Aprendizagem 2022/23

Homework IV – Group 041

I. Pen-and-paper

1)

30 de outubro de 2022 18:21

1- $\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$

	x_1	x_2
$\underline{x}_1$	1	2
$\underline{x}_2$	-1	1
$\underline{x}_3$	1	0

$\mu_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ $\quad \Sigma_1 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

$\mu_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\quad \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

Priors

$\boxed{\pi_1 = 0{,}5 \quad \pi_2 = 0{,}5}$

$|\Sigma_1| = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 2\times 2 - 1\times 1 = 3$ $\quad \Sigma_1^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$

$|\Sigma_2| = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 2\times 2 = 4$ $\quad \Sigma_2^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$

$K = 2$

E - Step $\quad$ Expectation

Likelihood

$$P(x_n \mid C_k = 1) = \mathcal{N}(x_n \mid \mu_k, \Sigma_k) = \frac{1}{(2.\pi)^{D/2}} \cdot \frac{1}{|\Sigma_k|^{1/2}} \cdot \exp\left(-\frac{1}{2} \cdot (x_n - \mu_k)^T \, \Sigma_k^{-1} \circ (x_n - \mu_k)\right)$$

$$P(x_1 \mid C_1 = 1) = \frac{1}{(2\pi)^{2/2}} \cdot \frac{1}{3^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (1-2,\ 2-2) \cdot \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \cdot \begin{bmatrix} 1-2 \\ 2-2 \end{bmatrix}\right)$$

$$= \frac{1}{2\pi} \times \frac{1}{\sqrt{3}} \times e^{\left(\left[\frac{1}{2},\ 0\right] \cdot \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 0 \end{bmatrix}\right)} =$$

$$= \frac{1}{2\pi\sqrt{3}} e^{\left(\left[\frac{1}{3},\ -\frac{1}{6}\right] \cdot \begin{bmatrix} -1 \\ 0 \end{bmatrix}\right)} = \frac{1}{2\pi\sqrt{3}} e^{-\frac{1}{3}} = 0{,}0658$$

$$P(x_1 \mid C_2 = 1) = \frac{1}{2\pi} \times \frac{1}{\sqrt{4}} \times e^{\left(-\frac{1}{2} \times (1-0,\ 2-0) \circ \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} 1-0 \\ 2-0 \end{pmatrix}\right)} =$$

$$= \frac{1}{4\pi} \times e^{\left(\left(-\frac{1}{2},\ -1\right) \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right)} =$$

$$= \frac{1}{4\pi} e^{\left(-\frac{1}{4}\ \ -\frac{1}{2}\right) \begin{pmatrix} 1 \\ 2 \end{pmatrix}} = \frac{1}{4\pi} e^{-\frac{5}{4}} = 0{,}02279$$

Aprendizagem 2022/23

Homework IV – Group 041

$$P(x_2 \mid c_1=1) = \frac{1}{2\pi} \times \frac{1}{\sqrt{3}} \times e^{\left(-\frac{1}{2} \times (-1-2,\ 1-2) \times \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \times \begin{pmatrix} -1-2 \\ 1-2 \end{pmatrix}\right)} =$$

$$= \frac{1}{2\pi\sqrt{3}} \times e^{\left(-\frac{1}{2} \times (-3,\ -1) \times \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} -3 \\ -1 \end{bmatrix}\right)}$$

$$= \frac{1}{2\pi\sqrt{3}} e^{\left(\left(+\frac{3}{2},\ +\frac{1}{2}\right) \times \begin{bmatrix} -\frac{5}{3} \\ \frac{1}{3} \end{bmatrix}\right)} = \frac{1}{2\pi\sqrt{3}} \times e^{-\frac{7}{3}} = 0{,}00891$$

$$P(x_2 \mid c_2=1) = \frac{1}{2\pi} \times \frac{1}{\sqrt{4}} \times e^{\left(-\frac{1}{2} \times (-1-0,\ 1-0) \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} -1-0 \\ 1-0 \end{bmatrix}\right)} =$$

$$= \frac{1}{4\pi} e^{\left(\left(\frac{1}{2},\ -\frac{1}{2}\right) \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right)} = \frac{1}{4\pi} \times e^{\left(\left(\frac{1}{2},\ -\frac{1}{2}\right) \times \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}\right)} = \frac{1}{4\pi} \times e^{-\frac{1}{2}} = 0{,}04827$$

