



**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
GRADUAÇÃO EM ESTATÍSTICA**

Uso do Spline para Reconhecimento de Assinaturas

Byanka Karoline Matias Da Silva

Recife – PE

2022



Byanka Karoline Matias da Silva

Uso do spline para reconhecimento de assinaturas

Trabalho de Conclusão de Curso apresentado no Departamento de Estatística da UFPE como requisito básico para a conclusão do Curso de Graduação em Estatística.

Orientador (a): Prof. Dr. Manoel Raimundo de Sena Jr.

Recife

2022

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Silva, Byanka Karoline Matias da.

Uso do Spline para Reconhecimento de Assinaturas / Byanka Karoline
Matias da Silva. - Recife, 2023.

18 p.

Orientador(a): Manoel Raimundo de Sena Jr

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de
Pernambuco, Centro de Ciências Exatas e da Natureza, Estatística -
Bacharelado, 2023.

1. Spline. 2. Dynamic time warping. 3. Comparação de assinaturas. I. Sena
Jr, Manoel Raimundo de. (Orientação). II. Título.

310 CDD (22.ed.)

RESUMO

O trabalho a seguir busca comparar assinaturas falsas e verdadeiras, a partir da aplicação dos métodos Spline e DTW que serviram para ajustar os dados para a obtenção de uma assinatura média, para então encontrar um intervalo de banda que irá mostrar a proporção de aceitação de uma assinatura verdadeira e rejeição de uma assinatura falsa.

Palavras-chave: DTW; Spline; Comparação de assinaturas

ABSTRACT

The following monograph want to compare false and true signatures, using the Spline and DTW methods that served to adjust the data to obtain an average signature, and then find a bandwidth interval that will show the proportion of acceptance of a true signature and rejection of a false signature.

Key-words: DTW; Spline; Signatures comparison

SUMÁRIO

RESUMO	3
ABSTRACT	4
INTRODUÇÃO	6
OBJETIVOS	7
OBJETIVO GERAL	7
OBJETIVO ESPECÍFICO	7
METODOLOGIA	8
1. Ajuste dos dados	8
1.1 Modelagem de uma Spline	8
1.2 Dynamic Time Warping (DTW)	10
6. Erro médio da distância	11
7. Banda de confiança do erro médio	12
7.1 Taxa do erro	12
8. O conjunto de Dados	12
RESULTADOS	13
1. Modelagem de uma assinatura	13
2. Criação da assinatura média	15
CONCLUSÃO	17
REFERÊNCIAS BIBLIOGRÁFICAS	18

INTRODUÇÃO

Os Sistemas de Reconhecimentos de Padrões (RP), tem como objetivo decidir se um novo objeto, supostamente desconhecido, pertence a um padrão. Em geral, é associada alguma medida de distância como regra de decisão, isto é, se a distância entre um ponto do novo objeto e um ponto do objeto padrão for menor que um limiar, o novo objeto pode ser considerado como pertencente ao mesmo grupo do objeto padrão.

A comparação de assinaturas é uma técnica utilizada para determinar se duas assinaturas foram feitas pela mesma pessoa. Para fazer essa comparação de forma objetiva, é possível utilizar métodos estatísticos para avaliar as semelhanças e diferenças entre as assinaturas.

É importante ressaltar que a comparação de assinaturas com métodos estatísticos não é uma técnica infalível e pode apresentar limitações, principalmente quando as assinaturas são muito diferentes ou quando há interferências externas, como tremores nas mãos ou mudanças na posição do papel.

Com a construção de uma assinatura padrão, podemos obter uma medida de distância entre ela e uma nova, supostamente desconhecida, e compará-las. A distribuição dessa distância pode ser levada em conta na construção da regra de decisão, sendo o limiar de referência fixado de modo a se obter um nível de significância adequado.

Numa aplicação em verificação de assinaturas, Sena Jr têm discutidos algumas medidas de distância com base na distância euclidiana e a de Mahalanobis., principalmente em casos de assinaturas off-line, que é o caso em que não temos informações sobre como a assinatura foi coletada, tendo apenas como informação a imagem da mesma e nada mais, enquanto a assinatura on-line tem informações sobre a coleta dessa assinatura, como o tempo em que ela foi realizada, a pressão utilizada na caneta etc. Já o Zhong tem usado a curva spline para descrever o comportamento contínuo do modelo de Cox, em curvas de Análise de Sobrevida. Essa ideia pode ser aplicada ao caso on-line, em que o objetivo é descrever cada curva da assinatura, quando comparadas a cada um dos eixos (X, Y), tendo como referência o tempo. A curva spline corrige os pontos de descontinuidade inerente ao processo de captura da assinatura, composta de pontos igualmente espaçados.

OBJETIVOS

OBJETIVO GERAL

O presente trabalho tem como principal objetivo analisar a diferença entre assinaturas verdadeiras e falsas, de modo que seja possível diferenciar ambas as assinaturas, através da curva spline e do dynamic time warping.

