Aprendizagem 2023/24
Homework II

I. Pen-and-paper [13v]

Consider the following dataset:

D	y_1	y_2	y_3	y_4	y_5	y_6
$\mathbf{x}_1$	0.24	0.36	1	1	0	A
$\mathbf{x}_2$	0.16	0.48	1	0	1	A
$\mathbf{x}_3$	0.32	0.72	0	1	2	A
$\mathbf{x}_4$	0.54	0.11	0	0	1	B
$\mathbf{x}_5$	0.66	0.39	0	0	0	B
$\mathbf{x}_6$	0.76	0.28	1	0	2	B
$\mathbf{x}_7$	0.41	0.53	0	1	1	B
$\mathbf{x}_8$	0.38	0.52	0	1	0	A
$\mathbf{x}_9$	0.42	0.59	0	1	1	B

1. Consider $\mathbf{x}_1$–$\mathbf{x}_7$ to be training observations, $\mathbf{x}_8$–$\mathbf{x}_9$ to be testing observations, y_1–y_5 to be input variables and y_6 to be the target variable.
   Hint: you can use *scipy.stats.multivariate_normal* for multivariate distribution calculus

   a. [3.5v] Learn a Bayesian classifier assuming: i) $\{y_1, y_2\}$, $\{y_3, y_4\}$ and $\{y_5\}$ sets of independent variables (e.g., $y_1 \perp\!\!\!\perp y_3$ yet $y_1 \not\perp\!\!\!\perp y_2$), and ii) $y_1 \times y_2 \in \mathbb{R}^2$ is normally distributed. Show all parameters (distributions and priors for subsequent testing).

Homework 2:

Bayesian and Lazy Learning

I. Pen-and-Paper

Exercício 1)

· Alínea a)

Bayes Rule: $p(h|D) = \frac{p(D|h) \cdot p(h)}{p(D)}$

$p(h)$: priors $\rightarrow p(A) = \frac{3}{7} \quad p(B) = \frac{4}{7}$

$p(D|h)$:

· Probability Mass Functions

· $\{y_3, y_4\}$ and $\{y_5\}$ are sets of independent variables.

$p(y_3=0; y_4=0 | A) = 0 \qquad p(y_3=0; y_4=0 | B) = \frac{1}{2} \qquad p(y_5=0|A) = \frac{1}{3}$

$p(y_3=0; y_4=1 | A) = \frac{1}{3} \qquad p(y_3=0; y_4=1 | B) = \frac{1}{4} \qquad p(y_5=1|A) = \frac{1}{3}$

$p(y_3=1; y_4=0 | A) = \frac{1}{3} \qquad p(y_3=1; y_4=0 | B) = \frac{1}{4} \qquad p(y_5=2|A) = \frac{1}{3}$

$p(y_3=1; y_4=1 | A) = \frac{1}{3} \qquad p(y_3=1; y_4=1 | B) = 0 \qquad p(y_5=0|B) = \frac{1}{4}$

$p(y_5=1|B) = \frac{1}{2}$

$p(y_5=2|B) = \frac{1}{4}$

· Probability Density Functions

· $\{y_1, y_2\}$ are dependent variables and $y_1 \cdot y_2 \in \mathbb{R}^2$ is normally distributed.

$$\mu_{\{y_1,y_2\}/h=A} = \frac{1}{3}\left(\begin{bmatrix}0{,}24\\0{,}36\end{bmatrix} + \begin{bmatrix}0{,}16\\0{,}48\end{bmatrix} + \begin{bmatrix}0{,}32\\0{,}72\end{bmatrix}\right) = \begin{bmatrix}0{,}24\\0{,}52\end{bmatrix} \rightarrow$$

$$\rightarrow \Sigma_{\{y_1,y_2\}/h=A} = \begin{bmatrix}\Sigma_{00} & \Sigma_{10}\\ \Sigma_{01} & \Sigma_{11}\end{bmatrix}$$

$\cdot\ \Sigma_{00} = \frac{1}{3-1}\left((0{,}24-0{,}24)^2 + (0{,}16-0{,}24)^2 + (0{,}32-0{,}24)^2\right) = 0{,}0064$

$\cdot\ \Sigma_{10} = \Sigma_{01} = \frac{1}{3-1}\Big((0{,}24-0{,}24)(0{,}36-0{,}52) + (0{,}16-0{,}24)(0{,}48-0{,}52) +$
$+ (0{,}32-0{,}24)(0{,}72-0{,}52)\Big) = 0{,}0096$

$\cdot\ \Sigma_{11} = \frac{1}{3-1}\left((0{,}36-0{,}52)^2 + (0{,}48-0{,}52)^2 + (0{,}72-0{,}52)^2\right) = 0{,}0336$

$$\Sigma_{\{y_1,y_2\}/h=A} = \begin{bmatrix}0{,}0064 & 0{,}0096\\ 0{,}0096 & 0{,}0336\end{bmatrix}$$

$$\mu_{\{y_1,y_2\}/h=B} = \frac{1}{4}\left(\begin{bmatrix}0{,}54\\0{,}11\end{bmatrix} + \begin{bmatrix}0{,}66\\0{,}39\end{bmatrix} + \begin{bmatrix}0{,}76\\0{,}28\end{bmatrix} + \begin{bmatrix}0{,}41\\0{,}53\end{bmatrix}\right) = \begin{bmatrix}0{,}5925\\0{,}3275\end{bmatrix}$$

$$\Sigma_{\{y_1,y_2\}/h=B} = \begin{bmatrix}\Sigma_{00} & \Sigma_{10}\\ \Sigma_{01} & \Sigma_{11}\end{bmatrix}$$

