00
03	97,25	1,50	28	86,15	1,50	03	97,25	1,50	28	86,15	1,50
04	92,50	1,50	29	98,20	1,50	04	92,50	1,50	29	98,20	1,50
05	91,15	1,00	30	83,50	1,50	05	92,20	2,00	30	83,50	1,50
06	74,65	1,50	31	97,20	1,00	06	74,65	1,50	31	97,25	2,00
07	81,45	1,50	32	92,45	1,50	07	81,45	1,50	32	92,45	1,50
08	80,45	1,50	33	87,45	1,00	08	80,45	1,50	33	87,80	2,00
09	81,10	1,00	34	97,05	1,50	09	82,10	2,00	34	97,05	1,50
10	74,30	1,00	35	72,15	1,50	10	74,80	2,00	35	72,15	1,50
11	89,25	1,50	36	96,00	1,50	11	89,25	1,50	36	96,00	1,50
12	72,00	1,50	37	90,00	1,50	12	72,00	1,50	37	90,00	1,50
13	89,50	1,50	38	76,20	1,50	13	89,50	1,50	38	76,20	1,50
14	96,55	1,50	39	99,00	1,50	14	96,55	1,50	39	99,00	1,50
15	72,95	1,50	40	78,95	1,50	15	72,95	1,50	40	78,95	1,50
16	87,30	1,50	41	78,20	1,50	16	87,30	1,50	41	78,20	1,50
17	85,25	1,00	42	85,05	1,50	17	86,35	2,00	42	85,05	1,50
18	88,50	1,50	43	79,15	1,00	18	88,50	1,50	43	80,75	2,00
19	70,10	1,00	44	76,05	1,50	19	72,70	2,00	44	76,05	1,50
20	77,40	1,00	45	83,45	1,00	20	79,40	2,00	45	86,25	2,00
21	80,80	1,50	46	91,30	1,50	21	80,80	1,50	46	91,30	1,50
22	85,15	1,00	47	87,25	1,50	22	86,15	2,00	47	87,25	1,50
23	77,20	1,50	48	75,35	1,00	23	77,20	1,50	48	76,90	2,00
24	70,50	2,00	49	81,25	1,50	24	68,50	1,00	49	81,25	1,50
25	95,60	1,00	50	84,50	1,50	25	95,75	2,00	50	84,50	1,50

Médias		
	CdLPP $r_n(1)$	CdLPP $r_n(2)$
Acertos	84,45	84,73
<i>Ranks</i>	1,37	1,63
+ / ~ / -	1/35/14	14/35/1

Essa tabela mostra que, para a maioria das bases, ambas configurações do CdLPP r_n atingem mesmas taxas de acerto, o que torna semelhantes seus valores médios de acerto e *rank*. Contudo, em 14 dessas bases o CdLPP $r_n(2)$ obtém valores maiores do que o CdLPP $r_n(1)$. Essa maior frequência nos melhores resultados faz com que o CdLPP $r_n(2)$ tenha, segundo teste de Friedman à 95% de confiança, desempenho estatisticamente superior ao CdLPP $r_n(1)$, como mostra a Figura 42. Nesse figura, os intervalos de confiança do CdLPP $r_n(1)$ e CdLPP $r_n(2)$ não possuem interseção, indicando o melhor desempenho da técnica com maior *rank* médio, no caso, CdLPP $r_n(2)$.

Figura 42 – Comparação do desempenho das variações do CdLPP r_n com o teste de Friedman.

Fonte: Autoria própria.

Desse modo, as variações do CdLPP que são comparadas entre si são: CdLPP, CdLPP r_b ($r_b = 1$) e CdLPP r_n ($r_n = 2$). A Tabela 15 apresenta os desempenhos obtidos com essas variações. Vê-se, nessa tabela, que para 22 das bases de dados as três variações comparadas obtiveram iguais taxas de acerto. Além disso, para cada base de dados, a diferença absoluta entre as duas maiores taxas de acerto não atingiu 1,50%. Conseqüentemente, são próximos os valores médios das taxas de acerto para as três variações comparadas.

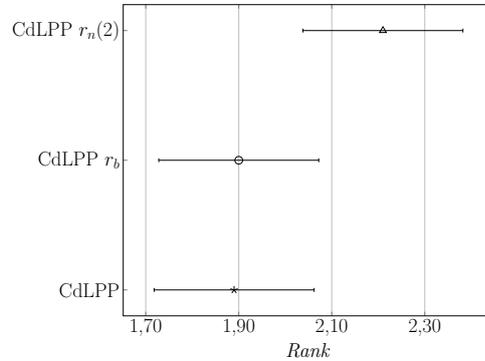
Tabela 15 – Comparação entre os desempenhos das variações CdLPP, CdLPP r_b e CdLPP $r_n(2)$.

CdLPP			CdLPP r_b						CdLPP $r_n(2)$											
<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>			
01	80,05	3,00	26	79,05	2,50	01	79,35	1,50	26	78,00	1,00	01	79,35	1,50	26	79,05	2,50			
02	89,25	2,00	27	87,80	1,00	02	89,25	2,00	27	88,40	3,00	02	89,25	2,00	27	88,30	2,00			
03	97,25	2,00	28	85,60	1,00	03	97,25	2,00	28	86,15	2,50	03	97,25	2,00	28	86,15	2,50			
04	92,50	2,00	29	97,70	1,00	04	92,50	2,00	29	98,20	2,50	04	92,50	2,00	29	98,20	2,50			
05	92,20	2,50	30	83,50	1,50	05	91,75	1,00	30	84,00	3,00	05	92,20	2,50	30	83,50	1,50			
06	74,65	2,00	31	96,95	1,00	06	74,65	2,00	31	97,30	3,00	06	74,65	2,00	31	97,25	2,00			
07	81,45	2,00	32	92,45	2,00	07	81,45	2,00	32	92,45	2,00	07	81,45	2,00	32	92,45	2,00			
08	80,40	2,00	33	87,20	1,00	08	80,45	2,00	33	87,45	2,00	08	80,45	2,00	33	87,80	3,00			
09	82,10	2,50	34	97,20	3,00	09	81,55	1,00	34	97,05	1,50	09	82,10	2,50	34	97,05	1,50			
10	74,80	2,50	35	72,15	2,00	10	74,30	1,00	35	72,15	2,00	10	74,80	2,50	35	72,15	2,00			
11	89,25	2,50	36	95,30	1,00	11	87,50	1,00	36	96,00	2,50	11	89,25	2,50	36	96,00	2,50			
12	72,00	1,50	37	90,00	2,00	12	73,25	3,00	37	90,00	2,00	12	72,00	1,50	37	90,00	2,00			
13	89,50	2,50	38	76,20	2,00	13	89,30	1,00	38	76,20	2,00	13	89,50	2,50	38	76,20	2,00			
14	96,55	2,00	39	99,00	2,00	14	96,55	2,00	39	99,00	2,00	14	96,55	2,00	39	99,00	2,00			
15	72,95	2,00	40	78,95	2,00	15	72,95	2,00	40	78,95	2,00	15	72,95	2,00	40	78,95	2,00			
16	87,30	2,00	41	78,20	2,00	16	87,30	2,00	41	78,20	2,00	16	87,30	2,00	41	78,20	2,00			
17	85,90	2,00	42	83,65	1,00	17	85,10	1,00	42	85,05	2,50	17	86,35	3,00	42	85,05	2,50			
18	86,90	1,00	43	79,95	2,00	18	88,50	2,50	43	79,00	1,00	18	88,50	2,50	43	80,75	3,00			
19	71,70	2,00	44	76,05	2,00	19	69,75	1,00	44	76,05	2,00	19	72,70	3,00	44	76,05	2,00			
20	78,85	2,00	45	86,25	2,50	20	78,10	1,00	45	84,00	1,00	20	79,40	3,00	45	86,25	2,50			
21	80,80	2,00	46	91,10	1,00	21	80,80	2,00	46	91,30	2,50	21	80,80	2,00	46	91,30	2,50			
22	85,00	1,00	47	87,25	2,00	22	86,15	2,50	47	87,25	2,00	22	86,15	2,50	47	87,25	2,00			
23	77,20	2,00	48	76,95	3,00	23	77,20	2,00	48	76,90	1,50	23	77,20	2,00	48	76,90	1,50			
24	68,50	2,00	49	80,70	1,00	24	68,50	2,00	49	82,70	3,00	24	68,50	2,00	49	81,25	2,00			
25	95,70	2,00	50	84,50	2,00	25	95,55	1,00	50	84,50	2,00	25	95,75	3,00	50	84,50	2,00			
Médias																				
CdLPP			CdLPP r_b						CdLPP $r_n(2)$											
Acertos			84,53						84,73											
<i>Ranks</i>			1,89						2,21											
3,0/2,5/2,0/1,5/1,0			3/7/27/2/11						5/7/23/3/12						6/14/25/5/0					

Entretanto, a análise dos *ranks* dessas variações mostra que o CdLPP $r_n(2)$: (i) alcança maior quantidade de primeiras posições do que as demais variações, (ii) não atinge nenhum desempenho mínimo absoluto, conseqüentemente, (iii) em 44 das bases, o CdLPP $r_n(2)$ atinge ou disputa a segunda posição de melhor desempenho. Desse modo, como mostra a Figura 43, o *rank* médio do CdLPP $r_n(2)$ é distintamente superior ao *rank* médio das

demais variações.

Figura 43 – Comparação dos desempenhos entre CdLPP, CdLPP r_b e CdLPP $r_n(2)$ com o teste de Friedman.



Fonte: Autoria própria.

Nessa figura, através do teste de Friedman à 95% de confiança, vê-se que o desempenho do CdLPP $r_n(2)$ é superior ao desempenho do CdLPP. Contudo, como há interseção entre os intervalos de confiança do CdLPP r_b e CdLPP $r_n(2)$, essas variações são consideradas estatisticamente equivalentes. Todavia, como o CdLPP $r_n(2)$ tende a obter os melhores posicionamentos quanto a taxa de acerto, essa variação é utilizada como representante do CdLPP nas comparações realizadas na Seção 4.3.3.2.

Comparação entre as variações do CdMD

CdMD pode ser configurada de modo a conter: apenas seu procedimento obrigatório (CdMD), procedimento obrigatório mais dispersão entreclasses (CdMD r_b) e procedimento obrigatório mais dispersão negativa (CdMD r_n). De modo similar ao CdLPP, a variação CdMD r_b contém apenas uma possibilidade de configuração, dado que todas as bases possuem duas classes. Entretanto, a variação CdMD r_n pode ser variada com $r_n = \{1, 2\}$, uma vez que essas bases são tridimensionais. Devido a isso, antes de determinar a variação do CdMD melhor adaptada a cenários com multimodalidade intraclasse, é preciso determinar qual das configurações do CdMD r_n melhor representa essa variação nesse tipo de cenário. Isso é feito através da análise dos resultados da Tabela 16.

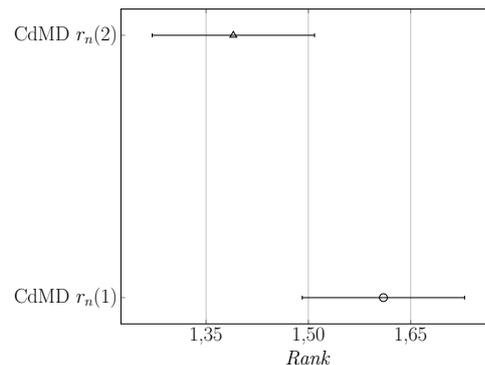
Os resultados nessa tabela mostram que o CdMD $r_n(1)$ alcança, com maior frequência, maiores taxas de acerto do que o CdMD $r_n(2)$. Além disso, a média da diferença absoluta entre as taxas de acerto dessas configurações quando: (i) CdMD $r_n(1)$ é menor do que CdMD $r_n(2)$, é de 1,18%; (ii) CdMD $r_n(1)$ é maior do que CdMD $r_n(2)$, é de 5,16%. A tendência do CdMD $r_n(1)$ em alcançar as maiores taxas de acerto com maior frequência do que o CdMD $r_n(2)$ é reforçada pelos maiores valores médios de taxa de acerto e *rank* que essa primeira configuração alcança. Entretanto, como mostra a Figura 44, há interseção entre os intervalos de confiança dos *rank*s dessas configurações. Desse modo, segundo o

Tabela 16 – Comparação entre os desempenhos das configurações do CdMD r_n .