$$P(x_3 \mid c_1=1) = \frac{1}{2\pi} \times \frac{1}{\sqrt{3}} \times e^{\left(-\frac{1}{2} \times (1-2,\ 0-2) \times \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \times \begin{bmatrix} 1-2 \\ 0-2 \end{bmatrix}\right)} =$$

$$= \frac{1}{2\pi\sqrt{3}} e^{\left(\left(\frac{1}{2},\ 1\right) \times \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \times \begin{bmatrix} -1 \\ -2 \end{bmatrix}\right)} = \frac{1}{2\pi\sqrt{3}} e^{\left(\left(\frac{1}{2},\ 1\right) \times \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right)} = \frac{1}{2\pi\sqrt{3}} e^{-1} = 0{,}09189$$

$$P(x_3 \mid c_2=1) = \frac{1}{2\pi} \times \frac{1}{\sqrt{4}} \times e^{\left(-\frac{1}{2} \times (1-0,\ 0-0) \times \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \times \begin{bmatrix} 1-0 \\ 0-0 \end{bmatrix}\right)} = \frac{1}{4\pi} e^{\left(\left(-\frac{1}{2},\ 0\right) \times \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)} = \frac{1}{4\pi} e^{\left(\left(-\frac{1}{4},\ 0\right) \times \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)} = \frac{1}{4\pi} e^{-\frac{1}{4}} = 0{,}06198$$

Então probabilidade conjunta,

$$P(c_k=1,\ x_n) = \pi_k \times N(x_n \mid \mu_k, \Sigma_k)$$

$$P(c_1=1, x_1) = \pi_1 \times P(x_1 \mid c_1=1) = 0{,}5 \times 0{,}0658 = 0{,}0329$$

$$P(c_1=1, x_2) = \pi_1 \times P(x_2 \mid c_1=1) = 0{,}5 \times 0{,}00891 = 4{,}455 \times 10^{-3}$$

$$P(c_1=1, x_3) = \pi_1 \times P(x_3 \mid c_1=1) = 0{,}5 \times 0{,}09189 = 0{,}045945$$

$$P(c_2=1, x_1) = \pi_2 \times P(x_1 \mid c_2=1) = 0{,}5 \times 0{,}02279 = 0{,}011395$$

$$P(c_2=1, x_2) = \pi_2 \times P(x_2 \mid c_2=1) = 0{,}5 \times 0{,}04827 = 0{,}024135$$

$$P(c_2=1, x_3) = \pi_2 \times P(x_3 \mid c_2=1) = 0{,}5 \times 0{,}06198 = 0{,}03099$$

TÉCNICO LISBOA

Aprendizagem 2022/23

Homework IV – Group 041

$$P(x_m) = \sum_{k=1}^{K} P(c_k=1, x_m) = \sum_{k=1}^{K} \pi_k \times N(x_m \mid \mu_k, \Sigma_k)$$

$$P(x_1) = \pi_1 \times P(x_1 \mid c_1=1) + \pi_2 \times P(x_1 \mid c_2=1) =$$
$$= 0{,}0329 + 0{,}011335 = 0{,}044235$$

$$P(x_2) = \pi_1 \times P(x_2 \mid c_1=1) + \pi_2 \, P(x_2 \mid c_2=1)$$
$$= 4{,}455 \times 10^{-3} + 0{,}024135$$
$$= 0{,}02859$$

$$P(x_3) = \pi_1 \times P(x_3 \mid c_1=1) + \pi_2 \times P(x_3 \mid c_2=1) =$$
$$= 0{,}045345 + 0{,}03099 = 0{,}076335$$

Usando a normalização de Bayes

$$\gamma(c_{mk}) = P(c_k=1 \mid x_m) = \frac{P(c_k=1, x_m)}{P(x_m)}$$

$$\gamma(c_{11}) = P(c_1=1 \mid x_1) = \frac{P(c_1=1, x_1)}{P(x_1)} = \frac{0{,}0329}{0{,}044235} = 0{,}74275$$

$$\gamma(c_{12}) = P(c_2=1 \mid x_1) = \frac{P(c_2=1, x_1)}{P(x_1)} = \frac{0{,}011335}{0{,}044235} = 0{,}25725$$

$$\gamma(c_{21}) = P(c_1=1 \mid x_2) = \frac{P(c_1=1, x_2)}{P(x_2)} = \frac{4{,}455 \times 10^{-3}}{0{,}02859} = 0{,}15582$$

$$\gamma(c_{22}) = P(c_2=1 \mid x_2) = \frac{P(c_2=1, x_2)}{P(x_2)} = \frac{0{,}024135}{0{,}02859} = 0{,}8442$$

