

**I. Pen-and-paper**

$$\begin{aligned}
 1) (a) \text{ Posterior: } & P(Y_6 = \text{class} | Y_1 = v_1, Y_2 = v_2, Y_3 = v_3, Y_4 = v_4, Y_5 = v_5) \stackrel{\text{T. Bayes}}{=} \\
 & = \frac{P(Y_6 = \text{class}) P(Y_1 = v_1, Y_2 = v_2, Y_3 = v_3, Y_4 = v_4, Y_5 = v_5 | Y_6 = \text{class})}{P(Y_1 = v_1, Y_2 = v_2, Y_3 = v_3, Y_4 = v_4, Y_5 = v_5)} = \\
 & = \frac{P(Y_6 = \text{class})}{\underbrace{P(Y_1 = v_1, Y_2 = v_2, Y_3 = v_3, Y_4 = v_4, Y_5 = v_5)}_{\substack{(1) \\ (2) \\ (3) \\ (4)}}} \cdot \underbrace{P(Y_1 = v_1, Y_2 = v_2 | Y_6 = \text{class})}_{(2')} \cdot \underbrace{P(Y_3 = v_3, Y_4 = v_4 | Y_6 = \text{class})}_{(3')} \cdot \underbrace{P(Y_5 = v_5 | Y_6 = \text{class})}_{(4')} \\
 & \quad \cdot \underbrace{P(Y_1 = v_1, Y_2 = v_2)}_{(2)} \cdot \underbrace{P(Y_3 = v_3, Y_4 = v_4)}_{(3)} \cdot \underbrace{P(Y_5 = v_5)}_{(4)}
 \end{aligned}$$

Dado uma nova observação  $D$ , nós temos classes de classificação calculando  $P(\text{class}|D)$  para todos os classes e escolhermos a classe com maior probabilidade.

$$\hat{z} = \underset{C}{\operatorname{argmax}} P(C|D) = \underset{C}{\operatorname{argmax}} \frac{P(Y_1, Y_2|C) P(Y_3, Y_4|C) P(Y_5|C)}{P(Y_1, Y_2) P(Y_3, Y_4) P(Y_5)} = \underset{C}{\operatorname{argmax}} P(Y_1, Y_2|C) P(Y_3, Y_4|C) \\
 \times P(Y_5|C) \quad \text{igual para todas as } C$$

$$\textcircled{1} \text{ Brusos: } \bullet P(Y_6 = A) = \frac{3}{7} \quad \bullet P(Y_6 = B) = \frac{4}{7}$$

$$\textcircled{2} \bullet \hat{\mu}_{Y_6=A} = \begin{bmatrix} \frac{\sum_{i=1}^3 Y_{1,i}}{3} \\ \frac{\sum_{i=1}^3 Y_{2,i}}{3} \end{bmatrix} = \begin{bmatrix} \frac{0.24 + 0.16 + 0.32}{3} \\ \frac{0.36 + 0.48 + 0.72}{3} \end{bmatrix} = \begin{bmatrix} 0.24 \\ 0.52 \end{bmatrix}$$

Como  $Y_1 \times Y_2 \in \mathbb{R}^2$  tem distribuição normal, então:

$$P(Y_1, Y_2) \sim N(Y_1, Y_2 | \mu = ?, \Sigma = ?)$$

$$\bullet \text{Var}(Y_1 | Y_6 = A) = \frac{\sum (y_{1,i} - \bar{y})^2}{m-1} = \frac{\sum (y_{1,i})^2 - m\bar{y}^2}{m-1} = \frac{0.24^2 + 0.16^2 + 0.32^2 - 3 \times 0.24^2}{2} \approx 0.0064$$

$$\forall C, P(Y_1, Y_2 | C) \sim N(Y_1, Y_2 | \mu_C = ?, \Sigma_C = ?)$$

$$\bullet \text{Var}(Y_2 | Y_6 = A) = \frac{\sum (y_{2,i} - \bar{y})^2}{m-1} = \frac{\sum (y_{2,i})^2 - m\bar{y}^2}{m-1} = \frac{0.36^2 + 0.48^2 + 0.72^2 - 3 \times 0.52^2}{2} \approx 0.0336$$