$\cdot\ \Sigma_{00} = \frac{1}{4-1}\left((0{,}54-0{,}5925)^2 + (0{,}66-0{,}5925)^2 + (0{,}76-0{,}5925)^2 + (0{,}41-0{,}5925)^2\right) =$
$= 0{,}02289$

$\cdot\ \Sigma_{10} = \Sigma_{01} = \frac{1}{4-1}\Big((0{,}54-0{,}5925)(0{,}11-0{,}3275) + (0{,}66-0{,}5925)(0{,}39-0{,}3275) +$
$+ (0{,}76-0{,}5925)(0{,}28-0{,}3275) + (0{,}41-0{,}5925)(0{,}53-0{,}3275)\Big) = -0{,}009758$

$\cdot\ \Sigma_{11} = \frac{1}{4-1}\left((0{,}11-0{,}3275)^2 + (0{,}39-0{,}3275)^2 + (0{,}28-0{,}3275)^2 + (0{,}53-0{,}3275)^2\right) =$
$= 0{,}03149$

Aprendizagem 2023/24
Homework II

$$\mathcal{N}\left(\mu_{\{y_1,y_2\}/h=A}=\begin{bmatrix}0{,}24\\0{,}52\end{bmatrix},\Sigma_{\{y_1,y_2\}/h=A}=\begin{bmatrix}0{,}0064 & 0{,}0096\\0{,}0096 & 0{,}0336\end{bmatrix}\middle| A\right)$$

$$\mathcal{N}\left(\mu_{\{y_1,y_2\}/h=B}=\begin{bmatrix}0{,}5925\\0{,}3275\end{bmatrix},\Sigma_{\{y_1,y_2\}/h=B}=\begin{bmatrix}0{,}02289 & -0{,}009758\\-0{,}009758 & 0{,}03149\end{bmatrix}\middle| B\right)$$

$p(D)$:

$p(y_3=0;y_4=0)=\frac{2}{7}$ $p(y_3=0;y_4=1)=\frac{2}{7}$ $p(y_3=1;y_4=0)=\frac{2}{7}$ $p(y_3=1;y_4=1)=\frac{1}{7}$

$p(y_5=0)=\frac{2}{7}$ $p(y_5=1)=\frac{3}{7}$ $p(y_5=2)=\frac{2}{7}$

$$\mu_{\{y_1,y_2\}}=\frac{1}{7}\left(\begin{bmatrix}0{,}24\\0{,}36\end{bmatrix}+\begin{bmatrix}0{,}16\\0{,}48\end{bmatrix}+\begin{bmatrix}0{,}32\\0{,}72\end{bmatrix}+\begin{bmatrix}0{,}54\\0{,}11\end{bmatrix}+\begin{bmatrix}0{,}66\\0{,}39\end{bmatrix}+\begin{bmatrix}0{,}76\\0{,}28\end{bmatrix}+\begin{bmatrix}0{,}41\\0{,}53\end{bmatrix}\right)=\begin{bmatrix}0{,}4414\\0{,}41\end{bmatrix}$$

$$\Sigma_{\{y_1,y_2\}}=\begin{bmatrix}\Sigma_{00} & \Sigma_{10}\\\Sigma_{01} & \Sigma_{11}\end{bmatrix}$$

- $\Sigma_{00}=\frac{1}{7-1}\big((0{,}24-0{,}4414)^2+(0{,}16-0{,}4414)^2+(0{,}32-0{,}4414)^2+(0{,}54-0{,}4414)^2+(0{,}66-0{,}4414)^2+(0{,}76-0{,}4414)^2+(0{,}41-0{,}4414)^2\big)=0{,}04908$
- $\Sigma_{10}=\Sigma_{01}=\frac{1}{7-1}\big((0{,}24-0{,}4414)(0{,}36-0{,}41)+(0{,}16-0{,}4414)(0{,}48-0{,}41)+(0{,}32-0{,}4414)(0{,}72-0{,}41)+(0{,}54-0{,}4414)(0{,}11-0{,}41)+(0{,}66-0{,}4414)(0{,}39-0{,}41)+(0{,}76-0{,}4414)(0{,}28-0{,}41)+(0{,}41-0{,}4414)(0{,}53-0{,}41)\big)=-0{,}02107$
- $\Sigma_{11}=\frac{1}{7-1}\big((0{,}36-0{,}41)^2+(0{,}48-0{,}41)^2+(0{,}72-0{,}41)^2+(0{,}11-0{,}41)^2+(0{,}39-0{,}41)^2+(0{,}28-0{,}41)^2+(0{,}53-0{,}41)^2\big)=0{,}03753$

$$\Sigma_{\{y_1,y_2\}}=\begin{bmatrix}0{,}04908 & -0{,}02107\\-0{,}02107 & 0{,}03753\end{bmatrix}$$

$$\mathcal{N}\left(\mu_{\{y_1,y_2\}}=\begin{bmatrix}0{,}4414\\0{,}41\end{bmatrix},\Sigma_{\{y_1,y_2\}}=\begin{bmatrix}0{,}04908 & -0{,}02107\\-0{,}02107 & 0{,}03753\end{bmatrix}\right)$$

b. [2.5v] Under a MAP assumption, classify each testing observation showing all your calculus.