TÉCNICO LISBOA

Aprendizagem 2022/23

Homework IV – Group 041

$$\gamma(c_{31}) = P(c_1 = 1 \mid x_3) = \frac{P(c_1 = 1, x_3)}{P(x_3)} = \frac{0{,}045945}{0{,}076935} = 0{,}59719$$

$$\gamma(c_{32}) = P(c_2 = 1 \mid x_3) = \frac{P(c_2 = 1, x_3)}{P(x_3)} = \frac{0{,}03089}{0{,}076935} = 0{,}40281$$

M - step Maximização

Atualizamos:

$$N_k = \sum_{n=1}^{N} \gamma(c_{nk})$$

$$N_1 = \gamma(c_{11}) + \gamma(c_{21}) + \gamma(c_{31}) = 0{,}74275 + 0{,}15582 + 0{,}59719 = 1{,}49576$$

$$N_2 = \gamma(c_{12}) + \gamma(c_{22}) + \gamma(c_{32}) = 0{,}25725 + 0{,}8442 + 0{,}40281 = 1{,}50426$$

Determinar a Média de valores (Centróides)

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(c_{nk}) \cdot x_n$$

$$\mu_1 = \frac{1}{1{,}49576} \cdot \left(0{,}74275 \times \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 0{,}15582 \times \begin{pmatrix} -1 \\ 1 \end{pmatrix} + 0{,}59719 \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) =$$

$$= \frac{1}{1{,}49576} \cdot \left(\begin{pmatrix} 0{,}74275 - 0{,}15582 + 0{,}59719 \\ 1{,}4855 + 0{,}15582 + 0 \end{pmatrix} \right) =$$

$$= \frac{1}{1{,}49576} \cdot \begin{pmatrix} 1{,}18412 \\ 1{,}64132 \end{pmatrix} = \begin{pmatrix} 0{,}79165 \\ 1{,}09732 \end{pmatrix}$$

Aprendizagem 2022/23

Homework IV – Group 041

$$\mu_2 = \frac{1}{1{,}50426} \times \left(0{,}25725 \times \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 0{,}8442 \times \begin{pmatrix} -1 \\ 1 \end{pmatrix} + 0{,}40281 \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)$$

$$= \frac{1}{1{,}50426} \times \begin{pmatrix} 0{,}25725 - 0{,}8442 + 0{,}40281 \\ 0{,}5145 + 0{,}8442 + 0 \end{pmatrix} = \frac{1}{1{,}50426} \begin{pmatrix} -0{,}18414 \\ 1{,}3587 \end{pmatrix} = \begin{pmatrix} -0{,}12241 \\ 0{,}90323 \end{pmatrix}$$

determinar a nova matriz de covariância

$$\Sigma_k = \frac{1}{N_k} \cdot \sum_{n=1}^{N} \gamma(c_{nk}) \cdot (x_n - \mu_k) \cdot (x_n - \mu_k)^T$$

$$\Sigma_1 = \frac{1}{1{,}43576} \times \left(0{,}74275 \cdot \begin{pmatrix} 1 - 0{,}79165 \\ 2 - 1{,}09732 \end{pmatrix} \cdot \begin{pmatrix} 1 - 0{,}79165 & 2 - 1{,}09732 \end{pmatrix} + \right.$$

$$0{,}15582 \times \begin{pmatrix} -1 - 0{,}79165 \\ 1 - 1{,}09732 \end{pmatrix} \begin{pmatrix} -1 - 0{,}79165 & 1 - 1{,}09732 \end{pmatrix} +$$

$$\left. 0{,}59719 \times \begin{pmatrix} 1 - 0{,}79165 \\ 0 - 1{,}09732 \end{pmatrix} \cdot \begin{pmatrix} 1 - 0{,}79165 & 0 - 1{,}09732 \end{pmatrix} \right)$$