$$\bullet \text{Cov}(Y_1, Y_2 | Y_6 = A) = \frac{\sum (y_{1,i} - \bar{y}_1)(y_{2,i} - \bar{y}_2)}{m-1} = \frac{1}{2} ((0.24 - 0.24)(0.36 - 0.52) + (0.16 - 0.24) \times \\
 \times (0.48 - 0.52) + (0.32 - 0.24)(0.72 - 0.52)) = 0.0096$$

$$\bullet \sum_{Y_6=A} = \begin{bmatrix} \text{Var}(Y_1 | Y_6 = A) & \text{Cov}(Y_1, Y_2 | Y_6 = A) \\ \text{Cov}(Y_1, Y_2 | Y_6 = A) & \text{Var}(Y_2 | Y_6 = A) \end{bmatrix} = \begin{bmatrix} 0.0064 & 0.0096 \\ 0.0096 & 0.0336 \end{bmatrix}$$

$$\bullet \hat{\mu}_{Y_6=B} = \begin{bmatrix} \frac{\sum_{i=1}^4 Y_{1,i}}{4} \\ \frac{\sum_{i=1}^4 Y_{2,i}}{4} \end{bmatrix} = \begin{bmatrix} \frac{0.54 + 0.66 + 0.76 + 0.41}{4} \\ \frac{0.11 + 0.39 + 0.28 + 0.53}{4} \end{bmatrix} = \begin{bmatrix} 0.5925 \\ 0.3275 \end{bmatrix}$$

$$\begin{aligned} \bullet \text{Var}(Y_1 | Y_6 = B) &= \frac{\sum (Y_i - \bar{Y})^2}{m-1} = \frac{\sum (Y_i)^2 - m\bar{Y}^2}{m-1} = \frac{0.54^2 + 0.66^2 + 0.76^2 + 0.41^2 - 4 \times 0.5925^2}{3} \approx 0.02289 \\ \bullet \text{Var}(Y_2 | Y_6 = B) &= \frac{\sum (Y_i - \bar{Y})^2}{m-1} = \frac{\sum (Y_i)^2 - m\bar{Y}^2}{m-1} = \frac{0.11^2 + 0.39^2 + 0.28^2 + 0.53^2 - 4 \times 0.3275^2}{3} \approx 0.03149 \\ \bullet \text{Cov}(Y_1, Y_2 | Y_6 = B) &= \frac{\sum (Y_{1,i} - \bar{Y}_1)(Y_{2,i} - \bar{Y}_2)}{m-1} = \frac{1}{3} ((0.54 - 0.5925)(0.11 - 0.3275) + (0.66 - 0.5925)(0.39 - 0.3275) + (0.76 - 0.5925)(0.28 - 0.3275) + (0.41 - 0.5925)(0.53 - 0.3275)) \approx -0.0098 \end{aligned}$$

$$\bullet \sum_{Y_6=B} = \begin{bmatrix} \text{Var}(Y_1 | Y_6 = B) & \text{Cov}(Y_1, Y_2 | Y_6 = B) \\ \text{Cov}(Y_1, Y_2 | Y_6 = B) & \text{Var}(Y_2 | Y_6 = B) \end{bmatrix} = \begin{bmatrix} 0.02289 & -0.0098 \\ -0.0098 & 0.03149 \end{bmatrix}$$

$$\textcircled{2) } \quad \boxed{P(Y_1 = 0, Y_2 = 0)} \quad \bullet \hat{\mu} = \begin{bmatrix} \frac{\sum Y_{1,i}}{7} \\ \frac{\sum Y_{2,i}}{7} \end{bmatrix} = \begin{bmatrix} \frac{0.24 + 0.16 + 0.32 + 0.54 + 0.66 + 0.76 + 0.41}{7} \\ \frac{0.36 + 0.48 + 0.72 + 0.11 + 0.39 + 0.28 + 0.53}{7} \end{bmatrix} = \begin{bmatrix} 0.4414 \\ 0.4100 \end{bmatrix}$$

$$\begin{aligned} \bullet \text{Var}(Y_1) &= \frac{\sum (Y_i - \bar{Y})^2}{m-1} = \frac{\sum (Y_i)^2 - m\bar{Y}^2}{m-1} = \frac{0.24^2 + 0.16^2 + 0.32^2 + 0.54^2 + 0.66^2 + 0.76^2 + 0.41^2 - 7 \times 0.4414^2}{6} \approx 0.0491 \\ \bullet \text{Var}(Y_2) &= \frac{\sum (Y_i - \bar{Y})^2}{m-1} = \frac{\sum (Y_i)^2 - m\bar{Y}^2}{m-1} = \frac{0.36^2 + 0.48^2 + 0.72^2 + 0.11^2 + 0.39^2 + 0.28^2 + 0.53^2 - 7 \times 0.41^2}{6} \approx 0.0375 \end{aligned}$$

