UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

João Pedro Coelho Amorim de Lavor

Explorando Técnicas de Aprendizado de Máquina para a Classificação de Formas Tridimensionais

Recife

2024

João Pedro Coelho Amorim de Lavor

Explorando Técnicas de Aprendizado de Máquina para a Classificação de Formas Tridimensionais

Dissertação apresentada ao Programa de Pós-graduação em estatística do Departamento de estatística da Universidade Federal de Pernambuco como requisito parcial para a obtenção do grau de Mestre em estatística

Área de Concentração: Estatística Aplicada

Orientador: Getúlio José Amorim do Amaral

Recife

2024

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Lavor, João Pedro Coelho Amorim de.
Explorando técnicas de aprendizado de máquina para a classificação de formas tridimensionais / João Pedro Coelho Amorim de Lavor. - Recife, 2024.
70f.: il.

Dissertação (Mestrado) - Universidade Federal de Pernambuco, Centro de Ciências Exatas e da Natureza, Pós-Graduação em Estatística, 2024.
Orientação: Getúlio José Amorim do Amaral.
Coorientação: Renata Maria Cardoso Rodrigues de Souza.
Inclui referências.

1. Estatística Aplicada; 2. Classificação supervisionada; 3. Dados tridimensionais; 4. K-vizinhos mais próximos; 5. Análise discriminante linear; 6. Análise discriminante quadrática. I. Amaral, Getúlio José Amorim do. II. Souza, Renata Maria Cardoso Rodrigues de. III. Título.

UFPE-Biblioteca Central

JOÃO PEDRO COELHO AMORIM DE LAVOR

Explorando Técnicas de Aprendizado de Máquina para a Classificação de Formas Tridimensionais

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovado em: 06 de agosto de 2024.

BANCA EXAMINADORA

Prof. Dr. Getúlio José Amorim do Amaral
Presidente, UFPE

Prof. Dr. Alex Dias Ramos
Examinador Interno, UFPE

Profª Drª Carla Almeida Vivacqua
Examinador Externo, UFRN

Aos meus pais, irmãos e amigos.

AGRADECIMENTOS

Agradeço à minha mãe, Ana Leopoldina, por todo o apoio que me deu. Agradeço em memória ao meu pai, que sempre me incentivou a estudar e a seguir esse caminho. Também agradeço aos meus irmãos, Carlos e Igor pelo apoio e pelos momentos divertidos de descontração. Agradeço também à minha namorada, Jadiane, por todo o carinho e ajuda durante esse processo, sempre me incentivando e apoiando. Gostaria de agradecer aos meus orientadores, professor Getúlio e professora Renata, por sempre me ajudarem nas reuniões a melhorar meu trabalho. Por último, quero agradecer à Universidade Federal de Pernambuco e a todo seu corpo docente.

RESUMO

Esta dissertação tem como objetivo propor novos métodos supervisionados de classificação para dados de pré-forma, considerando dados tridimensionais para a classificação de dois ou mais grupos conhecidos. Os novos métodos são baseados em técnicas discriminantes e modelos de aprendizado de máquina já estabelecidos no contexto de classificação, como K vizinhos mais próximos, análise discriminante linear e análise discriminante quadrática. No entanto, a inovação deste trabalho reside na adaptação dessas técnicas para lidar com dados em espaços não euclidianos, utilizando distâncias Procrustes para trabalhar com dados de pré-forma, algo que ainda não havia sido explorado nesse contexto. Isso exigiu a modificação dos modelos tradicionais para trabalhar com dados matriciais, permitindo uma análise mais precisa e robusta em cenários onde as relações espaciais e geométricas dos dados são fundamentais para a classificação. Para dados simulados, gerados a partir de uma distribuição normal multivariada, propomos um cenário de classificação utilizando a taxa de acerto, ou acurácia, como métrica para avaliar o desempenho dos algoritmos. Nesses testes, todos os modelos alcançaram bons resultados de acurácia, com destaque para o K vizinhos mais próximos e o Discriminante Linear. Ao analisarmos os dados de landmarks faciais, com o objetivo de classificar entre três classes distintas, verificamos que o modelo discriminante quadrático se mostrou superior, pois apresentou melhores resultados de acurácia na classificação, especialmente em cenários com mais de duas variáveis. Esse modelo foi mais eficaz em capturar as nuances dos dados tridimensionais, obtendo uma taxa de acerto mais elevada quando comparado ao K vizinhos mais próximos e ao Discriminante Linear. Isso nos leva a concluir que, em ambos os cenários, os novos métodos conseguiram classificar com precisão as classes dos dados.

Palavras-chaves: Classificação supervisionada; Dados tridimensionais; marcos anatômicos; K-vizinhos mais próximos; Análise discriminante linear; Análise discriminante quadrática

ABSTRACT

This dissertation aims to propose new supervised classification methods for pre-form data, considering three-dimensional data for the classification of two or more known groups. The new methods are based on discriminant techniques and machine learning models already established in the context of classification, such as K-nearest neighbors, linear discriminant analysis, and quadratic discriminant analysis. However, the innovation of this work lies in the adaptation of these techniques to handle data in non-Euclidean spaces, using Procrustes distances to work with pre-form data, something that had not yet been explored in this context. This required the modification of traditional models to work with matrix data, allowing for a more precise and robust analysis in scenarios where the spatial and geometric relationships of the data are fundamental for classification. For simulated data, generated from a multivariate normal distribution, we propose a classification scenario using the hit rate, or accuracy, as a metric to evaluate the performance of the algorithms. In these tests, all models achieved good accuracy results, with emphasis on K-nearest neighbors and Linear Discriminant Analysis. When analyzing facial landmark data with the objective of classifying into three distinct classes, we observed that the quadratic discriminant model performed better, as it presented higher accuracy results in classification, especially in scenarios with more than two variables. This model was more effective in capturing the nuances of three-dimensional data, achieving a higher hit rate when compared to K-nearest neighbors and Linear Discriminant Analysis. This leads us to conclude that, in both scenarios, the new methods were able to accurately classify the data classes.

Keywords: Supervised classification; Three-dimensional data; landmarks; K-nearest neighbors; Linear discriminant analysis; Quadratic discriminant analysis

LISTA DE TABELAS

LISTA DE FIGURAS

SUMÁRIO

1 INTRODUÇÃO

O conteido de forma, é bastante intuitivo. Contudo, torna-se não trivial quando propomos uma descrição que proporcione uma análise matemática e estatística. Durante os últimos trinta anos, tivemos alguns avanços tecnológicos, como por exemplo, na captura de imagens bidimensionais e tridimensionais. Demandando-se um melhor entendimento destes conceitos nas várias áreas que eles estão inseridos. A teoria das formas, na maneira que abordarem teve origens distintas, em (KENDALL, 1977) abordado a forma no contexto da arqueologia e astronomia, e em (BOOKSTEIN, 1984) através do estudo problemas teóricos de forma no contexto da zoologia.

A análise de formas, ou *Statistical Shape Analysis*, é um campo de pesquisa que utiliza métodos estatísticos com finalidade de entender a variabilidade e os padrões das formas de objetos ou organismos. Esta análise é bastante significativa em áreas como: biologia, medicina, antropologia e visão computacional. Permitindo-se obter identificação, categorização e comparação precisas de formas. Um dos focos centrais destes tipo de estudo esta nos modelos de classificação, fundamentais na identificação de padrões em dados morfológicos, aplicados em diagnóstico médico, biologia evolutiva e visão computacional, esses modelos automatizam processos e aumentam a precisão das análises.

A morfometria, base da análise de formas, estuda quantitativamente a forma, tamanho e proporções dos organismos. Com abordagens tradicional e geométrica, a morfometria compara formas entre diferentes grupos e analisa variações morfológicas. Dentro da análise de formas, distingue-se a análise de forma pura, de tamanho e a conjunta de tamanho e forma , cada uma com foco em aspectos específicos e essenciais para estudos de morfometria e análise de forma estatística.

Temos uma vasta gama de trabalhos em análise de formas os quais se concentra em dados bidimensionais, como em (CARVALHO; AMARAL, 2020), que utiliza métodos específicos para classificar formas bidimensionais, esses métodos nem sempre podem ser generalizados para dados tridimensionais. Muitos dos algoritmos que temos hoje em dia são baseados em reconhecimento de marcos, como reconhecimento facial, ou seja, utilizam propriedades geométricas de tamanho e forma. Inspirado pelo artigo de

(VINUÉ; SIMÓ; ALEMANY, 2016), nosso trabalho propõe novos métodos de análise de formas no contexto supervisionado em dados tridimensionais. Marcos tridimensionais, isto é, pontos de referência para representar a geometria e estrutura de um objeto no espaço, são fundamentais em medicina, engenharia e visão computacional.

Com o avanço da tecnologia, a importância dos marcos tridimensionais tem crescido significativamente. Um exemplo disso é a rede neural FaceNet do Google (SCHROFF; KALENICHENKO; PHILBIN, 2015), que utiliza redes neurais para gerar marcos faciais com o propósito de detecção, identificação e autenticação precisa de indivíduos. Marcos, ou *landmarks*, são pontos de referência específicos em uma estrutura, usados para identificar e alinhar características importantes, como olhos, nariz e boca no caso de rostos. Eles são fundamentais em aplicações como reconhecimento facial, onde servem para mapear a geometria do rosto de forma precisa e consistente. Além disso, essa tecnologia encontra aplicações em realidade aumentada e entretenimento. Dessa forma, este estudo visa explorar e expandir as metodologias de análise de formas tridimensionais, contribuindo para a evolução contínua deste campo interdisciplinar.

Embora a análise de tamanho e forma sejam comumente estudada, como no trabalho de (HONÓRIO; AMARAL, 2023), neste trabalho focaremos na análise dos marcos no espaço de Pré-Forma. No entanto, será detalhada toda a base teórica presente na análise estatística de formas, garantindo que os conceitos fundamentais e as metodologias clássicas sejam abordados. Com essa abordagem, pretendemos explorar as complexidades e potencialidades dos dados geométricos de forma aprofundada, proporcionando uma compreensão mais rica e detalhada das estruturas estudadas.

1.1 OBJETIVOS

Os objetivos gerais desse trabalho são: utilizar classificadores da literatura e os principais testes de classificação no contexto da análise estatística de forma tridimensional.

1.2 ORGANIZAÇÃO DA DISSERTAÇÃO

O trabalho tem a seguinte estrutura:

- No **Capítulo 2**, revisamos conceitos fundamentais e literaturas clássicas da Análise Estatística de Formas, com o objetivo de esclarecer a usabilidade de seus métodos e as mudanças tocantes à mudança de espaço e perspectiva presentes no seu domínio. Exploramos, também, o processo de estimação da forma média e algumas variações para o caso em que $m = 3$;

- No **Capítulo 3**, primeiro, iremos revisitar os algoritmos supervisionados, como K-Nearest Neighbours, Linear Discriminant Analysis e Quadratic Discriminant Analysis. Em particular, focamos na implementação do algoritmo no contexto de formas e na avaliação de sua performance.

- No **Capítulo 4**, consideramos uma aplicação dos desenvolvimentos apresentados nos capítulos anteriores, na área da classificação de marcos, que tem como objetivo identificar atraves dos marcos, posição dos rostos escaneados e em dados simulados.

- No **Capítulo 5**, Neste capítulo há o objetivo de proporcionar uma visão geral do trabalho, suas contribuições e conclusões. Também será ilustrado o que pode ainda ser feito para que mais conclusões possam ser desenvolvidas a partir do que já foi demonstrado ao longo desse trabalho e possívelmente propor trabalhos futuros.

1.3 SUPORTE COMPUTACIONAL

Para a realização desta dissertação, utilizamos diferentes ferramentas e linguagens de programação. Alguns algoritmos foram desenvolvidos em Python versão 3.10, enquanto outros foram implementados em R versão 4.3.2 (2023-10-31 ucrt), incluindo a geração de gráficos.

O ambiente computacional consistiu em um computador com sistema operacional Windows 11, 64 bits, equipado com 24GB de RAM, processador Intel i5 de 11ª geração e uma placa gráfica Nvidia 1050ti com 6GB de memória dedicada.

2 ANÁLISE ESTATÍSTICA DE FORMAS

Corriqueiramente usamos a palavra forma quando nos referimos à aparência de um objeto ou quando comparamos objetos semelhantes. Seguindo esse raciocínio, queremos traduzir essa comparação entre os objetos que estudamos, atribuindo a eles uma estrutura geométrica que nos permita analisar suas similaridades e variabilidade. No entanto, definir forma de maneira que seja suscetível à análise matemática e estatística não é uma tarefa trivial. Como mencionado no livro de Kendall (KENDALL et al., 2009), assumimos que a forma de um objeto é essencialmente capturada pela composição de um subconjunto finito de seus pontos. Embora isso possa parecer uma restrição severa, não há limite teórico para o número de pontos que consideramos. Essa abordagem tem a vantagem significativa de que as dimensões dos espaços de forma resultantes são sempre finitas e aumentam apenas linearmente com o número de pontos.

2.1 CONCEITOS FUNDAMENTAIS

Apresentamos uma definição formal sobre o que é forma. Ela, é baseada no estudo original de Kendall (KENDALL et al., 2009).

Definição 2.1.1. Forma *é toda informação geométrica que permanece quando os efeitos de locação, escala e rotação são removidos do objeto.*

A forma de um objeto é invariante sob as transformações de similaridade Euclidiana de translação, escala e rotação. Isso significa que a forma de um objeto não muda quando o objeto é movido, redimensionado ou girado. Translação envolve mover o objeto de um lugar para outro sem alterar sua orientação ou tamanho. Escala refere-se ao redimensionamento do objeto, aumentando ou diminuindo seu tamanho. Mesmo que um objeto seja ampliado ou reduzido, sua forma básica permanece a mesma. Rotação implica girar o objeto em torno de um ponto fixo. Quando você gira um objeto, a orientação dele muda, mas a forma permanece a mesma. Essas transformações são chamadas de "similaridade Euclidiana"porque mantêm as proporções relativas e os ângulos entre os pontos no objeto. Portanto, a "forma"de um objeto é considerada invariável (não

muda) sob essas operações, pois todas essas transformações não alteram a estrutura geométrica essencial do objeto.

Por exemplo, a forma de um crânio humano consiste em todas as propriedades geométricas do crânio que permanecem inalteradas quando ele é transladado, redimensionado ou rotacionado em um sistema de coordenadas arbitrário. Dois objetos têm a mesma forma se eles puderem ser transladados, redimensionados e rotacionados um em relação ao outro de modo que coincidam exatamente, ou seja, se os objetos forem semelhantes. Na Figura 1, os contornos de duas vértebras de camundongo têm a mesma forma. Na prática, estamos interessados em comparar objetos com formas diferentes e, portanto, precisamos de uma maneira de medir a forma, alguma noção de distância entre duas formas e métodos para a análise estatística da forma.

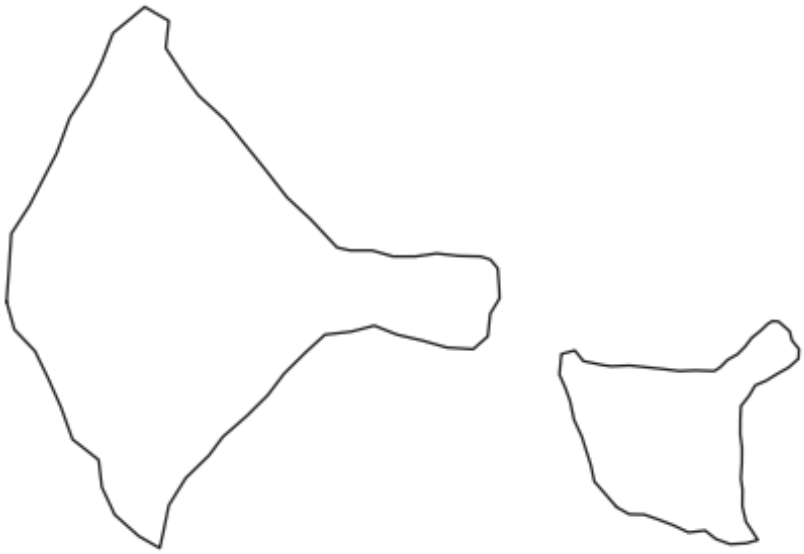


Figura 1 – Duas silhuetas da mesma segunda vértebra torácica (T2) de um camundongo, que possuem diferentes localizações, rotações e escalas, mas a mesma forma.