OBJETIVO ESPECÍFICO

- Modelar as curvas através do uso do spline e do dynamic time warping;
- Encontrar a assinatura padrão (assinatura média)
- Calcular o erro médio das assinaturas, que é a média da distância entre a assinatura padrão e a assinatura que se deseja comparar;
- Criar uma banda de confiança para avaliar se devemos ou não aceitar uma assinatura, a um nível de 95% de confiança;
- Minimizar o erro tipo I, de rejeitar uma assinatura quando ela é verdadeira e o erro tipo II de não rejeitar uma assinatura quando ela é falsa.

METODOLOGIA

Para a modelagem das assinaturas, serão utilizadas as técnicas Spline para a suavização das curvas e erros de continuidade no gráfico das assinaturas e dynamic time warping (DTW) para ajustar os tamanhos de cada assinatura, para que fiquem do mesmo tamanho ao criar seus gráficos, além de funções criadas manualmente para auxílio das análises.

1. Ajuste dos dados

Os dados foram suavizados por meio do Spline natural, para que cada ponto seja ligado de modo que não tenham pontas e as curvas sejam tão naturais ao fazer o gráfico quanto ao realizar a assinatura no papel e será também aplicado o DTW, para que

1.1 Modelagem de uma Spline

A curva Spline é uma aproximação linear que permite ao usuário criar uma curva que se ajusta a determinados pontos de dados que são chamados de “nós”, escolhendo os mesmos de forma distinta no intervalo de observações e assim retornando uma lista de pontos, de forma que se possa modelar curvas complexas por polinômios simples, podendo alterar a forma da curva apenas movendo os nós.

Uma curva tem a forma geral:

$$y = a + bx + cx^2 + dx^3 + \dots$$

O grau é definido pelo maior coeficiente não-zero, geralmente é mais utilizada na computação gráfica para construir curvas suaves de grau 3, conhecida como polinomial cúbica, porque é o menor grau que se pode construir uma inflexão, criando curvas interessantes e se comporta bem graficamente, por criar curvas suaves com poucas curvas.

Existem três tipos de métodos de interpolação:

- **Spline linear:** é definida através dos nós que fazem uma ligação dos conjuntos de polinômios de grau um (para um conjunto de pontos ordenados). Esse polinômio chamado $S_1(x)$ é calculado com base em x , $f(x)$, nos nós x_0, x_1, \dots, x_n , que faz com que os pontos sejam ligados por linhas retas, sem curvas, mas de forma contínua. Pode ser escrita em cada subintervalo $[x_{i-1}, x_i]$, $i = 1, 2, \dots, n$, como

$$s_{i(x)} = f(x_{i-1}) \frac{(x-x_i)}{x_i-x_{i-1}} + f(x_i) \frac{(x-x_{i-1})}{x_i-x_{i-1}}, \quad \forall x \in [x_{i-1}, x_i]$$

Exemplo:

Sejam os pontos de $(x, f(x))$: (1, 2), (-3, 1), (2, -5), (4, 2), (0, 5), (-4, 7) e (6, -8)

Temos a seguinte tabela:

x	-4	-3	0	1	2	4	6
f(x)	7	1	5	2	-5	2	-8

Para achar a função spline, precisamos interpolar $s_i(x)$, tal qual:

$$s_1(x) = f(x_0) \frac{(x_1-x)}{x_1-x_0} + f(x_1) \frac{(x-x_0)}{x_1-x_0} = 7 \frac{(-3-x)}{-3+4} + 1 \frac{(x+4)}{-3+4} = -6x - 17, x \in (-4, -3)$$

$$s_2(x) = f(x_1) \frac{(x_2-x)}{x_2-x_1} + f(x_2) \frac{(x-x_1)}{x_2-x_1} = \frac{(-x)}{3} + 5 \frac{(x+3)}{3} = \frac{4x+15}{3}, x \in (-3, 0)$$

