

I. Pen-and-paper

1)

a)

HW - 2

① as Bayesian Classifier $P(z|x) = \frac{P(x|z) \cdot P(z)}{P(x)}$

$\{y_1, y_2\}$ $\forall x, y_i \in \mathbb{R}^1 \sim$ Normally Distributed

$\mu_{1,A} = 0.24$ $\Sigma_A = \begin{bmatrix} 0.0064 & 0.0036 \\ 0.0036 & 0.0036 \end{bmatrix}$

$\mu_{1,B} = 0.5925$

$\mu_{2,A} = 0.52$ $\Sigma_B = \begin{bmatrix} 0.02289 & -0.003752 \\ -0.003752 & 0.02143 \end{bmatrix}$

$\mu_{2,B} = 0.3275$

$P(y_1/y_2 | A) \sim \text{Multivariate Normal}(\mu_{1,A}; \Sigma_A)$

$P(y_1/y_2 | B) \sim \text{Multivariate Normal}(\mu_{1,B}; \Sigma_B)$

$\{y_3, y_4\}$

Prob	$y_3=0, y_4=0$	$y_3=1, y_4=0$	$y_3=0, y_4=1$	$y_3=1, y_4=1$
A	0	1/3	1/3	1/3
B	1/2	1/4	1/4	0

$\{y_5\}$

Prob	$y_5=0$	$y_5=1$	$y_5=2$
A	1/3	1/3	1/3
B	1/4	1/2	1/4

b)

b) MAP $c_i = \{A, B\}$ // Nota: Divisão por $P(x)$ e simplificação

$y_8 \rightarrow \text{argmax} \{P(c_i) \times P(x_8 | c_i)\}$

$y_8 \rightarrow \text{argmax} \{P(c_i) \times P(x_8 | c_i)\}$

$P(A) = 3/7 //$

$P(B) = 4/7 //$

$x_8 = A$

$\mu_{A1} \sim \text{Normal}(0,17, 0,51, \Sigma_A) = 0,9847$

$P(y_1=0, y_2=0, y_3=1 | A) //$

$P(y_3=0, y_4=1 | A) = 1/3 //$

$P(y_5=0 | A) = 1/3$

$P(x_2 | A) = 0,1034$

$y_{8A} = 0,04699$

$x_8 = B$

$P(y_1=0, y_2=0, y_3=1 | B) = 1,26237$

$P(y_3=0, y_4=1 | B) = 1/4$

$P(y_5=0 | B) = \frac{1}{4}$

$P(x_2 | B) = 0,122648$

$y_{8B} = 0,07 //$

$y_8 = 0,07$, pois $0,07 > 0,04699 //$

$y_8 = A$ Me previsão

$X_B - A$
 $P(y_1 = 0,41; y_2 = 0,89 | A) \sim \text{Multivariate Normal}(\begin{pmatrix} 0,24; 0,72 \\ \Sigma_A \end{pmatrix})$
 $= 0,403079$
 $P(y_3 = 0; y_4 = 1 | A) = \frac{1}{2}$
 $P(y_5 = 1 | A) = \frac{1}{2}$
 $P(x_5 | A) = 0,0447857$
 $y_{5A} = P(A) \times P(x_5 | A) = 0,013134$
 $X_B - B$
 $P(y_1 = 0,41; y_2 = 0,89 | B) = 1,72857$
 $P(y_3 = 0; y_4 = 1 | B) = \frac{1}{4}$
 $P(y_5 = 1 | B) = \frac{1}{2}$
 $P(x_5 | B) = 0,21607$
 $y_{5B} = P(B) \times P(x_5 | B) = 0,1134693$
 $y_5 = 0,1134693$, pois $0,1134693 > 0,013134$
 $y_6 = B$ previsão correta

c)

c) Maximum Likelihood Assumption:

$$P(x|A) \approx P(A|x) \quad \text{Normalized: } P(A|x_2) =$$

$$P(x_2|A) = 0,1094$$

$$= \frac{P(x_2|A)}{P(x_2|A) + P(x_2|B)}$$

$$P(x_3|A) = 0,044957$$

$$= 0,47145$$

$$P(x_3|\theta) = \begin{cases} A, & P(A|x_3) > \theta \\ B, & \text{otherwise} \end{cases} \quad P(A|x_3) = \frac{P(x_3|A)}{P(x_3|A) + P(x_3|B)} =$$

$$\theta < 0,47145 \quad x_3 = 56 = A \quad = 0,1716876$$

$$P(x_3|\theta) = \begin{cases} A, & P(A|x_3) > \theta \\ \text{---} & \end{cases} \quad x_3 = 56 = B$$

$$\theta > 0,1716876 \quad R: [0,1716876; 0,47145[$$

2)

a)