$$\begin{aligned} \bullet \text{Cov}(Y_1, Y_2) &= \frac{\sum (Y_{1,i} - \bar{Y}_1)(Y_{2,i} - \bar{Y}_2)}{m-1} = \frac{1}{6} ((0.24 - 0.4414)(0.36 - 0.41) + (0.16 - 0.4414)(0.48 - 0.41) + (0.32 - 0.4414)(0.72 - 0.41) + (0.54 - 0.4414)(0.11 - 0.41) + (0.66 - 0.4414)(0.39 - 0.41) + (0.76 - 0.4414)(0.28 - 0.41) + (0.41 - 0.4414)(0.53 - 0.41)) \approx -0.0211 \end{aligned}$$

$$\bullet \sum = \begin{bmatrix} \text{Var}(Y_1) & \text{Cov}(Y_1, Y_2) \\ \text{Cov}(Y_1, Y_2) & \text{Var}(Y_2) \end{bmatrix} = \begin{bmatrix} 0.0491 & -0.0211 \\ -0.0211 & 0.0375 \end{bmatrix}$$

$$\textcircled{3) } \quad \boxed{P(Y_3 = 0, Y_4 = 1)} \quad \bullet P(Y_3 = 1, Y_4 = 1) = \frac{1}{7} \quad \bullet P(Y_3 = 1, Y_4 = 0) = \frac{2}{7} \\ \bullet P(Y_3 = 0, Y_4 = 0) = \frac{2}{7} \quad \bullet P(Y_3 = 0, Y_4 = 1) = \frac{2}{7}$$

$$\begin{aligned} \textcircled{3) } \quad \bullet P(Y_3 = 1, Y_4 = 1 | Y_6 = A) &= \frac{1}{3} & \bullet P(Y_3 = 1, Y_4 = 1 | Y_6 = B) &= \frac{0}{4} = 0 \\ \bullet P(Y_3 = 1, Y_4 = 0 | Y_6 = A) &= \frac{1}{3} & \bullet P(Y_3 = 1, Y_4 = 0 | Y_6 = B) &= \frac{1}{4} \\ \bullet P(Y_3 = 0, Y_4 = 0 | Y_6 = A) &= \frac{0}{3} = 0 & \bullet P(Y_3 = 0, Y_4 = 0 | Y_6 = B) &= \frac{2}{4} = \frac{1}{2} \\ \bullet P(Y_3 = 0, Y_4 = 1 | Y_6 = A) &= \frac{1}{3} & \bullet P(Y_3 = 0, Y_4 = 1 | Y_6 = B) &= \frac{1}{4} \end{aligned}$$

$$\textcircled{4) } \quad \boxed{P(Y_5 = 0)} \quad \bullet P(Y_5 = 0) = \frac{2}{7} \quad \bullet P(Y_5 = 1) = \frac{3}{7} \quad \bullet P(Y_5 = 2) = \frac{2}{7}$$

$$\textcircled{4) } \quad \bullet P(Y_5 = 0 | Y_6 = A) = \frac{1}{3} \quad \bullet P(Y_5 = 0 | Y_6 = B) = \frac{1}{4} \quad \text{Agora, já tens todos os} \\ \bullet P(Y_5 = 1 | Y_6 = A) = \frac{1}{3} \quad \bullet P(Y_5 = 1 | Y_6 = B) = \frac{2}{4} = \frac{1}{2} \quad \text{parâmetros necessários para aplicar} \\ \bullet P(Y_5 = 2 | Y_6 = A) = \frac{1}{3} \quad \bullet P(Y_5 = 2 | Y_6 = B) = \frac{1}{4} \quad \text{o classificador Bayesiano às novas} \\ \text{observações.}$$

**Homework 2 – Group 075**

(ist1103860, ist1103883)