- Alínea b)

X_8: $p(y_1=0{,}38;y_2=0{,}52;y_3=0;y_4=1;y_5=0|A)=$

$=p(y_1=0{,}38;y_2=0{,}52|A)\cdot p(y_3=0;y_4=1|A)\cdot p(y_5=0|A)\cdot p(A)$,

- $p(y_3=0;y_4=1|A)=\frac{1}{3}$, $p(y_5=0|A)=\frac{1}{3}$, $p(A)=\frac{3}{7}$ and
- $p(y_1=0{,}38;y_2=0{,}52|A)=\mathcal{N}\left(\begin{bmatrix}0{,}38\\0{,}52\end{bmatrix}\middle|\mu_{\{y_1,y_2\}/h=A},\Sigma_{\{y_1,y_2\}/h=A}\right)$,
- Using the calculations of the previous exercise:

$$\left|\Sigma_{\{y_1,y_2\}/h=A}\right|=0{,}0064\cdot 0{,}0336-0{,}0096^2=1{,}2288\times10^{-4}$$

$$\Sigma^{-1}_{\{y_1,y_2\}/h=A}=\frac{1}{1{,}2288\times10^{-4}}\begin{bmatrix}0{,}0336 & -0{,}0096\\-0{,}0096 & 0{,}0064\end{bmatrix}=\begin{bmatrix}273{,}4375 & -78{,}125\\-78{,}125 & 52{,}0833\end{bmatrix}$$

- Multivariate Gaussian distribution in m-dimensional spaces:

$$p(x|\mu,\Sigma)=\mathcal{N}(x|\mu,\Sigma)=\frac{1}{(2\pi)^{\frac{m}{2}}\sqrt{|\Sigma|}}e^{\left(-\frac{1}{2}(x-\mu)^T\cdot\Sigma^{-1}\cdot(x-\mu)\right)}$$

- $p(y_1=0{,}38;y_2=0{,}52|A)=\frac{1}{(2\pi)^{\frac{2}{2}}\sqrt{1{,}2288\times10^{-4}}}e^{\left(-\frac{1}{2}\left(\begin{bmatrix}0{,}38\\0{,}52\end{bmatrix}-\begin{bmatrix}0{,}24\\0{,}52\end{bmatrix}\right)^T\cdot\Sigma^{-1}_{\{y_1,y_2\}/h=A}\left(\begin{bmatrix}0{,}38\\0{,}52\end{bmatrix}-\begin{bmatrix}0{,}24\\0{,}52\end{bmatrix}\right)\right)}=$

$$=\frac{1}{2\pi\sqrt{1{,}2288\times10^{-4}}}e^{\left(-\frac{1}{2}([0{,}14\ \ 0])\begin{bmatrix}273{,}4375 & -78{,}125\\-78{,}125 & 52{,}0833\end{bmatrix}\begin{bmatrix}0{,}14\\0\end{bmatrix}\right)}=$$

$$=\frac{1}{2\pi\sqrt{1{,}2288\times10^{-4}}}e^{\left([-0{,}07\ \ 0]\begin{bmatrix}38{,}28125\\-10{,}9375\end{bmatrix}\right)}\simeq 0{,}989\text{, therefore:}$$

- $p(y_1=0{,}38;y_2=0{,}52;y_3=0;y_4=1;y_5=0|A)=0{,}989\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{3}{7}\simeq\boxed{0{,}04689}$
- $p(y_1=0{,}38;y_2=0{,}52;y_3=0;y_4=1;y_5=0|B)=$

$=p(y_1=0{,}38;y_2=0{,}52|B)\cdot p(y_3=0;y_4=1|B)\cdot p(y_5=0|B)\cdot p(B)$,

- $p(y_3=0;y_4=1|B)=\frac{1}{4}$, $p(y_5=0|B)=\frac{1}{4}$, $p(B)=\frac{4}{7}$ and
- $p(y_1=0{,}38;y_2=0{,}52|B)=\mathcal{N}\left(\begin{bmatrix}0{,}38\\0{,}52\end{bmatrix}\middle|\mu_{\{y_1,y_2\}/h=B},\Sigma_{\{y_1,y_2\}/h=B}\right)$,
- Using the calculations of the previous exercise:

$$\left|\Sigma_{\{y_1,y_2\}/h=B}\right|=0{,}02289\cdot 0{,}03149-(-0{,}009758)^2=6{,}2559\times10^{-4}$$

$$\Sigma^{-1}_{\{y_1,y_2\}/h=B}=\frac{1}{6{,}2559\times10^{-4}}\begin{bmatrix}0{,}03149 & 0{,}009758\\0{,}009758 & 0{,}02289\end{bmatrix}=\begin{bmatrix}50{,}3367 & 15{,}5981\\15{,}5981 & 36{,}5896\end{bmatrix},$$

- Using scipy.stats.multivariate_normal:

$$p(y_1=0{,}38;y_2=0{,}52|B)=\mathcal{N}\left(\begin{bmatrix}0{,}38\\0{,}52\end{bmatrix}\middle|\begin{bmatrix}0{,}5925\\0{,}3275\end{bmatrix},\begin{bmatrix}0{,}02289 & -0{,}009758\\-0{,}009758 & 0{,}03149\end{bmatrix}\right)\simeq 1{,}962,$$

therefore:

- $p(y_1=0{,}38;y_2=0{,}52;y_3=0;y_4=1;y_5=0|B)=1{,}962\cdot\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{4}{7}\simeq\boxed{0{,}07008}$
- Therefore X_8 is classified as B.