CdLPP $r_n(1)$						CdLPP $r_n(2)$					
<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>
01	76,40	2,00	26	77,15	1,00	01	49,60	1,00	26	79,10	2,00
02	88,40	2,00	27	89,80	1,50	02	86,60	1,00	27	89,80	1,50
03	97,15	1,00	28	88,15	1,50	03	97,25	2,00	28	88,15	1,50
04	93,75	2,00	29	98,30	2,00	04	84,95	1,00	29	98,20	1,00
05	92,30	2,00	30	82,60	2,00	05	92,25	1,00	30	74,45	1,00
06	76,40	1,50	31	97,55	1,50	06	76,40	1,50	31	97,55	1,50
07	79,75	1,50	32	92,75	1,50	07	79,75	1,50	32	92,75	1,50
08	81,55	2,00	33	87,65	1,00	08	81,40	1,00	33	89,45	2,00
09	78,60	2,00	34	97,40	2,00	09	70,20	1,00	34	89,60	1,00
10	76,85	1,00	35	70,95	1,00	10	79,00	2,00	35	72,55	2,00
11	88,60	2,00	36	96,05	2,00	11	76,50	1,00	36	81,80	1,00
12	69,05	1,00	37	90,55	1,00	12	73,05	2,00	37	90,60	2,00
13	89,80	1,00	38	79,35	2,00	13	90,50	2,00	38	79,30	1,00
14	96,35	1,00	39	99,10	2,00	14	96,55	2,00	39	97,40	1,00
15	73,55	1,00	40	79,25	2,00	15	73,60	2,00	40	78,70	1,00
16	84,85	1,50	41	78,10	1,50	16	84,85	1,50	41	78,10	1,50
17	88,65	1,50	42	83,05	2,00	17	88,65	1,50	42	82,85	1,00
18	87,65	1,50	43	82,45	1,50	18	87,65	1,50	43	82,45	1,50
19	75,10	2,00	44	76,30	1,00	19	74,75	1,00	44	76,35	2,00
20	79,35	2,00	45	84,85	2,00	20	78,60	1,00	45	83,70	1,00
21	79,50	2,00	46	92,10	1,50	21	77,80	1,00	46	92,10	1,50
22	84,90	1,00	47	85,35	2,00	22	85,85	2,00	47	81,10	1,00
23	76,15	1,00	48	74,95	2,00	23	77,85	2,00	48	74,90	1,00
24	56,35	2,00	49	80,05	2,00	24	53,70	1,00	49	67,75	1,00
25	94,35	2,00	50	83,40	1,50	25	84,70	1,00	50	83,40	1,50

Médias		
	CdMD $r_n(1)$	CdMD $r_n(2)$
Acertos	84,25	82,08
Ranks	1,61	1,39
+ / ~ / -	24/13/13	13/13/24

teste de Friedman com 95% de confiança, não a evidência estatística que permita declarar CdMD $r_n(1)$ como configuração com melhor desempenho do que o CdMD $r_n(2)$. Porém, devido a evidência de que CdMD $r_n(1)$ alcança, com maior frequência, maiores taxas de acerto, então essa configuração é utilizada na comparação dos desempenhos das variações do CdMD.

Figura 44 – Comparação dos desempenhos entre CdMD r_n com o teste de Friedman.

Fonte: Autoria própria.

CdMD, CdMD r_b e CdMD $r_n(1)$ são as variações do CdMD comparadas na Tabela 17, a qual é utilizada para determinar quais dessas variações melhor representa o CdMD em cenários com multimodalidade intraclasse. Vê-se, nessa tabela, que CdMD r_b obtém, com maior frequência, os maiores valores de taxa de acerto, sendo seguido pelo CdMD $r_n(1)$. Esse comportamento é refletido nos *ranks* médios das variações comparadas, tendo a variação CdMD os menores valores médios de taxa de acerto e *rank*.

Como mostra a Figura 45, o teste de Friedman à 95% de confiança indica CdMD

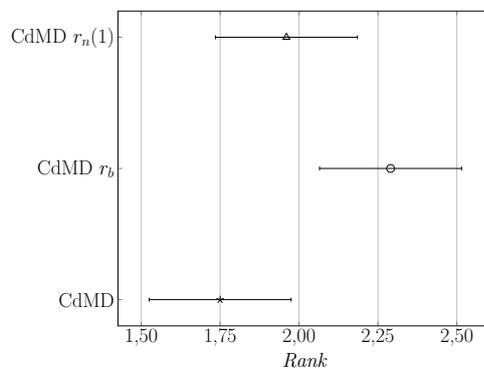
Tabela 17 – Comparação entre os desempenhos das variações CdMD, CdMD r_b e CdMD $r_n(1)$.

CdMD						CdMD r_b						CdMD $r_n(1)$					
<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>									
01	76,75	2,00	26	78,25	3,00	01	82,45	3,00	26	77,35	2,00	01	76,40	1,00	26	77,15	1,00
02	86,60	1,00	27	89,80	2,00	02	89,30	3,00	27	89,80	2,00	02	88,40	2,00	27	89,80	2,00
03	97,20	2,00	28	88,15	2,50	03	97,25	3,00	28	88,10	1,00	03	97,15	1,00	28	88,15	2,50
04	84,90	1,00	29	98,25	1,50	04	93,90	3,00	29	98,25	1,50	04	93,75	2,00	29	98,30	3,00
05	92,30	2,50	30	80,45	1,00	05	91,80	1,00	30	84,40	3,00	05	92,30	2,50	30	82,60	2,00
06	76,40	2,00	31	97,55	2,00	06	76,40	2,00	31	97,55	2,00	06	76,40	2,00	31	97,55	2,00
07	79,60	1,00	32	92,75	2,00	07	81,50	3,00	32	92,75	2,00	07	79,75	2,00	32	92,75	2,00
08	79,70	1,00	33	88,40	3,00	08	81,40	2,00	33	87,75	2,00	08	81,55	3,00	33	87,65	1,00
09	69,60	1,00	34	91,60	1,00	09	78,20	2,00	34	97,60	3,00	09	78,60	3,00	34	97,40	2,00
10	79,00	2,50	35	72,55	3,00	10	79,00	2,50	35	70,85	1,00	10	76,85	1,00	35	70,95	2,00
11	76,50	1,00	36	84,10	1,00	11	88,40	2,00	36	96,00	2,00	11	88,60	3,00	36	96,05	3,00
12	70,90	2,00	37	90,20	1,00	12	73,10	3,00	37	90,25	2,00	12	69,05	1,00	37	90,55	3,00
13	90,50	3,00	38	79,30	1,50	13	90,45	2,00	38	79,30	1,50	13	89,80	1,00	38	79,35	3,00
14	96,55	2,00	39	98,40	1,00	14	96,60	3,00	39	98,85	2,00	14	96,35	1,00	39	99,10	3,00
15	73,60	2,00	40	78,70	1,00	15	74,05	3,00	40	79,80	3,00	15	73,55	1,00	40	79,25	2,00
16	84,85	1,50	41	76,05	1,00	16	86,15	3,00	41	78,05	2,00	16	84,85	1,50	41	78,10	3,00
17	88,65	2,00	42	85,80	3,00	17	88,65	2,00	42	83,45	2,00	17	88,65	2,00	42	83,05	1,00
18	87,65	2,00	43	82,45	2,00	18	87,65	2,00	43	82,45	2,00	18	87,65	2,00	43	82,45	2,00
19	74,65	1,00	44	76,35	2,50	19	74,70	2,00	44	76,35	2,50	19	75,10	3,00	44	76,30	1,00
20	79,05	1,00	45	83,70	1,00	20	79,35	2,50	45	86,30	3,00	20	79,35	2,50	45	84,85	2,00
21	77,35	1,00	46	92,10	2,00	21	80,05	3,00	46	92,10	2,00	21	79,50	2,00	46	92,10	2,00
22	86,05	3,00	47	85,75	3,00	22	85,25	2,00	47	85,45	2,00	22	84,90	1,00	47	85,35	1,00
23	75,50	1,00	48	70,85	1,00	23	77,75	3,00	48	78,05	3,00	23	76,15	2,00	48	74,95	2,00
24	68,50	3,00	49	71,60	1,00	24	65,15	2,00	49	80,85	3,00	24	56,35	1,00	49	80,05	2,00
25	84,70	1,00	50	83,40	2,00	25	93,85	2,00	50	83,40	2,00	25	94,35	3,00	50	83,40	2,00

Médias			
	CdMD	CdMD r_b	CdMD $r_n(1)$
Acertos	83,07	84,95	84,25
Ranks	1,75	2,29	1,96
3,0/2,5/2,0/1,5/1,0	8/4/14/3/21	17/3/25/2/3	11/3/21/1/14

r_b com melhor desempenho do que CdMD. O mesmo não pode ser dito em relação ao CdMD $r_n(1)$, cujo intervalo de confiança possui interseção com ambos rivais.

Figura 45 – Comparação dos desempenhos entre CdMD com o teste de Friedman.



Fonte: Autoria própria.

Devido ao comportamento frequentemente mais bem sucedido, o CdMD r_b é a variação do CdMD que representa essa técnica nas comparações realizadas na Seção 4.3.3.2.

Comparação entre as variações do CdCMD

Na modelagem do CdCMD, é utilizado um limiar sobre o valor da evidência de multimodalidade retornada pela aplicação de MD sobre os dados (Seção 3.4.2, página 65). Esse limiar é utilizado para decidir o modo como manipular esses dados, ou seja, de forma multimodal ou não.

Como diferentes valores de limiar conduzem o CdCMD a diferentes desempenhos de classificação, esse parâmetro deve ser ajustado, para cada variação do CdCMD, de modo a determinar qual valor de limiar é mais adequado ao cenário de multimodalidade intraclasse expressos nas bases da Seção 4.3.2. Nesse trabalho, os valores do limiar de decisão para CdCMD e CdCDM foram variados em $\{10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%\}$. Para esses valores de limiar, as Tabelas 18-21 apresentam os valores médios das taxas de acerto e *rank* respectivamente para as variações CdCMD, CdCMD r_b , CdCMD $r_n(1)$ e CdCMD $r_n(2)$. Além disso, essas tabelas trazem as frequências dos valores de *rank* alcançados com cada valor de limiar.

Tabela 18 – Sumário dos desempenhos do CdCMD nos diferentes limiares de decisão.

Limiares (%)	Acerto médio (%)	<i>rank</i> médio	Frequências por <i>rank</i>																
			9,00	8,50	8,00	7,50	7,00	6,50	6,00	5,50	5,00	4,50	4,00	3,50	3,00	2,50	2,00	1,50	1,00
10	83,43	5,24	1	2	3	3	4	4	4	5	4	4	6	1	3	1	2	0	3
20	83,51	5,28	1	4	3	3	1	2	8	6	1	5	5	2	2	1	3	3	0
30	83,26	5,04	3	1	2	2	5	7	2	4	1	2	8	0	4	1	2	1	5
40	83,37	5,11	7	0	1	1	1	1	5	4	6	3	3	8	0	2	1	2	3
50	83,46	5,53	4	0	4	3	1	6	6	6	4	2	6	2	2	0	1	0	3
60	83,24	5,18	4	2	1	3	3	1	5	4	4	5	8	0	3	0	5	1	1
70	83,23	4,88	1	3	2	4	0	4	7	4	2	2	6	1	3	2	4	0	5
80	83,09	4,24	2	1	3	0	0	1	4	3	2	6	7	2	4	5	7	2	1
90	82,79	4,50	8	1	3	1	1	2	1	0	6	1	6	0	1	1	4	2	12

Na Tabela 18, são próximos os valores médios da taxa de acerto e *rank* para os diferentes valores de limiar do CdCMD. Entretanto, a configuração do CdCMD com limiar em 50% (CdCMD^{50%}) alcança o maior *rank* médio, indicando que essa configuração, com maior frequência, obtém melhores posicionamentos que as demais configurações de limiar.

Tabela 19 – Sumário dos desempenhos do CdCMD r_b nos diferentes limiares de decisão.

Limiares (%)	Acerto médio (%)	<i>rank</i> médio	Frequências por <i>rank</i>																
			9,00	8,50	8,00	7,50	7,00	6,50	6,00	5,50	5,00	4,50	4,00	3,50	3,00	2,50	2,00	1,50	1,00
10	84,92	5,17	3	1	1	1	8	2	3	5	6	4	4	1	2	5	3	0	1
20	84,84	5,24	2	1	2	2	6	4	5	5	3	2	8	0	3	0	7	0	0
30	84,80	5,40	2	2	3	3	6	4	3	6	5	2	4	0	3	1	2	0	4
40	84,89	5,28	2	3	5	2	5	2	4	3	3	2	6	1	4	2	2	1	3
50	84,78	5,28	6	1	1	2	2	4	5	4	8	4	3	0	1	2	1	1	5
60	84,74	5,21	4	1	3	2	3	4	2	6	8	1	2	1	5	2	3	0	3
70	84,65	5,26	2	3	6	1	5	1	3	4	4	3	4	2	4	2	3	0	3
80	84,51	4,32	2	2	1	2	3	1	4	5	2	3	0	2	5	5	7	2	4
90	84,24	3,84	6	2	2	1	3	0	0	0	2	1	5	1	1	3	4	2	17

Também na Tabela 19 os valores médios de taxa de acerto e *rank* são próximos para todos os valores de limiar utilizados. Contudo, a configuração com limiar de 30% conduziu o CdCMD r_b (CdCMD^{30%} r_b) ao maior valor de *rank* médio.

Tabela 20 – Sumário dos desempenhos do CdCMD $r_n(1)$ nos diferentes limiares de decisão.

Limiares (%)	Acerto médio (%)	<i>rank</i> médio	Frequências por <i>rank</i>																
			9,00	8,50	8,00	7,50	7,00	6,50	6,00	5,50	5,00	4,50	4,00	3,50	3,00	2,50	2,00	1,50	1,00
10	84,45	5,58	4	2	3	1	4	4	7	6	4	1	5	1	3	1	2	0	2
20	84,47	5,39	3	3	4	2	5	1	6	3	4	3	4	2	1	2	4	1	2
30	84,47	5,09	2	0	3	2	6	3	4	6	6	0	3	2	6	2	2	0	3
40	84,26	5,12	0	0	4	2	7	4	7	5	3	1	4	1	2	2	6	1	1
50	84,34	4,97	1	1	0	3	5	4	8	4	6	0	4	2	4	0	4	1	3
60	84,13	5,00	3	0	3	1	5	3	7	5	2	2	2	2	5	1	7	0	2
70	84,26	5,04	2	0	5	1	4	3	4	7	5	2	2	1	6	1	3	1	3
80	84,02	4,37	4	0	5	0	1	2	4	6	1	1	2	1	6	2	4	5	6
90	84,32	4,44	11	2	3	0	2	0	3	0	1	0	1	0	4	1	3	3	16

Novamente, os valores médios de taxa de acerto e *rank* são próximos para todos os valores de limiar utilizados na Tabela 20. Tendo a configuração com limiar de 10% conduzido o CdCMD $r_n(1)$ ao maior valor de *rank* médio.