(b) Sabendo que  $P(Y_6 = \text{class} | Y_1 = v_1, Y_2 = v_2, Y_3 = v_3, Y_4 = v_4, Y_5 = v_5) = \frac{P(Y_6 = \text{class}) P(Y_1 = v_1, Y_2 = v_2 | Y_6 = \text{class})}{P(Y_1 = v_1, Y_2 = v_2) \times \dots}$

$$\text{e } P(Y_3 = v_3, Y_4 = v_4 | Y_6 = \text{class}) P(Y_5 = v_5 | Y_6 = \text{class}) \cdot P(Y_3 = v_3, Y_4 = v_4) P(Y_5 = v_5)$$

Como  $P(Y_6 = A | \text{class}) + P(Y_6 = B | \text{class})$  devem ser 1, é preciso normalizar os valores

$$\bullet P(Y_6 = A | Y_1, Y_2, Y_3, Y_4, Y_5) = \frac{\frac{P(Y_1, Y_2 | Y_6 = A) P(Y_3, Y_4 | Y_6 = A) P(Y_5 | Y_6 = A) P(Y_6 = A)}{P(Y_1, Y_2) P(Y_3, Y_4) P(Y_5)}}{P(Y_1, Y_2 | Y_6 = A) P(Y_3, Y_4 | Y_6 = A) P(Y_5 | Y_6 = A) P(Y_6 = A) + P(Y_1, Y_2 | Y_6 = B) P(Y_3, Y_4 | Y_6 = B) P(Y_5 | Y_6 = B)}$$

$$\frac{P(Y_1, Y_2 | Y_6 = A) P(Y_3, Y_4 | Y_6 = A) P(Y_5 | Y_6 = A) P(Y_6 = A)}{P(Y_1, Y_2) P(Y_3, Y_4) P(Y_5)} = \frac{P(Y_1, Y_2 | Y_6 = A) P(Y_3, Y_4 | Y_6 = A) P(Y_5 | Y_6 = A) P(Y_6 = A)}{P(Y_1, Y_2 | Y_6 = A) P(Y_3, Y_4 | Y_6 = A) P(Y_5 | Y_6 = A) P(Y_6 = A) + P(Y_1, Y_2 | Y_6 = B) P(Y_3, Y_4 | Y_6 = B) P(Y_5 | Y_6 = B)}$$

$$x_8: \bullet P(Y_6 = A | x_8) = \frac{P(Y_1 = 0.38, Y_2 = 0.52 | Y_6 = A) P(Y_3 = 0, Y_4 = 1 | Y_6 = A) P(Y_5 = 0 | Y_6 = A)}{P(Y_1 = 0.38, Y_2 = 0.52 | Y_6 = A) P(Y_3 = 0, Y_4 = 1 | Y_6 = A) P(Y_5 = 0 | Y_6 = A) + P(Y_1 = 0.38, Y_2 = 0.52 | Y_6 = B)}$$

$$\frac{P(Y_3 = 0, Y_4 = 1 | Y_6 = A) P(Y_5 = 0 | Y_6 = B)}{P(Y_3 = 0, Y_4 = 1 | Y_6 = B) P(Y_5 = 0 | Y_6 = B) P(Y_6 = B)} \cong \frac{0.9847 \times \frac{1}{3} \times \frac{1}{3} \times \frac{3}{7}}{0.9847 \times \frac{1}{3} \times \frac{1}{3} \times \frac{3}{7} + 1.974 \times \frac{1}{4} \times \frac{1}{4} \times \frac{4}{7}} \cong 0.3994$$

$$\bullet P(B | x_8) = P(Y_6 = B | Y_1 = 0.38, Y_2 = 0.52, Y_3 = 0, Y_4 = 1, Y_5 = 0) = 1 - P(Y_6 = A | x_8) = 1 - 0.3994$$

$$x_9: \bullet P(Y_6 = A | x_9) = \frac{P(Y_1 = 0.42, Y_2 = 0.59 | Y_6 = A) P(Y_3 = 0, Y_4 = 1 | Y_6 = A) P(Y_5 = 1 | Y_6 = A) P(Y_6 = A)}{P(Y_1 = 0.42, Y_2 = 0.59 | Y_6 = A) P(Y_3 = 0, Y_4 = 1 | Y_6 = A) P(Y_5 = 1 | Y_6 = A) P(Y_6 = A) + P(Y_1 = 0.42, Y_2 = 0.59 | Y_6 = B)}$$