Aprendizagem 2023/24
Homework II

X_9: $\cdot\, \rho(y_1=0{,}42;\ y_2=0{,}59;\ y_3=0;\ y_4=1;\ y_5=1 \mid A) =$

$= \rho(y_1=0{,}42;\ y_2=0{,}59 \mid A) \cdot \rho(y_3=0;\ y_4=1 \mid A) \cdot \rho(y_5=1 \mid A) \cdot \rho(A)$,

$\cdot\, \rho(y_3=0;\ y_4=1 \mid A) = \frac{1}{3}$, $\rho(y_5=1 \mid A) = \frac{1}{3}$, $\rho(A) = \frac{3}{7}$ and

$\cdot\, \rho(y_1=0{,}42;\ y_2=0{,}59 \mid A) = \mathcal{N}\left(\begin{bmatrix}0{,}42\\0{,}59\end{bmatrix} \middle| \mu_{\{y_1,y_2\}/h=A}, \Sigma_{\{y_1,y_2\}/h=A}\right)$,

. Using the calculations of the previous ~~exercise~~ and this exercises:

$$\left|\Sigma_{\{y_1,y_2\}/h=A}\right| = 1{,}2288 \times 10^{-4} \quad \Sigma^{-1}_{\{y_1,y_2\}/h=A} = \begin{bmatrix}273{,}4375 & -78{,}125\\ -78{,}125 & 52{,}0833\end{bmatrix}$$

. Using scipy.stats.multivariate_normal:

$\rho(y_1=0{,}42;\ y_2=0{,}59 \mid A) = \mathcal{N}\left(\begin{bmatrix}0{,}42\\0{,}59\end{bmatrix} \middle| \begin{bmatrix}0{,}24\\0{,}52\end{bmatrix}, \begin{bmatrix}0{,}0064 & 0{,}0096\\0{,}0096 & 0{,}0336\end{bmatrix}\right) \simeq 0{,}4031$,

therefore:

$\cdot\, \rho(y_1=0{,}42;\ y_2=0{,}59;\ y_3=0;\ y_4=1;\ y_5=1 \mid A) = 0{,}403 \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{3}{7} \simeq \boxed{0{,}01919}$

$\cdot\, \rho(y_1=0{,}42;\ y_2=0{,}59;\ y_3=0;\ y_4=1;\ y_5=1 \mid B) =$

$= \rho(y_1=0{,}42;\ y_2=0{,}59 \mid B) \cdot \rho(y_3=0;\ y_4=1 \mid B) \cdot \rho(y_5=1 \mid B) \cdot \rho(B)$,

$\rho(y_3=0;\ y_4=1 \mid B) = \frac{1}{4}$, $\rho(y_5=1 \mid B) = \frac{1}{2}$, $\rho(B) = \frac{4}{7}$ and

$\cdot\, \rho(y_1=0{,}42;\ y_2=0{,}59 \mid B) = \mathcal{N}\left(\begin{bmatrix}0{,}42\\0{,}59\end{bmatrix} \middle| \mu_{\{y_1,y_2\}/h=B}, \Sigma_{\{y_1,y_2\}/h=B}\right)$,

. Using the calculations of the previous and this exercises:

$$\left|\Sigma_{\{y_1,y_2\}/h=B}\right| = 6{,}2559 \times 10^{-4} \quad \Sigma^{-1}_{\{y_1,y_2\}/h=B} = \begin{bmatrix}50{,}3367 & 15{,}5981\\ 15{,}5981 & 36{,}5896\end{bmatrix}$$

→

→ . Using scipy.stats.multivariate_normal:

$\rho(y_1=0{,}42;\ y_2=0{,}59 \mid B) = \mathcal{N}\left(\begin{bmatrix}0{,}42\\0{,}59\end{bmatrix} \middle| \begin{bmatrix}0{,}5425\\0{,}3275\end{bmatrix}, \begin{bmatrix}0{,}02289 & -0{,}009758\\-0{,}009758 & 0{,}03149\end{bmatrix}\right) \simeq 1{,}7286$,

therefore:

$\cdot\, \rho(y_1=0{,}42;\ y_2=0{,}59;\ y_3=0;\ y_4=1;\ y_5=1 \mid B) = 1{,}7286 \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{4}{7} \simeq \boxed{0{,}1235}$

. Therefore X_9 is classified as B.

c. [2v] Consider that the default decision threshold of $\theta = 0.5$ can be adjusted according to

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

Under a maximum likelihood assumption, what thresholds optimize testing accuracy?