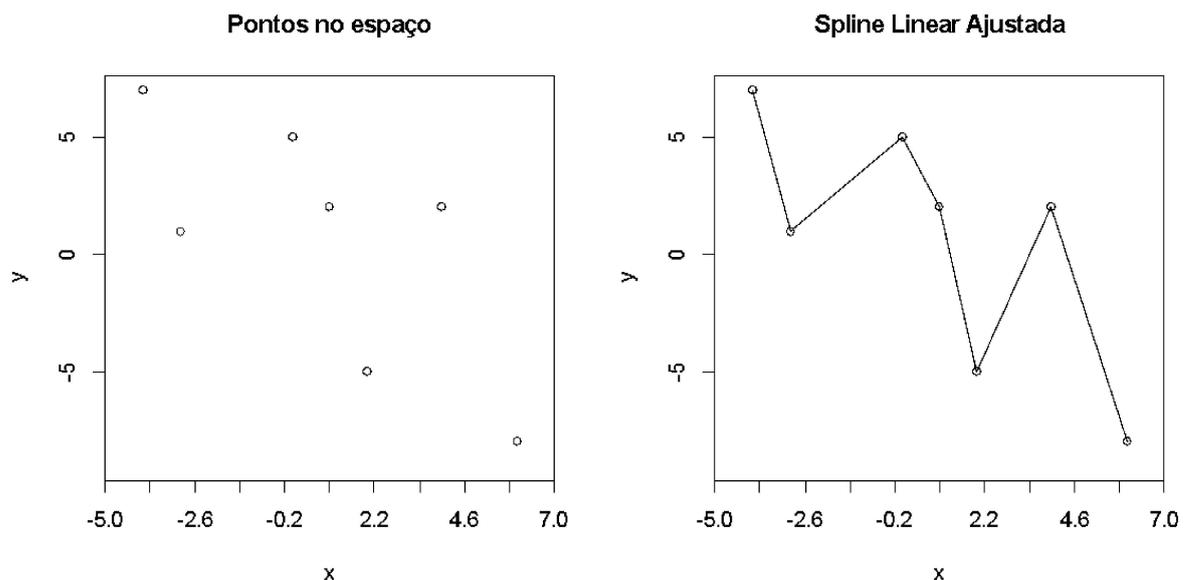
$$s_3(x) = f(x_2) \frac{(x_3-x)}{x_3-x_2} + f(x_3) \frac{(x-x_2)}{x_3-x_2} = 5 \frac{(1-x)}{1-0} + 2 \frac{(x-0)}{1-0} = -3x + 5, x \in (0, 1)$$

$$s_4(x) = f(x_3) \frac{(x_4-x)}{x_4-x_3} + f(x_4) \frac{(x-x_3)}{x_4-x_3} = 2 \frac{(2-x)}{2-1} - 5 \frac{(x-1)}{2-1} = -7x + 9, x \in (1, 2)$$

$$s_5(x) = f(x_4) \frac{(x_5-x)}{x_5-x_4} + f(x_5) \frac{(x-x_4)}{x_5-x_4} = -5 \frac{(4-x)}{4-2} + 2 \frac{(x-2)}{4-2} = \frac{7x}{2} - 12, x \in (2, 4)$$

$$s_6(x) = f(x_5) \frac{(x_6-x)}{x_6-x_5} + f(x_6) \frac{(x-x_5)}{x_6-x_5} = 2 \frac{(6-x)}{6-4} - 8 \frac{(x-4)}{6-4} = -3x + 22, x \in (4, 6)$$

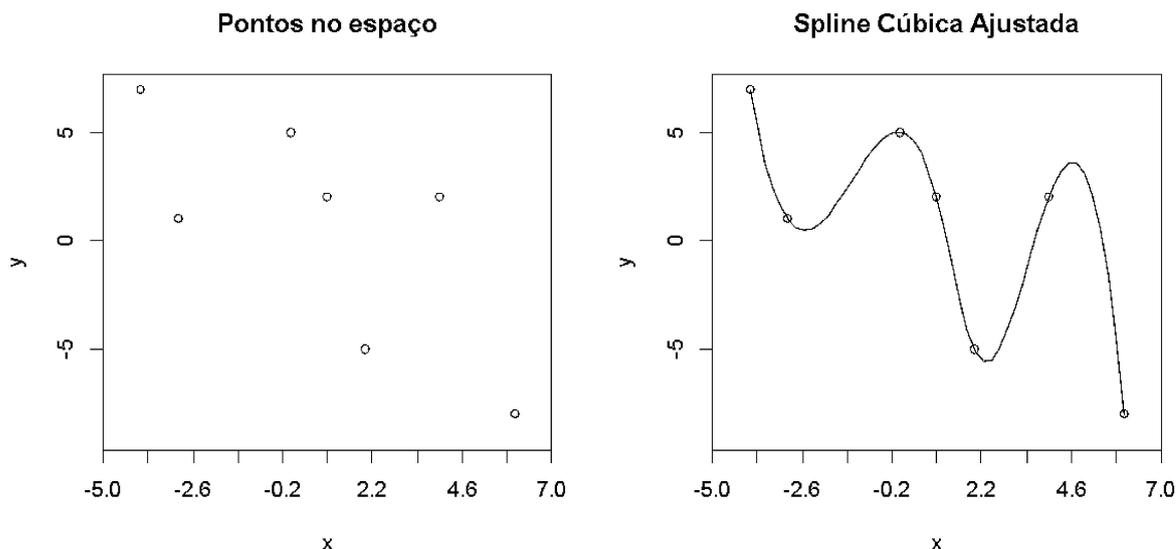
Fazendo uma representação gráfica com R, temos o seguinte gráfico:



- **Spline cúbica (ou Hermite):** Existe uma desvantagem da interpolação linear, que é o fato de que a primeira derivada é descontínua nos nós, já a spline quadrática tem derivada contínua somente até a primeira ordem, podendo trocar os pontos nas curvaturas, por isso é mais comum utilizar a spline cúbica, pois os nós são suaves o suficiente para não serem capazes de serem vistos a olho nu. Ela é dividida em duas categorias: Restrita ou também conhecido como spline natural (as caudas são modeladas pelas funções lineares) e a Irrestrita (as caudas não são modeladas pelas funções lineares);

Exemplo de ajuste de Spline Hermite:

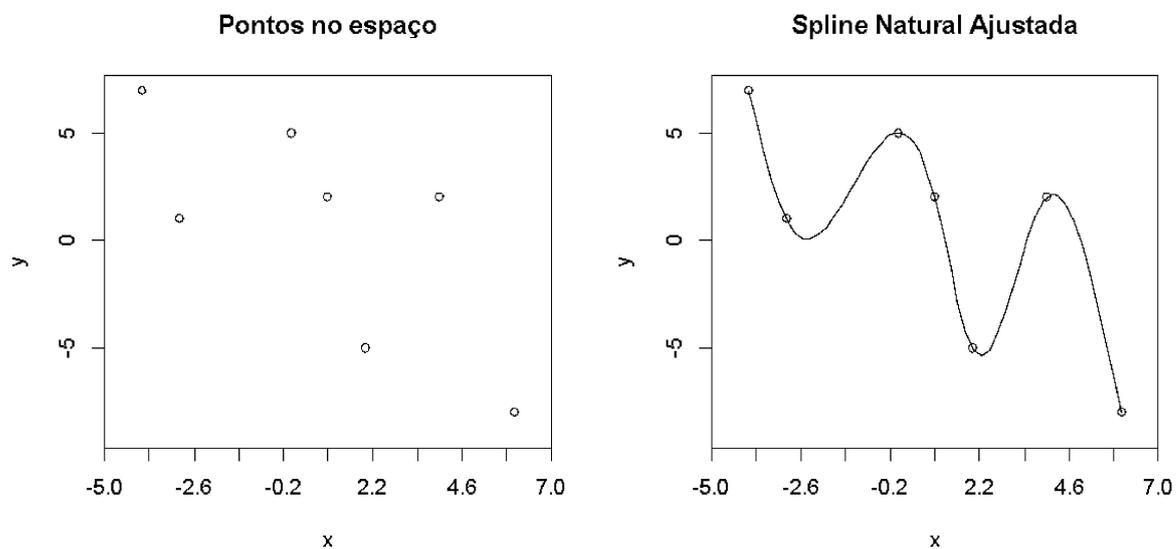
Vamos pegar os pontos do exemplo anterior (1, 2), (-3, 1), (2, -5), (4, 2), (0, 5), (-4, 7) e (6, -8) e ajustar a uma Spline Cúbica. Teremos o seguinte gráfico:



- **Spline natural (ou Spline cúbica restrita):** É o equivalente a supor que os polinômios cúbicos nos intervalos extremos ou são lineares ou próximos de funções lineares. É criada por partes que se unem suavemente em alguns locais ou nós pré-especificados, para evitar que a calda da curva oscile demais.

Exemplo de ajuste de Spline natural:

Sejam os pontos do primeiro exemplo (1, 2), (-3, 1), (2, -5), (4, 2), (0, 5), (-4, 7) e (6, -8) e ajustar a uma Spline Natural. Teremos o seguinte gráfico:



No trabalho a seguir, será utilizada a spline natural.

1.2 Dynamic Time Warping (DTW)

Dynamic time warping (DTW) é um algoritmo que serve para encontrar padrões entre duas curvas feitas em diferentes espaços de tempo, ajustando as mesmas para que tenham o mesmo espaço de tempo, no caso de vídeos, áudios etc. Um bom exemplo da utilização de DTW é a detecção do padrão da nossa voz ao falar com uma assistente virtual, independente de falarmos mais rápido ou mais lento, ela sempre entende o que é dito, outro bom exemplo é o de assinaturas feitas pela mesma pessoa que pode ficar mais desleixada se feita num período

menor ou mais bem escrita em um período mais longo, mesmo assim ambas as assinaturas serão reconhecidas como sendo da mesma pessoa.