Às vezes, estamos interessados em reter a informação de escala (tamanho) além da forma do objeto, e assim, desenvolvemos a análise conjunta de tamanho e forma (ou forma), o que é bastante relevante.

Definição 2.1.2. Tamanho e Forma *é toda informação geométrica que permanece quando os efeitos de locação e rotação são removidos do objeto.*

Análogo à definição 2.1.1, dizemos que dois objetos têm o mesmo tamanho e forma, se eles puderem ser transladados e rotacionados um ao outro, de modo que

correspondam exatamente, isto é, se os objetos forem transformações rígidas um do outro. Como explicamos anteriormente, assumimos que a forma de um objeto é essencialmente capturada pela forma de um subconjunto finito de seus pontos, que é uma abordagem muito utilizada em morfometria. Portanto, precisamos de uma maneira numérica para descrever estas estruturas usando esse conjunto de pontos, no qual ainda podemos entende-las de maneira visual, dessa forma.

Definição 2.1.3. *Um* ***Marco (Landmark)*** *é um ponto de correspondência em cada objeto que coincide entre e dentro da população.*

Essencialmente, os *marcos* são pontos de referência comuns que permitem a análise e comparação de formas entre diferentes objetos e populações, facilitando estudos de variação, simetria, e outras características geométricas. Esses pontos podem variar, podendo ser no espaço bidimensional, tridimensional, nos quais cada ponto, reflete uma característica particular do objeto de interesse. Originalmente, descrito no livro de Mardia (DRYDEN; MARDIA, 2016), existem três tipos principais de *marcos*, a saber: *marcos* anatômicos, *marcos* Matemáticos e Pseudo-*marcos*.

- ***marcos* Anatômicos**: Definido por um especialista no assunto, irá corresponder de forma biologica significativa que transcreva alguma característica do objeto de estudo.
- ***marcos* Matemáticos**: São pontos definidos em um objeto de acordo com propriedades matemáticas, tais elas como, pontos de curvatura, máximos, mínimos, etc.
- **Pseudo-*marcos***: São pontos construídos em um objeto, localizados no contorno ou entre os *marcos* anatômicos ou matemáticos. Mais utilizado para conceituar a forma do objeto.

Devido a evolução de áreas como visão computacional e com o avanço das tecnologias de escaneamento, o uso de pseudo-landmarks evoluiu, proporcionando novos *marcos* que podem ser frutos de estudos em dados atuais, como os definidos em Vinue e Brignell et al. (2010) (VINUÉ; SIMÓ; ALEMANY, 2016).

- ***marcos* Automáticos**: são calculados automaticamente com algoritmos do programa de scanner, baseados em aspectos geométricos do corpo.
- ***marcos* Manuais**: são pontos que não são refletidos na geometria externa do corpo; eles são localizados através de palpação por pessoal especializado e identificados por um marcador físico.
- ***marcos* Digitais**: são detectados na tela do computador na imagem escaneada em 3D. Eles não são robustos no cálculo automático, mas são fáceis de detectar na tela.

2.2 MEDIDAS DE TAMANHO E COORDENADAS DE FORMA

Até então, introduzimos todos os conceitos fundamentais sobre formas, proporcionando uma compreensão inicial de como estudar uma estrutura física. No entanto, agora precisamos traduzir todas essas informações para definições matemáticas, formalizando o conceito de *marcos* para iniciar uma análise estatística e matemática da estrutura física do objeto de estudo. Formalizar esses conceitos vai nos permitir aplicar métodos estatísticos já conhecidos para entender a variabilidade dentro de um conjunto de dados e introduzir algoritmos de aprendizado de máquina, a fim de tentar sumarizar as propriedades geométricas desses objetos. Isso torna a análise de formas uma ferramenta poderosa em diversas disciplinas, incluindo biologia, medicina, antropologia e visão computacional. A seguir, temos a definição que formaliza o uso de um conjunto de *marcos* para descrever a forma de um objeto.

Definição 2.2.1. *A* ***configuração*** *é o conjunto de marcos em um objeto particular. A matriz de configuração X é a matriz $k \times m$ de coordenadas cartesianas dos k marcos em m dimensões. O* ***espaço de configuração*** *é o espaço de todas as coordenadas dos landmarks.*

Para representar matematicamente a forma de um objeto, inicialmente k *marcos* são escolhidos para o contorno desse objeto. As coordenadas desses *marcos* são representados pela seguinte matriz de configuração:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{km} \end{pmatrix} \tag{2.1}$$

Na aplicação que faremos posteriormente, consideramos $k = 68$ e $m = 3$ dimensões, e o espaço de configuração é o espaço real das matrizes $k \times m$, representado por $\mathbb{R}^{km}$. Utilizaremos essa definição para determinar a forma do objeto, que conseguimos eliminando os efeitos de escala, locação e rotação. Para eliminar esses efeitos, algumas transformações devem ser realizadas na nossa matriz de configuração X. Como dito anteriormente, essa dissertação irá abordar o caso tridimensional, isto é, cada elemento de estudo será uma matriz da forma $k \geq 3$ e $m = 3$. Assim, a matriz de configuração X da expressão (2.1) resume-se a:

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} \end{pmatrix}$$

2.3 FILTRANDO EFEITOS

Para iniciar as transformações em busca da forma ou do tamanho-e-forma de um objeto, um dos pontos mais importantes é a proposição de um sistema de coordenadas. Esses sistemas têm a finalidade de obter a forma de um objeto por meio dos pontos dispostos através dos *marcos*. Os principais sistemas de coordenadas discutidos em Mardia (DRYDEN; MARDIA, 2016) são o sistema de coordenadas de Bookstein (BOOKSTEIN, 1984) e o de Kendall (KENDALL, 1984).

2.4 COORDENADAS DE BOOKSTEIN

Considere (x_j, y_j), $j = 1, \ldots, k$, sendo $k \geq 3$ *marcos* em um plano ($m = 2$ dimensões). Bookstein (1986, 1984) sugere remover as transformações de similaridade através de translação, rotação e escala, de forma que os *marcos* 1 e 2 sejam fixados em uma posição específica. Se o marco 1 é enviado para $(0, 0)$ e o marco 2 é enviado para $(1, 0)$, então as variáveis de forma adequadas são as coordenadas dos $k-2$ *marcos* restantes após essas operações. Para preservar a simetria, consideramos o sistema de coordenadas onde os *marcos* de base são enviados para $(-\frac{1}{2}, 0)$ e $(\frac{1}{2}, 0)$.

Definição 2.4.1. Coordenadas de Bookstein $(u_j^B, v_j^B)^T, j = 3, \ldots, k$, *são as coordenadas restantes de um objeto após a translação, rotação e escala da linha de base para* $(-\frac{1}{2}, 0)$ *e* $(\frac{1}{2}, 0)$, *de modo que*

$$\begin{aligned} u_j^B &= \frac{(x_2 - x_1)(x_j - x_1) + (y_2 - y_1)(y_j - y_1)}{D_{12}^2} - \frac{1}{2}, \\ v_j^B &= \frac{(x_2 - x_1)(y_j - y_1) - (y_2 - y_1)(x_j - x_1)}{D_{12}^2}, \end{aligned} \tag{2.2}$$

onde $j = 3, \ldots, k$, $D_{12}^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 > 0$ e $-\infty < u_j^B, v_j^B < \infty$.

Quando dizemos que os *marcos* 1 e 2 são "fixados", estamos nos referindo a um processo de normalização que envolve a remoção das transformações de similaridade (translação, rotação e escala) dos dados. Isso é feito para padronizar a forma e permitir comparações diretas entre diferentes configurações de *marcos*. O marco 1 é transladado para a origem do sistema de coordenadas, que é o ponto $(0, 0)$. Em seguida, o marco 2 é escalado e transladado para o ponto $(1, 0)$, de modo que a distância entre os *marcos* 1 e 2 seja igual a 1 unidade. A configuração é então rotacionada de forma que a linha que conecta os *marcos* 1 e 2 fique alinhada com o eixo x como podemos ver na figura 2. Isso facilita a análise, pois remove qualquer efeito de posicionamento inicial dos objetos, normaliza o tamanho dos objetos para que a comparação seja baseada apenas na forma e elimina as diferenças de orientação entre os objetos.

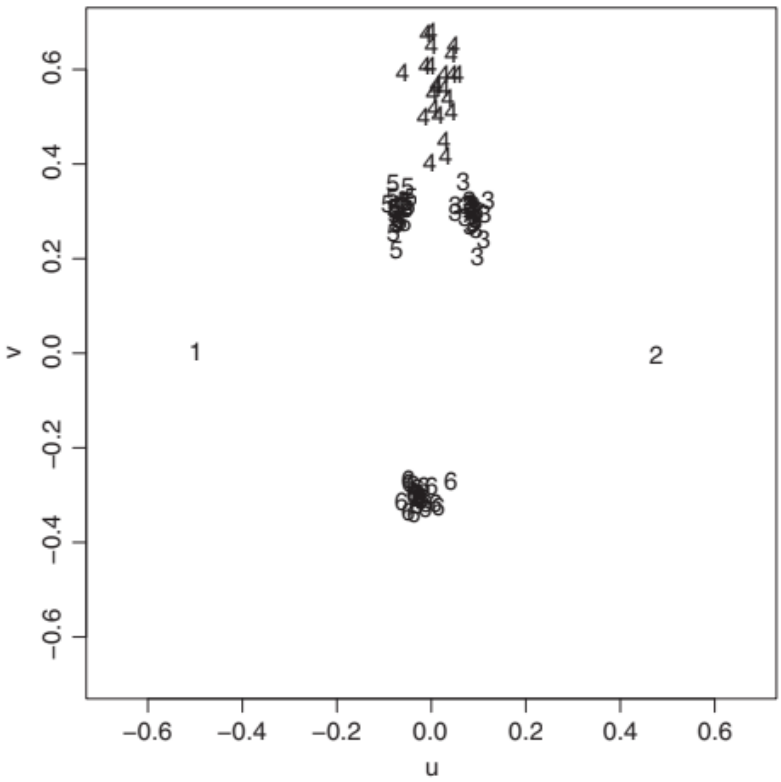


Figura 2 – Gráficos de dispersão das coordenadas de Bookstein das vértebras T2 Small registradas na linha de base 1, 2.

As transformações de similaridade também podem ser realizadas para dados tridimensionais, de forma semelhante às transformações nas coordenadas de Bookstein para dados bidimensionais. Considere k *marcos* $X_j = (x_{1j}, x_{2j}, x_{3j})^T \in \mathbb{R}^{3k}$, $j = 1, \ldots, k$. O número de variáveis de forma é $3k-7$, já que temos $3k$ coordenadas de *marcos*, mas devemos remover 3 parâmetros de localização, 1 de escala e 3 de rotação. As coordenadas de Bookstein são definidas como

$$u_j = (u_{1j}, u_{2j}, u_{3j})^T = \frac{1}{\|X_2 - X_1\|} A \left(X_j - \frac{(X_1 + X_2)}{2} \right), \quad j = 3, \ldots, k,$$

onde A é uma matriz de rotação 3×3 (uma função de (X_1, X_2, X_3)) e

$$X_1 \to (-1/2, 0, 0)^T, \quad X_2 \to (1/2, 0, 0)^T, \quad X_3 \to u_3 = (u_{13}, u_{23}, 0)^T,$$

com $u_{23} \geq 0$, $u_{33} = 0$ e $X_j \to u_j$ para $j = 4, \ldots, k$.

As setas $\to$ indicam a transformação dos *marcos* para um novo sistema de coordenadas, onde X_1 e X_2 são fixados nos pontos $(-1/2, 0, 0)^T$ e $(1/2, 0, 0)^T$, respectivamente,

e X_3 é projetado no plano $u_3 = (u_{13}, u_{23}, 0)^T$, de tal forma que $u_{23} \geq 0$. Para os demais pontos ($j = 4, \ldots, k$), a transformação os mapeia para novas coordenadas u_j.
Procedemos da seguinte maneira:

Calculamos a distância euclidiana entre os *marcos* X_1 e X_2 e normalizamos essa distância, como $\|X_2 - X_1\|$. O ponto médio entre X_1 e X_2, dado por $\frac{(X_1+X_2)}{2}$, é subtraído de cada marco X_j.

Aplicamos uma matriz de rotação A, que é uma função dos três primeiros *marcos* X_1, X_2, X_3, para alinhar os *marcos* na configuração padrão. Essa matriz alinha X_1 e X_2 nas posições $(-1/2, 0, 0)^T$ e $(1/2, 0, 0)^T$, estabelecendo uma linha de base no eixo x com comprimento unitário.

O terceiro marco X_3 é posicionado de forma que sua nova coordenada tenha a forma $(u_{13}, u_{23}, 0)^T$. A condição $u_{23} \geq 0$ garante que X_3 esteja no semiplano superior do plano xy, enquanto $u_{33} = 0$ fixa X_3 no plano xy, eliminando qualquer componente na direção z.

Ao fixar X_1 e X_2 nas posições específicas $(-1/2, 0, 0)$ e $(1/2, 0, 0)$, e X_3 no plano xy com $u_3 = (u_{13}, u_{23}, 0)$, estamos removendo 3 parâmetros de localização, 1 de escala e 3 de rotação. Assim, das $3k$ coordenadas iniciais dos *marcos* tridimensionais, restam $3k - 7$ variáveis de forma após a normalização.

2.5 COORDENADAS DE KENDALL

As coordenadas de Kendall (Kendall 1984) estão relacionadas às coordenadas de Bookstein, mas a localização é removida de maneira diferente. A fim de remover a localização, devemos definir a submatriz de Helmert **H**. Antes, vamos definir a matriz de Helmert $\mathbf{H^F}$ (LANCASTER, 1965).

Definição 2.5.1. *A* ***Matriz de Helmert Completa*** H^F *é uma matriz ortogonal quadrada* $k \times k$ *onde a primeira linha tem todos os elementos iguais a* $\dfrac{1}{\sqrt{k}}$*, e as demais linhas são ortogonais à primeira linha e entre si. As linhas da matriz* H^F *são definidas da seguinte maneira:*

$$H^F = \begin{pmatrix} \frac{1}{\sqrt{k}} & \frac{1}{\sqrt{k}} & \frac{1}{\sqrt{k}} & \cdots & \frac{1}{\sqrt{k}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{k(k-1)}} & \frac{1}{\sqrt{k(k-1)}} & \frac{1}{\sqrt{k(k-1)}} & \cdots & -\frac{k-1}{\sqrt{k(k-1)}} \end{pmatrix}$$

A j-ésima linha da submatriz de Helmert H é dada por:

$$(h_j, \ldots, h_j, -jh_j, 0, \ldots, 0), \quad h_j = -j(j+1)^{-1/2}$$

Assim, a j-ésima linha consiste em h_j repetido j vezes, seguido por $-jh_j$ e depois $k-j-1$ zeros, para $j = 1, \ldots, k-1$.