Tabela 21 – Sumário dos desempenhos do CdCMD $r_n(2)$ nos diferentes limiares de decisão.

Limiares (%)	Acerto médio (%)	<i>rank</i> médio	Frequências por <i>rank</i>																
			9,00	8,50	8,00	7,50	7,00	6,50	6,00	5,50	5,00	4,50	4,00	3,50	3,00	2,50	2,00	1,50	1,00
10	82,04	5,39	2	2	1	1	2	5	3	9	10	6	3	1	2	1	2	0	0
20	82,09	5,33	0	2	0	3	3	5	4	9	10	7	2	2	0	0	0	1	2
30	82,06	5,02	1	0	0	2	3	3	4	10	9	8	2	2	1	1	2	0	2
40	82,03	4,72	1	0	0	0	0	4	5	10	9	9	4	1	0	2	0	2	3
50	81,98	5,40	1	0	3	1	3	5	4	8	11	7	4	0	0	0	2	1	0
60	81,89	5,21	1	0	2	1	3	5	3	11	10	6	2	0	0	2	2	0	2
70	81,78	4,57	2	0	0	0	2	2	2	8	10	7	4	0	5	1	5	1	1
80	81,72	4,38	2	1	0	0	2	3	0	10	9	6	1	0	0	1	9	3	3
90	81,84	4,98	14	1	2	0	0	2	1	1	8	0	2	0	0	2	3	2	12

Na Tabela 21, são próximos os valores médios de taxas de acerto e *rank* obtidos com os diferentes valores de limiar para o CdCMD $r_n(2)$. Porém, a configuração com limiar em 50% obteve o maior valor de *rank* médio.

Referente a esses resultados, a comparação dos desempenhos das variações do CdCMD configuradas com diferentes valores de limiar possibilita a escolha pela configuração (valor de limiar) mais adequada ao cenário de multimodalidade intraclasse, para cada variação do CdCMD. As escolhas foram feitas pelas configurações com maior valor médio de *rank* em cada variação do CdCMD. Desse modo, as configurações escolhidas foram: CdCMD^{50%}, CdCMD^{30%} r_b , CdCMD^{10%} $r_n(1)$ e CdCMD^{50%} $r_n(2)$.

Todavia, até esse momento, a variação do CdCMD r_n possui duas configurações: CdCMD com $r_n = 1$ no limiar de 10% (CdCMD^{10%} $r_n(1)$) e CdCMD com $r_n = 2$ no limiar de 50% (CdCMD^{50%} $r_n(2)$). Desse modo, resta determinar qual dessas configurações melhor representa essa variação. A Tabela 22 mostra que CdCMD^{10%} $r_n(1)$ alcança maiores valores médios de taxa de acerto e *rank* do que o CdCMD^{50%} $r_n(2)$. Assim, CdCMD^{10%} $r_n(1)$ é a configuração do CdCMD r_n escolhida como melhor adaptada aos cenários de multimodalidade intraclasse com respeito as 50 bases de dados desse experimento.

Tabela 22 – Comparação entre os desempenhos das configurações do CdCMD r_n .

		CdCMD ^{10%} $r_n(1)$	CdCMD ^{50%} $r_n(2)$
Médias	Acertos	84,45	81,98
	<i>Ranks</i>	1,61	1,39
+ / ~ / -		23/15/12	12/15/23

A Tabela 23 apresenta os resultados experimentais das configurações melhor adaptadas para as variações do CdCMD. Nessa tabela, os maiores valores médios de taxa de acertos e *rank* foram alcançados por CdCMD^{30%} r_b , sugerindo essa configuração como melhor adaptada aos cenários de multimodalidade intraclasse dentre as demais configurações possíveis ao CdCMD. Entretanto, como mostra a Figura 46, o intervalo de confiança para a média dos *ranks* do CdCMD^{30%} r_b possui interseção com os intervalos de confiança das demais configurações. Desse modo, segundo o teste de Friedman à 95% de confiança, não é possível afirmar nenhuma das configurações comparadas como tendo melhor desempenho

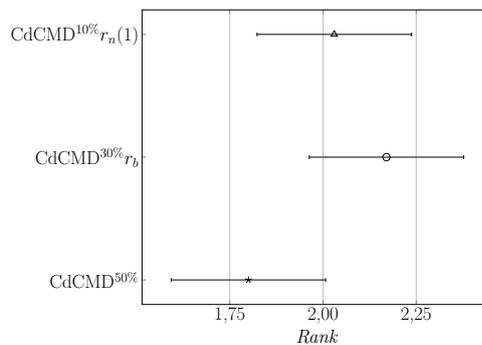
do que as demais. Assim, dado que $\text{CdCMD}^{30\%r_b}$ possui o maior valor médio de *rank*, essa configuração é utilizada para representar o CdCMD nas comparações realizadas na Seção 4.3.3.2.

Tabela 23 – Comparação entre os desempenhos das configurações $\text{CdCMD}^{50\%}$, $\text{CdCMD}^{30\%r_b}$ e $\text{CdCMD}^{10\%r_n(1)}$.

CdCMD ^{50%}						CdCMD ^{30%r_b}						CdCMD ^{10%r_n(1)}					
<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>
01	76.20	2.00	26	78.35	3.00	01	82.30	3.00	26	77.15	1.00	01	75.90	1.00	26	77.20	2.00
02	86.60	1.00	27	85.80	1.00	02	89.15	3.00	27	87.55	2.00	02	88.60	2.00	27	89.80	3.00
03	97.25	3.00	28	88.15	3.00	03	97.20	1.50	28	88.10	1.50	03	97.20	1.50	28	88.10	1.50
04	84.95	1.00	29	98.30	3.00	04	94.05	3.00	29	98.25	1.50	04	93.85	2.00	29	98.25	1.50
05	91.85	1.00	30	79.60	1.00	05	92.30	2.50	30	84.40	3.00	05	92.30	2.50	30	82.85	2.00
06	76.40	2.00	31	97.55	2.00	06	76.40	2.00	31	97.55	2.00	06	76.40	2.00	31	97.55	2.00
07	80.95	3.00	32	92.75	2.00	07	80.10	2.00	32	92.75	2.00	07	79.70	1.00	32	92.75	2.00
08	81.40	2.00	33	88.25	3.00	08	81.40	2.00	33	87.80	2.00	08	81.40	2.00	33	87.65	1.00
09	70.20	1.00	34	89.60	1.00	09	78.80	2.00	34	97.55	2.00	09	80.20	3.00	34	97.75	3.00
10	79.00	2.00	35	75.85	3.00	10	79.00	2.00	35	71.10	2.00	10	79.00	2.00	35	70.85	1.00
11	84.15	1.00	36	96.45	3.00	11	88.45	2.00	36	95.95	1.00	11	88.70	3.00	36	96.00	2.00
12	71.30	1.00	37	90.00	1.00	12	73.05	2.50	37	90.55	3.00	12	73.05	2.50	37	90.20	2.00
13	90.65	3.00	38	79.30	1.50	13	90.40	2.00	38	79.30	1.50	13	89.85	1.00	38	79.35	3.00
14	96.55	2.50	39	98.65	1.00	14	96.45	1.00	39	98.85	2.50	14	96.55	2.50	39	98.85	2.50
15	73.60	1.50	40	78.70	1.00	15	74.20	3.00	40	79.80	3.00	15	73.60	1.50	40	79.30	2.00
16	84.85	1.50	41	77.40	1.00	16	86.60	3.00	41	77.70	2.00	16	84.85	1.50	41	78.05	3.00
17	85.30	1.00	42	86.40	3.00	17	87.90	2.00	42	83.65	2.00	17	88.65	3.00	42	83.30	1.00
18	87.65	2.00	43	82.45	2.50	18	87.65	2.00	43	82.45	2.50	18	87.65	2.00	43	81.10	1.00
19	74.50	1.00	44	76.35	2.00	19	74.60	2.00	44	76.35	2.00	19	74.65	3.00	44	76.35	2.00
20	79.00	1.00	45	83.70	1.50	20	79.20	3.00	45	83.70	1.50	20	79.10	2.00	45	85.10	3.00
21	77.25	1.00	46	92.10	2.00	21	80.00	3.00	46	92.10	2.00	21	79.45	2.00	46	92.10	2.00
22	86.05	3.00	47	87.15	3.00	22	84.90	1.00	47	85.40	1.00	22	84.95	2.00	47	85.50	2.00
23	75.60	1.00	48	74.90	1.00	23	77.80	2.00	48	77.40	3.00	23	77.85	3.00	48	74.95	2.00
24	69.30	3.00	49	67.75	1.00	24	63.05	2.00	49	80.85	3.00	24	58.30	1.00	49	79.95	2.00
25	84.70	1.00	50	82.30	1.00	25	94.40	3.00	50	84.55	3.00	25	94.20	2.00	50	83.80	2.00

Médias			
	CdCMD ^{50%}	CdCMD ^{30%r_b}	CdCMD ^{10%r_n(1)}
Acertos	83,46	84,80	84,45
Ranks	1,80	2,17	2,03
3,0/2,5/2,0/1,5/1,0	13/2/9/4/22	14/4/22/5/5	10/4/23/5/8

Figura 46 – Comparação dos desempenho das configurações do CdCMD com o teste de Friedman.



Fonte: Autoria própria.

Comparação entre as variações do CdCDM

A determinação das configurações das variações do CdCDM, que são melhor adaptadas aos cenários com multimodalidade intraclasse, requer a definição dos valores de limiar que operam sobre essas variações. Desse modo as Tabelas 24-27 apresentam os desempenhos médios de cada variação do CdCDM para cada valor de limiar. Assim, essas tabelas são respectivamente associadas as seguintes variações: CdCDM, CdCDM r_b , CdCDM $r_n(1)$ e CdCDM $r_n(2)$.

Tabela 24 – Sumário dos desempenhos do CdCDM nos diferentes limiares de decisão.

Limiares (%)	Acerto médio (%)	$rank$ médio	Frequências por $rank$																
			9,00	8,50	8,00	7,50	7,00	6,50	6,00	5,50	5,00	4,50	4,00	3,50	3,00	2,50	2,00	1,50	1,00
10	83,03	5,02	3	3	1	1	2	3	2	8	9	2	4	1	3	0	3	0	5
20	83,04	5,35	2	2	1	0	4	4	3	10	10	4	4	1	2	0	1	0	2
30	82,99	5,69	4	0	5	2	5	4	2	6	9	2	4	1	3	0	2	0	1
40	82,74	5,07	2	1	1	1	2	3	6	9	8	4	3	1	2	1	4	1	1
50	82,88	5,38	2	3	4	1	3	2	3	8	9	2	5	0	2	0	4	0	2
60	82,91	5,29	4	1	1	0	4	3	4	9	7	3	8	0	0	0	4	1	1
70	82,55	4,41	0	0	3	0	2	3	2	10	6	4	6	0	1	1	6	1	5
80	82,82	4,70	4	0	1	1	2	3	2	7	9	4	2	0	3	0	6	3	3
90	82,49	4,09	10	0	3	0	0	1	0	1	3	1	5	0	3	0	5	4	14

Tabela 25 – Sumário dos desempenhos do CdCDM r_b nos diferentes limiares de decisão.

Limiares (%)	Acerto médio (%)	$rank$ médio	Frequências por $rank$																
			9,00	8,50	8,00	7,50	7,00	6,50	6,00	5,50	5,00	4,50	4,00	3,50	3,00	2,50	2,00	1,50	1,00
10	84,67	5,25	5	3	0	4	6	2	1	5	4	3	2	1	5	1	3	0	5
20	84,80	5,06	1	4	1	3	5	2	2	5	5	2	7	3	3	1	2	0	4
30	84,79	5,15	1	1	5	3	6	3	0	6	5	3	2	2	6	1	3	0	3
40	84,66	5,33	2	0	2	4	4	3	9	5	3	2	7	1	3	2	1	0	2
50	84,70	5,22	4	2	3	0	4	5	2	5	7	2	2	3	3	1	2	2	3
60	84,61	5,39	2	3	1	2	6	5	3	4	7	2	5	3	2	0	3	0	2
70	84,45	4,96	2	2	2	3	4	2	3	5	4	2	5	2	7	1	2	1	3
80	84,42	4,40	5	1	2	2	1	2	3	4	3	2	2	1	3	1	9	3	6
90	84,29	4,24	7	0	2	1	3	2	1	1	4	2	4	0	0	2	8	2	11

Tabela 26 – Sumário dos desempenhos do CdCDM $r_n(1)$ nos diferentes limiares de decisão.