$$\frac{P(Y_3 = 0, Y_4 = 1 | Y_6 = A) P(Y_5 = 1 | Y_6 = B)}{P(Y_3 = 0, Y_4 = 1 | Y_6 = B) P(Y_5 = 1 | Y_6 = B) P(Y_6 = B)} \cong \frac{0.4031 \times \frac{1}{3} \times \frac{1}{3} \times \frac{3}{7}}{0.4031 \times \frac{1}{3} \times \frac{1}{3} \times \frac{3}{7} + 1.5925 \times \frac{1}{4} \times \frac{1}{2} \times \frac{4}{7}} \cong 0.1444$$

$$\bullet P(Y_6 = B | x_9) = P(Y_6 = B | Y_1 = 0.42, Y_2 = 0.59, Y_3 = 0, Y_4 = 1, Y_5 = 1) = 1 - P(Y_6 = A | x_9) = 1 - 0.1444 \cong 0.8556$$

R: Desta forma, as observações de teste  $x_8$  e  $x_9$  são ambas classificadas com a classe B, pois  $P(Y_6 = A | x_8) < P(Y_6 = B | x_8)$  e  $P(Y_6 = A | x_9) < P(Y_6 = B | x_9)$ .

(c)  $\theta = 0.5$ 

$$f(x|y) = \begin{cases} A & P(A|x) > \theta \\ B & \text{c.c.} \end{cases}$$

Maximum Likelihood Assumption (todas as hipóteses têm a mesma probabilidade):  $h_{ML} = \arg\max_{y_6} P(x|y_6)$

Sabe-se que a observação de teste  $x_8$  deverá ser classificada como A e  $x_9$  como B, ~~para que~~ de modo a otimizar a accuracy tem-se:

$$\begin{cases} P(A|x_8) > \theta \\ P(A|x_9) \leq \theta \end{cases} \Leftrightarrow P(A|x_9) \leq \theta < P(A|x_8)$$

Tomar-se que  $P(x|h)$  é uma aproximação das posteriores, pelo que se tem:

$$\begin{aligned} P(x_8|A) &= \frac{P(A|x_8)P(x_8)}{P(A)} = \frac{0.3994 \times (P(Y_1=0.38, Y_2=0.52, Y_3=0, Y_4=1, Y_5=0))}{\cancel{\frac{3}{7}}} = \\ &= 0.3994 \times \frac{(P(Y_1=0.38, Y_2=0.52)P(Y_3=0, Y_4=1)P(Y_5=0))}{\cancel{\frac{3}{7}}} = \frac{0.3994 \times (3.625 \times \cancel{\frac{3}{7}} \times \cancel{\frac{2}{7}})}{\cancel{\frac{3}{7}}} \approx \cancel{0.27578} \end{aligned}$$

$$P(x_8|B) = \frac{P(B|x_8)P(x_8)}{P(A)} = \frac{0.6006 \times (3.625 \times \cancel{\frac{2}{7}} \times \cancel{\frac{2}{7}})}{\cancel{\frac{3}{7}}} \approx \cancel{0.311025}$$

Normalizando  $P(x_8|A)$ ,  $P(x_8|B) = \frac{P(x_8|A)}{P(x_8|A) + P(x_8|B)} = \frac{0.27578}{\cancel{0.27578} + \cancel{0.311025}} \approx \cancel{0.47}$

$$\begin{aligned} P(x_9|A) &= \frac{P(A|x_9)P(x_9)}{P(A)} = \frac{0.1444 \times P(Y_1=0.42, Y_2=0.59, Y_3=0, Y_4=1, Y_5=1)}{\cancel{\frac{4}{7}}} = \\ &= 0.1444 \times \frac{(P(Y_1=0.42, Y_2=0.59)P(Y_3=0, Y_4=1)P(Y_5=1))}{\cancel{\frac{4}{7}}} = \frac{0.1444 \times (2.126 \times \cancel{\frac{2}{7}} \times \cancel{\frac{3}{7}})}{\cancel{\frac{4}{7}}} \approx \cancel{0.0658} \end{aligned}$$

$$P(x_9|B) = \frac{P(B|x_9)P(x_9)}{P(B)} = \frac{0.8556 \times (2.126 \times \cancel{\frac{2}{7}} \times \cancel{\frac{3}{7}})}{\cancel{\frac{4}{7}}} \approx \cancel{0.3898}$$