Definição 2.5.2. *A **Submatriz de Helmert** H $(k-1) \times k$ é a matriz de Helmert completa H^F sem a primeira linha.*

Para $k = 4$, a matriz de Helmert completa é explicitamente

$$H^F = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & \frac{3}{\sqrt{12}} \end{pmatrix}$$

e a submatriz de Helmert é:

$$H = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & \frac{3}{\sqrt{12}} \end{pmatrix}$$

A primeira linha da matriz de Helmert completa tem todos os elementos iguais a $\frac{1}{\sqrt{k}}$. Quando esta linha é multiplicada por um vetor de dados, o resultado é proporcional à soma dos elementos desse vetor, ou seja, à localização média dos dados. Ao eliminar essa linha, removemos a média (ou localização) dos dados, o que é uma forma de normalização. A submatriz de Helmert, que é a matriz de Helmert completa sem a

primeira linha, retém apenas as componentes ortogonais que descrevem a forma relativa dos pontos.

Portanto, para remover o efeito de locação da matriz de configuração $\mathbf{X}$, basta multiplicar a matriz de configuração pela sub-matriz de Helmert $\mathbf{H}$ de dimensão $(k-1)\times k$. Com isso, a configuração Hermetizada é dada por:

$$X^H_{(k-1)\times m} = H_{(k-1)\times k} X_{k\times m} \in \mathbb{R}^{(k-1)m} \tag{2.3}$$

A partir dessa transformação, utilizando a matriz H, temos uma matriz de configuração, agora, centrada X^H. Após retirar o efeito de locação, desejamos retirar o efeito de escala do objeto. Com isso em mente, considere a seguinte definição.

Definição 2.5.3. *O* ***tamanho do centroide*** *é dado como a norma euclidiana da matriz de configuração centrada*

$$S(X) = \|CX\| = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij} - \bar{X}_j)^2}, \quad X \in \mathbb{R}^{km},$$

onde $\bar{X}_j$ é a média dos valores na j-ésima coluna da matriz X.

em que X_{ij} é a (i,j)-ésima entrada de X, a média aritmética na j-ésima dimensão é $\bar{X}_j = \frac{1}{k}\sum_{i=1}^{k} X_{ij}$,

$$C = I_k - \frac{1}{k}\mathbf{1}_k\mathbf{1}_k^T$$

é a matriz de centralização, e

$$\|X\| = \sqrt{tr(X^T X)}$$

é a norma euclidiana, I_k é a matriz identidade $k\times k$ e $\mathbf{1}_k$ é o vetor de uns $k\times 1$. Note que C também pode ser usada, como uma alternativa, para a retirada da locação do objeto, de modo que

$$X_C = CX$$

são as coordenadas centradas dos *marcos*. Vale notar também que na configuração Helmertizada, 2.3, perdemos a dimensão original dos dados. Este problema é corrigido

multiplicando pela submatriz de Helmert transposta $\mathbf{H^T}$. Por $\mathbf{H^F}$ ser ortogonal, temos a seguinte relação entre ambas as alternativas de centralização

$$H^T H = I_k - \frac{1}{k}\mathbf{1}_k\mathbf{1}_k^T = C \tag{2.4}$$

e

$$H^T X^H = H^T H X = CX$$

Com o objetivo de retirar o efeito de escala, nós padronizamos dividindo pela noção de tamanho da 2.5.3. Primeiro, vamos calcular a norma da matriz de configuração centralizada X^H. A norma é definida como:

$$\|X^H\| = \sqrt{tr(X^T H^T H X)}$$

Substituindo essa relação 2.4 na equação da norma, obtemos:

$$\|X^H\| = \sqrt{tr(X^T H^T H X)} = \sqrt{tr(X^T C X)}$$

Como C é uma matriz simétrica e idempotente (ou seja, $C^T = C$ e $C^2 = C$). Portanto, podemos escrever:

$$\|X^H\| = \sqrt{tr(X^T C X)} = \sqrt{tr(X^T C^T C X)} = \|CX\|$$

Portanto, a relação completa para a padronização da matriz de configuração centralizada é:

$$\|X^H\| = \|CX\| = S(X).$$

Isso garante que a centralização seja aplicada corretamente. Observamos também que $S(X) > 0$, pois não permitimos um conjunto de *marcos* completamente coincidentes.

2.6 ESPAÇO DE FORMA

Ao retirar o efeito de escala de uma matriz de configuração, temos o que é chamado como espaço de pré-forma.

Definição 2.6.1. *A* ***Pré-Forma*** *da matriz de configuração é dada por*

$$Z = \frac{X^H}{\|X^H\|} = \frac{HX}{\|HX\|}$$

a qual é invariante sob translação e o escalonamento da configuração original.

Uma representação alternativa da pré-forma é inicialmente centralizar a configuração e depois dividir pelo tamanho. A *pré-forma centrada* é dada por

$$Z_C = \frac{CX}{\|CX\|} = H^T Z \tag{2.5}$$

Entretanto, o uso de **Z** nos traz vantagem do número de linhas ser inferior ao definido em C. Por outro lado, a vantagem de trabalhar com a pré-forma centrada Z_C é que um gráfico das coordenadas cartesianas fornece uma visão geométrica correta da forma da configuração original. Vamos usar a notação S_m^k para denotar o espaço de pré-forma de k pontos em m dimensões.

Definição 2.6.2. *O* ***Espaço de Pré-Forma*** *é o espaço de todas as pré-formas. Formalmente, o espaço de pré-forma S_m^k é o espaço em órbita das configurações não coincidentes do conjunto de k pontos em $\mathbb{R}^m$ sob os efeitos de translação e escala.*

A "órbita" refere-se a todas as configurações que podem ser alcançadas a partir de uma configuração original através dessas transformações, e trabalhar nesse espaço facilita a análise de formas de maneira consistente e robusta.

O espaço de pré-forma S_m^k é um conjunto de configurações de pontos que foram helmetizados e normalizados. Esse espaço é uma hiperesfera de raio unitário em $(k-1)m$ dimensões reais. A equivalência $S_m^k \equiv S^{(k-1)m-1}$ destaca que esse espaço é uma hiperesfera de $(k-1)m-1$ dimensões. O fato de $\|Z\| = 1$ indica que todas as configurações no espaço de pré-forma têm norma igual a 1, garantindo que a comparação entre formas não dependa da escala.

Fica mais fácil entender essa hiperesfera se olharmos para o caso das pré-formas Z_C. Ao aplicarmos a centralização, os dados correspondem ao subespaço de $\mathbb{R}^{km}$:

$$\mathbb{F}^{km-m} = \left\{ (x_1, \ldots, x_k) \in \mathbb{R}^{km} : \sum_{j=1}^{k} x_j = 0 \right\}.$$

Isso significa que, após a centralização, a soma das coordenadas de cada dimensão é zero, removendo o efeito de translação.

Ao normalizar os pontos, todos eles terão norma igual a 1, o que significa que estarão a uma distância unitária da origem, formando assim uma esfera. Considerando a esfera unitária, temos:

$$S^{km-1} = \left\{ (x_1, \ldots, x_k) \in \mathbb{R}^{km} : \sum_{j=1}^{k} \|x_j\|^2 = 1 \right\}.$$

A interseção do espaço centralizado com a esfera unitária é dada por:

$$S_*^{km-(m-1)} = \mathbb{F}^{km-m} \cap S^{km-1}.$$

Essa interseção $S_*^{km-(m-1)}$ é uma esfera $((k-1)m-1)$-dimensional dentro do espaço $\mathbb{R}^{km}$. Vamos nos referir a esta esfera como o espaço de pré-forma. Para aprofundar-se no que foi discutido acima, consulte o livro Small (SMALL, 2012) ou o livro do Kendall (KENDALL, 1984).

Agora, podemos observar que a análise estatística de formas não é algo simples, como discutimos na introdução. À medida que nos aprofundamos no assunto, vemos que a matemática envolvida se torna cada vez mais refinada. A partir deste ponto, como estamos trabalhando em uma esfera, precisaremos de conceitos de geometria diferencial, que serão abordados em seguida. Tendo removido os efeitos de locação e escala, resta-nos remover o efeito de rotação da nossa pré-forma, para assim obtermos o que definimos em 2.1.1. A rotação de uma configuração sobre a origem é dada pela pós-multiplicação da matriz de configuração X ou X_C, ambas com dimensões $k \times m$, pela matriz de rotação Γ, de dimensão $m \times m$.

Definição 2.6.3. *Uma* ***Matriz de Rotação*** $m \times m$ *satisfaz* $\Gamma^T\Gamma = \Gamma\Gamma^T = I_m$ *e* $\det(\Gamma) = +1$. *Uma matriz de rotação é uma matriz ortogonal especial, ou seja, uma matriz ortogonal com determinante* $+1$. *O conjunto de todas as matrizes de rotação* $m \times m$ *é conhecido como o grupo especial* $SO(m)$.

Uma matriz de rotação tem $\frac{1}{2}m(m-1)$ graus de liberdade. Quando $m = 3$, são necessários três ângulos. Por exemplo, pode-se considerar uma rotação no sentido horário de um ângulo θ_1, $-\pi \leq \theta_1 < \pi$ em torno do eixo z, seguida por uma rotação de um ângulo θ_2, $-\pi/2 \leq \theta_2 \leq \pi/2$ em torno do eixo x, e finalmente uma rotação de um ângulo θ_3, $-\pi \leq \theta_3 < \pi$ em torno do eixo z. Esta parametrização é conhecida como a convenção x e é única, exceto por uma singularidade em $\theta_2 = -\pi/2$, que faz as rotações θ_1 e θ_3 não possam mais serem distinguidas, se tornando redundantes, levando à perda de um grau de liberdade. A matriz de rotação é:

$$\Gamma = \begin{pmatrix} \cos\theta_3 & \sin\theta_3 & 0 \\ -\sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_2 & \sin\theta_2 \\ 0 & -\sin\theta_2 & \cos\theta_2 \end{pmatrix} \begin{pmatrix} \cos\theta_1 & \sin\theta_1 & 0 \\ -\sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Existem muitas escolhas de representações de ângulos, e todas possuem singularidades (STUELPNAGEL, 1964). A presença dessas singularidades significa que, embora os ângulos sejam uma maneira conveniente de representar rotações, eles apresentam limitações em certos casos. Em termos de perda de grau de liberdade, uma das rotações se torna redundante, e a rotação não pode ser parametrizada de forma única. Além disso, em termos de dificuldade computacional, as singularidades podem causar problemas em cálculos e simulações, exigindo tratamentos especiais para evitar erros numéricos. Portanto, agora podemos fazer a seguinte definição.

Definição 2.6.4. *A* ***Forma*** *de uma matriz de configuração* $\mathbf{X}$ *é toda informação geométrica acerca de* $\mathbf{X}$ *que é invariante sob translação, escala e rotação. A forma pode ser representada pelo conjunto* $[X]$ *dado por*

$$[X] = \{Z\Gamma : \Gamma \in SO(m)\}$$

em que $SO(m)$ *é o grupo especial ortogonal de rotações e* $\mathbf{Z}$ *é a pré-forma de* $\mathbf{X}$*. A notação para o espaço de forma de* k *pontos e* m *dimensões é* Σ_m^k*.*

Definição 2.6.5. *O* ***Espaço de Forma*** *é o espaço de todas as formas. Formalmente, o espaço de forma* Σ_m^k *é o espaço em órbita das configurações não coincidentes do conjunto de* k *pontos em* $\mathbb{R}^m$ *sob efeitos de translação, rotação e escala.*

É importante observar que a dimensão do espaço de formas é dada por

$$q = km - m - 1 - \frac{m(m-1)}{2}.$$

Isto é devido ao fato de que, inicialmente, tínhamos $k \times m$ coordenadas e, em seguida, perdemos dimensões da seguinte maneira. m dimensões devido à translação, pois a centralização remove m graus de liberdade (um por cada dimensão). Perde 1 dimensão devido à escala, pois a normalização remove um grau de liberdade, ajustando o tamanho total. Perde $\frac{m(m-1)}{2}$ dimensões devido à rotação, pois para m dimensões existem $\frac{m(m-1)}{2}$ pares de eixos para os quais a ortogonalidade deve ser mantida.

Podemos alterar a ordem de remoção das transformações de similaridade ou remover apenas algumas das transformações. Por exemplo, se a locação e a rotação forem removidas, mas não a escala, então temos o tamanho-e-forma de X.

Definição 2.6.6. O tamanho-e-forma *de uma matriz de configuração X é toda a informação geométrica sobre X que é invariante sob locação e rotação (transformações rígidas), e isso pode ser representado pelo conjunto $[X]_S$ dado por:*

$$[X]_S = \{X_H\Gamma : \Gamma \in SO(m)\},$$

onde X_H são as coordenadas Helmertizadas. Usamos a notação $S\Sigma_m^k$ para denotar o espaço de tamanho-e-forma de k pontos em m dimensões.

Definição 2.6.7. O espaço de tamanho-e-forma *é o espaço de todas as formas e tamanhos. Formalmente, o espaço de tamanho-e-forma $S\Sigma_m^k$ é o espaço de órbitas de configurações de conjuntos de pontos em $\mathbb{R}^m$ sob a ação de translação e rotação.*

O tamanho-e-forma também é conhecido como **forma**, particularmente em biologia. Se a escala for removida do tamanho-e-forma (por exemplo, redimensionando para o tamanho unitário do centróide), então obtemos a forma de X,

$$[X] = [X]_S/S(X) = \{Z\Gamma : \Gamma \in SO(m)\}.$$

2.7 DISTÂNCIA NO ESPAÇO DE FORMA

Anteriormente, vimos que o espaço de pré-forma S_m^k é uma hiperesfera de raio unitário em $\mathbb{R}^{(k-1)m}$. Medir distâncias na esfera é diferente de medir distâncias no espaço euclidiano. É aí que surge a importância de uma nova métrica. Temos a necessidade de calcular distâncias na esfera, e tal necessidade é atendida por uma métrica riemanniana ou alguma distância topologicamente equivalente. No contexto da Análise Estatística de Formas, o espaço de pré-forma S_m^k é uma variedade de Riemann isto é essencial para definirmos medidas de distância entre dois pontos. Antes de introduzirmos essa métrica riemanniana, precisamos compreender o conceito de variedade.

Uma variedade é uma generalização do nosso entendimento de uma superfície curva em três dimensões. Costumamos pensar em uma superfície curva como um subconjunto do espaço euclidiano tridimensional $\mathbb{R}^3$ que herda as propriedades geométricas da estrutura geométrica do espaço euclidiano em que está inserida. A representação de um espaço como um subconjunto de outro espaço é formalmente chamada de imersão. No entanto, nossa intuição, sendo limitada a objetos em dimensões menores ou iguais a três, tem dificuldade em visualizar a curvatura de conjuntos ou espaços que não podem ser imersos no espaço euclidiano tridimensional. A definição formal de uma variedade diferenciável não tem essa limitação. Como grande parte do cálculo envolve construções locais, variedades diferenciáveis, que localmente se assemelham ao espaço euclidiano, tornam-se um domínio natural para operações como calcular o gradiente de uma função, calcular vetores tangentes e outras construções do cálculo multivariável. Exemplos de variedades diferenciáveis são comuns: um toro, uma esfera ou um plano.

Uma variedades Riemanniana é uma generalização dos conceitos métricos, diferenciais e topológicos do espaço euclidiano para objetos geométricos que, localmente, têm a mesma estrutura que o espaço euclidiano, mas globalmente podem representar formas "curvas". Com efeito, os exemplos mais simples de variedades de Riemann são precisamente superfícies curvas de $\mathbb{R}^3$ e subconjuntos abertos de $\mathbb{R}^n$. Para mais informações sobre o tópico, o livro do Small (SMALL, 2012) aborda os conceitos de variedades no espaço de formas e o livro de (CARMO M, 2008) é uma boa leitura como base para o assunto.