O seu algoritmo segue os seguintes passos:

1. Definição das curvas: Suponha que tenhamos duas curvas, A e B, com comprimentos M e N, respectivamente.
2. Cálculo da matriz de distâncias: Cria-se uma matriz de distâncias D de tamanho M x N, onde cada elemento $D[i][j]$ representa a distância entre os elementos $A[i]$ e $B[j]$. A distância pode ser calculada usando a distância euclidiana.
3. Inicialização da matriz de custos: Cria-se uma matriz de custos C de tamanho M x N, onde cada elemento $C[i][j]$ representa o custo mínimo para alinhar parcialmente as sequências $A[1:i]$ e $B[1:j]$.
4. Preenchimento da matriz de custos: Para cada elemento $C[i][j]$ da matriz, calcula-se o custo mínimo de alinhamento parcial até aquele ponto. O custo mínimo é dado por:
$$C[i][j] = D[i][j] + \min(C[i-1][j], C[i][j-1], C[i-1][j-1])$$

Ou seja, o custo mínimo é a soma da distância entre os elementos $A[i]$ e $B[j]$ com o menor custo entre as três posições vizinhas (acima, à esquerda e diagonalmente acima à esquerda).
5. Encontrar o alinhamento ótimo: Após preencher completamente a matriz de custos, o alinhamento ótimo é encontrado seguindo o caminho com menor custo da célula $C[M][N]$ até a célula $C[0][0]$. Esse caminho representa a correspondência entre os pontos das sequências A e B.

O DTW resulta em um alinhamento flexível entre as sequências, permitindo que elas sejam comparadas mesmo que tenham tamanhos diferentes ou variações temporais. O algoritmo é eficaz em capturar relações de similaridade que não seriam detectadas por uma comparação direta ponto a ponto.

Mais à frente, haverá o exemplo na prática do uso do DTW, na seção resultados.

6. Erro médio da distância

Para comparar as assinaturas, será considerado o erro médio da distância, calculado com base na distância euclidiana entre a assinatura média das referências e a assinatura que se deseja comparar, para verificar se a distância é aceitável.

A distância euclidiana entre um ponto da assinatura e um ponto da assinatura média é calculada da seguinte forma:

$$d[(x, y), (\bar{x}, \bar{y})] = \sqrt{(x - \bar{x})^2 + (y - \bar{y})^2} \quad (1)$$

em que (x, y) são as coordenadas das observações e (\bar{x}, \bar{y}) são as coordenadas da assinatura média

Esse pode ser considerado o erro da assinatura, uma vez que quanto mais distante uma assinatura for da média, mais diferente ela vai ser da mesma.

A média dos erros da distância pode ser calculada da seguinte forma:

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n \sqrt{\left(x_i - \bar{x}\right)^2 + \left(y_i - \bar{y}\right)^2}, \quad i = 1, \dots, n \quad (2)$$

em que (x, y) são as coordenadas da assinatura, (\bar{x}, \bar{y}) são as coordenadas da assinatura média e n é o número de pontos em cada assinatura.

7. Banda de confiança do erro médio

Será utilizado 95% de nível de confiança para avaliar se rejeita ou não uma nova assinatura. Para criar esse intervalo de confiança, serão utilizados os erros médios da referência, para encontrarmos a variância dos erros médios e a média dos erros médios, da seguinte forma:

$$IC(e; 0, 95) = \bar{e} \pm z_{0,025} \sqrt{Var(\bar{e})} \quad (3)$$

7.1 Taxa do erro

A taxa de erro é uma medida da precisão de uma estimativa ou previsão. Ela indica a proporção de vezes em que uma estimativa está incorreta em relação ao total de observações ou previsões feitas. Temos 2 tipos de erros:

- Erro tipo I: é o erro que cometemos ao rejeitar uma hipótese nula dada que ela é verdadeira
- Erro tipo II: é o erro que cometemos ao não rejeitar uma hipótese nula dado que ela é falsa

A taxa de erro é importante para avaliar a qualidade e confiabilidade das estimativas e previsões feitas em estudos estatísticos. Uma taxa de erro alta indica que as estimativas ou previsões podem ser incertas ou imprecisas.

	H_0 é verdadeira	H_0 é falso
Rejeitar H_0	Erro tipo I (rejeitar H_0 verdadeiro)	Decisão Correta
Não rejeitar H_0	Decisão Correta	Erro tipo II (deixar de rejeitar H_0 falsa)

Podemos calcular as taxas de erro da seguinte maneira:

$$P(\text{Erro tipo I}) = P(\text{Rejeitar } H_0 | H_0 \text{ verdadeira}) \quad (4)$$

$$P(\text{Erro tipo II}) = P(\text{Não rejeitar } H_0 | H_0 \text{ falsa}) \quad (5)$$