. <u>Alínea C)</u>

X_8: $\cdot\, \rho(A \mid y_1=0{,}38;\ y_2=0{,}52;\ y_3=0;\ y_4=1;\ y_5=0) \underset{\text{Using Bayes Rule}}{=}$

$= \dfrac{\rho(y_1=0{,}38;\ y_2=0{,}52;\ y_3=0;\ y_4=1;\ y_5=0 \mid A)}{\rho(y_1=0{,}38;\ y_2=0{,}52;\ y_3=0;\ y_4=1;\ y_5=0)} =$

$= \dfrac{\rho(y_1=0{,}38;\ y_2=0{,}52 \mid A) \cdot \rho(y_3=0;\ y_4=1 \mid A) \cdot \rho(y_5=0 \mid A) \cdot \rho(A)}{\rho(y_1=0{,}38;\ y_2=0{,}52) \cdot \rho(y_3=0;\ y_4=1) \cdot \rho(y_5=0)}$,

. From the previous exercises:

$\cdot\, \rho(y_1=0{,}38;\ y_2=0{,}52 \mid A) = 0{,}985$, $\rho(y_3=0;\ y_4=1 \mid A) = \frac{1}{3}$, $\rho(y_5=0 \mid A) = \frac{1}{3}$, ~~and~~

$\rho(A) = \frac{3}{7}$, $\rho(y_3=0;\ y_4=1) = \frac{2}{7}$ and $\rho(y_5=0) = \frac{2}{7}$

$\cdot\, \rho(y_1=0{,}38;\ y_2=0{,}52) = \mathcal{N}\left(\begin{bmatrix}0{,}38\\0{,}52\end{bmatrix} \middle| \mu_{\{y_1,y_2\}}, \Sigma_{\{y_1,y_2\}}\right) =$

$= \mathcal{N}\left(\begin{bmatrix}0{,}38\\0{,}52\end{bmatrix} \middle| \begin{bmatrix}0{,}4414\\0{,}41\end{bmatrix}, \begin{bmatrix}0{,}04908 & -0{,}02107\\-0{,}02107 & 0{,}03753\end{bmatrix}\right)$,

→

→ . Using scipy.stats.multivariate_normal:

$\cdot\, \rho(y_1=0{,}38;\ y_2=0{,}52) = \mathcal{N}\left(\begin{bmatrix}0{,}38\\0{,}52\end{bmatrix} \middle| \begin{bmatrix}0{,}4414\\0{,}41\end{bmatrix}, \begin{bmatrix}0{,}04908 & -0{,}02107\\-0{,}02107 & 0{,}03753\end{bmatrix}\right) \simeq 3{,}6229$,

therefore:

$\cdot\, \rho(A \mid y_1=0{,}38;\ y_2=0{,}52;\ y_3=0;\ y_4=1;\ y_5=0) = \dfrac{0{,}985 \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{3}{7}}{3{,}6229 \cdot \frac{2}{7} \cdot \frac{2}{7}} \simeq \boxed{0{,}1584}$

X_9: $\cdot\, \rho(A \mid y_1=0{,}42;\ y_2=0{,}59;\ y_3=0;\ y_4=1;\ y_5=1) \underset{\text{Using Bayes Rule}}{=}$

$= \dfrac{\rho(y_1=0{,}42;\ y_2=0{,}59;\ y_3=0;\ y_4=1;\ y_5=1 \mid A)}{\rho(y_1=0{,}42;\ y_2=0{,}59;\ y_3=0;\ y_4=1;\ y_5=1)} =$

$= \dfrac{\rho(y_1=0{,}42;\ y_2=0{,}59 \mid A) \cdot \rho(y_3=0;\ y_4=1 \mid A) \cdot \rho(y_5=1 \mid A) \cdot \rho(A)}{\rho(y_1=0{,}42;\ y_2=0{,}59) \cdot \rho(y_3=0;\ y_4=1) \cdot \rho(y_5=1)}$,

. From the previous exercises:

$\rho(y_1=0{,}42;\ y_2=0{,}59 \mid A) = 0{,}4031$, $\rho(y_3=0;\ y_4=1 \mid A) = \frac{1}{3}$, $\rho(y_5=1 \mid A) = \frac{1}{3}$,

$\rho(A) = \frac{3}{7}$, $\rho(y_3=0;\ y_4=1) = \frac{2}{7}$ and $\rho(y_5=1) = \frac{3}{7}$

$\cdot\, \rho(y_1=0{,}42;\ y_2=0{,}59) = \mathcal{N}\left(\begin{bmatrix}0{,}42\\0{,}59\end{bmatrix} \middle| \mu_{\{y_1,y_2\}}, \Sigma_{\{y_1,y_2\}}\right) =$

$= \mathcal{N}\left(\begin{bmatrix}0{,}42\\0{,}59\end{bmatrix} \middle| \begin{bmatrix}0{,}4414\\0{,}41\end{bmatrix}, \begin{bmatrix}0{,}04908 & -0{,}02107\\-0{,}02107 & 0{,}03753\end{bmatrix}\right)$,

. Using scipy.stats.multivariate_normal:

$\cdot\, \rho(y_1=0{,}42;\ y_2=0{,}59) = \mathcal{N}\left(\begin{bmatrix}0{,}42\\0{,}59\end{bmatrix} \middle| \begin{bmatrix}0{,}4414\\0{,}41\end{bmatrix}, \begin{bmatrix}0{,}04908 & -0{,}02107\\-0{,}02107 & 0{,}03753\end{bmatrix}\right) \simeq 2{,}5387$,

therefore:

$\cdot\, \rho(A \mid y_1=0{,}42;\ y_2=0{,}59;\ y_3=0;\ y_4=1;\ y_5=1) = \dfrac{0{,}4031 \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{3}{7}}{2{,}5387 \cdot \frac{2}{7} \cdot \frac{3}{7}} \simeq \boxed{0{,}0617}$

Aprendizagem 2023/24

Homework II

→ Conclusion:

. The thresholds that maximize test accuracy are the thresholds in the range [0,0617; 0,1584[.