Duas pré-formas Z_1 e Z_2 estão na mesma classe de equivalência $[X]$ se possuem

a mesma forma. Nesse caso, existe uma rotação tal que $\Gamma(Z_1) = Z_2$. No entanto, um conjunto de classes de equivalência tem pouco valor a menos que possamos compará-las e obter alguma intuição geométrica sobre Σ_m^k. Para isso, precisamos definir uma métrica em Σ_m^k. Se pensarmos em Σ_m^k como um espaço, então seus elementos podem ser considerados como pontos no espaço, para o qual queremos uma definição apropriada de distância. Uma forma de fazer isso é usar uma métrica entre órbitas no espaço de pré-formas. Como as pré-formas podem ser representadas como pontos na esfera, a distância entre as pré-formas é a distância geodésica [1], ou "great circle distance", entre elas.

2.8 DISTÂNCIA PROCRUSTES

O termo “Procrustes” é usado porque as operações de correspondência acima são idênticas às da análise de Procrustes, uma técnica comumente usada para comparar matrizes (até transformações) na análise multivariada.

Considere duas matrizes de configurações em $\mathbb{R}^{km}$. X_1 e X_2, com respectivas pré-formas Z_1 e Z_2. Descreveremos inicialmente as distâncias de Procrustes parcial e completa. Primeiro, minimizamos sobre rotações para encontrar a menor distância euclidiana entre Z_1 e Z_2.

Definição 2.8.1. *A distância de Procrustes parcial d_P é obtida ao corresponder as pré-formas Z_1 e Z_2 de X_1 e X_2 da forma mais próxima possível, considerando rotações. Assim,*

$$d_P(X_1, X_2) = \inf_{\Gamma \in SO(m)} \|Z_2 - Z_1\Gamma\|,$$

onde $Z_j = HX_j/\|HX_j\|$, $j = 1, 2$.
Aqui “inf” denota o infimo e iremos nos referir ao “sup” para supremo.

Teorema 2.8.2. *A distância de Procrustes parcial é dada por:*

$$d_P(X_1, X_2) = \sqrt{2}\left(1 - \sum_{i=1}^{m} \lambda_i\right)^{1/2},$$

[1] Uma geodésica é o caminho mais curto entre dois pontos em uma variedade Riemanniana, generalizando o conceito de linha reta em espaços curvos.

onde $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{m-1} \geq |\lambda_m|$ *são as raízes quadradas dos autovalores de* $Z_1^T Z_2 Z_2^T Z_1$*, e o menor valor* λ_m *é a raiz quadrada negativa se, e somente se,* $\det(Z_1^T Z_2) < 0$.

Definição 2.8.3. *A distância de Procrustes completa entre* X_1 *e* X_2 *é*

$$d_F(X_1, X_2) = \inf_{\Gamma \in SO(m), \beta \in \mathbb{R}^+} \|Z_2 - \beta Z_1 \Gamma\|,$$

onde $Z_r = HX_r/\|HX_r\|$, $r = 1, 2$.

Teorema 2.8.4. *A distância de Procrustes completa é*

$$d_F(X_1, X_2) = \left\{1 - \left(\sum_{i=1}^{m} \lambda_i\right)^2\right\}^{1/2}.$$

em que $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{m-1} \geq |\lambda_m|$ *são as raízes quadradas dos autovalores de* $Z_1^T Z_2 Z_2^T Z_1$*, e o menor valor* λ_m *é a raiz quadrada negativa se, e somente se,* $\det(Z_1^T Z_2) < 0$.

Utilizaremos distância parcial de Procrustes no espaço de pré-forma para fins práticos. As distâncias de Procrustes são tipos de distâncias extrínsecas, pois são realmente medidas em uma imersão do espaço de pré-formas. Elas consideram como o espaço de formas está posicionado dentro de um espaço maior e medem a distância com base nessa imersão. Em contraste, outro tipo de distância é a distância intrínseca, que é definida diretamente no espaço das formas. As distâncias intrínsecas levam em conta apenas a geometria interna do próprio espaço de formas, sem referência a um espaço maior. Portanto, enquanto as distâncias de Procrustes são úteis para comparar formas considerando sua posição em um espaço maior, as distâncias intrínsecas são mais adequadas para entender a estrutura interna do espaço de formas.

2.9 DISTÂNCIA DE RIEMANN

Definição 2.9.1. *A distância de Riemann* $\rho(X_1, X_2)$ *é a menor distância de círculo máximo entre* Z_1 *e* Z_2 *na esfera de pré-formas, onde* $Z_j = HX_j/\|HX_j\|$, $j = 1, 2$. *A minimização é realizada sobre rotações.*

Teorema 2.9.2. *A distância de Riemann ρ é*

$$\rho(X_1, X_2) = \arccos\left(\sum_{i=1}^{m} \lambda_i\right),$$

onde $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{m-1} \geq |\lambda_m|$ são as raízes quadradas dos autovalores de $Z_1^T Z_2 Z_2^T Z_1$, e o menor valor λ_m é a raiz quadrada negativa se, e somente se, $\det(Z_1^T Z_2) < 0$.

As demonstrações dos teoremas 2.8.2, 2.8.4, 2.9.2 podem ser vistas em (DRYDEN; MARDIA, 2016). A distância Riemanniana definida em 2.9.1 é uma métrica Riemanniana. Embora a distância de Procrustes completa definida em 2.8.3 não forneça a métrica Riemanniana no espaço de formas, essas duas distâncias são topologicamente equivalentes como podemos ver em (KENDALL et al., 2009) e a maioria dos resultados de distribuição assintótica de médias amostrais intrínsecas em variedades Riemannianas são herdados por ela. Existe relações entre d_F, d_P e ρ, onde

$$d_F(X_1, X_2) = \sin\rho,$$

$$d_P(X_1, X_2) = 2\sin(\rho/2).$$

Note que a distância de Riemann ρ pode ser considerada como o menor ângulo (com respeito às rotações das pré-formas) entre os vetores correspondentes a Z_1 e Z_2 na esfera de pré-formas. Em (3) temos uma representação para dar uma ideia de como podemos verificar nosso estudo tridimensional. No contexto de análise de formas, uma **fibra** se refere ao conjunto de todas as configurações que correspondem a mesma forma sob transformação de rotação, isto é, a pré-forma Z rotacionado na esfera de pré-formas é chamado de fibra do espaço de pré-formas S_m^k. As fibras $[X_1]$ e $[X_2]$ representam todos os pontos na esfera de pré-formas que são obtidos rotacionando Z_1 e Z_2, respectivamente, ou seja, $[X_1]$ é o conjunto de todas as pré-formas que podem ser obtidas a partir de Z_1 por qualquer rotação, e $[X_2]$ é o conjunto de todas as pré-formas que podem ser obtidas a partir de Z_2 por qualquer rotação.

Portanto, em termos de distâncias, a distância de **Menor Grande Círculo (ρ)**, é a menor distância de arco entre dois pontos na superfície da esfera de pré-formas. Essa distância considera o caminho mais curto ao longo da superfície da esfera. No

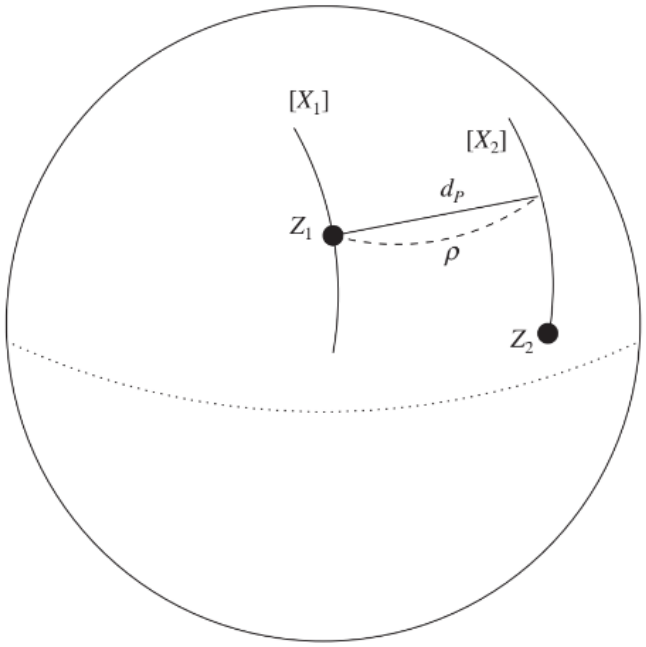


Figura 3 – Uma visão esquemática diagramática de duas fibras $\mathcal{F}_{[Z_1]}$ e $\mathcal{F}_{[Z_2]}$ na esfera de pré-formas, que correspondem às formas das matrizes de configuração originais X_1 e X_2 que possuem pré-formas Z_1 e Z_2. Também são exibidas o menor grande círculo ρ e a distância cordal d_P entre as fibras. (DRYDEN; MARDIA, 2016)

contexto das fibras $\mathcal{F}_{[Z_1]}$ e $\mathcal{F}_{[Z_1]}$, ρ representa a menor distância de grande círculo entre qualquer ponto em $\mathcal{F}_{[Z_1]}$ e qualquer ponto em $\mathcal{F}_{[Z_1]}$.

Já a **Distância procruster parcial (d_P)** É a distância direta entre dois pontos na esfera de pré-formas, medida através do espaço tridimensional em que a esfera está imersa. No contexto das fibras $\mathcal{F}_{[Z_1]}$ e $\mathcal{F}_{[Z_2]}$, d_P representa a menor distância cordal entre qualquer ponto em $\mathcal{F}_{[Z_1]}$ e qualquer ponto em $\mathcal{F}_{[Z_2]}$.

Já a **Distância de Procrustes completa (d_F)** é a menor distância entre os pontos correspondentes das formas ajustadas por rotação e escala. Isso pode ser visualizado como encontrar o ponto na fibra $[X_1]$ (considerando rotações e mudanças de escala) que está mais próximo de um ponto na fibra $\mathcal{F}_{[Z_2]}$.

Para formas próximas, não será visto muita diferença entre comparações das distâncias de formas que vimos nas definições2.9.1, 2.8.1 e 2.8.3. Portanto, para muitos conjuntos de dados práticos com pequena variabilidade, há pouca diferença nas análises quando se utilizam diferentes distâncias Procrustes. A tabela 1 dispõe um resumo das medidas de distância no espaço de forma e os intervalos nas quais elas assumem valores.

Distância	**Notação**	**Fórmula**	**Intervalo**
Distância de Procrustes Completa	d_F	$\left\{1-\left(\sum_{i=1}^{m}\lambda_i\right)^2\right\}^{1/2}$	$0 \leq d_F \leq 1$
Distância de Procrustes Parcial	d_P	$\sqrt{2(1-\sum_{i=1}^{m}\lambda_i)^{1/2}}$	$0 \leq d_P \leq \sqrt{2}$
Distância de Riemann	ρ	$\arccos\left(\sum_{i=1}^{m}\lambda_i\right)$	$0 \leq \rho \leq \pi/2$

Tabela 1 – Distâncias no espaço de formas.

2.10 FORMA MÉDIA

Como vimos nas seções anteriores, o espaço de formas não são espaços lineares. Em termos simples, podemos dizer que todos são espaços métricos, organizados naturalmente em diferentes níveis, onde cada nível é uma variedade Riemanniana. Para entender os problemas que podem surgir quando tentamos definir uma média na ausência de uma estrutura linear dada, basta considerar as mais simples variedades não lineares. Para uma medida de probabilidade em um arco de um círculo, poderíamos mapear o arco isometricamente em um intervalo da linha real, isto é, criar uma correspondência entre esses dois espaços que preserva as distâncias. Se dois pontos estão a uma certa distância no arco, a mesma distância será mantida na reta. Iremos realizar nosso cálculo da média lá da maneira usual e, em seguida, transferi-la de volta para o arco. O resultado certamente estaria de acordo com nossa concepção intuitiva de onde a média deveria estar. No entanto, o que aconteceria se a medida de probabilidade fosse suportada em um círculo inteiro? Não há como mapear isso isometricamente para a linha real. Poderíamos, de maneira bastante artificial e certamente não intrínseca, escolher uma imersão isométrica do círculo no plano e realizar o cálculo clássico para uma distribuição planar. No entanto, isso nos daria uma média que não está no círculo. Se isso fosse um cálculo de forma, então teríamos uma forma média que não é uma forma.

Agora estamos em posição de introduzir o conceito de forma média de um dado conjunto de realizações de forma. Enfrentamos novamente o problema de que, em espaços não euclidianos, não existe um único conceito de média que corresponda, como nos espaços euclidianos, à média aritmética. Em nosso procedimento, precisamos usar uma média do tipo Fréchet (Fréchet 1948). Três dos mais bem-sucedidos são as noções

de médias de Fréchet de variáveis aleatórias (cf. Ziezold, 1977), de baricentros de medidas (cf. Emery, 1989) e de centros de massa Riemannianos (cf. Karcher, 1977). Essas médias são todas definidas em um espaço métrico.

Definição 2.10.1. (A Média de Procrustes Completa) *Dado um conjunto de matrizes de configuração $X_1, \ldots, X_n$, a média de Procrustes completa em Σ_m^k é dada por $[\hat{\mu}]$, onde*

$$[\hat{\mu}] = \arg \inf_{\mu : S(\mu)=1} \sum_{i=1}^{n} d_F^2(X_i, \mu) \tag{2.6}$$

No entanto, para $m \geq 3$, o procedimento de correspondência não é uma expressão linear e requer um algoritmo iterativo. Mas antes, precisamos fazer algumas definições.

Existe uma abordagem de mínimos quadrados para encontrar uma estimativa de $[\mu]$, que é a análise de Procrustes generalizada (GPA). Veremos que o GPA fornece um método prático de calcular a média completa de procrustes de um conjunto de Formas, como definidos em 2.10.1.

Definição 2.10.2. *O método do GPA completo envolve transladar, redimensionar e rotacionar as configurações em relação umas às outras de modo a minimizar uma soma total de quadrados*

$$G(X_1, \ldots, X_n) = \sum_{i=1}^{n} \left\| \beta_i X_i \Gamma_i + 1_k \gamma_i^T - \mu \right\|^2 \tag{2.7}$$

com respeito a $\beta_i, \Gamma_i, \gamma_i$, $i = 1, \ldots, n$ e μ, sujeito a uma restrição de tamanho geral. A restrição sobre os tamanhos pode ser escolhida de várias maneiras. O ajuste de Procrustes generalizado completo envolve o registro de todas as configurações em posições ótimas, isto é, rotacionando e redimensionando cada figura para minimizar a soma das distâncias euclidianas ao quadrado entre cada figura transformada e a média estimada.

Definição 2.10.3. *O ajuste de Procrustes completo de cada uma das matrizes de configuração X_i é dado por:*

$$X_i^P = \hat{\beta}_i X_i \hat{\Gamma}_i + 1_k \hat{\gamma}_i^T, \quad i = 1, \ldots, n.$$

onde $\hat{\Gamma}_i \in SO(m)$ *(matriz de rotação),* $\hat{\beta}_i > 0$ *(parâmetro de escala),* $\hat{\gamma_i}^T$ *(parâmetros de localização),* $i = 1, \ldots, n$*, são os parâmetros minimizadores.*

Um algoritmo para estimar os parâmetros de transformação $(\gamma_i, \beta_i, \Gamma_i)$ é descrito abaixo. Os parâmetros $(\Gamma_i, \beta_i, \gamma_i)$.

Teorema 2.10.4. *O ponto no espaço de forma correspondente à média aritmética dos ajustes de Procrustes,*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i^P, \quad (7.15)$$

tem a mesma forma que a média completa de Procrustes, que foi definida na Equação 2.10.1.

Após uma coleção de objetos ter sido ajustada em posições ótimas de Procrustes completas em relação uns aos outros, o cálculo da forma média de Procrustes completa é simples; é calculado tomando as médias aritméticas de cada coordenada. Para dados bidimensionais, uma solução explícita de autovetor para o problema de otimização na definição da média de Procrustes completa está disponível, como discutido no livro de Mardia.