Limiares (%)	Acerto médio (%)	$rank$ médio	Frequências por $rank$																
			9,00	8,50	8,00	7,50	7,00	6,50	6,00	5,50	5,00	4,50	4,00	3,50	3,00	2,50	2,00	1,50	1,00
10	84,50	5,49	4	3	2	0	4	4	3	8	7	3	2	3	1	0	5	0	1
20	84,36	5,14	2	2	5	1	4	5	2	7	2	2	1	4	5	1	2	0	5
30	84,45	4,99	4	1	1	1	1	5	5	6	6	3	3	4	2	2	0	1	5
40	84,39	5,42	1	2	5	0	5	4	5	8	5	3	2	3	1	0	4	0	2
50	84,33	5,37	5	1	2	1	3	5	3	7	4	3	5	3	3	1	2	0	2
60	84,32	5,61	5	1	4	2	3	3	2	11	4	4	3	2	1	0	2	0	3
70	84,08	4,47	2	0	1	1	2	6	3	6	3	3	2	2	6	2	8	1	2
80	84,21	4,66	5	0	3	0	3	3	2	8	3	2	2	1	1	2	8	2	5
90	83,78	3,85	5	0	4	0	2	1	1	1	4	1	3	2	4	2	1	4	15

Tabela 27 – Sumário dos desempenhos do CdCDM $r_n(2)$ nos diferentes limiares de decisão.

Limiares (%)	Acerto médio (%)	$rank$ médio	Frequências por $rank$																
			9,00	8,50	8,00	7,50	7,00	6,50	6,00	5,50	5,00	4,50	4,00	3,50	3,00	2,50	2,00	1,50	1,00
10	82,10	5,11	1	2	3	1	2	1	7	7	9	7	0	1	1	4	0	0	4
20	82,10	5,34	0	3	2	1	2	1	6	9	12	7	2	2	0	1	1	0	1
30	82,07	5,06	0	0	1	1	4	0	7	10	11	7	2	2	1	2	1	0	1
40	82,07	5,39	2	0	4	1	2	1	8	8	10	7	1	2	0	2	0	0	2
50	81,95	5,28	0	1	1	1	4	2	6	10	11	6	4	2	0	0	1	0	1
60	81,88	4,94	1	1	0	0	2	2	5	10	13	7	3	0	0	0	3	0	3
70	81,82	4,92	3	0	0	0	1	3	4	9	11	7	1	2	4	1	3	0	1
80	81,69	4,37	2	1	1	0	2	2	1	7	10	6	2	0	1	0	8	3	4
90	81,83	4,59	13	0	1	1	1	0	1	0	10	0	1	1	0	0	4	3	14

A escolha por uma configuração é realizada segundo o valor de limiar com maior valor médio de $rank$. Desse modo, para as variações do CdCDM, as configurações escolhidas são: CdCDM^{30%}, CdCDM^{60%} r_b , CdCDM^{60%} $r_n(1)$ e CdCDM^{40%} $r_n(2)$.

Contudo, até esse momento, a variação do CdCDM r_n possui duas configurações: CdCDM com $r_n = 1$ no limiar de 60% (CdCDM^{60%} $r_n(1)$) e CdCDM com $r_n = 2$ no limiar de 40% (CdCDM^{40%} $r_n(2)$). Desse modo, para determinar qual dessas configurações melhor representa essa variação, a Tabela 28 apresenta os valores médios dos desempenhos obtidos com essas configurações.

Essa tabela mostra que CdCDM^{60%} $r_n(1)$ alcança maiores valores médios de taxa de acerto e $rank$ do que o CdCDM^{40%} $r_n(2)$. Portanto, CdCDM^{60%} $r_n(1)$ é a configuração do CdCDM

Tabela 28 – Comparação entre os desempenhos das configurações do CdCDM r_n .

		CdCDM ^{60%} $r_n(1)$	CdCDM ^{40%} $r_n(2)$
Médias	Acertos	84,32	82,07
	Ranks	1,62	1,38
+ / ~ / -		24/14/12	12/14/24

r_n escolhida como melhor adaptada aos cenários de multimodalidade intraclasse com respeito as 50 bases de dados desse experimento.

Escolhidas as configurações melhor adaptadas para cada variação do CdCDM, elas são comparadas entre si de modo a possibilitar identificar qual dentre elas é a melhor adaptada para cenários com multimodalidade intraclasse. A Tabela 29 apresenta os resultados experimentais obtidos com as configurações escolhidas, a saber: CdCDM^{30%}, CdCDM^{60%} r_b e CdCDM^{60%} $r_n(1)$.

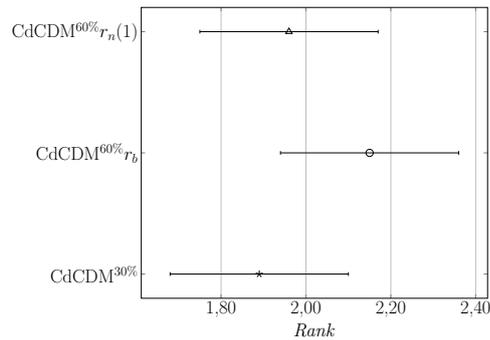
Tabela 29 – Comparação entre os desempenhos das configurações CdCDM^{30%}, CdCDM^{60%} r_b e CdCDM^{60%} $r_n(1)$.

CdCDM ^{30%}					CdCDM ^{60%} r_b					CdCDM ^{60%} $r_n(1)$							
<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>	<i>Toy</i>	Acertos (%)	<i>Rank</i>			
01	77,20	1,00	26	78,35	2,00	01	81,95	3,00	26	77,90	1,00	01	77,35	2,00	26	79,10	3,00
02	87,15	3,00	27	88,65	3,00	02	86,60	2,00	27	87,65	2,00	02	84,65	1,00	27	87,30	1,00
03	97,15	1,00	28	88,15	2,50	03	97,25	2,50	28	88,15	2,50	03	97,25	2,50	28	88,10	1,00
04	84,90	1,00	29	98,25	1,00	04	89,95	2,00	29	98,30	2,50	04	91,00	3,00	29	98,30	2,50
05	92,30	3,00	30	74,45	1,00	05	92,25	2,00	30	82,50	3,00	05	91,50	1,00	30	80,45	2,00
06	76,40	2,00	31	97,60	3,00	06	76,40	2,00	31	97,55	1,50	06	76,40	2,00	31	97,55	1,50
07	80,10	1,00	32	92,75	2,00	07	81,00	2,50	32	92,75	2,00	07	81,00	2,50	32	92,75	2,00
08	80,85	1,00	33	89,45	3,00	08	81,40	2,50	33	87,95	2,00	08	81,40	2,50	33	87,85	1,00
09	70,05	1,00	34	89,60	1,00	09	80,55	3,00	34	97,55	2,00	09	80,40	2,00	34	97,75	3,00
10	79,00	2,00	35	72,55	2,50	10	79,00	2,00	35	70,05	1,00	10	79,00	2,00	35	72,55	2,50
11	76,50	1,00	36	81,80	1,00	11	89,75	2,00	36	95,80	2,50	11	90,05	3,00	36	95,80	2,50
12	70,20	2,00	37	90,55	2,00	12	71,15	3,00	37	90,55	2,00	12	69,50	1,00	37	90,55	2,00
13	90,50	1,00	38	79,30	1,50	13	90,75	2,50	38	79,30	1,50	13	90,75	2,50	38	79,35	3,00
14	96,65	3,00	39	97,40	1,00	14	96,55	1,50	39	99,15	3,00	14	96,55	1,50	39	99,00	2,00
15	73,60	3,00	40	78,70	1,00	15	72,80	2,00	40	80,45	3,00	15	72,55	1,00	40	79,75	2,00
16	84,85	1,00	41	77,40	1,00	16	85,45	2,50	41	77,90	3,00	16	85,45	2,50	41	77,85	2,00
17	87,90	3,00	42	86,20	3,00	17	83,85	1,50	42	84,25	1,00	17	83,85	1,50	42	84,45	2,00
18	87,75	3,00	43	82,45	2,00	18	87,65	1,50	43	82,45	2,00	18	87,65	1,50	43	82,45	2,00
19	75,05	2,00	44	76,35	2,00	19	75,10	3,00	44	76,35	2,00	19	74,95	1,00	44	76,35	2,00
20	79,20	2,00	45	84,60	1,00	20	78,05	1,00	45	85,95	3,00	20	79,40	3,00	45	85,75	2,00
21	78,80	1,00	46	92,10	2,00	21	80,45	3,00	46	92,10	2,00	21	79,95	2,00	46	92,10	2,00
22	86,00	3,00	47	86,45	3,00	22	85,25	2,00	47	80,95	1,00	22	85,00	1,00	47	82,35	2,00
23	77,80	3,00	48	77,45	3,00	23	76,30	2,00	48	77,35	2,00	23	76,15	1,00	48	76,35	1,00
24	65,80	3,00	49	67,75	1,00	24	65,20	2,00	49	81,75	2,00	24	57,90	1,00	49	82,80	3,00
25	84,70	1,00	50	81,00	1,00	25	95,95	2,00	50	85,45	3,00	25	96,25	3,00	50	83,40	2,00

Médias			
	CdCDM ^{30%}	CdCDM ^{60%} r_b	CdCDM ^{60%} $r_n(1)$
Acertos	82,99	84,61	84,32
Ranks	1,89	2,15	1,96
3,0/2,5/2,0/1,5/1,0	15/2/11/1/21	11/8/21/5/5	8/8/18/4/12

Esses resultados mostram que a configuração CdCDM^{60%} r_b alcança os maiores valores médios de taxa de acerto e *rank*, indicando melhor adaptação a esse cenário do que as demais configurações experimentadas. Entretanto, como mostra a Figura 47, essa configuração é estatisticamente equivalente as demais, não podendo ser estabelecida como tendo melhor desempenho, segundo teste de Friedman com 95% de confiança. Contudo, dado que CdCDM^{60%} r_b possui o maior valor médio de *rank*, essa configuração é utilizada para representar o CdCDM nas comparações realizadas na Seção 4.3.3.2.

Figura 47 – Comparação dos desempenho das configurações do CdCDM com o teste de Friedman.



Fonte: Autoria própria.

4.3.3.2 Comparação entre as técnicas

Nessa seção são comparados os desempenhos de classificação com as TECs mencionadas nesse trabalho, a saber: *Principal Component Analysis* (PCA), *Linear Discriminant Analysis* (LDA), *Locality Preserving Projections* (LPP), *Classwise Locality Preserving Projections* (CLPP), *Local Fisher Discriminant Analysis* (LFDA), *Classwise Principal Component Analysis* (CPCA), *Class-dependent Locality Preserving Projections* (CdLPP), *Class-dependent Multimodal Decomposition* (CdMD), *Class-dependent Conditional Multimodal Decomposition* (CdCMD) e *Class-dependent Conditional Detachment of Modes* (CdCDM).

Dado que as técnicas *classwise* admitem mais de uma configuração, na Seção 4.3.3.1 essas configurações foram avaliadas de modo a determinar aquelas melhores adaptadas a cenários com multimodalidade intraclasse. Essas melhores configurações estão sumarizadas na Tabela 30 por técnica *classwise*.

Tabela 30 – Configurações melhores adaptadas para cada uma das técnicas *classwise*.

Técnica <i>classwise</i>	Melhor configuração
CPCA	CPCA
CdLPP	CdLPP $r_n(2)$
CdMD	CdMD r_b
CdCMD	CdCMD ^{30%r_b}
CdCDM	CdCDM ^{60%r_b}

O protocolo metodológico definido na Seção 4.3.1 encontra para cada configuração das TECs a melhor redução dimensional. Essa melhor redução é aquela que obtém taxa de acerto médio (validação cruzada) equivalente a redução com maior taxa de acerto médio (teste-*t*) e que possui menor desvio padrão em suas taxas. Desse modo, a Tabela 31 apresenta a frequência com que cada TEC obteve as melhores reduções possíveis (uni-, bi- ou tridimensional).

Como as bases de dados utilizadas possuem apenas duas classes, o LDA fica restrito a encontrar soluções unidimensionais. O CdLPP $r_n(2)$ determina soluções que obrigatoriamente utilizam duas componentes da matriz de dispersão negativa. Devido a isso, suas

Tabela 31 – Frequência das melhores reduções obtidas pelas técnicas de extração de características comparadas.

Reduções	PCA	LDA	LPP	CLPP	LFDA	CPCA	CdLPP $r_n(2)$	CdMD r_b	CdCMD ^{30%} r_b	CdCDM ^{60%} r_b
1D	2	50	4	5	5	1	0	0	0	0
2D	6	0	11	10	10	6	0	28	27	27
3D	42	0	35	35	35	43	50	22	23	23

soluções são exclusivamente tridimensionais. As configurações das técnicas *classwise* com investigação de multimodalidade que foram comparadas utilizam obrigatoriamente uma componente da matriz de dispersão entreclasses. Assim, as soluções dessas configurações podem variar entre bi ou tridimensionais.

Como a obtenção de melhores reduções tridimensionais geram matrizes de transformação que apenas rotacionam os padrões no espaço tridimensional, as soluções dessa ordem conduzem à desempenhos de classificação idênticos aos obtidos sem uso de qualquer TEC. Desse modo, as frequências apresentadas na Tabela 31 mostram que as técnicas *classwise* com investigação de multimodalidade alcançam com maior frequência as reduções não tridimensionais (com exceção do LDA).

A Tabela 32 apresenta as taxas de acerto obtidas por cada TEC segundo o protocolo metodológico definido na Seção 4.3.1.