$$= \frac{1}{1{,}43576} \times \left(0{,}74275 \times \begin{pmatrix} 0{,}0434 & 0{,}18807 \\ 0{,}18807 & 0{,}81483 \end{pmatrix} + 0{,}15582 \times \begin{pmatrix} 3{,}21001 & 0{,}17436 \\ 0{,}17436 & 0{,}00947 \end{pmatrix} \right.$$

$$\left. + 0{,}59719 \times \begin{pmatrix} 0{,}04341 & -0{,}22863 \\ -0{,}22863 & 1{,}20411 \end{pmatrix} \right) =$$

$$= \frac{1}{1{,}43576} \times \left(\begin{pmatrix} 0{,}03224 & 0{,}13969 \\ 0{,}13969 & 0{,}60521 \end{pmatrix} + \begin{pmatrix} 0{,}5001 & 0{,}02717 \\ 0{,}02717 & 0{,}00148 \end{pmatrix} + \begin{pmatrix} 0{,}02592 & -0{,}13654 \\ -0{,}13654 & 0{,}71908 \end{pmatrix} \right)$$

$$= \frac{1}{1{,}43576} \begin{pmatrix} 0{,}55826 & 0{,}03032 \\ 0{,}03032 & 1{,}32577 \end{pmatrix} =$$

$$= \begin{pmatrix} 0{,}37323 & 0{,}02027 \\ 0{,}02027 & 0{,}88635 \end{pmatrix}$$

$$\Sigma_2 = \frac{1}{1{,}50426} \times \left(0{,}25725 \times \begin{pmatrix} 1+0{,}12241 \\ 2-0{,}90323 \end{pmatrix} \begin{pmatrix} 1+0{,}12241 & 2-0{,}90323 \end{pmatrix} + \right.$$

$$0{,}8442 \times \begin{pmatrix} -1+0{,}12241 \\ 1-0{,}90323 \end{pmatrix} \begin{pmatrix} -1+0{,}12241 & 1-0{,}90323 \end{pmatrix} +$$

$$\left. 0{,}40281 \times \begin{pmatrix} 1+0{,}12241 \\ 0-0{,}90323 \end{pmatrix} \begin{pmatrix} 1+0{,}12241 & 0-0{,}90323 \end{pmatrix} \right) =$$

$$= \frac{1}{1{,}50426} \times \left(0{,}25725 \times \begin{pmatrix} 1{,}25980 & 1{,}23103 \\ 1{,}23103 & 1{,}20290 \end{pmatrix} + 0{,}8442 \times \begin{pmatrix} 1{,}25880 & -0{,}10862 \\ -0{,}10862 & 0{,}00936 \end{pmatrix} \right.$$

$$\left. + 0{,}40281 \times \begin{pmatrix} 1{,}25980 & -1{,}01379 \\ -1{,}01379 & 0{,}81582 \end{pmatrix} \right) =$$

$$= \frac{1}{1{,}50426} \times \left(\begin{pmatrix} 0{,}32408 & 0{,}31668 \\ 0{,}31668 & 0{,}30945 \end{pmatrix} + \begin{pmatrix} 1{,}06352 & -0{,}0917 \\ -0{,}0917 & 0{,}00790 \end{pmatrix} + \begin{pmatrix} 0{,}50746 & -0{,}40836 \\ -0{,}40836 & 0{,}32862 \end{pmatrix} \right) =$$

$$= \frac{1}{1{,}50426} \times \begin{pmatrix} 1{,}89506 & -0{,}18338 \\ -0{,}18338 & 0{,}64597 \end{pmatrix} = \begin{pmatrix} 1{,}25979 & -0{,}12191 \\ -0{,}12191 & 0{,}42943 \end{pmatrix}$$

Novo parametro de mistura

$$\pi_k = P(C_k=1) = \frac{N_k}{N}$$

$$\pi_1 = P(C_1=1) = \frac{1{,}49571}{3} = 0{,}49859$$

$$\pi_2 = P(C_2=1) = \frac{1{,}50426}{3} = 0{,}50142$$

Aprendizagem 2022/23

Homework IV – Group 041

2) A)

