Homework 2 - Aprendizagem

Francisco Silva 110409
Joana Cardoso 109864

1. Pen and Paper

1)

a.

Training observation: $x_1 \rightarrow x_6$

Priors:

$$P(Y_{out}=N) = \frac{3}{6} = \frac{1}{2} \qquad P(Y_{out}=P) = \frac{3}{6} = \frac{1}{2}$$

Since the variables are split into independent sets, then $P(X|Y_{out})$ factors into 3 parts

$$P(X|Y_{out}) = P(Y_1,Y_2|Y_{out})\,P(Y_3,Y_4|Y_{out})\,P(Y_5|Y_{out})$$

i) $P(Y_1,Y_2|Y_{out})$: normally distributed so calculate mean vector ($\vec{\mu}$) and covariance matrix (Σ) for both N and P } continuous/numeric variable (Y_1,Y_2)

ii) $P(Y_3,Y_4|Y_{out})$: Calculate joint probability for all combinations of $(Y_3,Y_4) \in \{(0,0);(0,1);(1,0);(1,1)\}$ for both N and P

iii) $P(Y_5|Y_{out})$: Calculate marginal probability for $Y_5 \in \{0,1\}$ for both N and P

} discrete variables (Y_3,Y_4,Y_5)

i) Class N

$$\vec{\mu}_N = \begin{pmatrix} \frac{0,52+0,53+0,42}{3} \\ \frac{0.8+0.92+0.48}{3} \end{pmatrix} \simeq \begin{pmatrix} 0,49 \\ 0,73 \end{pmatrix}$$

$$Var(Y_1) = \frac{(0,52-0,49)^2+(0,53-0,49)^2+(0,42-0,49)^2}{3-1} \simeq 0,0037$$

$$Var(Y_2) = \frac{(0.8-0.73)^2+(0.92-0.73)^2+(0.48-0.73)^2}{3-1} \simeq 0,0518$$

$$cov(Y_1,Y_2) = \frac{(0,52-0,49)(0.8-0.73)+(0.53-0.49)(0.92-0.73)+(0.42-0.49)(0.48-0.73)}{3-1} = 0,0136$$

$$\Sigma_N = \begin{bmatrix} 0,0037 & 0,0136 \\ 0,0136 & 0,0518 \end{bmatrix}$$

Class P

$$\vec{\mu}_P = \begin{bmatrix} \frac{0,49+0,62+0,44}{3} \\ \frac{0.58+0.31+0.38}{3} \end{bmatrix} \simeq \begin{pmatrix} 0,51(6) \\ 0,42(3) \end{pmatrix}$$

$$Var(Y_1) = \frac{(0,49-0,517)^2+(0,62-0,517)^2+(0,44-0,517)^2}{3-1} = 0,0086$$

$$Var(Y_2) = \frac{(0.58-0.423)^2+(0.31-0.423)^2+(0.38-0.423)^2}{3-1} = 0,0196$$

$$cov(Y_1,Y_2) = \frac{(0.49-0.517)(0.58-0.423)+(0.62-0.517)(0.31-0.423)+(0.44-0.517)(0.38-0.423)}{3-1} = -0,0063$$

$$\Sigma_P = \begin{bmatrix} 0,0086 & -0,0063 \\ -0,0063 & 0,0196 \end{bmatrix}$$

ii)

(Y_3,Y_4)	N (3 casos)	P (3 casos)
(0,0)	x_2 1/3	x_5,x_6 2/3
(0,1)	x_1,x_3 2/3	0
(1,0)	0	x_4 1/3
(1,1)	0	0

iii)

Y_5	N (3 casos)	P (3 casos)
0	x_2 1/3	0
1	x_1,x_3 2/3	x_4,x_5,x_6 3/3

b. MAP : $V_{yout} = \{N, P\}$ $X = \{x_7, x_8\}$

$$\underset{\mu \in V_{yout}}{\text{argmax}} \; P(Y_{out} = \mu \mid X) = \text{argmax}(P(X \mid Y_{out}) . P(Y_{out} = \mu))$$