Normalizando  $P(x_9|A)$ ,  $P(x_9|B) = \frac{P(x_9|A)}{P(x_9|A) + P(x_9|B)} = \frac{0.0658}{\cancel{0.0658} + \cancel{0.3898}} \approx \cancel{0.1444}$

Threshold ( $\theta$ )	# identificações corretamente	Accuracy
$\theta = 0.5$	1	1/2
$\theta = 0.444$	2	2/2
$\theta = 0.2$	2	2/2
$\theta = 0.4$	2	2/2
$\Theta \in [0.14, 0.47]$		

R: Os thresholds que otimizam a testing accuracy são thresholds entre  $P(A|x_9) \leq \theta < P(A|x_8)$ , ou seja no intervalo  $\Theta$ .

Q

$$(a) D|y_2 = [0, 1]$$

$$b_0: [0, 0.5[ \Rightarrow 0$$

$$b_1: [0.5, 1] \Rightarrow 1$$

Após a binarização:

D	$y_a$
$x_1$	0
$x_2$	0
$x_3$	1
$x_4$	0
$x_5$	0
$x_6$	0
$x_7$	1
$x_8$	1
$x_9$	1

Fld 1

D	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$
$x_1$	0.24	0	1	1	0	A
$x_2$	0.16	0	1	0	1	A
$x_3$	0.32	1	0	1	2	A
$x_4$	0.54	0	0	0	1	B
$x_5$	0.66	0	0	0	0	B
$x_6$	0.76	0	1	0	2	B
$x_7$	0.41	1	0	1	1	B
$x_8$	0.38	1	0	1	0	A
$x_9$	0.42	1	0	1	1	B

Fld 2

Fld 3

1º interrogatório ; Training set: Fld 1 + Fld 2

Testing set: Fld 3

(b) Usando o 1º El.: k=3

$x_7$	$d$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
		4	4	②	②	③	4

$$\cdot d(x_7, x_1) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 4 \quad \cdot d(x_7, x_2) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 4 \quad \cdot d(x_7, x_3) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 2$$

$$\cdot d(x_7, x_4) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} = 2 \quad \cdot d(x_7, x_5) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} = 3 \quad \cdot d(x_7, x_6) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = 4$$

$$\hat{y}_1(x_7) = \text{Weighted mean } \left( \frac{\frac{1}{2} \times 0.32 + \frac{1}{2} \times 0.54 + \frac{1}{3} \times 0.66}{\frac{1}{2} + \frac{1}{2} + \frac{1}{3}} \right) \approx 0.4875$$

$x_8$	$d$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
		4	4	①	4	③	5

$$\cdot d(x_8, x_1) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 2 \quad \cdot d(x_8, x_2) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 4 \quad \cdot d(x_8, x_3) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 1$$

$$\cdot d(x_8, x_4) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} = 4 \quad \cdot d(x_8, x_5) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} = 3 \quad \cdot d(x_8, x_6) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = 5$$

$$\hat{y}_1(x_8) = \text{Weighted mean } \left( \frac{\frac{1}{2} \times 0.24 + \frac{1}{1} \times 0.32 + \frac{1}{3} \times 0.66}{\frac{1}{2} + \frac{1}{1} + \frac{1}{3}} \right) \approx 0.36$$

$x_9$	$d$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
		4	4	②	②	③	4

$$\cdot d(x_9, x_1) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = 4 \quad \cdot d(x_9, x_2) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = 4 \quad \cdot d(x_9, x_3) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = 2$$

$$\cdot d(x_9, x_4) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} = 2 \quad \cdot d(x_9, x_5) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} = 3 \quad \cdot d(x_9, x_6) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = 4$$

$$\hat{y}_1(x_9) = \text{Weighted mean } \left( \frac{\frac{1}{2} \times 0.32 + \frac{1}{2} \times 0.54 + \frac{1}{3} \times 0.66}{\frac{1}{2} + \frac{1}{2} + \frac{1}{3}} \right) \approx 0.4875$$

$$\cdot MAE = \frac{1}{m} \sum_i |\text{RES}_i| = \frac{1}{3} \left( |0.41 - 0.4875| + |0.38 - 0.36| + |0.42 - 0.4875| \right) \approx 0.055$$

