3^a 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:

$$(\mathbf{X}| \in 1) \sim N_2\left(\boldsymbol{\mu}_0, \sum\nolimits_0\right) = N_2\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 4\end{bmatrix}\right)$$

and

$$(\mathbf{X}| \in 2) \sim N_2\left(\boldsymbol{\mu}_1, \sum\nolimits_1\right) = N_2\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}4 & 0\\0 & 1\end{bmatrix}\right)$$

Suponha que as duas classes são igualmente frequentes na população. Isto é, que $\pi_1 = \mathbb{P}(\mathbf{X} \in 0)$ seja igual a $\pi_2 = \mathbb{P}(\mathbf{X} \in 1)$.

Suponha também que os custos c_1 e c_2 de má classificação sejam iguais onde c_1 é o custo de classificar como 1 um item da classe 2 e c_2 é o custo de classificar como 2 um item da classe 1.

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 decisšao 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 1 aparece 3 vezes mais frequentemente que a classe 2 (isto é, que $\pi_1 = 3\pi_2$). Faça uma terceira figura supondo adicionalmente que os custos também são diferentes, com $c_1 = 5c_2$.

Solução: Seja R_1 a região dos pontos $\mathbf{x} = (x_1, x_2)$ que são alocados à classe 1 (veja que troquei de classes 0 e 1 para classes 1 e 2 neste gabarito para aproveitar minhas fórmulas já digitadas). Pela regra de classificação ótima de Bayes, temos R_1 composto por todos os pontos $\mathbf{x}$ tais que:

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{c(1| \in \pi_2)}{c(2| \in \pi_1)} \; \frac{\mathbb{P}(\pi_2)}{\mathbb{P}(\pi_1)}.$$

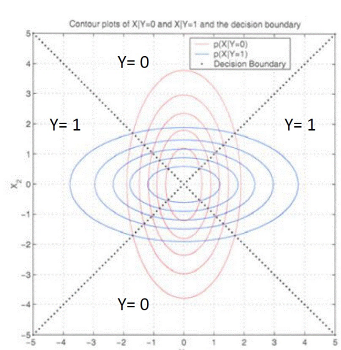

Figura 1: Curvas de nível de $\mathbf{X}$ em cada uma das classes e fronteira de decisão com custos iguais e probabiliades a priori iguais.

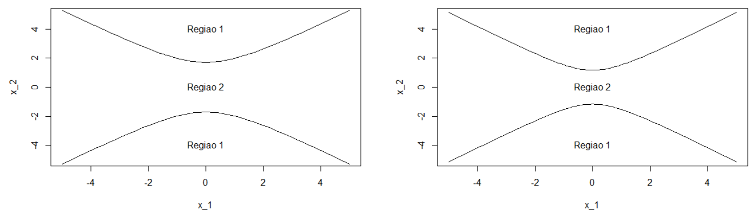

Figura 2: Fronteira de decisão com $\pi_2 = 3\pi$ (esquerda) e com $c_1 = 5c_2$, além de $\pi_2 = 3\pi$ (direita).

Com $c(1| \in \pi_2) = c(2| \in \pi_1)$ e com $\pi_1 = \pi_2$, temos $R_1 = \{\mathbf{x} \text{ tais que } f_1(\mathbf{x}) \geq f_2(\mathbf{x})\}$. Com as densidades gaussianas, temos

$$\begin{aligned}
\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &= \frac{(2\pi)^{-2/2}|\mathbf{\Sigma}_1|^{-1/2}\exp\left(-\frac{1}{2}\mathbf{x}^t\mathbf{\Sigma}_1^{-1}\mathbf{x}\right)}{(2\pi)^{-2/2}|\mathbf{\Sigma}_2|^{-1/2}\exp\left(-\frac{1}{2}\mathbf{x}^t\mathbf{\Sigma}_2^{-1}\mathbf{x}\right)} \qquad \text{pois } |\mathbf{\Sigma}_1| = |\mathbf{\Sigma}_2| \\
&= \exp\left(-\frac{1}{2}\mathbf{x}^t\begin{bmatrix} 1-1/4 & 0 \\ 0 & 1/4-1 \end{bmatrix}^{-1}\mathbf{x}\right) \\
&= \exp\left(-\frac{1}{2}\left(\frac{3}{4}x_1^2 - \frac{3}{4}x_2^2\right)\right)
\end{aligned}$$

Assim, $f_1(\mathbf{x})/f_2(\mathbf{x})$ se, e somente se, $x_1^2 \leq x_2^2$. Isto é, se, e somente se, $|x_1| \leq |x_2|$. Ou ainda, se e somente $-|x_2| \leq x_1 \leq |x_2|$. Esta região é determinada pelas duas retas $x_2 = x_1$ e $x_2 = -x_1$.