- $P(Y_{out} = \mu \mid x_7) = \text{argmax}(P(x_7 \mid N) . P(Y_{out} = N); P(x_7 \mid P) . P(Y_{out} = P))$
- $P(Y_{out} = \mu \mid x_8) = \text{argmax}(P(x_8 \mid N) . P(Y_{out} = N); P(x_8 \mid P) . P(Y_{out} = P))$

Para $Y_{out} = N$

x_7:

$$\rightarrow P(x_7 \mid N) = P(Y_1 = 0.45; Y_2 = 0.8 \mid \vec{\mu}_N \, \Sigma_N) . \overbrace{P(Y_3 = 0; Y_4 = 0 \mid N) . P(Y_5 = 1 \mid N)}^{\text{apuramos facilmente pelas tabelas de 1a.}} =$$

$$= \frac{1}{2\pi\sqrt{\det(\Sigma_N)}} \cdot \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu}_N)^T \Sigma_N^{-1} (\vec{x} - \vec{\mu}_N)\right) \cdot \frac{1}{3} \cdot \frac{2}{3}$$

Precisamos:

- $\det(\Sigma_N) = 0.0037 \times 0.0518 - (0.0136)^2 = 6.7 \times 10^{-6}$
- $\Sigma_N^{-1} = \frac{10^6}{6,7} \begin{bmatrix} 0,0518 & -0,0136 \\ -0,0136 & 0,0037 \end{bmatrix} = \begin{bmatrix} 7731,343 & -2029,850 \\ -2029,850 & 552,239 \end{bmatrix}$
- $(\vec{x} - \vec{\mu}_N) = \begin{bmatrix} 0,45 \\ 0.8 \end{bmatrix} - \begin{bmatrix} 0.49 \\ 0.73 \end{bmatrix} = \begin{bmatrix} -0,04 \\ 0,0667 \end{bmatrix}$

Assim:

$$P(x_7 \mid N) = \frac{1}{2\pi\sqrt{6.7 \times 10^{-6}}} \cdot \exp\left(-\frac{1}{2}[-0,04 \;\; 0,0667] \begin{bmatrix} 7731,343 & -2029,85 \\ -2029,850 & 552,239 \end{bmatrix} \begin{bmatrix} -0,04 \\ 0,0667 \end{bmatrix}\right) \cdot \frac{1}{3} \cdot \frac{2}{3} =$$

$$= 61.487 . \exp\left(-\frac{1}{2}[-0,04 \;\; 0,0667] \begin{bmatrix} -444,588 \\ 118,028 \end{bmatrix}\right) \cdot \frac{2}{9} = 61,487 \times \exp\left(-\frac{25,656}{2}\right) \cdot \frac{2}{9} =$$

$$= 1,65 \times 10^{-4} \times \frac{2}{9}$$

$$\rightarrow P(x_7 \mid N) . P(Y_{out} = N) = 1,65 \times 10^{-4} \times \frac{1}{9} = 1,834 \times 10^{-5}$$

x_8:

$$\rightarrow P(x_8 \mid N) = P(Y_1 = 0.5; Y_2 = 0.7 \mid \vec{\mu}_N \, \Sigma_N) . \underbrace{P(Y_3 = 0, Y_4 = 1 \mid N) . P(Y_5 = 1 \mid N)}_{\frac{2}{3} \cdot \frac{2}{3}} =$$

Já temos: $\det(\Sigma_N)$; Σ_N^{-1}

Precisamos:

- $(x - \vec{\mu}_N) = \begin{bmatrix} 0.5 \\ 0.7 \end{bmatrix} - \begin{bmatrix} 0.49 \\ 0,733 \end{bmatrix} = \begin{bmatrix} 0,01 \\ -0,03 \end{bmatrix}$

Assim:

$$P(x_8 \mid N) = 61,487 \cdot \exp\left(-\frac{1}{2}[0,01 \;\; -0,03] \begin{bmatrix} 7731,343 & -2029,85 \\ -2029,85 & 552,239 \end{bmatrix} \begin{bmatrix} 0,01 \\ -0,03 \end{bmatrix}\right) \cdot \frac{4}{9} =$$

$$= 61.487 \times \exp\left(-\frac{1}{2}[0,01 \;\; -0,03] \begin{bmatrix} 138,209 \\ -36,866 \end{bmatrix}\right) \times \frac{4}{9} = 61,487 \times \exp\left(-\frac{2.488}{2}\right) \cdot \frac{4}{9} =$$

$$= 17.722 \times \frac{4}{9}$$

$$\rightarrow P(X_8 \mid N) . P(Y_{out} = N) = 17.722 \times \frac{2}{9} = 3,938$$