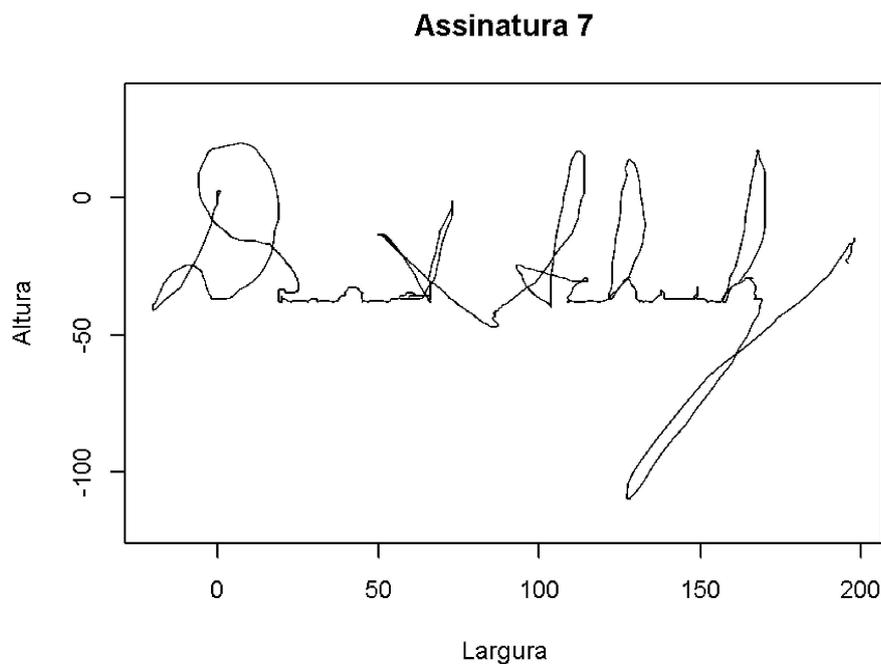
8. O conjunto de Dados

Foram utilizados 75 conjuntos de dados, sendo eles 50 assinaturas verdadeiras e 25 assinaturas falsas. Cada um dos 75 conjuntos de dados é formado por 3 colunas:

- V1 – O tempo em que a assinatura foi escrita;
- V2 – A altura em que o ponto foi marcado;
- V3 – A largura em que o ponto foi marcado.

RESULTADOS

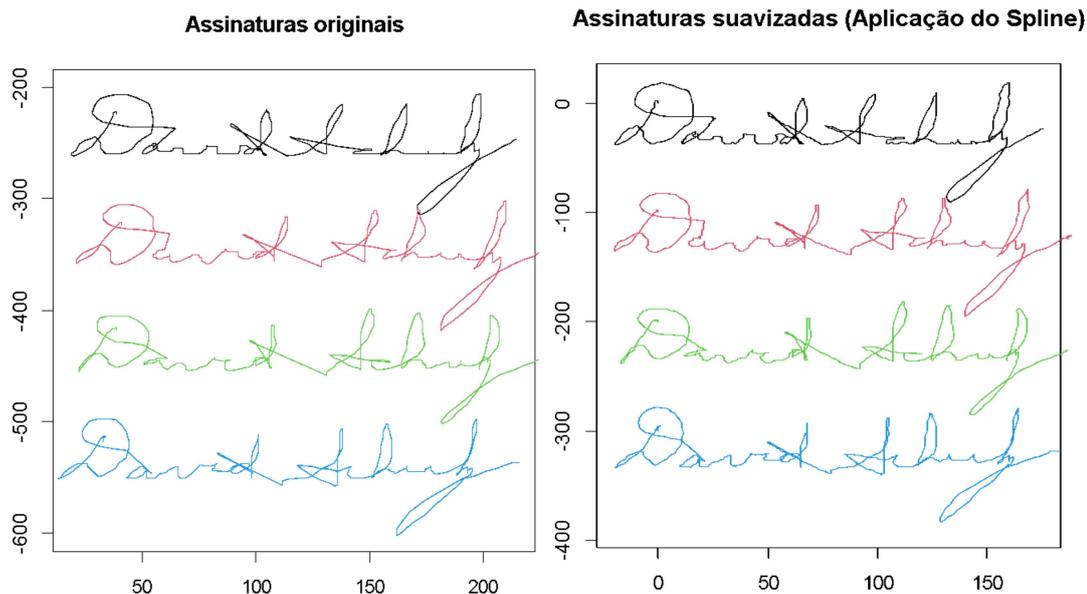
Dentre as 50 assinaturas verdadeiras coletadas, foram utilizadas as 12 primeiras, que representam aproximadamente 25% das mesmas como referência para ajuste do Intervalo de Confiança. Dentre as 12 primeiras assinaturas, foi escolhida a assinatura nº7 como referência para a aplicação do DTW, para ajustar as outras 11 assinaturas ao tempo da assinatura 7, por ser a assinatura de maior tempo entre as selecionadas, esse critério foi decidido pela autora, por conveniência.