Se a classe 1 aparece 3 vezes mais frequentemente que a classe 2, basta modificar os cálculos anteriores. Neste caso, R_1 é o conjunto de pontos $\mathbf{x} = (x_1, x_2)$ tais que $f_1(\mathbf{x})/f_2(\mathbf{x}) \geq 1/3$. Isto é,

$$\exp\left(-\frac{1}{2}\left(\frac{3}{4}x_1^2 + -\frac{3}{4}x_2^2\right)\right) \geq \frac{1}{3}$$

ou seja,

$$x_2^2 \geq x_1^2 + 2.93$$

implicando em $x_2 > \sqrt{x_1^2 + 2.93}$ e $x_2 < -\sqrt{x_1^2 + 2.93}$. O lado esquerdo da Figura 2 mostra a região R_1.

Script R:

```
x = seq(-5, 5, by=0.1)
y1 = sqrt( x^2 - 8*log(1/3)/3)
y2 = -sqrt( x^2 - 8*log(1/3)/3)
plot(x, y1, type="l", ylim=c(-5,5)); lines(x, y2)
text(0, 4, "Regiao 1"); text(0, 0, "Regiao 2"); text(0, -4, "Regiao 1")
```

Adicionando custos diferentes, devemos ter $f_1(\mathbf{x})/f_2(\mathbf{x}) \geq 5/3$ o que implica em $x_2 > \sqrt{x_1^2 + 1.36}$ e $x_2 < -\sqrt{x_1^2 + 1.36}$. O lado direito da Figura 2 mostra a nova região R_1.

Script R:

```
x = seq(-5, 5, by=0.1)
y1 = sqrt( x^2 + 8*log(5/3)/3)
y2 = -sqrt( x^2 + 8*log(5/3)/3)
plot(x, y1, type="l", ylim=c(-5,5),  xlab="x_1", ylab="x_2");
lines(x, y2)
text(0, 4, "Regiao 1"); text(0, 0, "Regiao 2"); text(0, -4, "Regiao 1")
```

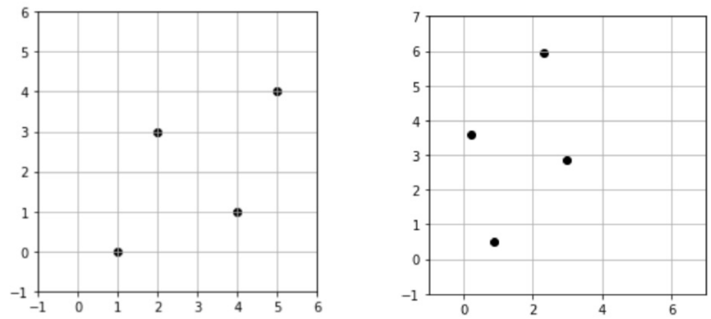

Figura 3: Quatro pontos amostrais.

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.

- Obtenha as direções dos dois componentes principais (ambos com comprimento 1) e indique qual deles é o primeiro.
- O gráfico da esquerda na Figura 3 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 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 3 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.

Solução: Subtraímos a média de cada variável (cada coluna) para obter a matriz centralizada $\tilde{\mathbf{X}}$:

$$\tilde{\mathbf{X}} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 2 & 2 \\ -2 & -2 \end{bmatrix}$$

e calculamos a matriz de covariância

$$\frac{1}{4}\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \begin{bmatrix} 10/4 & 6/4 \\ 6/4 & 10/4 \end{bmatrix}$$

produzindo o primeiro PC $\mathbf{v}_1 = [1/\sqrt{2}, 1/\sqrt{2}]$ com autovalor $\lambda_1 = 16$ e o segundo PC $\mathbf{v}_2 = [1/\sqrt{2}, -1/\sqrt{2}]$ com autovalor $\lambda_2 = 4$.

Os pontos projetados estão no lado esquerdo da Figura 4 e eles possuem coordenadas ao longo do PC dadas por $[1, 5, 5, 9]/\sqrt{2}$. Já o caso rotacionado está no lado direito e possui as mesmas coordenadas $[1, 5, 5, 9]/\sqrt{2}$.

---

3. **5 PONTOS** Suponha que os pontos da amostra venham de uma distribuição normal gaussiana multivariada p-dimensional com vetor esperado $\boldsymbol{\mu}$ e com uma matriz de covariância $p \times p$ representada por Σ.

- Mostre que, se Σ for uma matriz diagonal, a densidade conjunta será o produto de p densidades gaussianas univariadas. Pode assumir $\boldsymbol{\mu} = \mathbf{0}$ para simplificar suas contas.

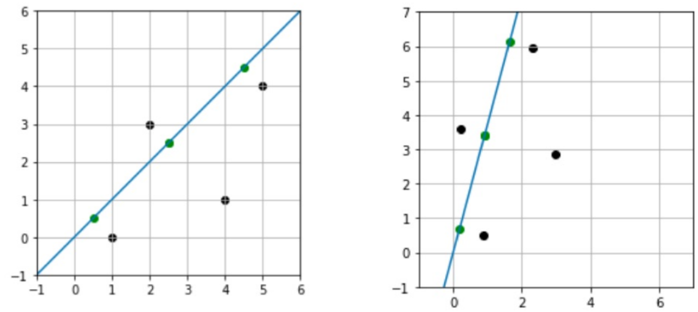

Figura 4: Quatro pontos amostrais projetados no primeiro PC.