$\boxed{\text{Para } Y_{out}=P}$

- Para x_7

$\rightarrow P(x_7|P) = P(\vec{x}_{Y_1,Y_2} | \vec{\mu}_p \Sigma_p) \cdot P(\vec{x}_{Y_3,Y_4}|P) \cdot P(\vec{x}_{Y_5}|P) =$

$= P(\vec{x}_{Y_1,Y_2} | \vec{\mu}_p \Sigma_p) \cdot P((0,0)|P) \cdot P(1|P) = P(Y_1=0.45; Y_2=0.8 | \vec{\mu}_p \Sigma_p) \cdot \frac{2}{3} \cdot 1$

- $P(Y_1=0.45; Y_2=0.8 | \vec{\mu}_p \Sigma_p)$ precisamos:

- $\det(\Sigma) = 0{,}0086 \cdot 0{,}0196 - (-0{,}0063)^2 = 1{,}3 \times 10^{-4}$ $\quad \bullet\ \Sigma^{-1} = \frac{10^4}{1{,}3}\begin{bmatrix} 0{,}0196 & 0{,}0063 \\ 0{,}0063 & 0{,}0086 \end{bmatrix}$

- $(\vec{x} - \vec{\mu}_p) = \begin{bmatrix} 0{,}45 \\ 0.80 \end{bmatrix} - \begin{bmatrix} 0.517 \\ 0{,}423 \end{bmatrix} = \begin{bmatrix} -0{,}067 \\ 0{,}377 \end{bmatrix}$ $\quad \Sigma^{-1}(\vec{x} - \vec{\mu}) = \begin{bmatrix} 150{,}769 & 48{,}46 \\ 48{,}46 & 66{,}41 \end{bmatrix} \begin{bmatrix} -0{,}067 \\ 0{,}377 \end{bmatrix} = \begin{bmatrix} 8{,}17 \\ 21{,}79 \end{bmatrix}$

- $(\vec{x} - \vec{\mu}_p)^T (\Sigma^{-1}(\vec{x} - \vec{\mu}_p)) = [-0{,}067 \ \ 0{,}377] \begin{bmatrix} 8{,}17 \\ 21{,}79 \end{bmatrix} \approx 7{,}6674$

substituindo:

$P(Y_1=0.45, Y_2=0.8 | \vec{\mu}_p \Sigma_p) = \frac{1}{\sqrt{2\pi}\sqrt{1{,}3 \times 10^{-4}}} \exp\left(-\frac{7{,}6674}{2}\right) = 0.302$

Assim temos que $P(x_7|P) = \underset{0.302}{\cancel{\ }} \times \frac{2}{3} = 0{,}201$

$\rightarrow P(Y_{out}=P) = \frac{1}{2}$

$\Rightarrow P(x_7|P) P(Y_{out}=P) = \underset{0{,}201}{\cancel{\ }} \times \frac{1}{2} \simeq 0.1$

- Para x_8

$\rightarrow P(x_8|P) = P(Y_1=0.5, Y_2=0.7 | \vec{\mu}_p \Sigma_p) \cdot P((0,1)|P)\, P(Y_5=1|P) = P(Y_1=0.5, Y_2=0.7|\vec{\mu}_p \Sigma_p) \cdot 0 \cdot 1 = 0$

$\rightarrow P(Y_{out}=P) = \frac{1}{2}$

$\Rightarrow P(x_8|P) \cdot P(Y_{out}=P) = 0$

Pelo MAP (posteriums maximum):

Para: $P(Y_{out}=\mu | x_7) = \text{argmax}\left(P(x_7|N)P(Y_{out}=N), P(x_7|P)P(Y_{out}=P)\right) = \text{argmax}(1{,}834 \times 10^{-3}; 0{,}1)$

$= 0{,}1 \Rightarrow$ classe de Yout de x_7: Positive (P)