1)

$$\Sigma_1 = \begin{pmatrix} 0,37323 & 0,02027 \\ 0,02027 & 0,88635 \end{pmatrix} \qquad \mu_1 = \begin{pmatrix} 0,78165 \\ 1,09732 \end{pmatrix} \qquad x_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad x_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\Sigma_2 = \begin{pmatrix} 1,25579 & -0,12191 \\ -0,12191 & 0,42943 \end{pmatrix} \qquad \mu_2 = \begin{pmatrix} -0,12241 \\ 0,90323 \end{pmatrix} \qquad x_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\pi_1 = 0,49858 \qquad |\Sigma_1| = 0,33040 \qquad \Sigma_1^{-1} = \begin{pmatrix} 2,68265 & -0,06135 \\ -0,06135 & 1,12963 \end{pmatrix}$$

$$\pi_2 = 0,50142 \qquad |\Sigma_2| = 0,52613 \qquad \Sigma_2^{-1} = \begin{pmatrix} 0,81621 & 0,23171 \\ 0,23171 & 2,39445 \end{pmatrix}$$

a) E-step

$$p(x_1 | c_1 = 1) = \frac{1}{2\pi \times \sqrt{0,33040}} \times e^{\left(-\frac{1}{2} \times (1 - 0,78165,\ 2 - 1,09732) \cdot \begin{pmatrix} 2,68265 & -0,06135 \\ -0,06135 & 1,12963 \end{pmatrix} \cdot \begin{pmatrix} 1 - 0,78165 \\ 2 - 1,09732 \end{pmatrix}\right)}$$

$$= \frac{1}{2\pi\sqrt{0,33040}} \times e^{\left(-\frac{1}{2} \times 1,0138\right)} = 0,16679$$

$$p(x_2 | c_1 = 1) = \frac{1}{2\pi \times \sqrt{0,33040}} \times e^{\left(-\frac{1}{2} \times (-1 - 0,78165,\ 1 - 1,09732) \cdot \begin{pmatrix} 2,68265 & -0,06135 \\ -0,06135 & 1,12963 \end{pmatrix} \cdot \begin{pmatrix} -1 - 0,78165 \\ 1 - 1,09732 \end{pmatrix}\right)}$$

$$= \frac{1}{2\pi\sqrt{0,33040}} \times e^{\left(-\frac{1}{2} \times 8,6006\right)} = 0,00376$$

$$P(x_3 | c_1 = 1) = \frac{1}{2\pi \times \sqrt{0,33040}} \times e^{\left(-\frac{1}{2} \times (1 - 0,78165,\ 0 - 1,09732) \cdot \begin{pmatrix} 2,68265 & -0,06135 \\ -0,06135 & 1,12963 \end{pmatrix} \cdot \begin{pmatrix} 1 - 0,78165 \\ 0 - 1,09732 \end{pmatrix}\right)}$$

$$= \frac{1}{2\pi \times \sqrt{0,33040}} \times e^{\left(-\frac{1}{2} \times 1,5047\right)} = 0,13048$$

$$P(x_1 | c_2 = 1) = \frac{1}{2\pi\sqrt{0,52613}} \times e^{\left(-\frac{1}{2} \times (1 + 0,12241,\ 2 - 0,90323) \cdot \begin{pmatrix} 0,81621 & 0,23171 \\ 0,23171 & 2,39445 \end{pmatrix} \cdot \begin{pmatrix} 1 + 0,12241 \\ 2 - 0,90323 \end{pmatrix}\right)}$$

$$= \frac{1}{2\pi\sqrt{0,52613}} \times e^{\left(-\frac{1}{2} \times 4,4790\right)} = 0,02337$$

$$P(x_2 \mid C_2=1) = \frac{1}{2\pi\sqrt{0,52613}} \times e^{\left(-\frac{1}{2} \times (-1+0,12241,\ 1-0,90323) \cdot \begin{pmatrix} 0,81621 & 0,23171 \\ 0,23171 & 2,39445 \end{pmatrix} \cdot \begin{pmatrix} -1+0,12241 \\ 1-0,90323 \end{pmatrix}\right)}$$

$$= \frac{1}{2\pi\sqrt{0,52613}} \times e^{\left(-\frac{1}{2} \times 0,61768\right)} = 0,16160$$

$$P(x_3 \mid C_2=1) = \frac{1}{2\pi\sqrt{0,52613}} \times e^{\left(-\frac{1}{2} \times (1+0,12241,\ 0-0,90323) \cdot \begin{pmatrix} 0,81621 & 0,23171 \\ 0,23171 & 2,39445 \end{pmatrix} \cdot \begin{pmatrix} 1+0,12241 \\ 0-0,90323 \end{pmatrix}\right)}$$

$$= \frac{1}{2\pi\sqrt{0,52613}} \times e^{\left(-\frac{1}{2} \times 2,5119\right)} = 0,06249$$