- Suponha agora que Σ não é diagonal mas ainda é simétrica e definida positiva. Faaa a decomposição espectral de Σ 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 Σ.

Solução: O caso diagonal é muito simples pois, com $\boldsymbol{\mu} = \mathbf{0}$, a distância estatística é dada por

$$\mathbf{x}^t \Sigma^{-1} \mathbf{x} = \mathbf{x}^t \begin{bmatrix} \sigma_1^2 & & \dots \\ & \ddots & \\ \dots & & \sigma_p^2 \end{bmatrix}^{-1} \mathbf{x} = \sum_{i=1}^{p} \frac{x_i^2}{\sigma_i^2}$$

O determinante é igual a $|\Sigma| = \prod_{i=1}^{p} \sigma_i^2$. Portanto, a densidade $f(\mathbf{x})$ da gaussiana multivariada com $\boldsymbol{\mu} = \mathbf{0}$ e Σ diagonal fica igual a

$$f(\mathbf{x}) = \frac{(2\pi)^{-p/2}}{\sqrt{\prod_{i=1}^{p} \sigma_i^2}} \exp\left(-\frac{1}{2}\mathbf{x}^t \Sigma^{-1} \mathbf{x}\right) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\frac{x_i^2}{\sigma_i^2}\right)$$

A última expressão é o produto de gaussianas univariadas com valor esperado 0 e variância σ_i^2.

Pelo teorema da deomcposição espectral, $\sigma = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$ onde $\mathbf{P} = [\mathbf{v}_1 \vdots \dots \vdots \mathbf{v}_p]$ é uma matriz $p \times p$ cujas colunas são os autovetores de $\boldsymbol{\Sigma}$ e $\boldsymbol{\Lambda}$ é uma diagonal com os correspondentes autovalores. Além disso, temos $\mathbf{P}\mathbf{P}^{prime} = \mathbf{P}^{prime}\mathbf{P} = \mathbf{I}$ (isto é, a inversa de $\mathbf{P}$ é $\mathbf{P}'$. Assim,

$$\begin{aligned} \mathbf{x}^t \Sigma^{-1} \mathbf{x} &= \mathbf{x}^t (\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}')^{-1})^{-1} \mathbf{x} \\ &= \mathbf{x}^t (\mathbf{P}')^{-1} \boldsymbol{\Lambda}^{-1} \mathbf{P}^{-1} \mathbf{x} \\ &= \mathbf{x}^t \mathbf{P} \boldsymbol{\Lambda}^{-1} \mathbf{P}' \mathbf{x} \\ &= (\mathbf{P}'\mathbf{x})' \boldsymbol{\Lambda}^{-1} \mathbf{P}' \mathbf{x} \\ &= \mathbf{y}' \boldsymbol{\Lambda}^{-1} \mathbf{y} \end{aligned}$$

onde $\mathbf{y} = \mathbf{P}'\mathbf{x}$. Como $\boldsymbol{\Lambda}^{-1}$ é uma matriz diagonal, usando o resultado do primeiro item teremos um produto de gaussianas univariadas com

De fato, como $|\boldsymbol{\Sigma}| = |\boldsymbol{\Lambda}| = \prod_i \lambda_i$, teremos

$$\begin{aligned} f(\mathbf{x}) &= (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}^t \boldsymbol{\Sigma}^{-1} \mathbf{x}\right) \\ &= (2\pi)^{-p/2} \prod_i |\lambda_i|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{y}' \boldsymbol{\Lambda}^{-1} \mathbf{y}\right) \\ &= \prod_i \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left(-\frac{1}{2}\frac{y_i^2}{\lambda_i}\right) \end{aligned}$$

Assim, a densidade conjunta gaussiana multivariada de p variáveis pode ser vista como a densidade de p gaussianas univariadas $y_i = \mathbf{v}_i'\mathbf{x}$ com variância λ_i.

---

4. **5 PONTOS** Suponha que você tenha uma distribuição normal multivariada com uma matriz de covariância definida positiva Σ. 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 θ é um certo ângulo. Você aprendeu como as curvas de nível da desnidade de probabilidade estão associados com os autovetores da matriz de covariância.

   - Os autovetores da nova matriz de covariância são rotacionados pelo ângulo θ? 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?

   Solução: Seja $\mathbf{v}$ um autovetor de $\mathbf{\Sigma}$. Isto é, $\mathbf{\Sigma v} = \lambda \mathbf{v}$. Então, $\cos(\theta)\mathbf{\Sigma v} = \cos(\theta)\lambda\mathbf{v}$. Assim, $\mathbf{v}$ também é um autovetor de $\cos(\theta)\mathbf{\Sigma}$ e os autovetores são os mesmos. Os autovalores passam a ser $\cos(\theta)\lambda$.

   Como $0 < \cos(\theta) \leq 1$ e os autovetores são os mesmos, as curvas de nível são as mesmas.