② a)

	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	0,34	0	0	0	1	B
x_5	0,66	0	0	0	0	B
x_6	0,26	0	1	0	2	B
x_7	0,41	1	0	1	1	B
x_8	0,39	1	0	1	0	A
x_9	0,42	1	0	1	1	B

Bin 1 ($< 0,5$) $y_F = 0$
 Bin 2 ($\geq 0,5$) $y_F = 1$

Fold 1 (Iteration 1):
 Testing: x_2, x_8, x_9 (Bin 2)
 Training: x_1, x_3, x_4, x_5, x_6 (Bin 1) x_7 (Bin 2)

Fold 2 (Iteration 2):
 Testing: x_4, x_5, x_6 (Bin 1)
 Training: x_1, x_2 (Bin 1) x_3, x_7, x_8, x_9 (Bin 2)

Fold 3 (Iteration 3):
 Testing: x_1, x_2 (Bin 1) x_3 (Bin 2)
 Training: x_4, x_5, x_6 (Bin 1) x_7, x_8, x_9 (Bin 2)

b)

b) 1st Iteration

Training: $x_1; x_2; x_3; x_4; x_5; x_6$

Testing: $x_7; x_8; x_9$

$x_1 = 0110 \text{ A}$	4	2	4
$x_2 = 0101 \text{ A}$	4	4	4
$x_3 = 1012 \text{ A}$	2	1	2
$x_4 = 0001 \text{ B}$	2	4	2
$x_5 = 0000 \text{ B}$	3	3	3
$x_6 = 0102 \text{ B}$	4	5	4
	$x_7 (1011 \text{ B})$	$x_8 (1010 \text{ A})$	$x_9 (1011 \text{ B})$

$K=3$ – Least Values

$$\hat{f}(x_{\text{new}}) \leftarrow \frac{\sum_{i=1}^K w_i f(x_i)}{\sum_{i=1}^K w_i} \quad w(x_{\text{new}}) = \frac{1}{d}$$

$$\hat{f}(x_{7,\text{new}}) = \frac{\frac{1}{2} \times 0,31 + \frac{1}{2} \times 0,54 + \frac{1}{3} \times 0,66}{4/3} = 0,4875$$