## II. Programming and critical analysis

1) (a)

```
from sklearn.naive_bayes import GaussianNB
from sklearn.neighbors import KNeighborsClassifier

from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.model_selection import StratifiedKFold

from scipy.io.arff import loadarff
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Reading the ARFF file
data = loadarff('column_diagnosis.arff')
df = pd.DataFrame(data[0])
df['class'] = df['class'].str.decode('utf-8')

X = df.drop('class', axis=1)
y = df['class']

# kNN
neigh = KNeighborsClassifier(n_neighbors=5)
neigh.fit(X, y)

# Bayesian
gnb = GaussianNB()
gnb.fit(X, y)

cv = StratifiedKFold(n_splits=10, shuffle=True, random_state=0)

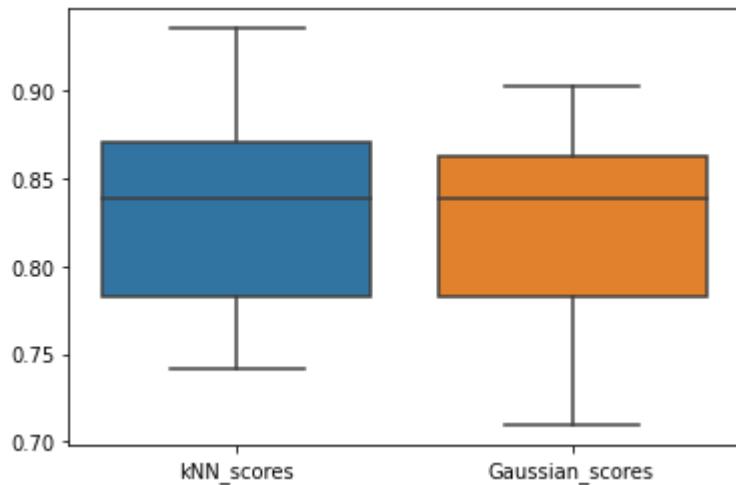
# Cross-validation for kNN
knn_scores = cross_val_score(neigh, X, y, cv=cv)

# Cross-validation for Gaussian Naive Bayes
gnb_scores = cross_val_score(gnb, X, y, cv=cv)

# Boxplot
data = pd.DataFrame({
    'label1': "Knn",
    'kNN_scores': knn_scores,
```

```
'label2': "Gnb",
'Gaussian_scores': gnb_scores
})

sns.boxplot(data=data)
plt.show()
```



(b)

Para descobrir se kNN é estatisticamente superior, utilizamos um teste de hipóteses, considerando as seguintes hipóteses:

H0: Accuracy do kNN igual à accuracy do GNB

H1: Accuracy do kNN maior do que a accuracy do GNB

```
from scipy import stats

res = stats.ttest_rel(knn_scores, gnb_scores, alternative='greater')
print("knn > gnb ? p-value=", res.pvalue)
```

knn > gnb ? p-value= 0.19042809062064092

Baseado neste p-value, não podemos rejeitar a hipótese nula para níveis de significância 1%, 5% e 10%. Concluímos assim que não há uma diferença estatisticamente significativa entre os modelos kNN e naive Bayes e não podemos afirmar com confiança que kNN é estatisticamente superior ao naive Bayes em termos de accuracy.