Para: $P(Y_{out}=\mu | x_8) = \text{argmax}\left(P(x_8|N)P(Y_{out}=N), P(x_8|P)P(Y_{out}=P)\right) = \text{argmax}(3{,}938; 0)$

$= 3{,}938 \Rightarrow$ classe de Yout de x_8 : Negative (N)

R:. Under a MAP assumption, we classified x_7 as P and x_8 as N

2. Hamming distance : # different variables

neighbours \ left out	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	-	2	0	2	1	1	1	0
x_2	2	-	2	2	1	1	1	2
x_3	0	2	-	2	1	1	1	0
x_4	2	2	2	-	1	1	1	2
x_5	1	1	1	1	-	0	0	1
x_6	1	1	1	1	0	-	0	1
x_7	1	1	1	1	0	0	-	1
x_8	0	2	0	2	1	1	1	-
K	2	3	2	3	2	2	2	2
$\hat{z}$	N	$\boxed{P}$	N	P	P	P	P	N
Z	N	$\boxed{N}$	N	P	P	P	P	N

K: K used (3 or 2)

$\hat{z}$: output predicted by knn algorithm : mode of the K variable's output

z: True output

$$\text{Accuracy} = \frac{N^\circ \text{ of correct previsions}}{N^\circ \text{ of observations}} = \frac{7}{8} = 0.875 \Rightarrow 87{,}5\%$$

3.a. MAP : escolhe a class c/ maior probabilidade à posteriori

$$\theta_{Bayes}(x) = \arg\max P(\theta = \mu \mid X = x) = \arg\max(P(\theta=0 \mid X=x); P(\theta=1 \mid X=x)) =$$

$$= \arg\max(P(X=x \mid \theta=0) \cdot P(\theta=0); P(X=x \mid \theta=1) P(\theta=1)) \underset{P(X=x\mid\theta=0)=P(X=x\mid\theta=1)}{=}$$

$$= \arg\max(P(\theta=0); P(\theta=1)) \underset{\substack{P(\theta=1)=1-P(\theta=0)=1-p \\ P(\theta=0)=p}}{=} \arg\max(p, 1-p)$$

R: como $p \in]1/2, 1] \Rightarrow p > 1-p$, logo o classificador Bayesiano vai sempre escolher a classe 0. Assim, podemos concluir que o erro ocorre quando a classe verdadeira é 1, o que acontece com a probabilidade $1-p$

$$E_{Bayes} = P(\theta=1) = 1-p$$

b. Em 1nn o erro ocorre (E_{1NN}) quando o vizinho mais próximo tem uma classe diferente da verdadeira de x.

$$E_{1NN} = 2 \times P(\theta=0 \mid X=x) \times P(\theta=1 \mid X=x) \underset{\text{Regra de Bayes}}{=}$$

$$= 2 \times \frac{P(X=x \mid \theta=0) \cdot P(\theta=0)}{P(X=x)} \times \frac{P(X=x \mid \theta=1) \cdot P(\theta=1)}{P(X=x)} \underset{\substack{P(X=x\mid\theta=0)=P(X=x\mid\theta=1)=f(x) \\ P(\theta=0)=p;\ P(\theta=1)=1-p}}{=}$$

$$= 2 \times \frac{f^2(x)\, p(1-p)}{P^2(X=x)} =$$

Lei da probabilidade Total

$$P(X=x) = P(X=x \mid \theta=0) P(\theta=0) + P(X=x \mid \theta=1) P(\theta=1) =$$
$$= f(x)p + f(x)(1-p) =$$
$$= f(x)p + f(x) - f(x)p = f(x)$$

$$= 2 \times \frac{f^2(x)\, p(1-p)}{f^2(x)}$$

$$= 2p(1-p) \quad \text{c.q.d}$$

c. na alínea a) provamos que $E_{Bayes} = 1-p$

na alínea b) provamos que $E_{1NN} = 2p(1-p)$

$E_{Bayes} = 1-p \Leftrightarrow p = 1 - E_{Bayes} \quad \wedge \quad 1-p = E_{Bayes}$

Substituindo

$$E_{1NN} = 2p(1-p) = 2(1-E_{Bayes})E_{Bayes} = 2E_{Bayes}(1-E_{Bayes})$$

c.q.d