A esses resultados foram acrescentadas as taxas de acerto obtidas sem uso de qualquer TEC (coluna *Original*). Isso permite, por base de dados, verificar se a aplicação de uma TEC possibilita maior discriminação de seus padrões ou não. De modo a facilitar essa verificação, ao lado de cada taxa de acerto obtida por TEC segue um dos símbolos: +, ~ e -. Respectivamente, esses símbolos indicam se a TEC alcançou taxa de acerto superior, igual ou inferior ao correspondente valor na coluna *Original*. Esses resultados salientam que, para essas 50 bases de dados, os valores de taxa de acerto obtidos pelas diferentes TECs são próximos. Essa proximidade dos resultados é globalmente salientada através dos valores médios das taxas de acerto obtidas pelas TECs. Ordenando as TECs pelos valores médios de taxa de acerto, as cinco TECs com os maiores valores foram: CdMD r_b , CdCMD^{30%} r_b , CdLPP $r_n(2)$, PCA e CdCDM^{60%} r_b . Esse posicionamento reforça a validade da utilização de técnicas *classwise* em cenários com multimodalidade intraclasse, dado que quatro dentre as cinco TECs com maiores taxas de acerto médio são *classwise*. Além disso, as três TECs propostas estão entre as cinco com maiores taxas de acerto médio, tendo CdMD r_b e CdCMD^{30%} r_b alcançado respectivamente os dois primeiros lugares. Por sua vez, esses resultados reforçam a validade da utilização de técnicas *classwise* com investigação de multimodalidade nesse tipo de cenário.

Mesmo tendo alcançado valores médios de taxa de acerto que são próximos do valor médio da taxa de acerto obtida sem utilização de TEC, são distintas as frequências com que as TECs comparadas alcançam taxas superiores (+), iguais (~) ou inferiores (-) a correspondente taxa na coluna *Original*. Essas frequências estão sumarizadas na

Tabela 32 – Taxas de acerto (%) obtidas com as técnicas de extração de características. Os símbolos +, ~ e – indicam se a TEC alcançou taxa de acerto superior, igual ou inferior ao correspondente valor na coluna *Original*.

<i>Tog</i>	Original	PCA	LDA	LPP	CLPP	LFDA	CPCA	CdLPP $r_n(2)$	CdMD r_b	CdCMD ^{30%} r_b	CdCMD ^{60%} r_b
01	79.35	79.35~	81.30+	80.30+	80.25+	79.35~	79.35~	79.35~	82.45+	82.30+	81.95+
02	89.25	89.25~	87.90–	89.25~	89.25~	89.25~	89.25~	89.25~	89.30+	89.15–	86.60–
03	97.25	97.25~	96.70–	97.25~	97.25~	97.25~	97.25~	97.25~	97.25~	97.20–	97.25~
04	92.50	92.50~	91.40–	92.50~	92.50~	92.50~	92.50~	92.50~	93.90+	94.05+	89.95–
05	92.20	92.20~	91.85–	92.20~	92.20~	92.20~	92.20~	92.20~	91.80–	92.30+	92.25+
06	74.65	74.65~	61.80–	74.65~	74.65~	74.65~	74.65~	74.65~	76.40+	76.40+	76.40+
07	81.45	81.45~	80.05–	81.45~	81.45~	81.45~	81.45~	81.45~	81.50+	80.10–	81.00–
08	80.45	80.45~	81.25+	80.30–	80.80+	80.45~	79.45–	80.45~	81.40+	81.40+	81.40+
09	82.10	80.50–	73.55–	80.65–	80.75–	80.80–	80.70–	82.10~	78.20–	78.80–	80.55–
10	74.80	74.80~	69.80–	74.80~	74.80~	74.80~	74.80~	74.80~	79.00+	79.00+	79.00+
11	89.25	89.25~	88.90–	86.65–	87.10–	89.55+	89.25~	89.25~	88.40–	88.45–	89.75+
12	72.00	72.00~	72.40+	72.00~	70.20–	73.00+	72.00~	72.00~	73.10+	73.05+	71.15–
13	89.50	89.50~	87.60–	89.50~	89.50~	89.50~	89.50~	89.50~	90.45+	90.40+	90.75+
14	96.55	96.55~	96.75+	96.55~	96.55~	96.55~	96.55~	96.55~	96.60+	96.45–	96.55~
15	72.95	72.95~	70.10–	72.95~	72.95~	72.95~	72.95~	72.95~	74.05+	74.20+	72.80–
16	87.30	87.30~	88.15+	87.30~	87.30~	87.30~	87.30~	87.30~	86.15–	86.60–	85.45–
17	86.35	86.35~	84.65–	86.35~	86.35~	86.35~	86.35~	86.35~	88.65+	87.90+	83.85–
18	88.50	88.50~	84.00–	88.50~	88.50~	88.50~	88.50~	88.50~	87.65–	87.65–	87.65–
19	72.70	72.70~	60.40–	69.15–	70.80–	72.70~	72.70~	72.70~	74.70+	74.60+	75.10+
20	79.40	79.40~	77.15–	79.40~	79.40~	79.40~	78.10–	79.40~	79.35–	79.20–	78.05–
21	80.80	80.80~	80.95+	80.80~	80.80~	80.80~	80.80~	80.80~	80.05–	80.00–	80.45–
22	86.15	86.15~	85.35–	86.15~	84.55–	84.20–	86.15~	86.15~	85.25–	84.90–	85.25–
23	77.20	77.20~	75.05–	77.20~	77.20~	77.20~	77.20~	77.20~	77.75+	77.80+	76.30–
24	68.50	68.90+	67.20–	67.25–	68.50~	70.00+	68.50~	68.50~	65.15–	63.05–	65.20–
25	95.75	95.55–	95.70–	95.50–	95.50–	95.40–	95.75~	95.75~	93.85–	94.40–	95.95+
26	79.05	79.05~	77.35–	76.05–	75.90–	76.10–	79.05~	79.05~	77.35–	77.15–	77.90–
27	88.30	87.25–	87.80–	88.30~	88.30~	87.45–	88.30~	88.30~	88.30+	87.55–	87.65–
28	86.15	86.15~	72.00–	86.15~	86.15~	86.15~	84.85–	86.15~	88.10+	88.10+	88.15+
29	98.20	98.20~	96.95–	98.20~	98.20~	98.20~	98.20~	98.20~	98.25+	98.25+	98.30+
30	83.50	83.50~	83.20–	83.50~	83.50~	84.05+	83.50~	83.50~	84.40+	84.40+	82.50–
31	97.25	97.25~	97.25~	97.10–	97.10–	97.10–	97.25~	97.25~	97.55+	97.55+	97.55+
32	92.45	92.45~	91.00–	92.45~	92.45~	92.45~	92.45~	92.45~	92.75+	92.75+	92.75+
33	87.80	87.90+	86.85–	87.75–	87.80~	87.90+	87.80~	87.80~	87.75–	87.80~	87.95+
34	97.05	97.50+	97.35+	97.30+	97.30+	97.35+	97.05~	97.05~	97.60+	97.55+	97.55+
35	72.15	72.15~	70.55–	70.00–	70.20–	72.15~	70.80–	72.15~	70.85–	71.10–	70.05–
36	96.00	96.05+	96.25+	96.00~	96.00~	96.25+	96.00~	96.00~	96.00~	95.95–	95.80–
37	90.00	90.00~	89.00–	90.00~	90.00~	90.00~	90.00~	90.00~	90.25+	90.55+	90.55+
38	76.20	76.20~	69.70–	76.20~	76.20~	76.20~	76.20~	76.20~	79.30+	79.30+	79.30+
39	99.00	99.00~	98.75–	99.00~	99.00~	99.00~	99.00~	99.00~	98.85–	98.85–	99.15+
40	78.95	78.95~	79.85+	78.95~	78.95~	78.95~	78.95~	78.95~	79.80+	79.80+	80.45+
41	78.20	78.20~	73.90–	78.20~	78.20~	78.20~	78.20~	78.20~	78.05–	77.70–	77.90–
42	85.05	85.05~	85.50+	82.90–	85.05~	85.05~	85.05~	85.05~	83.45–	83.65–	84.25–
43	80.75	80.75~	78.50–	78.80–	79.05–	80.75~	80.75~	80.75~	82.45+	82.45+	82.45+
44	76.05	76.05~	73.20–	76.05~	76.05~	74.15–	76.05~	76.05~	76.35+	76.35+	76.35+
45	86.25	86.25~	83.80–	86.25~	83.50–	83.45–	86.10–	86.25~	86.30+	83.70–	85.95–
46	91.30	91.30~	91.20–	91.30~	91.30~	91.30~	91.30~	91.30~	92.10+	92.10+	92.10+
47	87.25	87.25~	81.60–	87.25~	87.25~	87.25~	87.25~	87.25~	85.45–	85.40–	80.95–
48	76.90	76.90~	73.05–	76.90~	76.90~	76.90~	76.90~	76.90~	78.05+	77.40+	77.35+
49	81.25	82.40+	82.65+	81.45+	81.80+	82.45+	81.25~	81.25~	80.85–	80.85–	81.75+
50	84.50	84.50~	85.00+	84.50~	84.50~	84.50~	84.50~	84.50~	83.40–	84.55+	85.45+
Médias taxa acerto	84.73	84.72	82.64	84.38	84.40	84.59	84.60	84.73	84.95	84.80	84.61
+ / ~ / –		5/42/3	12/1/37	3/35/12	4/35/11	8/35/7	0/44/6	0/50/0	29/2/19	26/1/23	25/2/23
Balanco		+2	-25	-9	-7	+1	-6	0	+10	+3	+2

linha + / ~ / – para cada TEC. Desse modo, na linha *Balanco* são salientadas as relações entre os desempenhos das TECs com os desempenhos obtidos sem uso de extração. Isso é feito computando, para cada TEC, a diferença entre suas quantidades de + e –. Esses resultados mostram que, para a maioria das TECs não-*classwise*, os valores da taxa de acerto foram com maior frequência inferiores aos valores de taxa obtidos sem uso de extração (balanco negativo). Além disso, a ordenação das TECs com os cinco maiores valores de balanco é: CdMD r_b , CdCMD^{30%} r_b , PCA, CdCMD^{60%} r_b e LFDA. Novamente as três TECs propostas estão entre as cinco técnicas com maiores valores, indicando a validade do emprego de técnicas *classwise* com investigação de multimodalidade para cenários com multimodalidade intraclasse.

Dados os valores das taxas de acerto das TECs para as 50 bases de dados utilizadas, é possível calcular o posicionamento relativo de cada técnica (DEMŠAR, 2006). Desse modo, a Tabela 33 apresenta os valores dos posicionamentos (*ranks*) de cada TEC em

cada base de dados. Nessa tabela, os menores valores de *rank* indicam as menores taxas de acerto obtidas para uma base de dados; enquanto que os maiores valores de *rank* indicam as maiores taxas de acerto obtidas. Através desses valores, é possível posicionar as TECs segundo seus desempenhos considerando todas as bases de dados utilizadas no experimento. Esses posicionamentos globais são apresentados na linha *Rank médio* dessa tabela. Assim, essa linha contém, para cada TEC, o valor médio dos *ranks* obtidos pela técnica para todas as bases de dados. Desse modo, as cinco TECs melhor posicionadas, considerando todas as 50 bases de dados, são: CdMD r_b , CdCDM^{60%} r_b , CdCDM^{30%} r_b , PCA e CdLPP $r_n(2)$. Esses melhores posicionamentos são justificados por essas técnicas obterem, com maior frequência, os maiores valores de *rank*.

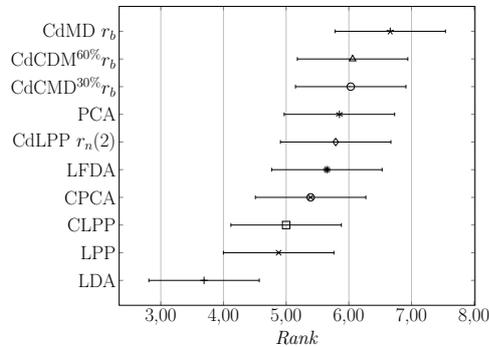
Tabela 33 – *Ranks* obtidos com as técnicas de extração de características calculados sobre os valores de taxa de acerto da Tabela 32.