2)

```
from sklearn.metrics import confusion_matrix
import numpy as np

# Create models
knn1 = KNeighborsClassifier(n_neighbors=1, weights='uniform', metric='euclidean')

knn5 = KNeighborsClassifier(n_neighbors=5, weights='uniform', metric='euclidean')

# Initialize SKFold
cv = StratifiedKFold(n_splits=10, shuffle=True, random_state=0)

# Initialize confusion matrices arrays
confusion_matrices_knn1 = []
confusion_matrices_knn5 = []

# Cross-validation
for train_index, test_index in cv.split(X, y):
    X_train, X_test = X.iloc[train_index], X.iloc[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]

    knn1.fit(X_train, y_train)
    knn1_pred = knn1.predict(X_test)

    knn5.fit(X_train, y_train)
    knn5_pred = knn5.predict(X_test)

    # Add the values to the confusion matrix
    confusion_matrices_knn1.append(confusion_matrix(y_test, knn1_pred))
    confusion_matrices_knn5.append(confusion_matrix(y_test, knn5_pred))

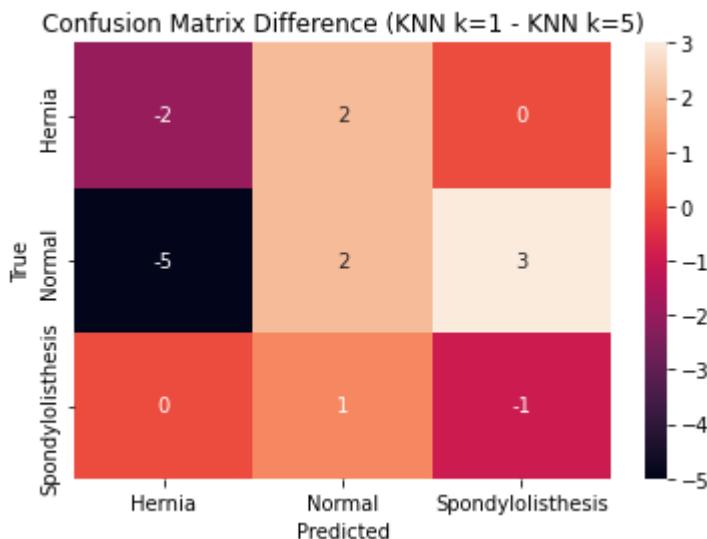
# Calculate the final confusion matrices for each model
confusion_matrix_knn1 = np.sum(confusion_matrices_knn1, axis=0)
confusion_matrix_knn5 = np.sum(confusion_matrices_knn5, axis=0)

# Result confusion matrix
confusion_matrix = confusion_matrix_knn1 - confusion_matrix_knn5

# Plotting
labels = ['Hernia', 'Normal', 'Spondylolisthesis']
sns.heatmap(confusion_matrix, annot=True, fmt='g', xticklabels=labels, yticklabels=labels)

plt.title("Confusion Matrix Difference (KNN k=1 - KNN k=5)")
plt.xlabel("Predicted")
plt.ylabel("True")
```

```
plt.show()
```



Este gráfico mostra a diferença entre as matrizes cumulativas dos preditores de kNN, com  $k=1$  e  $k=5$ . Nas secções com valor positivo, o modelo com  $k=1$  é superior, enquanto nos valores negativos  $k=5$  é mais preciso.

Observando o gráfico podemos ver que a matriz do preditor de KNN com  $k = 1$ , previu mais casos errados em que a pessoa não tinha doenças e na realidade sofria de hérnia ou de spondylolisthesis. Por outro lado, também previu mais vezes que o paciente tinha spondylolisthesis quando não apresentava doença.

Podemos ver também que o preditor de KNN com  $k = 5$ , previu mais casos em que a pessoa sofria de uma hérnia, que na realidade eram falsos positivos, visto que as pessoas não deviam ser diagnosticadas com hérnia.

Observando a diagonal, que identifica os **True Positives**, podemos ver que o preditor de kNN com  $k=1$  identifica mais casos de doenças corretamente, que o preditor com  $k = 5$ . Podemos observar também que os dois preditores, foram idênticos nos casos em que se identificou uma doença e na realidade era outra (previu-se que o indivíduo sofria de hérnia mas na realidade sofria de spondylolisthesis, e vice-versa).

Concluindo, podemos ver que o preditor com  $k = 1$  é superior em algumas secções, e que o preditor com  $k = 5$  é mais preciso noutros casos. Deste modo, não podemos afirmar que um é universalmente melhor que o outro.

3)

1 - O dataset possui mais casos de Spondylolisthesis que do resto das classes (150 de Spondylolisthesis contra 60 casos de Hérnia e 100 casos normais). Uma vez que o naive Bayes assume que os parâmetros são condicionalmente independentes dentro de cada classe, uma classe dominante

pode levar a que o naive Bayes tenha um "bias" e favoreça essa mesma classe incorretamente nas previsões.

2 - Naive Bayes assume que todos os parâmetros são independentes entre si. Em datasets clínicos como este, há uma grande possibilidade de alguns elementos estarem relacionados. Deste modo, ao ignorar relações entre parâmetros, o modelo pode produzir previsões imprecisas.

3 - O naive Bayes precisa de bastantes dados para aproximar com exatidão as suas funções de massa e densidade de probabilidade. Uma vez que o dataset tem uma dimensão média, de 310 elementos, a aproximação pode não ser a mais adequada, o que pode ser a causa da menor precisão deste modelo.

**END**