2. Let y_1 be the target numeric variable, y_2-y_6 be the input variables where y_2 is binarized under an equal-width (equal-range) discretization. For the evaluation of regressors, consider a 3-fold cross-validation over the full dataset ($\mathbf{x}_1$- $\mathbf{x}_9$) without shuffling the observations.

   a. [1v] Identify the observations and features per data fold after the binarization procedure.

Exercício 2)

. Alínea a)

. Binarization of Y_2:

$.\min Y_2 = 0,11 \quad \max Y_2 = 0,72 \qquad \frac{0,11+0,72}{2} = 0,415, \; Y_2 = \begin{cases} 0, X_i < 0,415 \\ 1, X_i > 0,415 \end{cases}$

$Y_2 = \{0,1,1,0,0,0,1,1,1\}$

. On Fold1 the training data subset is from X_1 to X_6 and the testing data subset is from X_7 to X_9;

. On Fold2 the training data subset is from X_1 to X_3 and X_7 to X_9. The testing data subset is from X_4 to X_6;

. On Fold3 the training data subset is from X_4 to X_9 and the testing data subset is from X_1 to X_3.

. Therefore:

D	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Fold1	Fold2	Fold3
X_1	0,24	0	1	1	0	A	Training data Subset	Training data Subset	Testing data Subset
X_2	0,16	1	1	0	1	A			
X_3	0,32	1	0	1	2	A			
X_4	0,54	0	0	0	1	B		Testing data Subset	Training data Subset
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	Testing data Subset	Training data Subset	
X_8	0,38	1	0	1	0	A			
X_9	0,42	1	0	1	1	B			

. The features are the input variables. Therefore, features are from Y_2 to Y_6.

b. [4v] Consider a distance-weighted kNN with $k = 3$, Hamming distance (d), and $1/d$ weighting. Compute the MAE of this kNN regressor for the 1st iteration of the cross-validation (i.e. train observations have the lower indices).

Alínea b)

· If we want to calculate the MAE of this KNN regressor for the 1st iteration of the cross-validation then, our training data subset is from X_1 to X_6 and our testing data subset is from X_7 to X_9.

$$H(X, Y) = \sum_{i=1}^{n} d(x_i, y_i)$$

X_7:

$H(X_7, X_i)$	X_1	↓ X_2	↓ X_3	↓ X_4	X_5	X_6
X_7	4	3	2	2	3	4

$d(X_7, X_1) = d(1,0) + d(0,1) + d(1,1) + d(1,0) + d(B,A) = 4$

$d(X_7, X_2) = d(1,1) + d(0,1) + d(1,0) + d(1,1) + d(B,A) = 3$

$d(X_7, X_3) = d(1,1) + d(0,0) + d(1,1) + d(1,2) + d(B,A) = 2$

$d(X_7, X_4) = d(1,0) + d(0,0) + d(1,0) + d(1,1) + d(B,B) = 2$

$d(X_7, X_5) = d(1,0) + d(0,0) + d(1,0) + d(1,0) + d(B,B) = 3$

$d(X_7, X_6) = d(1,0) + d(0,1) + d(1,0) + d(1,2) + d(B,B) = 4$

→ X_8:

$H(X_8, X_i)$	↓ X_1	↓ X_2	↓ X_3	X_4	X_5	X_6
X_8	2	3	1	4	3	5

$d(X_8, X_1) = d(1,0) + d(0,1) + d(1,1) + d(0,0) + d(A,A) = 2$

$d(X_8, X_2) = d(1,1) + d(0,1) + d(1,0) + d(0,1) + d(A,A) = 3$

$d(X_8, X_3) = d(1,1) + d(0,0) + d(1,1) + d(0,2) + d(A,A) = 1$

$d(X_8, X_4) = d(1,0) + d(0,0) + d(1,0) + d(0,1) + d(A,B) = 4$

$d(X_8, X_5) = d(1,0) + d(0,0) + d(1,0) + d(0,0) + d(A,B) = 3$

$d(X_8, X_6) = d(1,0) + d(0,1) + d(1,0) + d(0,2) + d(A,B) = 5$

X_9:

$H(X_9, X_i)$	X_1	↓ X_2	↓ X_3	↓ X_4	X_5	X_6
X_9	4	3	2	2	3	4

$d(X_9, X_1) = d(1,0) + d(0,1) + d(1,1) + d(1,0) + d(B,A) = 4$

$d(X_9, X_2) = d(1,1) + d(0,1) + d(1,0) + d(1,1) + d(B,A) = 3$

$d(X_9, X_3) = d(1,1) + d(0,0) + d(1,1) + d(1,2) + d(B,A) = 2$

$d(X_9, X_4) = d(1,0) + d(0,0) + d(1,0) + d(1,1) + d(B,B) = 2$

$d(X_9, X_5) = d(1,0) + d(0,0) + d(1,0) + d(1,0) + d(B,B) = 3$

$d(X_9, X_6) = d(1,0) + d(0,1) + d(1,0) + d(1,2) + d(B,B) = 4$

· For X_7 and X_9, the neighbors X_2, X_3 and X_4 were chosen. In both observations, the distance from the neighbor X_5 is equal to the distance of X_2, so a random choice was made.