Também será necessário desenvolver um entendimento da matriz de covariância para a análise de formas, de modo a adaptar posteriormente algoritmos de análise multivariada ao contexto de pré-formas. A covariância multivariada captura a relação linear entre múltiplas variáveis em um conjunto de dados, formando uma matriz de covariância essencial para técnicas estatísticas multivariadas. Para dados de formas, adaptamos este conceito usando a média de Fréchet. Utilizando as formas centralizadas, a matriz de covariância Σ é calculada como:

$$\Sigma = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu)(X_i - \mu)^T \quad (2.8)$$

Onde X_i é a nossa pré-forma e μ é a média Fréchet. Esse método permite aplicar a análise estatística multivariada a dados de formas, garantindo que a estrutura geométrica dos dados seja respeitada e capturada corretamente.

Algoritmo 1: O Algoritmo da Média de Procrustes Completa

Entrada: Conjunto de dados $D = \{X_1, X_2, \ldots, X_N\}$, em que $X_i \in \mathbb{R}^{m \times k}$

Passo 1: Centralize as configurações para remover a localização. Seja $X_i, i = 1, \ldots, n$ as configurações centralizadas.

Inicialmente, deixe

$$X_i^P = X_i, \quad i = 1, \ldots, n$$

Passo 2: Calcule $G = \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\| X_i^P - X_j^P \right\|^2$.

Para a i-ésima configuração, seja

$$\bar{X}_{(i)} = \frac{1}{n-1} \sum_{j \neq i} X_i^P.$$

Otimize $\left\| \bar{X}_{(i)} - X_i^P \Gamma \right\|^2$ sobre rotações.

Defina $X_i^P = X_i^P \hat{\Gamma}$, onde $\hat{\Gamma}$ é a matriz de rotação ótima. Repita para todos os i.

Calcule o novo valor de G.

Este processo é repetido até que G não possa ser mais reduzido.

Passo 3: Para a i-ésima configuração, calcule $\hat{\beta}_i = \left(\frac{\sum_{k=1}^{n} \left\| X_k^P \right\|^2}{\left\| X_i^P \right\|^2} \right)^{\frac{1}{2}} \phi_i$, onde ϕ_i é o i-ésimo componente do autovetor ϕ correspondente ao maior autovalor da matriz de correlação Φ do vec $\left(X_i^P \right)$.

Defina $X_i^P = \hat{\beta}_i X_i^P$.

Repita para todos os i.

Calcule o novo valor de G.

Passo 4: Repita os passos (ii) e (iii) até que G não possa ser mais reduzido.

Passo 5: $[\hat{\mu}] = \frac{1}{n} \sum_i^n X_i^P$.

2.11 VARIÂNCIA GENERALIZADA

Com uma única variável, a variância amostral é utilizada para descrever a quantidade de variação nas medições dessa variável. Quando temos p variáveis, a variação é descrita pela matriz de covariância S. A matriz de covariância S é definida como:

$$S = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$$

Segundo Johnson (JOHNSON; WICHERN, 2007), o conceito de variância generalizada é definido como o determinante da matriz de covariância. Geometricamente, a

raiz quadrada do determinante da matriz de covariância é proporcional ao volume do elipsoide que circunda a dispersão dos dados centrados na média.

$$|S|^{1/2} \propto \text{Volume do Elipsoide} \tag{2.9}$$

A variância total amostral, por outro lado, é dada pelo traço da matriz de covariância S:

$$\text{tr}(S) = \sum_{j=1}^{p} S_{jj}$$

Onde $\text{tr}(S)$ representa a soma das variâncias das variáveis individuais.

3 APRENDIZADO SUPERVISIONADO

Muitos problemas em análise de formas focam em modelos de classificação bidimensionais. No entanto, há poucos trabalhos quando se trata de formas tridimensionais. Dentro dessa abordagem, um trabalho que se destaca é o de Vinué (VINUÉ; SIMÓ; ALEMANY, 2016), que emprega técnicas de clustering não supervisionado para explorar conjuntos de dados de formas humanas 3D. Neste estudo, exploraremos uma categoria de modelos voltados para tarefas de classificação. O objetivo principal dos modelos de classificação é utilizar uma matriz de entrada X, composta por n exemplos e m características, e classificá-la em uma das K classes discretas, denotadas como C_k, onde $k = 1, 2, \ldots, K$. Essas classes são geralmente consideradas mutuamente exclusivas, o que significa que cada exemplo de entrada pertence a uma única classe. Em outras palavras, para cada vetor de características x_i na matriz X, o modelo de classificação irá atribuir uma única classe C_k, garantindo que não haja sobreposição entre as classes. O processo de classificação envolve o treinamento de um modelo com dados rotulados, permitindo que ele aprenda a distinguir entre as diferentes classes com base nos padrões observados nas características fornecidas.

Como vimos anteriormente, a análise de formas lida com matrizes de dimensões $m \times k$. Portanto, para nosso propósito, o modelo de classificação deve receber como entrada uma matriz $m \times k$ e atribuí-la a uma das K classes discretas. Para torná-las compatíveis com métodos de classificação padrão, uma estratégia comum é vetorizar as matrizes de configuração em vetores unidimensionais de dimensão $m \cdot k$. Por exemplo, uma matriz 68×3 seria transformada em um vetor unidimensional de 204 elementos; em outras palavras, para cada matriz de configuração, geraríamos um vetor de características x_i. No entanto, essa abordagem simplifica excessivamente os dados e resulta na perda de importantes propriedades geométricas dos objetos em questão. Portanto, neste trabalho, apresentamos algoritmos clássicos adaptados para trabalhar diretamente com as matrizes de configuração, preservando assim sua estrutura geométrica.

3.1 DIFERENÇA DE MODELAGEM ENTRE Σ_2^k E Σ_3^k

Embora a análise de formas bidimensionais ou planas seja amplamente explorada, o estudo de formas tridimensionais ainda é relativamente limitado. Isso se deve ao fato de que, para formas planas, nossas matrizes de configuração podem ser vetorizadas. A seguir, vamos demonstrar como isso pode ser feito para formas planas e explicar por que essa transformação não é aplicável a formas tridimensionais.

Para formas planas, podemos resumir a matriz apresentada em 2.1 para modelar formas planas, tomando $k \geq 3$ e $m = 2$, o que corresponde às formas planas.

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{k1} & x_{k2} \end{pmatrix}$$

Podemos escrever a matriz Y como um vetor complexo definido por:

$$z^0 = (x_{11} + i \cdot x_{12}, \ldots, x_{k1} + i \cdot x_{k2})^T = (z^0_{(1)}, \ldots, z^0_{(k)})^T$$

onde

$$(z^0)^T = \begin{pmatrix} x_{11} + i \cdot x_{12} \\ \vdots \\ x_{k1} + i \cdot x_{k2} \end{pmatrix}$$

Vamos ter $k \geq 3$ marcos em $\mathbb{C}$, $z^0 = (z^0_{(1)}, \ldots, z^0_{(k)})^T$ que não são todos coincidentes, onde

$$z^0_{(j)} = x^0_{(j)} + iy^0_{(j)}, \quad j = 1, \ldots, k, \quad i = \sqrt{-1}.$$

A localização é removida pela pré-multiplicação pela submatriz de Helmert H, fornecendo os marcos Helmertizados complexos

$$z_H = Hz^0,$$

onde H é definido em 2.5.1. O tamanho do centróide é:

$$S(z^0) = \{(z^0)^* C z^0\}^{1/2} = \|z_H\| = \sqrt{(z_H)^* z_H},$$

onde $(z^0)^*$ denota o conjugado complexo da transposta de z^0 e C é a matriz de centralização de 2.5.3. Assim, a forma pré-complexa z é obtida dividindo os marcos Helmertizados pelo tamanho do centróide,

$$z = z_H / S(z^0), \quad z \in S_2^k,$$

onde vemos que o espaço de pré-forma S_2^k é a esfera complexa em $k-1$ dimensões complexas:

$$CS^{k-2} = \{z : z^* z = 1, z \in \mathbb{C}^{k-1}\},$$

que é o mesmo que a esfera real de raio unitário em $2k-2$ dimensões reais, S^{2k-3}. Para remover a rotação, identificamos todas as versões rotacionadas de z umas com as outras, de modo que a forma de z^0 é:

$$[z^0] = \{ze^{i\theta} : 0 \leq \theta < 2\pi\}.$$

A esfera complexa CS^{k-2} que possui pontos z identificados com $ze^{i\theta}$ $(0 \leq \theta < 2\pi)$ é o espaço projetivo complexo CP^{k-2}.

Para entender mais sobre formas planas, indico continuar a leitura pelo livro de (DRYDEN; MARDIA, 2016) e trabalhos em modelos de classificação em formas planares de (CARVALHO; AMARAL, 2020) e (AMARAL LUIZ H. DORE; STOSIC, 2010).

Para formas tridimensionais, não existe uma transformação direta para o espaço projetivo complexo. Portanto, para utilizar modelos de aprendizado supervisionado com formas tridimensionais, é necessário adaptar os modelos de classificação para trabalharem diretamente com matrizes de configuração tridimensionais como dados de entrada. Nas seções seguintes, detalharemos como essa adaptação será realizada para cada modelo.

3.2 K - VIZINHOS MAIS PRÓXIMOS

O método dos K-Vizinhos mais Próximos (KNN) é uma técnica de aprendizado supervisionado usada para classificação e regressão como descrito em (COVER; HART, 1967). Este método baseia-se na ideia de que uma observação pode ser classificada ou seu valor predito com base nos valores das observações mais próximas a ela. O

KNN é simples e eficaz, especialmente em situações onde as relações entre os dados são altamente não lineares.

Para utilizar o KNN, escolhemos um valor de k, que representa o número de vizinhos mais próximos a serem considerados na decisão. A escolha do valor de k é crucial no método dos K-Vizinhos mais Próximos (KNN). Um valor pequeno de k, como 1, torna o modelo muito sensível ao ruído nos dados, o que pode levar a uma alta variância. Por exemplo, se considerarmos apenas o vizinho mais próximo, um outlier pode influenciar significativamente a classificação, resultando em predições incorretas. Por outro lado, um valor grande de k pode suavizar excessivamente as fronteiras de decisão, levando a um alto viés. Isso ocorre porque o modelo considera muitos vizinhos para tomar uma decisão, o que pode fazer com que ele não capture as nuances locais dos dados.

Portanto, é essencial encontrar um equilíbrio ao escolher k. Valores intermediários frequentemente funcionam melhor, e a validação cruzada é uma técnica comum para determinar o valor ideal de k. A abordagem adotada foi testar diferentes valores de k e utilizar a validação cruzada para determinar o valor de k que oferece a melhor performance.

Existem muitos trabalhos referentes ao KNN (COOMANS; MASSART, 1982), geralmente utilizando distâncias como a distância euclidiana, ou até mesmo a distância de Mahalanobis. Mas para o seguimento desse trabalho, iremos trabalhar com a distância procruster parcial 2.8.1 e a distância riemanniana 2.9.1.

Para classificação, a classe de uma nova observação é determinada pela classe mais comum entre os k vizinhos mais próximos. Formalmente, isso pode ser expresso como:

$$C_f = \text{mode}\{C_{n1}, C_{n2}, \ldots, C_{nk}\}$$

onde C_{ni} são as classes dos k vizinhos mais próximos.

Portanto, iremos escolher k vizinhos mais próximos das nossas matrizes de configuração, referente a distância procrustes e olhar o valor mais comum desses vizinhos e designar a nossa matriz de configuração a essa classe.

Algoritmo 2: K-Vizinhos mais Próximos com Distâncias Procrustes

Entrada: Conjunto de dados $D = \{(X_1, y_1), (X_2, y_2), \ldots, (X_N, y_N)\}$, em que $X_i \in \mathbb{R}^{m \times k}$ e $y_i \in \{C_1, C_2, \ldots, C_k\}$

Saída: Rótulo de classe C para uma nova matriz de configuração X

Passo 1: Para cada matriz de configuração X_i em D, compute a distância $d_p(X, X_i)$ usando a métrica de distância Procrustes:

$$d_p(X, X_i) = \text{dist}_{\text{Procrustes}}(X, X_i)$$

Passo 2: Ordene as distâncias computadas $d_p(X, X_i)$ em ordem crescente e deixe D' ser a lista ordenada de pares $(d_p(X, X_i), y_i)$.

Passo 3: Selecione os primeiros k elementos de D', denotados por D_k.

Passo 4: Conte as ocorrências de cada rótulo de classe em D_k, denotado por $n(C_i, D_k)$:

$$n(C_i, D_k) = \sum_{(d,y) \in D_k} \delta(y, C_i)$$

onde $\delta(a, b) = 1$ se $a = b$ e 0 caso contrário.

Passo 5: Atribua X à classe C com a contagem máxima:

$$C = \arg\max_{C_i} n(C_i, D_k)$$

3.3 ANÁLISE DE DISCRIMINANTE LINEAR

Como vemos no livro do Jhonson (JOHNSON; WICHERN, 2007) discriminação e classificação são técnicas multivariadas propostas com a finalidade de separar conjuntos distintos de objetos e com a alocação de novos objetos a grupos previamente definidos. Como um procedimento separativo, é frequentemente empregada uma única vez para investigar diferenças observadas quando as relações causais não são bem compreendidas. Procedimentos de classificação são menos exploratórios no sentido de que levam a regras bem definidas, que podem ser usadas para atribuir novas observações. A classificação normalmente exige mais estrutura de problema do que a discriminação. Dentre os objetivos que existe em análise de discriminante, queremos descrever algebricamente, as características diferenciais dos objetos de estudo. Queremos tentar encontrar "discriminantes" cujos valores numéricos sejam tais que as coleções estejam

separadas tanto quanto possível, essa terminologia foi introduzida por R. A. Fisher na primeira abordagem moderna de problemas separativos. Também procuramos classificar objetos em duas ou mais classes rotuladas. A ênfase está em derivar uma regra que possa ser usada para atribuir de forma ideal novos objetos às classes rotuladas, que é o nosso objetivo de classificação.

Aqui adaptaremos a técnica de Análise Discriminante Linear (LDA), que é uma abordagem estatística amplamente utilizada para classificar dados em grupos, identificando padrões nas variáveis que melhor diferenciam as classes. Seu objetivo principal é encontrar o eixo ou hiperplano que maximize a separação entre os grupos, minimizando a sobreposição entre eles. Isso permite uma classificação mais precisa de novas observações, ao enfatizar as diferenças entre as classes.

Para $m = 2$ classes, o objetivo do LDA é criar um novo eixo no qual os dados sejam projetados de forma a maximizar a separação entre essas duas categorias. O LDA determina esse novo eixo com base em dois critérios simultâneos.

Quando os dados são projetados no novo eixo, queremos maximizar a distância entre as médias das classes. No entanto, isso não é suficiente por si só, pois pode haver sobreposição significativa entre as classes em algumas regiões, como discutido no livro de Christopher (BISHOP, 2016).

O segundo critério é minimizar a variância dentro de cada classe. Portanto, o objetivo do LDA é maximizar a função $J(W)$, onde W é o vetor que define a direção do novo eixo, e $J(W)$ é definida como:

$$J(W) = \frac{(\mu_1 - \mu_2)^2}{S_1^2 + S_2^2} \tag{3.1}$$

onde μ_1 e μ_2 são as médias das classes 1 e 2 projetadas no eixo definido por W, e S_1 e S_2 são as variâncias dentro de cada classe projetadas no eixo definido por W.