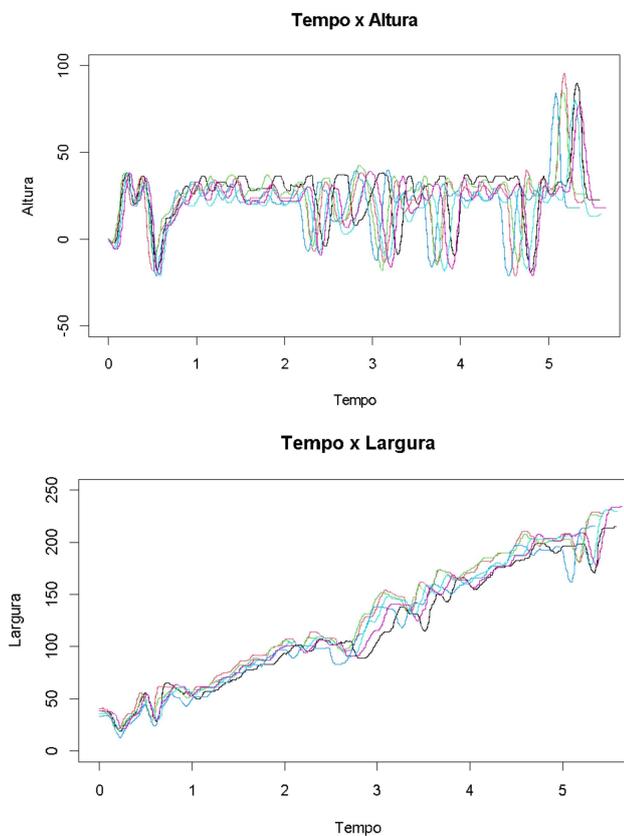
1. Modelagem de uma assinatura

Foram utilizadas as técnicas Spline e DTW (Dynamic Time Warping) para modelar as assinaturas, suavizando e deixando todas no mesmo tamanho.

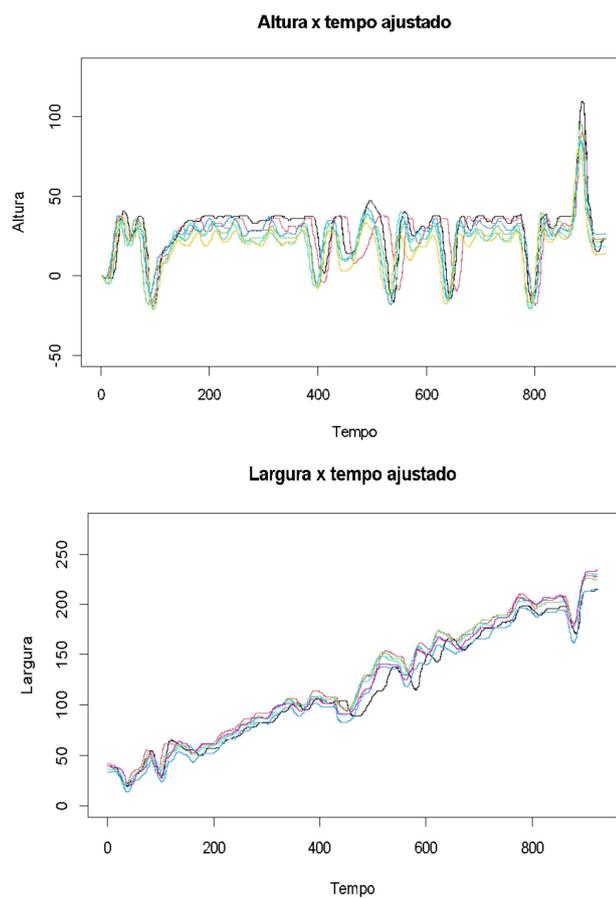
Ao aplicar o Spline, as curvas da assinatura ficam mais suaves.



mas as assinaturas foram escritas em diferentes velocidades, o que faz com que os gráficos não fiquem alinhados, como é mostrado nos gráficos de altura contra o tempo e largura contra o tempo:



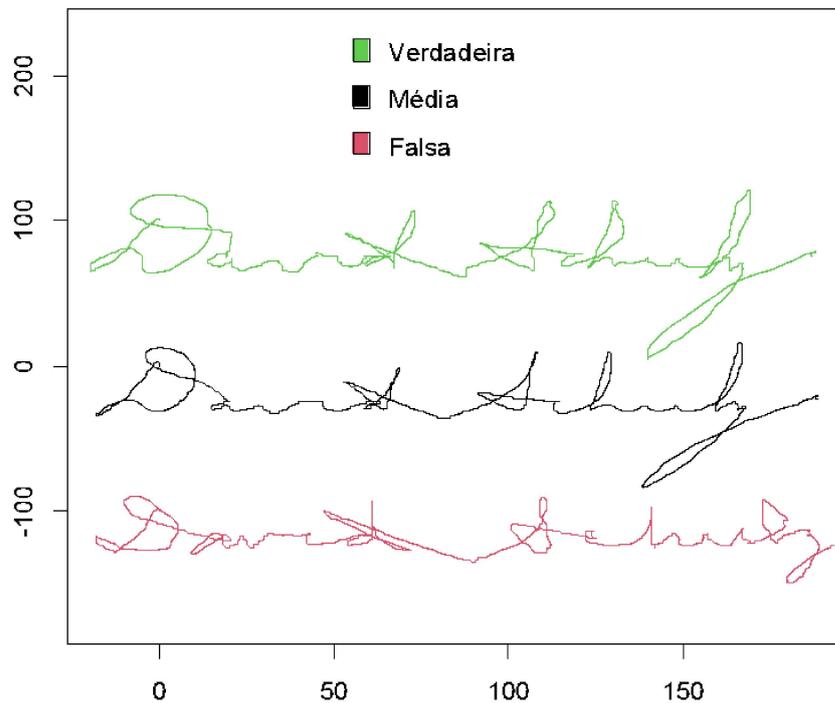
Aplicando o DTW, foi possível obter um ajuste tal qual todas as assinaturas tiveram o mesmo intervalo de tempo, com base na assinatura de maior tempo.