For X_8, the neighbors X_1, X_2 and X_3 were chosen. In this observation, the distance from the neighbor X_5 is equal to the distance of X_2, so a random choice was also made.

→

→

· $\hat{z}_1(X_7) = \text{weighted-mean}\left(\frac{1}{3} \cdot 0{,}16; \frac{1}{2} \cdot 0{,}32; \frac{1}{2} \cdot 0{,}54\right) =$

$$= \frac{\frac{1}{3} \cdot 0{,}16 + \frac{1}{2} \cdot 0{,}32 + \frac{1}{2} \cdot 0{,}54}{\frac{1}{3} + \frac{1}{2} + \frac{1}{2}} \simeq 0{,}36$$

· $\hat{z}_1(X_8) = \text{weighted-mean}\left(\frac{1}{2} \cdot 0{,}24; \frac{1}{3} \cdot 0{,}16; \frac{1}{1} \cdot 0{,}32\right) =$

$$= \frac{\frac{1}{2} \cdot 0{,}24 + \frac{1}{3} \cdot 0{,}16 + \frac{1}{1} \cdot 0{,}32}{\frac{1}{2} + \frac{1}{3} + \frac{1}{1}} \simeq 0{,}269$$

· $\hat{z}_1(X_9) = \text{weighted-mean}\left(\frac{1}{3} \cdot 0{,}16; \frac{1}{2} \cdot 0{,}32; \frac{1}{2} \cdot 0{,}54\right) =$

$$= \frac{\frac{1}{3} \cdot 0{,}16 + \frac{1}{2} \cdot 0{,}32 + \frac{1}{2} \cdot 0{,}54}{\frac{1}{3} + \frac{1}{2} + \frac{1}{2}} \simeq 0{,}36, \text{ therefore:}$$

z_1	0,41	0,38	0,42
$\hat{z}_1$	0,36	0,269	0,36

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |z_i - \hat{z}_i|, \text{ therefore:}$$

$$MAE = \frac{1}{3}\left(|0{,}41 - 0{,}36| + |0{,}38 - 0{,}269| + |0{,}42 - 0{,}36|\right) \simeq \boxed{0{,}0737}$$

Aprendizagem 2023/24
Homework II

I. Programming and critical analysis [7v]

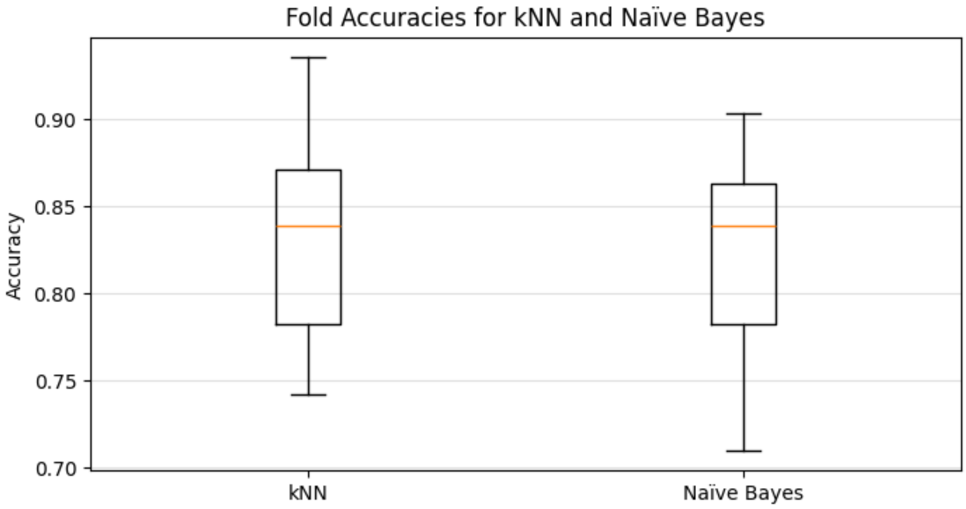
Considering the *column_diagnosis.arff* dataset available at the course webpage's homework tab. Using *sklearn*, apply a 10-fold stratified cross-validation with shuffling (random_state=0) for the assessment of predictive models along this section.

1) [3v] Compare the performance of kNN with $k = 5$ and naïve Bayes with Gaussian assumption (consider all remaining parameters for each classifier as *sklearn's* default):

```
import pandas as pd
from scipy.io.arff import loadarff
from sklearn.model_selection import StratifiedKFold
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score

# Read 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']

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

knn_accuracies, nb_accuracies = [], []
knn_classifier = KNeighborsClassifier(n_neighbors=5)
nb_classifier = GaussianNB()

# Iterate through each fold of stratified cross-validation
for train_index, test_index in stratified_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]

    # Train both classifiers
    knn_classifier.fit(X_train, y_train)
    nb_classifier.fit(X_train, y_train)

    # Predict and evaluate kNN
    knn_pred = knn_classifier.predict(X_test)
    knn_accuracy = accuracy_score(y_test, knn_pred)
    knn_accuracies.append(knn_accuracy)

    # Predict and evaluate Gaussian Naïve Bayes
    nb_pred = nb_classifier.predict(X_test)
    nb_accuracy = accuracy_score(y_test, nb_pred)
    nb_accuracies.append(nb_accuracy)
```