Iremos explorar neste estudo tanto o caso para $C = 2$ quanto para $C = 3$. No entanto, o caso $C = 2$ pode ser derivado do caso $C \geq 3$, como mostraremos adiante. Portanto, iremos dar foco no caso onde $C \geq 3$. Nesse cenário, o objetivo original do LDA para duas classes não se aplica diretamente, pois temos mais do que duas médias. Agora, nosso objetivo é maximizar a distância das médias de cada classe em relação a um ponto central representativo de todos os dados. Esse ponto central pode ser a média

global. Assim, buscamos maximizar a variância no conjunto das médias das classe e ao mesmo tempo, buscamos minimzar a variância dentro de cada classe. Sendo assim, a fórmula que apresentamos para duas classes 3.1 pode ser extendida da seguinte forma

$$J(W) = \frac{d_1^2 + d_2^2 + d_3^2}{S_1^2 + S_2^2 + S_3^2} \tag{3.2}$$

onde $d_i = (\mu_i - \mu)$, μ_i é a média da classe i projetada no eixo definido por W, μ é a média global projetada no eixo W, e S_i é a variância dentro da classe i projetada no eixo definido por W.

$$\mathbf{S}_W = \sum_{C=1}^{K} \mathbf{S}_C$$

onde

$$\mathbf{S}_C = \sum_{n \in \mathcal{C}_k} (\mathbf{x}_n - \mu_c)(\mathbf{X}_n - \mu_c)^T$$

e μ_c é a projeção da média Frechet definida em 2.10.1 calculada para cada classe.

Agora, vamos derivar a fórmula para a distância entre as classes projetadas. Para expressar a distância entre as médias das classes de forma geral, utilizamos a seguinte fórmula

$$\sum_{j=1}^{C} N_j(\mu_j - \mu)^2$$

onde μ_j é a média da classe j projetada no novo eixo e $\mathbf{m}_j$ é a média de Fréchet da classe j. No entanto, o modelo original faz uso da média ponderada, levando em consideração o tamanho de cada classe N_j. Portanto,

$$\begin{aligned}\sum_{j=1}^{C} N_j(\mu_j - \mu)^2 &= \sum_{j=1}^{C} N_j(W^T\mathbf{m}_j - W^T\mathbf{m})^2 \\ &= \sum_{j=1}^{C} N_j(W^T(\mathbf{m}_j - \mathbf{m}))^2 \\ &= \sum_{j=1}^{C} W^T N_j(\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^T W \\ &= W^T \left(\sum_{j=1}^{C} N_j(\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^T \right) W \\ &= W^T S_B W\end{aligned}$$

De forma geral, somando as contribuições de todas as classes, temos

$$\sum_{j=1}^{C} N_j(\mu_j - \mu)^2 = W^T S_B W$$

onde

$$\mathbf{S}_B = \sum_{k=1}^{K} N_k(\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})^T$$

é a matriz de covariância entre as classes, com $\mathbf{m}_k$ sendo a média Frechet definida em 2.10.1 para classe k, $\mathbf{m}$ sendo a média Frechet global, e N_k sendo o número de amostras na classe k. Agora, vamos observar a variância projetada em W

$$\begin{aligned}S_j^2 = \sum_{X_i \in C_j} (A_i - \mu_j)^2 &= \sum_{X_i \in C_j} (W^T X_i - W^T m_j)^2 \\ &= \sum_{X_i \in C_j} W^T (X_i - m_j)(X_i - m_j)^T W \\ &= W^T \left(\sum_{X_i \in C_j} (X_i - m_j)(X_i - m_j)^T \right) W \\ &= W^T S_j W\end{aligned}$$

onde A_i é a projeção das nossas matrizes de configuração X_i em W. Onde

$$S_j = \sum_{X_i \in C_j} (X_i - m_j)(X_i - m_j)^T$$

é chamada de matriz de dispersão dentro da classe. Dessa forma, fazendo de forma análoga para todos S_j, temos

$$\sum_{j=1}^{C} S_j = \sum_{j=1}^{C} w^T S_j w = w^T \left(\sum_{j=1}^{C} S_j \right) w = w^T S_W w$$

Juntando tudo, podemos escrever a função $J(w)$ como

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

Nosso objetivo é encontrar w que maximize a expressão acima.

Agora, tomando $C = 2$, vamos poder verificar em $\sum_{j=1}^{C} N_j(\mu_j - \mu)^2$ que quando existem apenas duas classes, as ponderações são apenas múltiplos escalares das definições não ponderadas. Portanto o LDA multiclasse $\frac{\sum_{j=1}^{C} N_j(\mu_j - \mu)^2}{\sum_{j=1}^{C} S_j}$ é uma generalização natural para LDA com apenas duas classes visto em 3.1.

Para maximizar $J(w)$, precisamos encontrar a derivada de $J(w)$ em relação a w e igualá-la a zero.

Teorema 3.3.1. *Se* $\mathbf{A}$ *não é uma função de* $\mathbf{u}$*, e* $\mathbf{A}$ *é simétrica, então* $\frac{\partial}{\partial \mathbf{u}} \left(\mathbf{u}^\top \mathbf{A} \mathbf{u} \right) = 2\mathbf{A}\mathbf{u}$.

A demonstração do teorema acima pode ser encontrado no libro (MAGNUS; NEUDECKER, 2019). Vamos reescrever $J(w)$ como uma razão de duas funções.

$$J(w) = \frac{f(w)}{g(w)}$$

onde $f(w) = w^T S_B w$ e $g(w) = w^T S_W w$.
Derivada de $f(w) = w^T S_B w$

$$\frac{d}{dw} f(w) = \frac{d}{dw}(w^T S_B w)$$

Usando o fato de que S_B é uma matriz simétrica, a derivada é

$$\frac{d}{dw}(w^T S_B w) = 2 S_B w$$

Derivada de $g(w) = w^T S_W w$:

$$\frac{d}{dw} g(w) = \frac{d}{dw}(w^T S_W w)$$

Usando o fato de que S_W é uma matriz simétrica, a derivada é

$$\frac{d}{dw}(w^T S_W w) = 2S_W w$$

Agora, substituímos as derivadas na regra do quociente

$$\frac{d}{dw}J(w) = \frac{(w^T S_W w)(2S_B w) - (w^T S_B w)(2S_W w)}{(w^T S_W w)^2}$$

Para maximizar $J(w)$, igualamos a derivada a zero

$$(w^T S_W w)S_B w - (w^T S_B w)S_W w = 0$$

Rearranjando, obtemos

$$(w^T S_W w)S_B w = (w^T S_B w)S_W w$$

Não nos importamos com a magnitude de w, apenas com sua direção. Portanto, podemos agrupar os fatores escalares $(w^T S_W w)$ e $(w^T S_B w)$, resultando em

$$S_B w = \lambda S_W w$$

onde λ é um escalar. O problema acima é um problema de autovalores e autovetores. Para encontrar w, resolvemos:

$$S_W^{-1} S_B w = \lambda w$$

Os autovetores correspondentes aos maiores autovalores λ são as direções que maximizam $J(W)$. Dessa forma, obtemos a matriz W composta pelos autovetores, que pode ser usada para projetar os dados originais em uma dimensão menor, maximizando a separabilidade das classes. No entanto, no nosso caso, manteremos a dimensão atual por razões computacionais. Embora isso possa incluir dimensões que não contribuem significativamente para a separação entre as classes, resultando em redundância e menor eficiência, optamos por essa abordagem para simplificar a implementação. A seguir, vamos detalhar o método de classificação que utilizaremos para atribuir novas observações às classes existentes.

Para cada classe C_j, a média da classe é projetada no novo espaço usando a matriz de projeção W

$$\mu_{\text{proj},j} = W^T \mu_j$$

onde μ_j é a média da classe C_j. Para cada ponto matriz de configuração nos dados de teste $\mathbf{X}_i$, o ponto é projetado no novo espaço usando a matriz de projeção W

$$\mathbf{X}_{\text{proj},i} = W^T \mathbf{X}_i$$

Para cada ponto de teste projetado, a distância até as médias projetadas das classes $\mu_{\text{proj},j}$ é calculada usando as distâncias procrustes definida anteriormente em 2.9.2 $d_P(\cdot, \cdot)$

$$d_{ij} = d(\mathbf{X}_{\text{proj},i}, \mu_{\text{proj},j})$$

Para cada matriz de configuração no conjunto de teste projetado, a classe é determinada com base na menor distância até as médias projetadas das classes. A classe C_j é atribuída ao ponto $\mathbf{x}_i$ se

$$j = \arg\min_k d_{ik}$$

Isto é, o índice da classe j com a menor distância d_{ij}. Finalmente, a acurácia é calculada como a proporção de pontos de teste que foram classificados corretamente

Algoritmo 3: Análise Discriminante Linear de Fisher Multiclasse (LDA)

Entrada: Conjunto de dados $D = \{(X_1, y_1), (X_2, y_2), \ldots, (X_N, y_N)\}$, onde $X_i \in \mathbb{R}^{m \times k}$ e $y_i \in \{C_1, C_2, \ldots, C_m\}$

Saída: Matriz de projeção W

Passo 1: Calcular as matrizes de configuração médios para cada classe:

$$m_i = \arg \inf_{\mu : S(\mu)=1} \sum_{i=1}^{N} d_F^2(M_i, \mu)$$

Passo 2: Calcular a matriz de dispersão dentro das classes S_W:

$$S_W = \sum_{i=1}^{m} S_i$$

$$S_i = \sum_{x \in C_i} (x - m_i)(x - m_i)^T$$

Passo 3: Calcular a matriz de dispersão entre as classes S_B:

$$S_B = \sum_{i=1}^{m} N_i (m_i - m)(m_i - m)^T$$

Passo 4: Resolver o problema de autovalores generalizado:

$$S_W^{-1} S_B w = \lambda w$$

Passo 5: Ordenar os autovalores em ordem decrescente.

Passo 6: Projetar os dados no novo espaço:

$$X' = XW$$

Passo 7: Classificação

$$d_{ij} = d(\mathbf{X}_{\text{proj},i}, \mu_{\text{proj},j})$$

$$C = \arg\min_k d_{ik}$$

onde C é a classe escolhida que tem a menor distância d_{ij}.

3.4 ANÁLISE DE DISCRIMINANTE QUADRÁTICO

Quadratic Discriminant Analysis (QDA) é uma técnica de classificação que assume que cada classe segue uma distribuição normal multivariada com matrizes de

covariância diferentes. Essa abordagem permite que o QDA se adapte melhor a dados cujas classes apresentam formas de distribuição distintas. Ao trabalharmos com QDA, assumimos que cada uma das k classes segue uma distribuição normal multivariada com média μ_k e matriz de covariância Σ_k

$$\mathbf{x} \sim \mathcal{N}(\mu_k, \Sigma_k)$$

Para uma observação $\mathbf{x}$ dada a classe k, a função de densidade de probabilidade é

$$P(\mathbf{x}|y = k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_k)^T \Sigma_k^{-1}(\mathbf{x} - \mu_k)\right)$$

Onde d é a dimensionalidade dos dados (número de variáveis), μ_k é o vetor de médias da classe k e Σ_k é a matriz de covariância da classe k. A probabilidade a priori de cada classe $P(y = k)$ é a proporção da classe k no conjunto de dados D.

Para determinar a qual classe uma nova observação pertence, precisamos calcular a probabilidade de que essa observação pertença a cada uma das classes possíveis. É aqui que o Teorema de Bayes se torna necessário. Usando o Teorema de Bayes, podemos combinar a função de densidade de probabilidade $P(\mathbf{x}|y = k)$ e a proporção a priori $P(y = k)$ para calcular a probabilidade a posteriori, que nos diz a probabilidade de uma observação pertencer a uma classe específica, dado os seus valores observados. O Teorema de Bayes é expresso da seguinte forma

$$P(y = k|\mathbf{x}) = \frac{P(\mathbf{x}|y = k)P(y = k)}{P(\mathbf{x})}$$

Como $P(\mathbf{x})$ é a mesma para todas as classes, podemos ignorá-la ao comparar as classes. Assim, temos

$$P(y = k|\mathbf{x}) \propto P(\mathbf{x}|y = k)P(y = k)$$

Vamos consider no momento o caso com apenas duas classes, cada uma tendo uma densidade Gaussiana condicional à classe, e suponha que temos um conjunto de dados $D = \{x_n, y_n\}$ onde $n = 1, \ldots, k$. Aqui $y_n = 1$ denota a classe C_1 e $y_n = 0$ denota a classe C_2. Denotamos a probabilidade a priori da classe $P(C_1) = \pi$, de modo que $P(C_2) = 1 - \pi$. Para um ponto de dados x_n da classe C_1, temos $t_n = 1$ e, portanto,

$$P(y = C_1|x_n) = P(C_1)p(x_n|y = C_1) = \pi\mathcal{N}(x_n|\mu_1, \Sigma_1).$$

De forma similar, para a classe C_2, temos $y_n = 0$ e, portanto,

$$P(y = C_2|x_n) = P(C_2)p(x_n|y = C_2) = (1 - \pi)\mathcal{N}(x_n|\mu_2, \Sigma_2).$$

Agora, queremos calcular os parâmetros π, μ e Σ de forma que possamos estimar quão bem uma nova observação se encaixa nesses parâmetros. Para isso, utilizamos a função de verossimilhança. Assim, a função de verossimilhança é dada por

$$L(X, \pi, \mu_1, \mu_2, \Sigma_1, \Sigma_2|y) = \prod_{n=1}^{N} [\pi\mathcal{N}(x_n|\mu_1, \Sigma_1)]^{t_n}[(1 - \pi)\mathcal{N}(x_n|\mu_2, \Sigma_2)]^{1-t_n}. \quad (3.3)$$

Para mais de duas classes, podemos generalizar a abordagem usando uma variável indicadora y_n, onde y_n, é definida como um vetor de indicadores para cada classe. Supondo m classes, uma observação x_n, o vetor de indicador y_n de comprimento m, onde $y_{nk} = 1$ se x_n pertence a classe C_k e 0 caso contrário. Poranto, para cada classe m

$$P(y = C_k|x_n) = P(C_k)p(x_n|y = C_k) = \pi_k\mathcal{N}(x_n|\mu_k, \Sigma_k).$$

onde π_k é a probabilidade a priori da classe k e $\mathcal{N}(x_n|\mu_k, \Sigma_k)$ é a função densidade de probabilidade da classe k. Usando vetor t_n, podemos escrever 3.3

$$L(X, \pi_1, \ldots, \pi_m, \mu_1, \ldots \mu_m, \Sigma_1, \ldots \Sigma_m|y) = \prod_{n=1}^{N}\prod_{k=1}^{m} \pi^{y_{nk}} P(x_n|y = k, \mu_k, \Sigma_k)^{y_{nk}}. \quad (3.4)$$

Para simplificar a notação, seja θ denotando todos os priors das classes, vetores médios específicos das classes e matrizes de covariância $\{\pi_1, \ldots, \pi_m, \mu_1, \ldots, \mu_m, \Sigma_1, \ldots, \Sigma_m\}$. Como sabemos, maximizar a verossimilhança é equivalente a maximizar a log-verossimilhança. A log-verossimilhança é

$$
\begin{aligned}
\ln L(X, \theta|y) &= \ln \prod_{n=1}^{N} \prod_{k=1}^{m} \pi_k^{y_{nk}} P(x_n|y = k, \mu_k, \Sigma_k)^{y_{nk}} \\
&= \sum_{n=1}^{N} \sum_{k=1}^{m} y_{nk} \left(\ln \pi_k + \ln P(x_n|y = k, \mu_k, \Sigma_k) \right) \\
&= \sum_{n=1}^{N} \sum_{k=1}^{m} y_{nk} \left(\ln \pi_k + \ln \left(\frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} \exp \left(-\frac{1}{2} (\mathbf{x}_n - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x}_n - \mu_k) \right) \right) \right) \\
&= \sum_{n=1}^{N} \sum_{k=1}^{m} y_{nk} \left(\ln \pi_k - \frac{d}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_k| - \frac{1}{2} (\mathbf{x}_n - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x}_n - \mu_k) \right)
\end{aligned}
$$

Expandindo a equação acima será útil nas derivadas subsequentes

$$
\ln L(X, \theta|y) = \sum_{n=1}^{N} \sum_{k=1}^{m} y_{nk} \left(\ln \pi_k - \frac{d}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_k| - \frac{1}{2} (\mathbf{x}_n - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x}_n - \mu_k) \right)
$$

$$
= \sum_{n=1}^{N} \sum_{k=1}^{m} y_{nk} \left(\ln \pi_k - \frac{d}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_k| - \frac{1}{2} (\mathbf{x}_n - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x}_n - \mu_k) \right) \tag{3.5}
$$

