

### 3ª Prova - FECD

Renato Assunção - DCC-UFMG

Setembro de 2021

1. **5 PONTOS** Suponha que o vetor aleatório  $\mathbf{X} = (X_1, X_2)$  seja observado em vários indivíduos ou itens que compõem uma amostra. Existem duas classes de itens, classe 0 e classe 1. Em cada classe, a densidade de probabilidade de  $\mathbf{X}$  segue uma Gaussiana bivariada:

$$(\mathbf{X} | \in 0) \sim N_2 \left( \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 \right) = N_2 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \right)$$

and

$$(\mathbf{X} | \in 1) \sim N_2 \left( \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1 \right) = N_2 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

Suponha que as duas classes são igualmente frequentes na população. Isto é, que  $\pi_0 = \mathbb{P}(\mathbf{X} \in 0)$  seja igual a  $\pi_1 = \mathbb{P}(\mathbf{X} \in 1)$ .

Suponha também que os custos  $c_0$  e  $c_1$  de má classificação sejam iguais onde  $c_0$  é o custo de classificar como 0 um item da classe 1 e  $c_1$  é o custo de classificar como 1 um item da classe 0.

Produza um código R ou python que desenhe algumas curvas de nível das densidades de  $\mathbf{X}$  em cada classe e esboce a fronteira de decisão determinada pela regra ótima de Bayes. O resultado deverá ser uma imagem como a da Figura 1.

A seguir, refaça a figura assumindo que a classe 0 aparece 3 vezes mais frequentemente que a classe 1 (isto é, que  $\pi_0 = 3\pi_1$ ). Faça uma terceira figura supondo adicionalmente que os custos também são diferentes, com  $c_0 = 5c_1$ .

2. **5 PONTOS** Considere a seguinte tabela de dados, representando uma amostra composta por quatro pontos amostrais, cada um dos pontos  $\mathbf{X} = (X_1, X_2)$  com dois atributos ou features:

$$\begin{bmatrix} 4 & 1 \\ 2 & 3 \\ 5 & 4 \\ 1 & 0 \end{bmatrix}$$

Queremos representar os dados em apenas uma dimensão usando PCA.

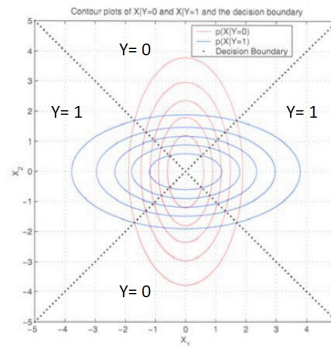


Figura 1: Curvas de nível de  $\mathbf{X}$  em cada uma das classes e fronteira de decisão com custos iguais e probabilidades a priori iguais.

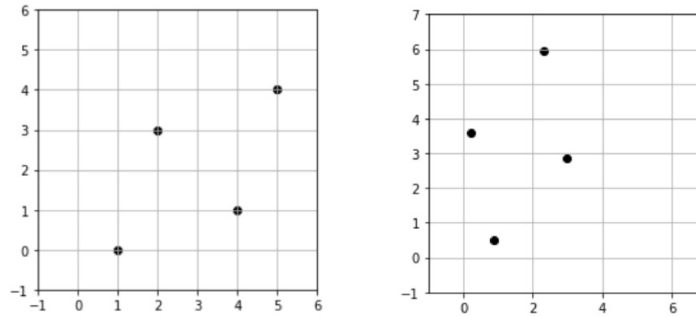


Figura 2: Quatro pontos amostrais.

- Obtenha as direções dos dois componentes principais (ambos com comprimento 1) e indique qual deles é o primeiro.
  - O gráfico da esquerda na Figura 2 mostra os quatro pontos amostrais. Desenhe a direção do componente principal como uma linha e as projeções de todos os quatro pontos de amostra no componente principal. Pode fazer o desenho a mão, fotografar e colar na sua prova como figura. Rotule cada ponto projetado com seu valor de coordenada principal (onde a coordenada principal de origem é zero).
  - O gráfico da direita na Figura 2 mostra os mesmos quatro pontos amostrais após sofrer uma rotação de 30 graus. Supondo que estes novos pontos constituem a amostra, como o PCA da primeira amostra relaciona-se com o PCA da segunda amostra? Justifique sua resposta.
3. **5 PONTOS** Suponha que os pontos da amostra venham de uma distribuição normal gaussiana multivariada  $p$ -dimensional com vetor esperado  $\mu$  e com uma matriz de covariância  $p \times p$  representada por  $\Sigma$ .
- Mostre que, se  $\Sigma$  for uma matriz diagonal, a densidade conjunta será o produto de  $p$  densidades gaussianas univariadas. Pode assumir  $\mu = \mathbf{0}$  para simplificar suas contas.
  - Suponha agora que  $\Sigma$  não é diagonal mas ainda é simétrica e definida positiva. Faça a decomposição espectral de  $\Sigma$  e mostre que podemos escrever a densidade gaussiana multivariada como um produto de densidades gaussianas univariadas, cada uma delas associada com um dos autovetores de  $\Sigma$ .
4. **5 PONTOS** Suponha que você tenha uma distribuição normal multivariada com uma matriz de covariância definida positiva  $\Sigma$ . Considere uma segunda distribuição gaussiana multivariada com o mesmo valor esperado mas cuja matriz de covariância seja  $\cos(\theta) \Sigma$ . Isto é, a matriz anterior multiplicada por um escalar (um número real) positivo  $\cos(\theta) > 0$  onde  $\theta$  é um certo ângulo. Você aprendeu como as curvas de nível da densidade de probabilidade estão associados com os autovetores da matriz de covariância.
- Os autovetores da nova matriz de covariância são rotacionados pelo ângulo  $\theta$ ? Justifique sua resposta.
  - Os autovalores da nova matriz de covariância são alterados? Justifique sua resposta.
  - Como serão alteradas as curvas de nível da segunda distribuição em comparação com aquelas da primeira distribuição?