2. Criação da assinatura média

A assinatura média foi criada a partir da equação 2, na seção de métodos do capítulo 6. Após obtenção da assinatura média, é possível fazer uma comparação visual entre uma assinatura verdadeira, a assinatura média e uma assinatura falsa:

Assinatura média



Porém, não basta somente compararmos de forma visual, é interessante ter alguma métrica que possa mostrar de forma mais precisa qual assinatura é verdadeira e qual de fato é falsa, pois visualmente algumas assinaturas falsas podem passar despercebidas. Para tal, foi utilizada a assinatura média como referência para calcular a distância entre a assinatura que se deseja comparar e a média, obtendo o erro médio, como citado anteriormente na parte dos métodos na equação (2).

Desta forma, foi possível obter um intervalo de banda com 95% de confiabilidade, para esses erros, que foi dado por:

$$IC(\bar{e}; 95\%) = [2, 2; 12, 9]$$

Ao calcular a proporção de que as assinaturas verdadeiras não pertencem ao intervalo de confiança, dividindo o número de assinaturas verdadeiras que estão fora do intervalo de confiança pelo total de assinaturas verdadeiras, tal qual mencionado na parte de métodos, na equação (4), foi obtido o valor de 0,08.

Ao calcular a probabilidade de as assinaturas falsas pertencerem ao intervalo de confiança, dividindo o número de assinaturas falsas que estão dentro do intervalo pelo total de assinaturas falsas, tal qual mencionado na parte de métodos, na equação (5), foi obtido o valor 0.

CONCLUSÃO

Conforme citado nos resultados, foi obtido uma probabilidade de cometer um erro de tipo I de 0,08, portanto, existe uma chance de rejeitar uma assinatura verdadeira, mesmo que baixa. No entanto, a probabilidade de cometer um erro do tipo II foi de 0, o que indica que é difícil a possibilidade de aceitar uma assinatura falsa, dado que todas as assinaturas que foram obtidas estavam fora do intervalo de confiança.

No caso em questão, após o teste de probabilidade, nenhum resultado indicou a aceitação de assinaturas falsas. Isso é um indicativo importante de que as assinaturas presentes na amostra são autênticas e possuem de fato um padrão entre elas.

Entretanto, é importante destacar que esse teste não é infalível. Apesar de não terem sido detectadas assinaturas falsas na amostra testada, é possível que haja outras assinaturas falsas que não sejam rejeitadas. Além disso, é importante lembrar que a análise estatística deve ser usada em conjunto com outras técnicas de verificação, como a análise de características da assinatura em si.

De qualquer forma, a utilização de testes estatísticos para análise de assinaturas é uma importante ferramenta para garantir a autenticidade de assinaturas. A partir dos resultados obtidos nesse trabalho, podemos ter maior segurança na validade de assinaturas.

REFERÊNCIAS BIBLIOGRÁFICAS

A Comparative Analysis of Signature Verification Techniques. **Kahn, Z., & Rosales, M. 2013.** s.l. : International Journal of Computer Applications,, 2013, Vol. 73(19).

Daubechies, Ingrid. 1992. *Ten Lectures on Wavelets*. Philadelphia : SIAM, 1992.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 1936, pp. 178-188.

Lee, L.L. and Berger, T. 1994. Reliable on-line human signature verification system for point-of-sales applications. *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3 - Conference C: Signal Processing.* 1994, pp. 19-23 vol.2.

Mardia, K.V., Kent, J.T. and Bibby, J.M. 1979. *Multivariate Analysis*. London : Academic - Press, 1979.

Nelson, W., Turin, W. and Hastie, T. 1994. Statistical methods for on-line signature verification. *International Journal of Pattern Recognition and Artificial Intelligence.* 1994, pp. 749-769.

Sena Jr, M. R., Nascimento, A. D. C., Cordeiro, G. M. and Barroso, L.P. 2013. Score type statistics in pattern classification. *Brazilian Journal of Probability and Statistics.* 2013, pp. 27, 210-226.

Zhong, C and Tibshirani. 2019. *Survival analysis as a classification problem.* 2019.