<i>Toy</i>	PCA	LDA	LPP	CLPP	LFDA	CPCA	CdLPP $r_n(2)$	CdMD r_b	CdCDM ^{30%} r_b	CdCDM ^{60%} r_b
01	2,50	7,00	6,00	5,00	2,50	2,50	2,50	10,00	9,00	8,00
02	6,50	2,00	6,50	6,50	6,50	6,50	6,50	10,00	3,00	1,00
03	6,50	1,00	6,50	6,50	6,50	6,50	6,50	6,50	2,00	6,50
04	5,50	2,00	5,50	5,50	5,50	5,50	5,50	9,00	10,00	1,00
05	5,50	2,00	5,50	5,50	5,50	5,50	5,50	1,00	10,00	9,00
06	4,50	1,00	4,50	4,50	4,50	4,50	4,50	9,00	9,00	9,00
07	6,50	1,00	6,50	6,50	6,50	6,50	6,50	10,00	2,00	3,00
08	4,00	7,00	2,00	6,00	4,00	1,00	4,00	9,00	9,00	9,00
09	4,00	1,00	6,00	8,00	9,00	7,00	10,00	2,00	3,00	5,00
10	4,50	1,00	4,50	4,50	4,50	4,50	4,50	9,00	9,00	9,00
11	7,00	5,00	1,00	2,00	9,00	7,00	7,00	3,00	4,00	10,00
12	4,50	7,00	4,50	1,00	8,00	4,50	4,50	10,00	9,00	2,00
13	4,50	1,00	4,50	4,50	4,50	4,50	4,50	9,00	8,00	10,00
14	5,00	10,00	5,00	5,00	5,00	5,00	5,00	9,00	1,00	5,00
15	5,50	1,00	5,50	5,50	5,50	5,50	5,50	9,00	10,00	2,00
16	6,50	10,00	6,50	6,50	6,50	6,50	6,50	2,00	3,00	1,00
17	5,50	2,00	5,50	5,50	5,50	5,50	5,50	10,00	9,00	1,00
18	7,50	1,00	7,50	7,50	7,50	7,50	7,50	3,00	3,00	3,00
19	5,50	1,00	2,00	3,00	5,50	5,50	5,50	9,00	8,00	10,00
20	8,00	1,00	8,00	8,00	8,00	3,00	8,00	5,00	4,00	2,00
21	6,50	10,00	6,50	6,50	6,50	6,50	6,50	2,00	1,00	3,00
22	8,50	6,00	8,50	2,00	1,00	8,50	8,50	4,50	3,00	4,50
23	5,50	1,00	5,50	5,50	5,50	5,50	5,50	9,00	10,00	2,00
24	9,00	4,00	5,00	7,00	10,00	7,00	7,00	2,00	1,00	3,00
25	6,00	7,00	4,50	4,50	3,00	8,50	8,50	1,00	2,00	10,00
26	9,00	5,50	2,00	1,00	3,00	9,00	9,00	5,50	4,00	7,00
27	1,00	5,00	7,50	7,50	2,00	7,50	7,50	10,00	3,00	4,00
28	5,00	1,00	5,00	5,00	5,00	2,00	5,00	8,50	8,50	10,00
29	4,50	1,00	4,50	4,50	4,50	4,50	4,50	8,50	8,50	10,00
30	5,00	2,00	5,00	5,00	8,00	5,00	5,00	9,50	9,50	1,00
31	5,00	5,00	1,50	1,50	5,00	5,00	5,00	9,00	9,00	9,00
32	4,50	1,00	4,50	4,50	4,50	4,50	4,50	9,00	9,00	9,00
33	8,50	1,00	2,50	5,50	8,50	5,50	5,50	2,50	5,50	10,00
34	7,00	5,50	3,50	3,50	5,50	1,50	1,50	10,00	8,50	8,50
35	9,00	4,00	1,00	3,00	9,00	5,00	9,00	6,00	7,00	2,00
36	8,00	9,50	5,00	5,00	9,50	5,00	5,00	5,00	2,00	1,00
37	4,50	1,00	4,50	4,50	4,50	4,50	4,50	8,00	9,50	9,50
38	4,50	1,00	4,50	4,50	4,50	4,50	4,50	9,00	9,00	9,00
39	6,50	1,00	6,50	6,50	6,50	6,50	6,50	2,50	2,50	10,00
40	3,50	9,00	3,50	3,50	3,50	3,50	3,50	7,50	7,50	10,00
41	7,50	1,00	7,50	7,50	7,50	7,50	7,50	4,00	2,00	3,00
42	7,00	10,00	1,00	7,00	7,00	7,00	7,00	2,00	3,00	4,00
43	5,50	1,00	2,00	3,00	5,50	5,50	5,50	9,00	9,00	9,00
44	5,00	1,00	5,00	5,00	2,00	5,00	5,00	9,00	9,00	9,00
45	8,00	4,00	8,00	2,00	1,00	6,00	6,00	10,00	3,00	5,00
46	4,50	1,00	4,50	4,50	4,50	4,50	4,50	9,00	9,00	9,00
47	7,50	2,00	7,50	7,50	7,50	7,50	7,50	4,00	3,00	1,00
48	4,50	1,00	4,50	4,50	4,50	4,50	4,50	10,00	9,00	8,00
49	8,00	10,00	5,00	7,00	9,00	3,50	3,50	1,50	1,50	6,00
50	4,50	9,00	4,50	4,50	4,50	4,50	4,50	1,00	8,00	10,00
<i>Rank médio</i>	5,85	3,69	4,88	5,00	5,65	5,39	5,79	6,66	6,03	6,06

A Figura 48 apresenta um gráfico dos intervalos de confiança obtidos com o teste de Friedman para comparação dos desempenhos entre as TECs. Essa figura mostra que à 95% de confiança: (i) o LDA possui desempenho inferior ao PCA, LFDA, CdLPP $r_n(2)$, CdMD r_b , CdCDM^{30%} r_b e CdCDM^{60%} r_b ; (ii) assim como o LPP possui desempenho inferior ao

CdMD r_b ; (iii) sendo os demais pares de TECs estatisticamente equivalentes entre si.

Figura 48 – Comparação do desempenho entre as técnicas de extração de características com o teste de Friedman.



Fonte: Autoria própria.

4.3.4 Conclusões do experimento

Na Seção 4.3.3 são apresentados os resultados experimentais do desempenho das TECs aplicadas a cenários de classificação com multimodalidade intraclasse. Esse experimento foi particionado nas Seções 4.3.3.1 e 4.3.3.2.

Na Seção 4.3.3.1 foram comparados os desempenhos entre as diversas configurações que cada técnica *classwise* pode assumir. O CPCA foi melhor configurado utilizando apenas seu procedimento obrigatório; enquanto que o CdLPP obteve melhor configuração utilizando duas componentes da matriz de dispersão negativa. Além de encontrar a configuração mais adequada a cenários com multimodalidade intraclasse para as técnicas *classwise* convencionais, nessa seção foram também encontradas as configurações mais adequadas para as técnicas *classwise* com investigação de multimodalidade propostas. Para essas TECs, verificou-se que as configurações mais adequadas utilizam uma componente da matriz de dispersão entreclasses ($r_b = 1$). Adicionalmente, os limiares de decisão utilizados com o CdCMD e CdCDM foram respectivamente 30% e 60%.

Essas melhores configurações tiveram seus desempenhos comparados com as demais TECs não-*classwise* na Seção 4.3.3.2. Nessa seção, as técnicas *classwise* com investigação de multimodalidade propostas alcançaram as melhores posições da comparação das taxas de acerto entre todas as TECs para as 50 bases de dados. Dentre as técnicas propostas, CdMD foi a que obteve o melhor posicionamento, alcançando também o maior valor médio de taxa de acerto e o segundo maior número de reduções dimensionais efetivas (uni ou bidimensionais).

5 Considerações finais

Técnicas de Extração de Características (TECs) que não garantem a preservação da estrutura multimodal da dispersão dos padrões podem causar sobreposição entre as modas no espaço de redução dimensional calculado. A sobreposição faz com que a região entre as modas, que no espaço original possuía baixa densidade de padrões, passe a ter alta densidade de padrões. Como consequência, ruídos encontrados nessa região passam a ter mais elevada chance de serem erroneamente classificados. Devido a isso, nesse trabalho é proposta uma metodologia que possibilita a criação de TECs adaptadas a cenários com multimodalidade intraclasse. As capacidades comuns às TECs desenvolvidas através dessa metodologia são: (i) estimar a multimodalidade de dispersões unidimensionais de padrões e (ii) decompor um espaço multidimensional segundo as direções ortogonais que possuem maior evidência de comportamento multimodal em relação a dispersão dos padrões. A realização dessas capacidades estão implementadas em duas outras técnicas também propostas nesse trabalho, respectivamente: *Index of Area-based Normal Intersection* (IANI) e *Multimodal Decomposition* (MD). A Tabela 34 apresenta um sumário das técnicas que foram propostas com o desenvolvimento desse trabalho.

Tabela 34 – Sumário das técnicas propostas.

Técnica	Descrição
Técnicas <i>classwise</i> com investigação de multimodalidade	Metodologia para criação de TECs supervisionadas capazes de preservar a multimodalidade presente na dispersão de padrões intraclasse. As TECs criadas com essa metodologia: (i) determinam uma matriz de transformação específica para cada classe do problema; (ii) estimam a multimodalidade na dispersão dos padrões de cada classe e (iii) decompõem o espaço multidimensional original segundo as direções com maiores evidências de conter comportamento multimodal (Seção 3).
<i>Multimodal Decomposition</i> (MD)	TEC não-supervisionada que, dada a dispersão de padrões em um espaço multidimensional, decompõe esse espaço segundo direções ortogonais que possuem maiores evidências de preservar o comportamento multimodal original (Seção 3.2).
<i>Index of Area-based Normal Intersection</i> (IANI)	Métrica de estimativa de multimodalidade baseada na área de interseção entre modas unidimensionais que seguem distribuição Normal. Os maiores valores dessa métrica indicam ocorrência de comportamento multimodal (Seção 3.2.3.5).
<i>Class-dependent Multimodal Decomposition</i> (CdMD)	Instância de Técnica <i>classwise</i> com investigação de multimodalidade. Trata-se de uma TEC supervisionada que decompõe o espaço da dispersão dos padrões de cada classe independente de quão evidente seja a multimodalidade nesse espaço (Seção 3.4.1).
<i>Class-dependent Conditional Multimodal Decomposition</i> (CdCMD)	Instância de Técnica <i>classwise</i> com investigação de multimodalidade. Trata-se de uma TEC supervisionada que determina a matriz de transformação de uma classe específica através de MD, se e somente se, houver suficiente evidência de comportamento multimodal na dispersão dos padrões dessa classe (Seção 3.4.2).
<i>Class-dependent Conditional Detachment of Modes</i> (CdCDM)	Instância de Técnica <i>classwise</i> com investigação de multimodalidade. É uma variação do CdCMD que, para o caso de haver suficiente evidência de comportamento multimodal na dispersão dos padrões da classe, adiciona ao cálculo da matriz de transformação objetivo de elevar a separação entre essas modas no espaço de redução dimensional calculado (Seção 3.4.3).

De acordo com o objetivo que cada técnica proposta busca cumprir, foram realizados, na Seção 4, experimentos que comparam as proposições com técnicas de referência existentes

na literatura. Assim, na Seção 5.1 são apresentadas as conclusões referentes aos resultados experimentais que foram obtidos com cada experimento.

Adicionalmente, na Seção 5.2 são discutidos rumos futuros para a continuidade da pesquisa que aqui foi desenvolvida.

5.1 Conclusões experimentais

As Seções 4.1-4.3 apresentam os experimentos que foram realizados para avaliar as técnicas propostas em relação às técnicas de referência existentes na literatura. Os resultados desses experimentos mostram a eficiência das técnicas propostas na realização de seus objetivos individuais. A Tabela 35 apresenta um sumário das avaliações de desempenho das técnicas propostas quando comparadas com suas respectivas técnicas rivais.

Tabela 35 – Sumário das avaliações de desempenho das técnicas propostas.

Técnica	Avaliação do desempenho
IANI	Alcançou, com maior frequência, os maiores valores médios de desempenho. Salientando-se nos cenários cujo desbalanceamento entre as modas é mais severo. Contudo, na comparação mais ampla (sem restrições quanto ao desbalanceamento), essa métrica é estatisticamente equivalente às métricas rivais.
MD	Obteve desempenho médio superior às TECs rivais. Conseguindo realizar correta classificação de ruídos mesmo quando a configuração do classificador é desfavorável a essa categorização.
CdMD	Atinge o primeiro melhor posicionamento médio quanto ao desempenho de classificação em relação as demais TECs comparadas. Possui desempenho superior ao LDA e LPP. Contudo, é estatisticamente equivalente as demais TECs rivais.
CdCMD	Atinge o segundo melhor posicionamento médio quanto ao desempenho de classificação em relação as demais TECs comparadas. Possui desempenho superior ao LDA. Contudo, é estatisticamente equivalente as demais TECs rivais.
CdCMD	Atinge o terceiro melhor posicionamento médio quanto ao desempenho de classificação em relação as demais TECs comparadas. Possui desempenho superior ao LDA. Contudo, é estatisticamente equivalente as demais TECs rivais.

Desse modo, dado que CdMD, CdCMD e CdCDM obtiveram os melhores posicionamentos na comparação dos desempenhos entre as TECs, é válida a aplicação da metodologia proposta para o desenvolvimento de TECs capazes de preservar a multimodalidade dos padrões intraclasses.

Na Seção 4.1 (página 71) são apresentados dois experimentos para a comparação do desempenho entre as métricas de estimativa de multimodalidade empregadas (Seção 3.2.3, página 51). O primeiro experimento avaliou o desempenho dessas métricas em auxiliar na determinação da direção com maior evidência de conter comportamento multimodal. A comparação mais ampla entre as métricas (considerando as 90 bases de dados utilizadas) mostra a equivalência estatística que há entre elas. Contudo, dado que o desbalanceamento entre as modas dificulta a percepção de comportamento multimodal, a comparação do desempenho das métricas como função desse desbalanceamento mostra que IANI obtém a maioria dos melhores desempenhos médios dentre os cenários em que o desbalanceamento é mais severo ($\rho \geq 35\%$ na Tabela 5, página 77). O segundo experimento avaliou o de-

sempenho de classificação obtido por MD utilizando as diferentes métricas de estimativa de multimodalidade como função de custo. Com respeito as 24 bases utilizadas nesse experimento, IANI é estatisticamente equivalente às métricas de estimativa de multimodalidade *dip* e *Silverman*. Entretanto, a maioria dos desempenhos médios de classificação são alcançados por IANI. Desse modo, ambos experimentos salientam a utilização de IANI como métrica de estimativa de multimodalidade mais adequada para busca pelas direções com maiores evidências de conter comportamento multimodal.