a. Plot two boxplots with the fold accuracies for each classifier.

```
import matplotlib.pyplot as plt

plt.figure(figsize=(8, 4))
plt.boxplot([knn_accuracies, nb_accuracies], labels=['kNN', 'Naïve Bayes'])
plt.title('Fold Accuracies for kNN and Naïve Bayes')
plt.ylabel('Accuracy')
plt.grid(axis='y', alpha=0.4)
plt.show()
```



b. Using *scipy*, test the hypothesis "kNN is statistically superior to naïve Bayes regarding accuracy", asserting whether is true.

```
from scipy import stats

# H0: kNN is statistically equal to Naïve Bayes regarding accuracy
# H1: kNN is statistically superior to Naïve Bayes regarding accuracy
t_statistic, p_value = stats.ttest_rel(knn_accuracies, nb_accuracies, alternative='greater')

alpha = 0.05  # Significance level
if p_value < alpha:
    print('Para níveis de signicância até 0.05, a hipótese nula (kNN é estatisticamente igual a Naïve Bayes em termos de precisão)'
          '\né rejeitada, ou seja, "kNN is statistically superior to Naïve Bayes regarding accuracy" confirma-se.')
else:
    print('Para níveis de signicância até 0.05, a hipótese nula (kNN é estatisticamente igual a Naïve Bayes em termos de precisão)',
          '\nnão pode ser rejeitada, ou seja, "kNN is statistically superior to Naïve Bayes regarding accuracy" é falso.')
```

Output:
Para níveis de signicância até 0.05, a hipótese nula (kNN é estatisticamente igual a Naïve Bayes em termos de precisão) não pode ser rejeitada, ou seja, "kNN is statistically superior to Naïve Bayes regarding accuracy" é falso.

Aprendizagem 2023/24
Homework II

2) [2.5v] Consider two kNN predictors with $k = 1$ and $k = 5$ (uniform weights, Euclidean distance, all remaining parameters as default). Plot the differences between the two cumulative confusion matrices of the predictors. Comment.

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

# Initialize confusion matrices
knn1_cumulative = np.array([[0,0,0], [0,0,0], [0,0,0]])
knn5_cumulative = np.array([[0,0,0], [0,0,0], [0,0,0]])
classes = ["Hernia", "Normal", "Spondylolisthesis"]

knn1 = KNeighborsClassifier(n_neighbors=1, weights='uniform', metric='euclidean')
knn5 = KNeighborsClassifier(n_neighbors=5, weights='uniform', metric='euclidean')

# Iterate through each fold of stratified cross-validation
for train_index, test_index in stratified_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]

    # Train both classifiers
    knn1.fit(X_train, y_train)
    knn5.fit(X_train, y_train)

    # Make predictions and calculate confusion matrix for kNN1
    knn1_pred = knn1.predict(X_test)
    knn1_cm = confusion_matrix(y_test, knn1_pred, labels=classes)
    knn1_cumulative += np.array(knn1_cm)

    # Make predictions and calculate confusion matrix for kNN5
    knn5_pred = knn5.predict(X_test)
    knn5_cm = confusion_matrix(y_test, knn5_pred, labels=classes)
    knn5_cumulative += np.array(knn5_cm)

# Calculate the difference between cumulative confusion matrices
difference_matrix = knn1_cumulative - knn5_cumulative

print("Difference between cumulative confusion matrices (kNN1 - kNN5):\n\n",
      pd.DataFrame(difference_matrix, index=classes, columns=classes),
      "\n\nNote:\tColumns are predicted values\n\tLines are test values")
```

Output:

```
Difference between cumulative confusion matrices (kNN1 - kNN5):

                   Hernia  Normal  Spondylolisthesis
Hernia                 -2       2                  0
Normal                 -5       2                  3
Spondylolisthesis       0       1                 -1

Note: Columns are predicted values
       Lines are test values
```

Comentário sobre os resultados obtidos:

- Para a classe 'Hernia', o modelo com k=5 previu corretamente mais 2 instâncias, enquanto o modelo com k=1 previu incorretamente mais 2 instâncias de 'Hernia' como 'Normal'.
- Para a classe 'Normal', o modelo com k=1 previu corretamente mais 2 instâncias e previu incorretamente mais 3 instâncias de 'Normal' como 'Spondylolisthesis', no entanto, o modelo com k=5 previu incorretamente mais 5 instâncias de 'Normal' como 'Hernia'.
- Para a classe 'Spondylolisthesis', o modelo com k=5 previu corretamente mais 1 instância, enquanto o modelo com k=1 previu incorretamente mais 1 instância de 'Spondylolisthesis' como 'Normal'.

Assim, pode-se concluir que o modelo com k=5 teve melhor performance para as classes 'Hernia' e 'Spondylolisthesis', enquanto o modelo com k=1 teve melhor performance para a classe 'Normal'.

3) [1.5v] Considering the unique properties of *column_diagnosis*, identify three possible difficulties of naïve Bayes when learning from the given dataset.

- Assume independência condicional entre as variáveis. Essa suposição pode ser muito restritiva, pois nem sempre é verdadeira.

- Assume que os dados seguem uma distribuição gaussiana (normal), o que pode não ser apropriado caso os dados não obedeçam a essa distribuição.

- O uso de um conjunto de observações limitado (pode levar a estimativas inadequadas ou probabilidades nulas).

END