Precisamos encontrar a solução de máxima verossimilhança para π_k, μ_k e Σ_k. Começando com π_k, precisamos tomar a derivada de 3.5, igualá-la a 0 e resolver para π_k, entretanto, precisamos manter a restrição $\sum_{k=1}^{m} \pi_k = 1$. Para isso, iremos utilizar multiplicadores de Lagrange para incorporar a restrição acima. Primeiro, definimos a função Lagrangiana como

$$
\mathcal{L} = \ln L(X, \theta|y) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)
$$

Para encontrar os pontos críticos, tomamos as derivadas parciais da função Lagrangiana em relação a π_k e λ

$$
\frac{\partial \mathcal{L}}{\partial \pi_k} = \frac{\partial}{\partial \pi_k} \left(\ln \Pr(\mathbf{t}|\theta) + \lambda \left(\sum_{k=1}^{m} \pi_k - 1 \right) \right) = 0
$$

$$
\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{k=1}^{m} \pi_k - 1 = 0
$$

A derivada da log-verossimilhança em relação a π_k é dada por

$$
\sum_{n=1}^{N} \frac{\partial}{\partial \pi_k} (y_{nk} \ln \pi_k) + \lambda \frac{\partial}{\partial \pi_k} (\pi_k - 1) = 0
$$

Simplificando, obtemos

$$
\sum_{n=1}^{N} \frac{y_{nk}}{\pi_k} + \lambda = 0
$$

Resultando em

$$\pi_k \lambda = -N_k$$

Sabemos que $\sum_{k=1}^{m} \pi_k = 1$, então

$$\sum_{k=1}^{m} \pi_k \lambda = -\sum_{k=1}^{m} N_k$$

Como $\sum_{k=1}^{m} N_k = N$, temos

$$\lambda = -N$$

Substituindo $\lambda = -N$ na equação $\pi_k \lambda = -N_k$

$$-\pi_k N = -N_k$$

Portanto

$$\pi_k = \frac{N_k}{N} \tag{3.6}$$

A Equação 3.6 nos diz que o prior da classe é simplesmente a proporção de pontos de dados que pertencem à classe, o que intuitivamente também faz sentido. Agora, iremos maximizar a log-verossimilhança com respeito a μ_k. Novamente, usando o resultado da Equação 3.5, é temos que tomar a derivada com respeito a μ_k, igualá-la a 0 e resolver para μ_k. Como as matrizes de covariância são sempre simétricas, e o inverso de uma matriz simétrica também é simétrica, podemos usar 3.3.1 epara obter

$$\begin{aligned}
\sum_{n=1}^{N} \frac{\partial}{\partial \mu_k} \left(-\frac{y_{nk}}{2} (\mathbf{x}_n - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x}_n - \mu_k) \right) &= 0 \\
\sum_{n=1}^{N} y_{nk} \Sigma_k^{-1} (\mathbf{x}_n - \mu_k) &= 0 \\
\sum_{n=1}^{N} y_{nk} \Sigma_k^{-1} \mathbf{x}_n &= \sum_{n=1}^{N} y_{nk} \Sigma_k^{-1} \mu_k \\
\sum_{n=1}^{N} y_{nk} \Sigma_k^{-1} \mathbf{x}_n &= N_k \Sigma_k^{-1} \mu_k
\end{aligned}$$

$$\frac{1}{N_k} \sum_{n=1}^{N} y_{nk} \mathbf{x}_n = \mu_k. \tag{3.7}$$

Como y_{nk} é igual a 1 apenas para os pontos de dados que pertencem à classe C_k. Isso significa que a soma no lado esquerdo da Equação 3.7 inclui apenas as variáveis de entrada $\mathbf{x}$ que pertencem à classe C_k. Depois, estamos dividindo essa soma de vetores pelo número de pontos de dados na classe N_k, o que é o mesmo que calcular a média. Isso significa que o vetor médio específico da classe μ_k é a média das variáveis de entrada $\mathbf{x}_n$ que pertencem à classe, ou seja, o vetor médio específico da classe é apenas a média dos vetores da classe. Por último, precisamos maximizar a log-verossimilhança em relação à matriz de covariância específica da classe Σ_k. Novamente, tomamos a derivada em relação a Σ_k usando o resultado da Equação 3.5, igualamos a zero e resolvemos:

$$\ln L(X, \theta|y) = \sum_{n=1}^{N} \sum_{k=1}^{m} y_{nk} \left(\ln \pi_k - \frac{1}{2} \ln \det \Sigma_k - \frac{1}{2} (\mathbf{x}_n - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x}_n - \mu_k) \right)$$

Focamos na parte que depende de Σ_k

$$\sum_{n=1}^{N} y_{nk} \left(-\frac{1}{2} \ln \det \Sigma_k - \frac{1}{2} (\mathbf{x}_n - \mu_c)^\top \Sigma_k^{-1} (\mathbf{x}_n - \mu_c) \right)$$

Primeiro, derivamos a parte $-\frac{1}{2} \ln \det \Sigma_k$

$$\frac{\partial}{\partial \Sigma_k} \left(-\frac{1}{2} \ln \det \Sigma_k \right) = -\frac{1}{2} \Sigma_k^{-1}$$

Agora, derivamos a parte $-\frac{1}{2} (\mathbf{x}_n - \mu_c)^\top \Sigma_k^{-1} (\mathbf{x}_n - \mu_k)$. Usando a identidade do cálculo matricial

$$\frac{\partial}{\partial \Sigma_k} \left((\mathbf{x}_n - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x}_n - \mu_k) \right) = -\Sigma_k^{-1} (\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^\top \Sigma_k^{-1}$$

Portanto, a derivada total é

$$\sum_{n=1}^{N} y_{nk} \left(-\frac{1}{2} \Sigma_k^{-1} + \frac{1}{2} \Sigma_k^{-1} (\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^\top \Sigma_k^{-1} \right)$$

Igualando a zero

$$\sum_{n=1}^{N} y_{nk} \left(-\frac{1}{2} \Sigma_k^{-1} + \frac{1}{2} \Sigma_k^{-1} (\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^\top \Sigma_k^{-1} \right) = 0$$

Multiplicando por Σ_k e rearranjando os termos

$$\sum_{n=1}^{N} y_{nk} \left(\Sigma_k - (\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^\top \right) = 0$$

Dividindo ambos os lados por $\sum_{n=1}^{N} y_{nk} = N_k$

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} y_{nk}(\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^\top$$

Assim como o vetor médio específico da classe é apenas a média dos vetores da classe, a matriz de covariância específica da classe é apenas a covariância dos vetores da classe, Agora que calculamos as otimizações para π, μ e Σ, podemos desenvolver o algoritmo para o QDA. Com o resultado 3.5, podemos simplificar, pois como $\log(2\pi)^{d/2}$ é constante para todas as classes, podemos ignorá-la, obtendo assim

$$\log p(y = k|\mathbf{x}) \propto -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(\mathbf{x} - \mu_k)^\top \Sigma_k^{-1}(\mathbf{x} - \mu_k) + \log \pi_k$$

Finalmente, a função discriminante quadrática é:

$$\delta_k(\mathbf{x}) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(\mathbf{x} - \mu_k)^\top \Sigma_k^{-1}(\mathbf{x} - \mu_k) + \log \pi_k$$

O modelo padrão QDA classifica a nova observação $\mathbf{x}$ na classe k que maximiza $\delta_k(\mathbf{x})$. No entanto, não seguiremos essa abordagem convencional. Como estamos trabalhando com dados de forma, estamos lidando com matrizes de configuração como nossos dados de entrada, matrizes de média como as mostradas em 2.10.1 e matrizes de covariância 2.8, Portanto, nosso resultado será uma matriz e não um número. Dessa forma, baseado no que vimos no capítulo anterior 2.9, implementaremos o método da Variância Generalizada e o método da Variância Total Amostral, dessa forma, desenvolvemos um método para designar a Matriz de configuração a classe mais apropriada, onde o determinante e o traço da matriz atuarão como hiperparâmetros do modelo adaptado. Nosso objetivo é

$$\hat{y} = \arg\max_k \det\left(-\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(\mathbf{x} - \mu_k)^\top \Sigma_k^{-1}(\mathbf{x} - \mu_k) + \log \pi_k\right) \tag{3.8}$$

$$\hat{y} = \arg\max_k \operatorname{tr}\left(-\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(\mathbf{x} - \mu_k)^\top \Sigma_k^{-1}(\mathbf{x} - \mu_k) + \log \pi_k\right) \tag{3.9}$$

Com base na teoria desenvolvida, o algoritmo que utilizaremos para calcular o QDA é definido a seguir

Algoritmo 4: Análise Discriminante Quadrático Multiclasse (QDA)

Entrada: Conjunto de dados $D = \{(X_1, y_1), (X_2, y_2), \ldots, (X_N, y_N)\}$, onde $X_i \in \mathbb{R}^{m \times k}$ e $y_i \in \{C_1, C_2, \ldots, C_m\}$

Saída: Classe predita $\hat{y}$

Passo 1: Calcular as médias específicas da classe μ_k

$$\mu_k = \frac{1}{N_k} \sum_{i:y_i=k}^{N} X_i$$

Passo 2: Calcular as matrizes de covariância específicas da classe Σ_k

$$\Sigma_k = \frac{1}{N_k} \sum_{i:y_i=k} (X_i - \mu_k)(X_i - \mu_k)^\top$$

Passo 3: Calcular os priors das classes π_k

$$\pi_k = \frac{N_k}{N}$$

Passo 4: Calcular a função discriminante quadrática $\delta_k(X)$

$$\delta_k(X) = -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2}(X - \mu_k)^\top \Sigma_k^{-1} (X - \mu_k) + \log \pi_k$$

Passo 5: Classificar a nova observação X

$$\hat{y} = \arg\max_k det(\delta_k(X))$$

Neste capítulo, apresentamos com certo detalhe os modelos KNN, LDA de Fisher e QDA com o objetivo de classificação, que serão utilizados neste trabalho. No próximo capítulo, discutiremos os resultados da aplicação desses algoritmos em dados reais e simulados, analisando suas eficiências.

4 RESULTADOS PRINCIPAIS

Até agora, desenvolvemos algoritmos para lidar com dados de formas tridimensionais, mas é necessário testá-los para verificar sua capacidade de classificar essas formas com precisão. O objetivo deste capítulo é aplicar os algoritmos desenvolvidos anteriormente e avaliar os resultados usando tanto dados simulados quanto dados reais. Para verificar a relevância desses resultados, utilizaremos a métrica de acurácia, que representa a proporção de previsões corretas entre todas as previsões realizadas.

$$\text{Acurácia} = \frac{\text{Número de Previsões Corretas}}{\text{Número Total de Previsões}} \times 100$$

O número de previsões corretas refere-se à alocação correta de um objeto à sua classe. O número total de previsões é a soma das previsões corretas e incorretas, onde um erro representa a alocação incorreta de um objeto à sua classe.

4.1 DADOS SIMULADOS

Após toda a fundamentação teórica, aplicaremos os algoritmos modificados propostos em uma base de dados artificial, e com ela verificaremos como os algoritmos se comportam em cada um dos cenários propostos. Analogamente ao estudo de desempenho realizado por Vinue utilizando o método de k-means no contexto da análise de formas tridimensionais, realizamos uma simulação numérica com dados controlados.Primeiramente, geramos dois conjuntos de pontos que representam duas figuras geométricas distintas (um cubo e um paralelepípedo), com um número diferente de marcos ($h = 8$). Em seguida, aplicamos transformações específicas a esses pontos, incluindo escalonamento e rotação. Para aumentar a variabilidade, introduzimos rotações com ângulos aleatórios entre 0 e 2π e selecionamos aleatoriamente o eixo de rotação (x, y ou z) para cada figura. Além disso, aplicamos fatores de escala aleatórios dentro de intervalos específicos para diversificar ainda mais as formas geradas. Os pontos para cada figura geométrica são gerados a partir de distribuição normal multivariada com vetores de médias tridimensionais específicos para cada vértice e uma matriz de covariância Σ. Finalmente, concatenamos os pontos gerados para formar um conjunto de dados

que inclui n_1 cubos e n_2 paralelepípedos, cada um associado a um rótulo que indica sua respectiva classe. O objetivo, é buscar tornar os dados mais realistas e variados, proporcionando um ambiente de teste mais robusto para os algoritmos propostos.

Na Tabela 2, apresentamos as coordenadas dos vetores de média para as formas geométricas com $k = 8$.

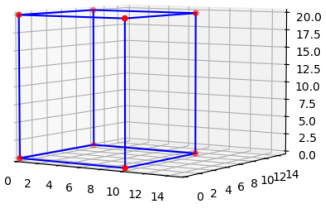
Tabela 2 – Coordenadas das formas médias (a) para o cubo e (b) para o paralelepípedo, no caso de $k = 8$

height**Landmark label**	**x**	**y**	**z**
(a)			
1	0	0	10
2	0	10	10
3	0	0	0
4	0	10	0
5	10	10	10
6	10	10	0
7	10	0	10
8	10	0	0
(b)			
1	0	0	10
2	10	0	10
3	0	0	0
4	10	0	0
5	10	20	10
6	10	20	0
7	0	20	10
8	0	20	0

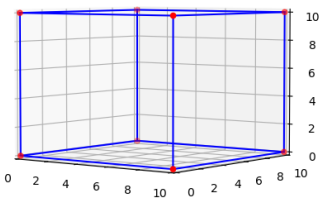
Optamos por omitir os vetores de médias para $k = 36$ devido ao seu comprimento. No entanto, as representações médias das classes para $k = 8$ marcos estão ilustradas na Figura 4, e na Figura 5 apresentamos a representação para $k = 36$ marcos.

As demais formas presentes em ambos os conjuntos para $k = 8$ e $k = 36$ correspondem a variações angulares e de escala referente as figuras acima.

Nesse primeiro cenário, construiremos uma tabela para apresentar os resultados da acurácia dos algoritmos discutidos ao longo desta dissertação, focados na classifica-

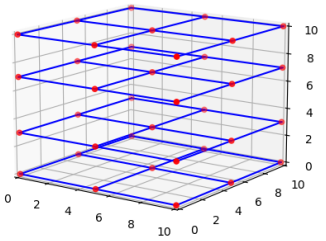


(a) Gráfico Landmarks para Cubo $k = 8$

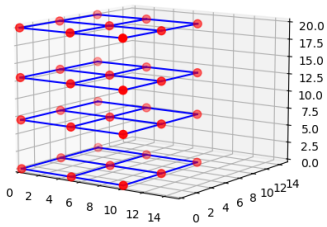


(b) Gráfico Landmarks para Paralelepípedo $k = 8$

Figura 4 – Gráficos dos marcos das formas médias (a) para o cubo e (b) para o paralelepipedo, no caso de $k = 8$.



(a) Gráfico Landmarks para Cubo $k = 36$



(b) Gráfico Landmarks para Paralelepípedo $k = 36$

Figura 5 – Gráficos dos marcos das formas médias (a) para o cubo e (b) para o paralelepipedo, no caso de $k = 36$.

ção binária usando nosso conjunto de formas geométricas simuladas para $k = 8$ e $k = 36$ marcos. Apresentaremos quatro resultados: K-Vizinhos Mais Próximos (KNN) usando distância Procrustes, K-Vizinhos Mais Próximos (KNN) usando distância Riemann, Análise Discriminante Linear (LDA) classificando baseado nas distâncias Riemann e Procrustes, e Análise Discriminante Quadrática (QDA) utilizando o determinante.