Na Seção 4.2 (página 82) o desempenho de classificação de MD foi comparado aos desempenhos do PCA e LPP em cenários com multimodalidade. Para o experimento realizado, ruídos foram colocados entre as médias das modas de cada base. Desse modo, o sucesso em corretamente classificar esses ruídos exprime o sucesso da TEC em preservar a multimodalidade. Os resultados experimentais atestam o superior desempenho de MD em relação as TECs rivais, em cenários com multimodalidade. Sendo essa superioridade verificada através de teste de Friedman à 95% de confiança e com limiar de decisão em 0,5. Demais resultados experimentais mostram que MD obtém elevado desempenho de classificação (em relação as TECs rivais) mesmo quando o limiar de decisão não favorece a correta classificação dos ruídos.

Na Seção 4.3 (página 88) são comparados os desempenhos de classificação com diversas TECs. Além das TECs de referência, foram comparadas as TECs desenvolvidas através da metodologia proposta: CdMD, CdCMD e CdCDM. O experimento apresentado nessa seção empregou 50 bases de dados com duas classes. Os padrões de cada classe seguem distribuição dada pela mistura de duas Normais, de modo a apresentarem diversas ocorrências de comportamento multimodal. Os resultados obtidos mostram que as melhores configurações das TECs propostas alcançam, com maior frequência, os maiores valores médios de desempenho. Tendo a melhor configuração do CdMD obtido a melhor posição na avaliação de desempenho dentre todas as demais TECs avaliadas.

5.2 Trabalhos futuros

Os resultados experimentais apresentados nessa trabalho evidenciam a eficiência das técnicas propostas quando aplicadas a cenários com multimodalidade. Contudo, mesmo havendo alcançado os melhores posicionamentos na comparação de desempenho, IANI, CdMD, CdCMD e CdCDM apresentam desempenhos que são estatisticamente equivalentes aos desempenhos de algumas de suas respectivas técnicas rivais. Desse modo, rumos futuros à continuação dessa pesquisa podem visar a melhoria dos desempenhos dessas técnicas e/ou a ampliação de seus escopos.

As Seções 5.2.1-5.2.3 apresentam três propostas para a continuação dessa pesquisa. Os desafios referentes ao emprego de bases reais para a realização dessa pesquisa são apre-

sentados na Seção 5.2.1. Na Seção 5.2.2 é discutido o desenvolvimento de um classificador que categoriza as dispersões quanto uni- ou multimodal. Enquanto que a Seção 5.2.3 trata da extensão de IANI para espaços multidimensionais.

5.2.1 Emprego de bases reais

O emprego de bases sintéticas possibilitam controle sobre as características da dispersão dos padrões, de modo a possibilitar inferências precisas quanto a observação dos resultados experimentais obtidos para específicos comportamentos. No caso dessa pesquisa, as bases sintéticas possibilitaram controle sobre a ocorrência da multimodalidade na dispersão dos padrões, o que permitiu avaliar as técnicas garantindo a ocorrência desse comportamento desejado. Todavia, o emprego de bases reais na avaliação das técnicas conferem às conclusões maior confiabilidade. Essa confiabilidade se deve tanto ao fato dessas bases representarem problemas de importância/aplicação real, quanto por usualmente serem largamente conhecidas e disponíveis ao público, o que possibilita replicação dos experimentos apresentados. Entretanto, alguns fatores impediram a utilização de bases reais nesse trabalho. Esses fatores são referentes a complexidade em estimar a multimodalidade em cenários não controlados, o que torna também complexa ou imprecisa a interpretação dos resultados obtidos. São exemplos desses fatores: (i) ocorrência de comportamento multimodal, (ii) adaptação das técnicas às diferentes funções de distribuição que podem reger a dispersão dos padrões e (iii) realizar estimativa de multimodalidade para dispersões compostas por mais do que duas modas.

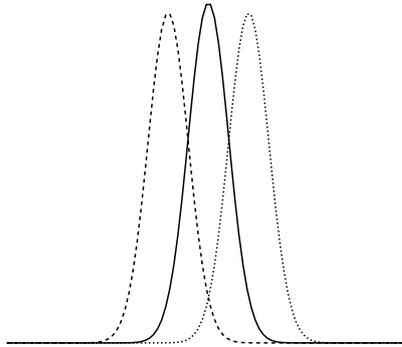
(i) É necessário inferir corretamente a ocorrência de multimodalidade na dispersão dos padrões, a fim de possibilitar correta interpretação dos desempenhos das técnicas propostas para operar sob esse tipo de cenário. Porém, anteriormente a esse trabalho, não foi encontrada na literatura técnica que possibilitasse a inferência de comportamento uni- ou multimodal para dispersão dos padrões em espaços multidimensionais. Sendo essa inferência usualmente realizada através da experiência de um analista utilizando procedimentos que possibilitam a visualização da dispersão; ou seja, projeção dos padrões para espaços bi- ou tridimensionais.

(ii) As técnicas propostas nesse trabalho assumem que a dispersão dos padrões segue distribuição Normal. Essa restrição não é verdadeira para muitas bases de dados reais. Além disso, nem sempre é correto/possível aproximar outras distribuições por distribuição Normal. Assim, implementações futuras dessas técnicas, aplicadas a bases reais, requerem suas modelagens para diferentes funções de distribuição. A complexidade matemática dessas novas modelagens se torna ainda maior se for considerada a possibilidade de cada moda seguir diferente função de distribuição.

(iii) Cenário de bimodalidade é uma restrição comum as técnicas propostas nesse trabalho. Essa restrição se deve a complexidade envolvida em estimar a multimodalidade de

dispersões com mais do que duas modas. Sendo essa complexidade ilustrada na Figura 49.

Figura 49 – Complexidade na estimativa de multimodalidade com mais de duas modas.



Fonte: Autoria própria.

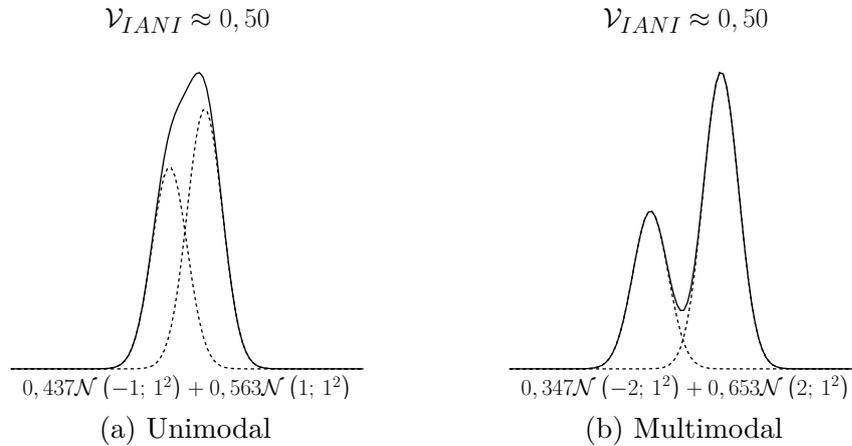
Nessa figura, é menor a evidência de comportamento multimodal entre os pares de Normais *sólida-tracejada* e *sólida-pontilhada* em relação ao comportamento multimodal expressado pelas Normais *tracejada-pontilhada*. Assim, como compor um valor de estimativa de multimodalidade para toda a dispersão dos padrões com base nos três valores de estimativa obtidos para os três pares de Normais da figura?

Assim, a continuação dessa pesquisa pode envolver o desenvolvimento de melhorias das técnicas propostas de modo a possibilitar suas aplicações em cenários menos restritos e, desse modo, possibilitando o emprego de bases reais.

5.2.2 Classificador de multimodalidade

Os resultados experimentais da Seção 4.3 (página 88) mostram que, a ordem decrescente das melhores posições quanto a comparação dos desempenhos de classificação foram atingidas respectivamente com CdMD, CdCDM e CdCMD. Contudo, essas TECs são estatisticamente equivalentes entre si. É possível justificar que o menor desempenho médio de CdCMD e CdCDM em relação a CdMD se deva ao critério de decisão adotado para inferência da categoria da dispersão dos padrões de cada classe (uni- ou multimodal). Tanto CdCMD quanto CdCDM realizam essa categorização através de limiarização sobre o valor de IANI, que é obtido para a primeira direção retornada por MD, a qual possui maior evidência de conter comportamento multimodal. Contudo, um mesmo valor de IANI pode ser obtido por diferentes configurações dos parâmetros dos modelos de mistura que aproximam a dispersão dos padrões. Assim, mesmo que duas dispersões possuam categorias distintas, elas podem ser assumidas como pertencentes a mesma categoria através do critério adotado por CdCMD e CdCDM. Essa situação é ilustrada na Figura 50. As Figuras 50a e 50b apresentam modelos de mistura com comportamentos respectivamente unimodal e multimodal. Todavia, os valores de IANI obtidos para seus modelos de mistura são aproximadamente iguais.

Figura 50 – Confusão na categorização de dispersões devido a limiarização sobre o valor de IANI. O modelo de mistura de 50a apresenta comportamento unimodal. Mesmo assim, o valor de IANI obtido para esse modelo é aproximadamente igual ao valor de IANI obtido para 50b, que apresenta comportamento multimodal.



Fonte: Autoria própria.

Assim, como mostram os resultados experimentais da Seção 4.1 (página 71), as métricas de estimativa de multimodalidade empregadas são estatisticamente equivalentes entre si. Contudo, esses resultados também mostraram que, sob diferentes condições, uma dada métrica pode obter melhores desempenhos do que as demais. Entretanto, mapear as possíveis condições de aplicação com a correspondente métrica de estimativa de multimodalidade melhor adaptada à essa condição não é uma tarefa trivial.

Desse modo, nessa seção é proposta a construção de um classificador de multimodalidade forte, a ser utilizado como critério de decisão quanto a categoria (uni- ou multimodal) a que pertence uma dada dispersão de padrões. Assim, dado que diferentes métricas estimam a multimodalidade com diferentes precisões sob diferentes condições, a modelagem desse classificador infere a categoria de uma dispersão a partir do conjunto dos valores (vetor de características) referentes aos valores dos parâmetros que aproximam a dispersão, as estimativas de multimodalidade realizadas por diferentes métricas e outros valores interessantes que podem favorecer a discriminação das dispersões entre essas categorias.

As tarefas iniciais associadas a criação desse classificador são: (i) modelagem do vetor de características, (ii) definição da arquitetura do classificador e (iii) construção/obtenção de base de dados para treinamento e teste do classificador.

(i) Sendo os parâmetros que aproximam a dispersão dos padrões, por modelo de mistura entre duas modas Normais, dados por: $\tilde{\mu}_1, \tilde{\mu}_2$ (médias das modas), S_1, S_2 (desvios padrão das modas) e ρ (valor do peso da moda 1 na mistura). Assim, o vetor de características, além de conter esses parâmetros, pode conter os valores das métricas de estimativa de multimodalidade para a dispersão associada. Adicionalmente, outros valores

podem ser adicionados ao vetor de característica. Dado que, mesmo que haja redundância de informação com uso de outros valores, um classificador forte pode minerar a informação útil do amontoado de características. Desse modo, as Equações 5.1-5.6 apresentam componentes adicionais ao vetor de características para o classificador proposto.

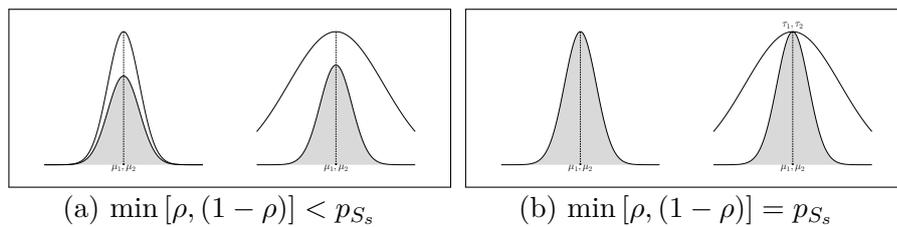
$$d_{\mu_s} = |\tilde{\mu}_1 - \tilde{\mu}_2| \quad (5.1) \quad \left| \quad p_{S_s} = \frac{\min(S_1, S_2)}{(S_1 + S_2)} \quad (5.2)\right.$$

$$d_b = \frac{d_{\mu_s}}{2(S_1 + S_2)} \quad (5.3) \quad \left| \quad a_i = \frac{\rho_i}{S_i \sqrt{2\pi}} \quad \text{com } i \in \{1, 2\} \quad (5.4)\right.$$

$$a_q = \frac{\min(a_1, a_2)}{\max(a_1, a_2)} \quad (5.5) \quad \left| \quad a_d = \frac{d_{\mu_s}}{a_1 + a_2} \quad (5.6)\right.$$

d_{μ_s} calcula a distância euclidiana entre as médias $\tilde{\mu}_1$ e $\tilde{\mu}_2$. Algumas configurações dos parâmetros da mistura fazem com que uma das modas esteja inteiramente contida sob a outra. Nesses casos, quando $\tilde{\mu}_1 = \tilde{\mu}_2$, a distinção entre os cenários para os quais não existem pontos de interseção entre as modas (Figura 51a) e os cenários com apenas um ponto de interseção entre as modas (Figura 51b) é dado por p_{S_s} , como mostra a Figura 51.