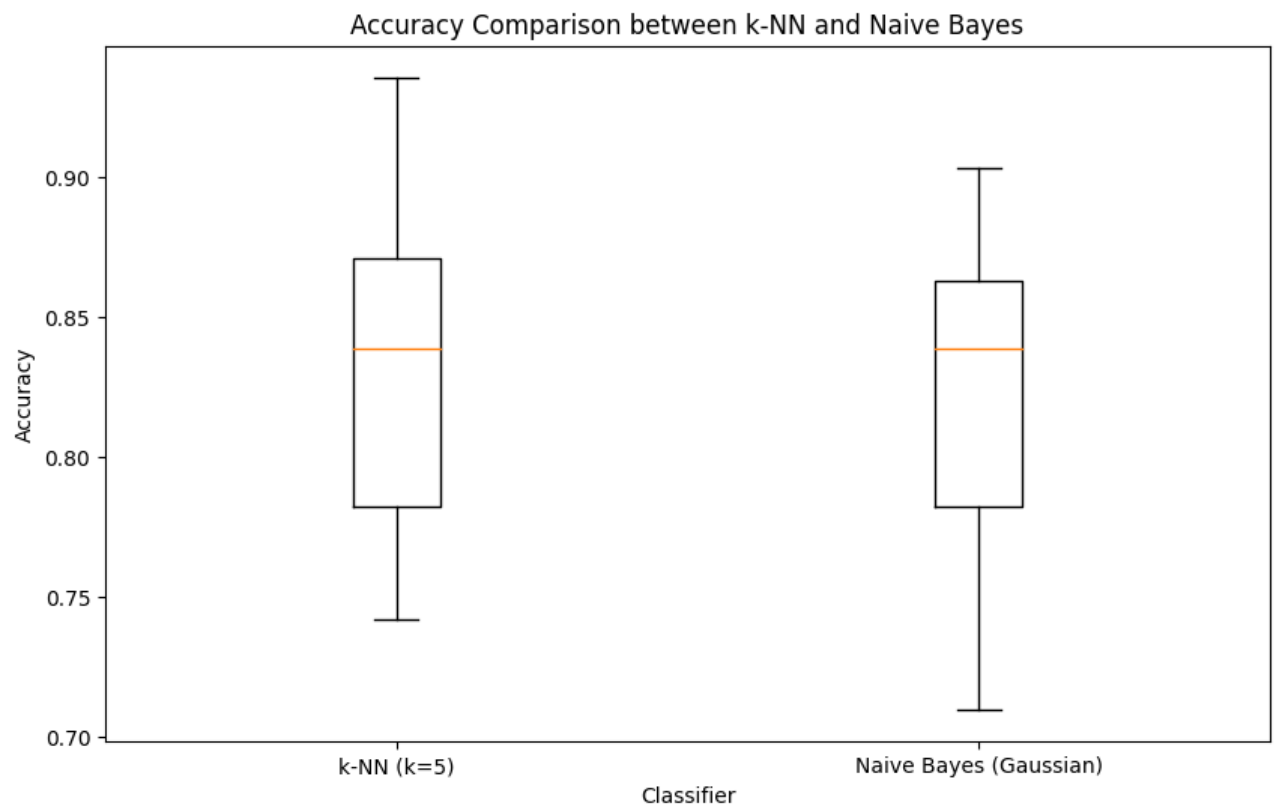
$$\hat{f}(x_{8,\text{new}}) = \frac{0,31 + \frac{1}{2} \times 0,24 + \frac{1}{3} \times 0,66}{11/6} = 0,36$$

$$\hat{f}(x_{9,\text{new}}) = \frac{\frac{1}{2} \times 0,31 + \frac{1}{2} \times 0,54 + \frac{1}{3} \times 0,66}{4/3} = 0,4875$$

$$\text{MAE} = \frac{(0,4875 - 0,41) + 0,38 \times 0,36 + 0,4875 - 0,42}{3} = 0,055$$