Todos os algoritmos exibem uma boa taxa de classificação, com destaque para o K-Vizinhos Mais Próximos (KNN) e a Análise Discriminante Linear (LDA), ambos apresentando 100% de acurácia nos dados simulados utilizando ambas as distâncias. A Análise Discriminante Quadrática (QDA) também apresenta uma boa adaptação, embora inferior ao KNN e ao LDA. A alta acurácia do LDA de Fisher, juntamente

Tabela 3 – Resultados de acurácia dos algoritmos de classificação das simulações

Algoritmo	**Acurácia com** $k = 8$	**Acurácia com** $k = 36$
KNN (Distância Procrustes)	100%	100%
KNN (Distância Riemann)	100%	100%
LDA (Distância Procrustes)	100%	100%
LDA (Distância Riemann)	100%	100%
QDA	80.00%	83.30%

com o kNN, indica que os dados são propícios para métodos que não exigem suposições complexas sobre a distribuição das classes. O sucesso do LDA de Fisher mostra que a separação linear em uma dimensão reduzida é suficiente para classificar os dados com precisão, sugerindo que as classes são bem separadas e linearmente discrimináveis. A menor acurácia do QDA indica que a suposição de covariâncias diferentes para cada classe não é necessária ou não é bem modelada nos dados simulados, o que justifica sua menor performance comparativa. Também podemos observar, como discutido em capítulos anteriores, que a distância Procrustes apresenta resultados semelhantes aos da distância Riemanniana. Isso ocorre porque, para muitos conjuntos de dados práticos com pequena variabilidade, as análises mostram resultados muito semelhantes ao usar diferentes distâncias de Procrustes. Isso reforça a ideia de que, em casos com pouca variabilidade, a escolha da métrica específica pode ter um impacto mínimo nos resultados finais.

4.2 DADOS REAIS

Para os experimentos desse trabalho, utilizaremos o conjunto de dados AFLW2000-3D (ZHU et al., 2015). Este conjunto de dados contém as faces 3D ajustadas das primeiras 2000 amostras do AFLW, especificamente projetado para avaliar métodos de alinhamento de face 3D. Os arquivos no conjunto de dados estão no formato .mat, e cada arquivo inclui uma variável chamada `pt3d.68`, que contém 68 marcos tridimensionais. Esses marcos são cruciais para o alinhamento de face, uma tarefa que ajusta um modelo de face a uma imagem e extrai significados semânticos dos pixels faciais. O alinhamento de face para grandes poses, até 90 graus, apresenta desafios significativos devido à variação na aparência da face e à dificuldade de rotular marcos para vistas de

perfil. O conjunto de dados AFLW2000-3D aborda esses desafios fornecendo um modelo de face 3D denso ajustado às imagens por meio de uma rede neural convolucional (CNN). Essa abordagem, conhecida como Alinhamento de Face 3D Denso (3DDFA), melhora a precisão do alinhamento em várias poses, como demonstrado por melhorias significativas de desempenho no desafiador banco de dados AFLW.

Nós categorizamos as imagens com base na posição da cabeça em três grupos: **Esquerda**, para imagens onde os sujeitos estão voltados para a esquerda; **Direita**, para imagens onde os sujeitos estão voltados para a direita; e **Frente**, para imagens onde os sujeitos estão olhando diretamente para frente. Excluímos imagens com outras posições da cabeça para manter a consistência e focar nessas orientações específicas. Na figura 6 podemos ver um exemplo em duas dimensões das formas que estamos lidando, onde temos sobreposição de marcos do tipo **Forward**, **Left** e **Right**.

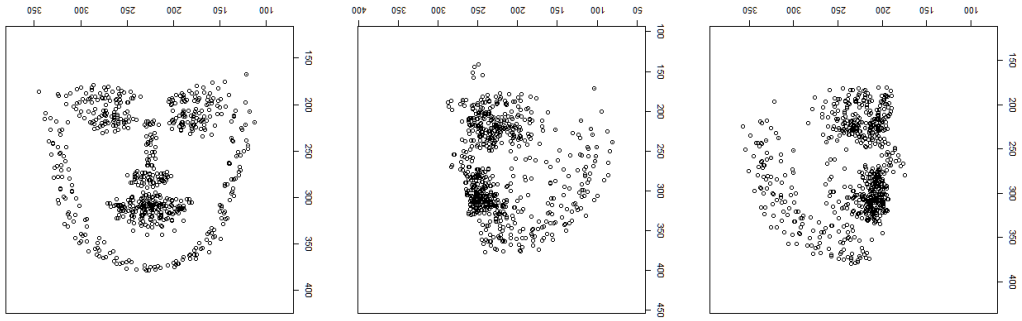


Figura 6 – Exemplo de landmarks faciais

O conjunto será composto por 495 exemplos da classe forward, 325 exemplos da classe left e 327 exemplos da classe right. Na sequência, apresentamos uma visualização em 3D dos marcos faciais para uma orientação frontal 7. Essa visualização tridimensional permite uma compreensão mais clara de como os marcos estão posicionados na face e como eles variam com a mudança de orientação da cabeça.

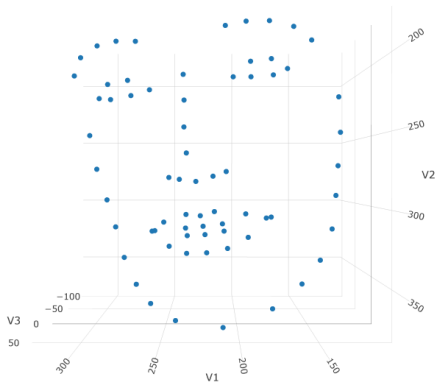


Figura 7 – Exemplo das landmarks faciais para uma label do tipo Forward

Vamos apresentar os resultados de forma mais detalhada. Nos dados simulados, obtivemos ótimas métricas com pouco esforço, não sendo necessário muitos testes para encontrar o melhor valor de k no KNN. No entanto, para os dados reais, foi necessário um estudo mais aprofundado para determinar o número ideal de vizinhos no KNN a fim de obter o melhor resultado, como podemos ver na figura 8.

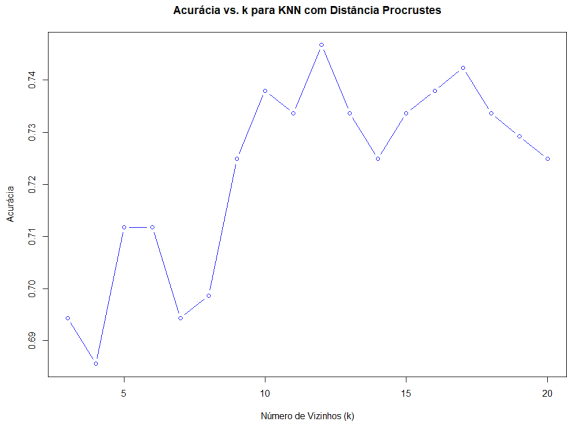


Figura 8 – Gráfico de Acurácia pelo tamanho de Vizinhos Próximos

Podemos observar que o melhor resultado foi obtido com $k = 12$, conforme mostrado na Tabela 4. Nessa tabela, apresentamos os resultados dos modelos previamente discutidos, aplicados aos marcos faciais modelados no espaço de pré-forma.

A acurácia de 75% tanto para o KNN com distância Procrustes quanto para o KNN com distância Riemann sugere que esses métodos são eficazes na classificação dos

Tabela 4 – Resultados de acurácia dos algoritmos de classificação nos marcos faciais

Algoritmo	**Acurácia com Marcos Faciais** $k = 68$
KNN (Distância Procrustes)	75.00%
KNN (Distância Riemann)	75.00%
LDA (Distância Procrustes)	70.00%
LDA (Distância Riemann)	70.00%
QDA	95%

dados reais, mas não são os mais ideais. Já a acurácia de 70% do LDA com ambas as distâncias indica que, embora a separação linear seja adequada, ela não é suficiente para capturar todas as nuances dos dados. Em contraste, o QDA alcançou uma acurácia impressionante de 95%, demonstrando que a consideração de covariâncias diferentes para cada classe é crucial para a modelagem adequada dos dados reais. Esse resultado sugere que os dados reais possuem uma estrutura mais complexa que é melhor capturada pelo QDA, justificando sua performance superior em comparação com os outros métodos.

Para entender melhor como o modelo está classificando cada classe, iremos analisar uma matriz de confusão. A matriz de confusão é uma ferramenta poderosa que nos permite visualizar o desempenho do modelo em termos de classificações corretas e incorretas para cada classe. Nos dados simulados, onde o controle é maior, a acurácia reflete bem o desempenho obtido. No entanto, para os dados reais, especialmente em um cenário multiclasse, é importante verificar se o modelo está enfrentando dificuldades em distinguir entre algumas classes específicas. A matriz de confusão 5 nos ajudará a identificar essas possíveis áreas de confusão, mostrando onde o modelo está acertando e onde está cometendo erros, o que pode fornecer insights valiosos para melhorar a classificação.

Matriz de Confusão do QDA			
Previsão \Real	**left**	**forward**	**right**
left	91	0	2
forward	5	71	1
right	5	0	55

Matriz de Confusão do KNN			
Previsão \Real	**left**	**forward**	**right**
left	44	4	12
forward	11	92	18
right	10	3	35

Matriz de Confusão do LDA de Fisher			
Previsão \Real	**left**	**forward**	**right**
left	45	3	17
forward	5	83	11
right	27	7	31

Tabela 5 – Matrizes de Confusão dos Modelos QDA, KNN e LDA de Fisher

Podemos observar que tanto o LDA de Fisher quanto o KNN tiveram alguma dificuldade em discernir entre as classes left e right, especialmente o LDA. Isso pode ser explicado pelo fato de essas duas classes terem médias próximas, o que faz com que o LDA de Fisher tenha dificuldade em encontrar uma projeção que as separe bem. Por outro lado, o KNN apresenta uma menor confusão, sugerindo que, com mais exemplos de teste, poderíamos obter uma melhor classificação. O QDA também apresenta uma confusão mínima entre as classes left e right, mas, em contrapartida, consegue distingui-las completamente da classe forward.

5 CONCLUSÃO E TRABALHOS FUTUROS

5.1 CONCLUSÃO

O capítul 5 desta dissertação tem como objetivo estabelecer conclusões sobre o trabalho realizado e sugerir recomendações para futuros estudos relacionados ao tema principal.

Nesta dissertação, foram apresentadas modificações em algoritmos de classificação, como Análise Discriminante Quadrática (QDA), K-Vizinhos Mais Próximos (KNN) e Análise Discriminante Linear de Fisher (LDA). Esses algoritmos foram adaptados para trabalhar com dados de pré-formas tridimensionais.

De forma geral, os resultados mostraram que os algoritmos propostos tiveram um bom desempenho ao realizar as classificações usando dados de pré-forma. Em relação aos dados simulados, onde foram gerados dados controlados de formas geométricas tridimensionais contendo duas classes, os algoritmos foram muito eficientes, especialmente o KNN e o LDA, que obtiveram total eficácia nesse cenário. Na avaliação dos resultados com dados reais, os algoritmos LDA e KNN apresentaram uma queda de desempenho, mas ainda mantiveram um bom poder de classificação. O modelo QDA, por sua vez, teve o melhor desempenho nesse cenário. Assim, este trabalho contribui para a literatura teórica dos métodos de classificação para dados de análise estatística de pré-forma, abordando um problema pouco explorado, que é o das formas tridimensionais.

Outro objetivo da dissertação foi abordar a teoria presente para a análise de formas tridimensionais, que difere um pouco da classificação de formas bidimensionais, geralmente o foco da literatura e de artigos publicados. Foram apresentados conceitos ligados à morfometria, formas e marcos anatômicos, matriz de Helmert, centralização, pré-formas, distâncias no espaço de pré-forma e métodos de classificação. Modelos padrão como QDA, LDA de Fisher e KNN foram propostos com adaptações para o uso de pré-formas.

5.2 TRABALHOS FUTUROS

Para trabalhos futuros, é interessante aplicar os mesmos classificadores a outros conjuntos de dados de marcos tridimensionais e comparar mais resultados. Além disso, vale a pena estender os algoritmos para incluir a análise de forma e até mesmo de tamanho e forma. Recomenda-se também introduzir novos classificadores, como Máquinas de Vetores de Suporte (SVM) e Redes Neurais. Por fim, é importante aprofundar a teoria para complementar possíveis lacunas em relação aos dados de formas tridimensionais.

REFERÊNCIAS

AMARAL LUIZ H. DORE, R. P. L. G. J. A.; STOSIC, B. k-means algorithm in statistical shape analysis. *Communications in Statistics - Simulation and Computation*, Taylor & Francis, v. 39, n. 5, p. 1016–1026, 2010. Disponível em: <https://doi.org/10.1080/03610911003765777>.

BISHOP, C. *Pattern Recognition and Machine Learning*. [S.l.]: Springer New York, 2016. (Information Science and Statistics). ISBN 9781493938438.

BOOKSTEIN, F. L. A statistical method for biological shape comparisons. *Journal of theoretical biology*, Elsevier, v. 107, n. 3, p. 475–520, 1984.

CARMO M, P. *Geometria Riemanniana*. [S.l.]: IMPA, Rio de Janeiro, 2008.

CARVALHO, J. B. de; AMARAL, G. J. A. do. Classification methods for planar shapes. *Expert Systems with Applications*, Elsevier, v. 151, p. 113320, 2020.

COOMANS, D.; MASSART, D. Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-nearest neighbour classification by using alternative voting rules. *Analytica Chimica Acta*, v. 136, p. 15–27, 1982. ISSN 0003-2670.

COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21–27, 1967.

DRYDEN, I.; MARDIA, K. *Statistical Shape Analysis: With Applications in R*. [S.l.]: Wiley, 2016. (Wiley Series in Probability and Statistics). ISBN 9780470699621.

HONÓRIO, J. B. d. N.; AMARAL, G. J. A. d. Unsupervised methods for size and shape. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, p. 1–16, 2023.

JOHNSON, R.; WICHERN, D. *Applied Multivariate Statistical Analysis*. [S.l.]: Pearson Prentice Hall, 2007. (Applied Multivariate Statistical Analysis). ISBN 9780131877153.

KENDALL, D.; BARDEN, D.; CARNE, T.; LE, H. *Shape and Shape Theory*. [S.l.]: Wiley, 2009. (Wiley Series in Probability and Statistics). ISBN 9780470317846.

KENDALL, D. G. The diffusion of shape. *Advances in applied probability*, Cambridge University Press, v. 9, n. 3, p. 428–430, 1977.

KENDALL, D. G. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London mathematical society*, Wiley Online Library, v. 16, n. 2, p. 81–121, 1984.

LANCASTER, H. O. The helmert matrices. *The American Mathematical Monthly*, Mathematical Association of America, v. 72, n. 1, p. 4–12, 1965. ISSN 00029890, 19300972. Disponível em: <http://www.jstor.org/stable/2312989>.

MAGNUS, J.; NEUDECKER, H. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. [S.l.]: Wiley, 2019. (Wiley Series in Probability and Statistics). ISBN 9781119541196.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015.

SMALL, C. *The Statistical Theory of Shape*. [S.l.]: Springer New York, 2012. (Springer Series in Statistics). ISBN 9781461240327.

STUELPNAGEL, J. On the parametrization of the three-dimensional rotation group. *SIAM Review*, v. 6, n. 4, p. 422–430, 1964. Disponível em: <https://doi.org/10.1137/1006093>.

VINUÉ, G.; SIMÓ, A.; ALEMANY, S. The k-means algorithm for 3d shapes with an application to apparel design. *Advances in Data Analysis and Classification*, Springer, v. 10, n. 1, p. 103–132, 2016.

ZHU, X.; LEI, Z.; LIU, X.; SHI, H.; LI, S. Z. Face alignment across large poses: A 3d solution. *CoRR*, abs/1511.07212, 2015. Disponível em: <http://arxiv.org/abs/1511.07212>.