Figura 51 – Controle do número de interseções quando uma das modas Normais está inteiramente contida sob a outra, sendo suas médias iguais.



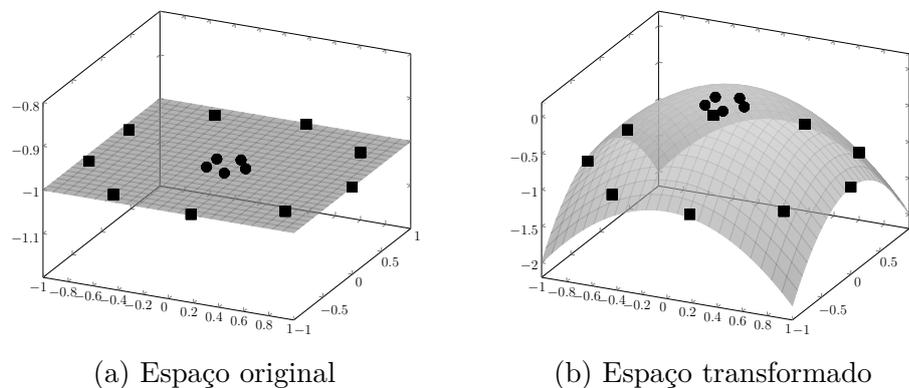
Fonte: Autoria própria.

Zhang, Mapes e Soden (2003) sugerem d_b como parâmetro para expressar a sobreposição entre as modas. a_i é a amplitude da moda $i \in \{1, 2\}$. a_q é obtido de forma similar a ρ_q e representa o desbalanceamento global entre as modas. Desse modo, $a_q \rightarrow 0$ implica em desbalanceamento acentuado entre os picos das modas e $a_q \rightarrow 1$ significa que seus picos tendem ao mesmo valor. a_d é modelada para relacionar as amplitudes com a distância entre as médias, segundo a premissa empírica de que o comportamento multimodal será acentuado para os maiores valores dessa razão.

(ii) Dado que a classificação de multimodalidade é um problema de duas classes, pode-se modelar esse classificador por *Support Vector Machine* (SVM). SVM (WANG, 2008) é um algoritmo supervisionado de aprendizado de máquina que busca por hiperplanos que separam os padrões das diferentes classes de forma ótima. Caso as classes do

problema apresentado ao SVM sejam linearmente separáveis, este encontra hiper plano que possibilita a classificação sem erros no conjunto de treinamento. Caso contrário, SVM busca um conjunto de hiper planos que juntos possibilitem uma precisa classificação. Faz parte da estratégia do SVM a transformação do espaço de características original em um espaço de características de maior dimensionalidade. Essa transformação é realizada segundo uma função de *kernel* e tem como objetivo facilitar a descoberta de hiper planos que melhor separem as classes. A Figura 52 ilustra esse procedimento.

Figura 52 – Transformação do espaço de características realizada pelo SVM. 52a apresenta o espaço de características onde é difícil a determinação de hiper planos de separação. 52b apresenta a transformação do espaço original segundo uma função polinomial.



Fonte: Autoria própria.

A Figura 52a apresenta o espaço de características original onde é difícil a determinação de hiperplanos de separação. Enquanto que na Figura 52b, o espaço original é transformado segundo uma função polinomial. Nesse novo espaço a obtenção de hiper plano que separa os padrões das classes ■ e ● são mais facilmente obtidos.

(iii) Como SVM utiliza método de aprendizado supervisionado, é necessário ao seu treinamento a apresentação de padrões previamente categorizados. Contudo, não se encontrou na literatura base que corresponda as características do problema em questão. Todavia, é simples a geração de padrões que representam modelos de mistura entre modas Normais. Entretanto, não há prévia rotulagem desses padrões gerados. Desse modo, é um problema em aberto a determinação de base de treinamento representativa, que possibilite suficiente grau de generalização por parte do classificador de multimodalidade.

5.2.3 Métrica de estimativa de multimodalidade multidimensional

IANI e as demais métricas de estimativa de multimodalidade utilizadas nesse trabalho estão restritas a cenários unidimensionais. Desse modo, antes de aplicar qualquer dessas métricas, faz-se necessário projetar os padrões do espaço original (multidimensional) para algum espaço unidimensional. Como reduções dimensionais envolvem perda de informação, não há garantia de igualdade entre a multimodalidade da dispersão dos

padrões no espaço original (sem perda de informação) e o máximo valor de estimativa de multimodalidade que pode ser encontrado para alguma direção nesse espaço. Portanto, a capacidade de estimar multimodalidade no espaço original da dispersão dos padrões é uma possibilidade a melhoria do desempenho na realização dessa tarefa.

Considerando que a dispersão dos padrões segue distribuição composta pela mistura de duas Normais em um espaço d -dimensional ($d > 1$), a modelagem de IANI pode ser estendida para cenários multidimensionais. Nesse caso, em analogia com a Equação 3.16 (página 55), a versão multidimensional de IANI pode ser expressa através da Equação 5.7.

$$\mathcal{V}_{IANI} = \rho_q \left(1 - \frac{V_{\cap}}{V_{\cap}^*} \right) \quad (5.7)$$

Sendo V_{\cap} o volume de interseção entre as modas e V_{\cap}^* o máximo volume de interseção entre as modas que pode ser obtido com o deslocamento de suas médias.

Cada Normal multidimensional é calculada através da Equação 5.8, sendo $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ respectivamente a média e a matriz de covariância que modelam essa distribuição no espaço d -dimensional.

$$\mathcal{N}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} e^{\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \right]} \quad (5.8)$$

Todavia, seguindo a analogia com o caso unidimensional, determinar o volume de interseção a partir da igualdade entre as Normais é um problema severamente complexo. Como alternativa, o intervalo de previsão para distribuições Normais multidimensionais possibilita a determinação de uma casca elipsoidal que expressa determinada probabilidade de conter os padrões produzidos segundo a respectiva função de distribuição Normal. A Equação 5.9 exprime o intervalo de previsão correspondente a distribuição Normal da Equação 5.8.

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \leq \chi_d^2(p) \quad (5.9)$$

Nessa equação, $\chi_d^2(p)$ é a função quantile para a probabilidade p da distribuição χ^2 com d graus de liberdade.

Desse modo, o cálculo de V_{\cap} com confiança igual a p consiste em determinar o volume de interseção entre os intervalos de previsão das distribuições Normais envolvidas na mistura.

Referências

- BALAKRISHNAMA, S.; GANAPATHIRAJU, A. Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, v. 18, 1998.
- BAPAT, R. B. *Linear Algebra and Linear Models (Universitext)*. 2. ed. [S.l.]: Springer, 2000.
- BEHBOODIAN, J. On the modes of a mixture of two normal distributions. *Technometrics*, v. 12, n. 1, p. 131–139, 1970.
- BELLMAN, R. E. *Adaptive control processes - A guided tour*. [S.l.]: Princeton University Press, 1961.
- BIERNACKI, C.; CELEUX, G.; GOVAERT, G. Choosing starting values for the em algorithm for getting the highest likelihood in multivariate gaussian mixture models. *Computational Statistics & Data Analysis*, v. 41, n. 3-4, p. 561–575, 2003.
- BILMES, J. *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*. [S.l.], 1998.
- CHAMBERS, L. D. *The Practical Handbook of Genetic Algorithms: Applications*. 2. ed. [S.l.]: CRC Press, 2001.
- CHAN, K. S.; TONG, H. Testing for multimodality with dependent data. *Biometrika*, v. 91, n. 1, p. 113–123, 2004.
- CHUNG, F. R. K. *Spectral Graph Theory*. [S.l.]: American Mathematical Society, 1997. (Regional conference series in mathematics, 92).
- DAS, K.; NENADIC, Z. An efficient discriminant-based solution for small sample size problem. *Pattern Recognition*, v. 42, n. 5, p. 857–866, 2009.
- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, v. 7, p. 1–30, 2006.
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2. ed. [S.l.]: Wiley-Interscience, 2001.
- EISENBERGER, I. Genesis of bimodal distributions. *Technometrics*, v. 6, n. 4, p. 357–363, 1964.
- FISCHER, N. I.; MAMMEN, E.; MARRON, J. S. Testing for multimodality. *Computational Statistics & Data Analysis*, v. 18, n. 5, p. 499–512, 1994.
- FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals Eugen.*, v. 7, p. 179–188, 1936.
- FRIEDMAN, J. H.; TUKEY, J. W. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, C-23, n. 9, p. 881–890, 1974.

-
- GNANADESIKAN, M.; GUPTA, S. S. A selection procedure for multivariate normal distributions in terms of the generalized variances. *Technometrics*, v. 12, n. 1, p. 103–117, 1970.
- HALL, P.; YORK, M. On the calibration of silverman’s test for multimodality. *Statistica Sinica*, v. 11, p. 515–536, 2001.
- HARTIGAN, J. A. *Clustering algorithms*. [S.l.]: Wiley, 1975. (Wiley series in probability and mathematical statistics: Applied probability and statistics).
- HARTIGAN, J. A.; HARTIGAN, P. M. The dip test of unimodality. *Annals of Statistics*, v. 13, n. 1, p. 70–84, 1985.
- HARTIGAN, J. A.; WONG, M. A. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society*, v. 28, n. 1, p. 100–108, 1979. Series C (Applied Statistics).
- HASSAN, R. et al. A comparison of particle swarm optimization and the genetic algorithm. In: *Multidisciplinary design optimization specialist conference*. Austin: [s.n.], 2005. p. 18–21.
- HE, X.; NIYOGI, P. Locality preserving projections. In: *Advances in Neural Information Processing Systems*. [S.l.]: MIT Press, 2003. p. 153–160.
- JOLLIFFE, I. *Principal component analysis*. [S.l.]: Wiley Online Library, 2002.
- JONES, M. C.; SIBSON, R. What is projection pursuit? *Journal of the Royal Statistical Society*, v. 150, n. 1, p. 1–37, 1987. Series A (General).
- KAMEYAMA, K. Particle swarm optimization - a survey. *IEICE Transactions on Information and Systems*, E92.D, n. 7, p. 1354–1361, 2009.
- KENNEDY, J.; EBERHART, R. Particle swarm optimization. In: *International Conference on Neural Networks*. Perth: IEEE, 1995. v. 4, p. 1942–1948.
- LEITE, B. d. B. et al. A learning-based eye detector coupled with eye candidate filtering and pca features. In: *Brazilian Symposium on Computer Graphics and Image Processing*. Minas Gerais: IEEE, 2007. p. 187–194.
- LI, J.-B.; PAN, J.-S.; CHU, S.-C. Kernel class-wise locality preserving projection. *Information Sciences*, v. 178, n. 7, p. 1825–1835, 2008.
- MINNOTTE, M. C. Nonparametric testing of the existence of modes. *The Annals of Statistics*, v. 25, n. 4, p. 1646–1660, 1997.
- NASON, G. P.; SIBSON, R. Measuring multimodality. *Statistics and Computing*, v. 2, n. 3, p. 153–160, 1992.
- NAZ, E.; FAROOQ, U.; NAZ, T. Analysis of principal component analysis-based and fisher discriminant analysis-based face recognition algorithms. In: *International Conference on Emerging Technologies*. Peshawar: IEEE, 2006. p. 121–127.
- RAY, S.; LINDSAY, B. G. The topography of multivariate normal mixtures. *The Annals of Statistics*, v. 33, n. 5, p. 2042–2065, 2005.

-
- SARMA, P. V. S.; RAO, K. S. S.; RAO, R. P. On a family of bimodal distributions. *Sankhya-: The Indian Journal of Statistics, Series B (1960-2002)*, v. 52, n. 3, p. 287–292, 1990.
- SAWITZKI, G. The excess mass approach and the analysis of multi-modality. In: *Annual Conference of the GfKl*. Heidelberg: Springer, 1996. p. 203–211.
- SEBER, G. A. *Multivariate Observations*. [S.l.]: John Wiley & Sons, 2004. v. 547.
- SILVA JR, E. R.; CAVALCANTI, G. D.; REN, T. I. Class-dependent locality preserving projections for multimodal scenarios. In: IEEE. *International Conference on Tools with Artificial Intelligence*. Athens, 2012. v. 1, p. 982–987.
- SILVA JR, E. R.; CAVALCANTI, G. D. C.; REN, T. I. Does the affinity matrix influence the performance of the locality preserving projection algorithm? In: IEEE. *International Conference on Systems Man and Cybernetics*. Istanbul, 2010. p. 4169–4175.
- SILVERMAN, B. W. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society*, v. 43, n. 1, p. 97–99, 1981. Series B (Methodological).
- SUGIYAMA, M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of Machine Learning Research*, v. 8, p. 1027–1061, 2007.
- WANG, G. A survey on training algorithms for support vector machine classifiers. In: *International Conference on Networked Computing and Advanced Information Management*. Gyeongju: IEEE, 2008. v. 1, p. 123–128.
- YE, J. et al. Feature extraction via generalized uncorrelated linear discriminant analysis. In: *International Conference on Machine Learning*. Banff: ACM, 2004. p. 113–120.
- ZHANG, C.; MAPES, B. E.; SODEN, B. J. Bimodality in tropical water vapour. *Quarterly Journal of the Royal Meteorological Society*, v. 129, n. 594, p. 2847–2866, 2003.