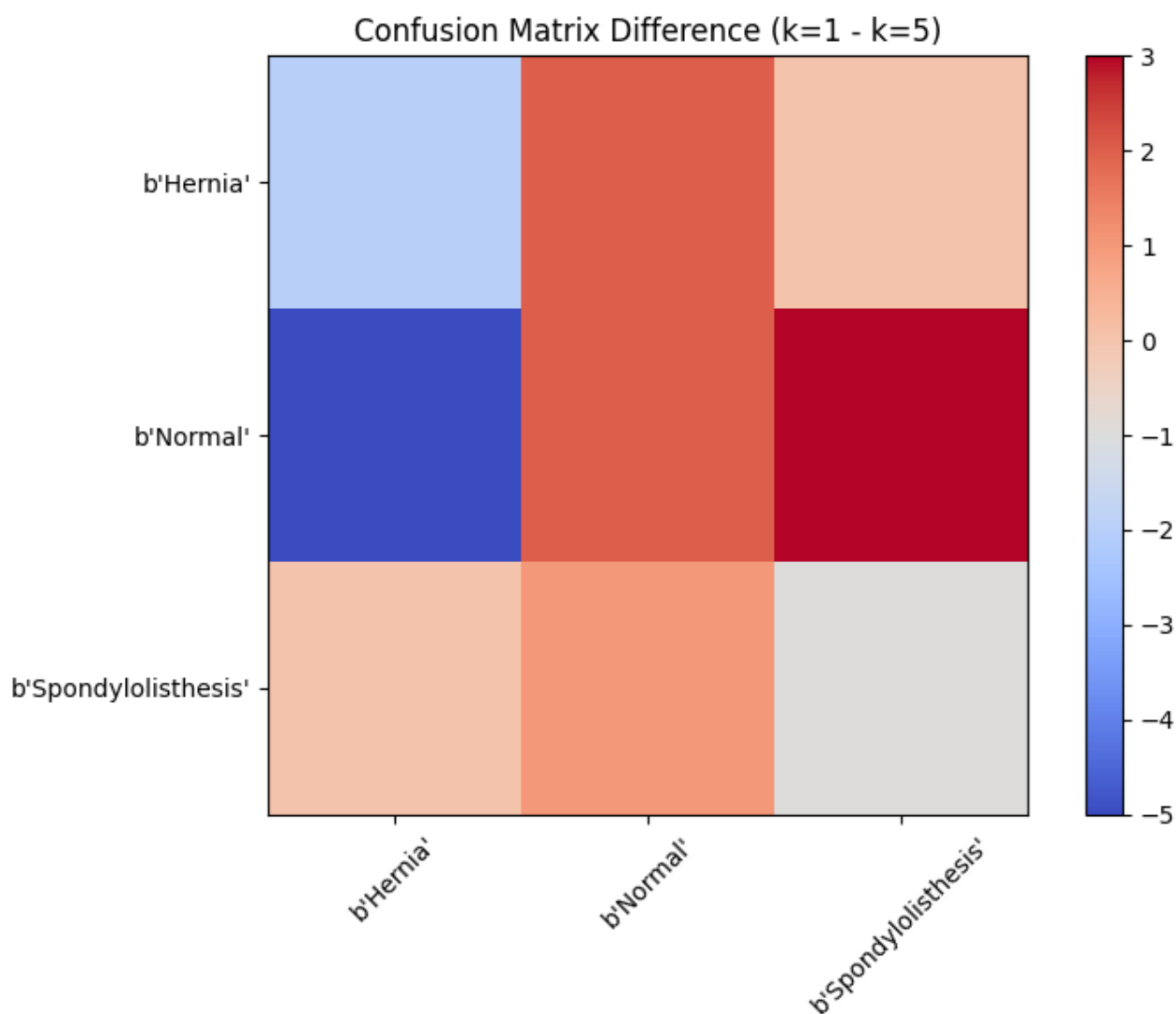
II. Programming and critical analysis

1. a)



b) False (fail to reject the null hypothesis): There is no statistically significant difference in accuracy between k-NN and Naive Bayes.

2.



Comentário:

Observando os quadrados da confusion matrix difference da diagonal, ou seja, os casos em que a previsão é correta podemos afirmar de acordo com o valor neste quadrado qual dos predicados foi o mais correto. Por exemplo, quando a previsão de Hernia é correta o valor nesta matriz por ser negativo (-2) diz-nos que o predicado K=5 preveu de forma mais acertada que o K=1, pois $K=5 > K=1$.

3.

Suposição de Independência das Características:

O Naïve Bayes assume que todas as características são condicionalmente independentes dadas as classes. No conjunto de dados "column_diagnosis," as características biomecânicas podem não ser totalmente independentes. Por exemplo, pode haver relações complexas entre essas características que violam a suposição de independência. Se essas relações forem cruciais para a classificação, o Naïve Bayes pode ter dificuldade em capturá-las de maneira eficaz.

Distribuição Desequilibrada de Classes:

Se o conjunto de dados tiver uma distribuição de classes altamente desequilibrada, em que uma classe supera significativamente as outras, o Naïve Bayes pode ter dificuldade em prever com precisão as classes minoritárias. Em conjuntos de dados de saúde, a distribuição de tipos de doenças pode não ser equilibrada. O Naïve Bayes pode ter dificuldade em aprender com as classes minoritárias devido à escassez de exemplos, o que pode levar a previsões incorretas/tendenciosas.

Sensibilidade a Valores Atípicos:

O Naïve Bayes pode ser sensível a valores atípicos (outliers) no conjunto de dados. Medidas biomecânicas no domínio da saúde podem conter valores atípicos devido a erros de medição ou casos extremos raros. Estes podem afetar significativamente as estimativas de média e variância usadas nos cálculos de probabilidade do Naïve Bayes, o que pode levar a resultados de classificação subótimos.



Homework2.ipynb

END