Então:

$$P(C_1=1, x_1) = 0,49859 \times 0,16679 = 0,08316$$

$$P(C_2=1, x_1) = 0,50142 \times 0,02337 = 0,01172$$

$$P(C_1=1, x_2) = 0,49859 \times 0,00376 = 0,00188$$

$$P(C_2=1, x_2) = 0,50142 \times 0,016160 = 0,00810$$

$$P(C_1=1, x_3) = 0,49859 \times 0,13048 = 0,065056$$

$$P(C_2=1, x_3) = 0,50142 \times 0,06248 = 0,03133$$

$$P(x_1) = 0,08316 + 0,01172 = 0,09488$$

$$P(x_2) = 0,00188 + 0,00810 = 0,00998$$

$$P(x_3) = 0,065056 + 0,03133 = 0,09639$$

Normalizando usando Bayes

MAP Assumption

$$P(C_1=1 \mid x_1) = \frac{P(C_1=1, x_1)}{P(x_1)} = \frac{0,08316}{0,09488} = 0,87648$$

$$P(C_2=1 \mid x_1) = \frac{P(C_2=1, x_1)}{P(x_1)} = \frac{0,01172}{0,09488} = 0,12352$$

$\rightarrow x_1 \in C_1$

$$P(C_1=1 \mid x_2) = \frac{P(C_1=1, x_2)}{P(x_2)} = \frac{0,00188}{0,00998} = 0,18838$$

$$P(C_2=1 \mid x_2) = \frac{P(C_2=1, x_2)}{P(x_2)} = \frac{0,00810}{0,00998} = 0,81162$$

$\rightarrow x_2 \in C_2$

Aprendizagem 2022/23
Homework IV – Group 041

$$P(C_1=1 \mid x_3) = \frac{P(C_1=1, x_3)}{P(x_3)} = \frac{0,065056}{0,09639} = 0,674925$$

$$P(C_2=1 \mid x_3) = \frac{P(C_2=1, x_3)}{P(x_3)} = \frac{0,03133}{0,09639} = 0,325034$$

$\Big\} \; x_3 \in C_1$

R: $x_1, x_3 \in C_1$ e $x_2 \in C_2$

b)

b)

Calcular a silhouette do cluster c_1 usando distâncias euclidianas

$x_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ $\quad x_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\quad$ distância euclidiana $\rightarrow$ $d(\underbrace{p}_{p(p_1,p_2)}, \underbrace{q}_{q(q_1,q_2)}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2}$

$x_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

Silhouette

$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & a(i) < b(i) \\ 0, & a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1, & a(i) > b(i) \end{cases}$$

$$a(i) = \frac{1}{|C_I| - 1} \times \sum_{j \in C_I, j \neq i} d(i,j)$$

$$b(i) = \min_{J \neq I} \frac{1}{|C_J|} \sum_{j \in C_J} d(i,j)$$

Para x_1:

$d(x_1, x_3) = \sqrt{(1-1)^2 + (0-2)^2} = 2$

$d(x_1, x_2) = \sqrt{(-1-1)^2 + (1-2)^2} = \sqrt{5}$

$a(i) = \frac{1}{1} \times d(x_1, x_3) = 2$

$b(i) = \frac{1}{1} \times \sqrt{5} = \sqrt{5}$

como $a(i) < b(i)$

$S(x_1) = 1 - \frac{a(i)}{b(i)} = 1 - \frac{2}{\sqrt{5}} = 0,10557$

Para x_3:

$d(x_3, x_1) = \sqrt{(1-1)^2 + (2-0)^2} = 2$

$d(x_3, x_2) = \sqrt{(-1-1)^2 + (1-0)^2} = \sqrt{5}$

$a(i) = \frac{1}{1} \times d(x_3, x_1) = 2$

$b(i) = \frac{1}{1} \times d(x_3, x_2) = \sqrt{5}$

como $a(i) < b(i)$

$S(x_3) = 1 - \frac{a(i)}{b(i)} = 1 - \frac{2}{\sqrt{5}} = 0,10557$

$$S(C_1) = \frac{S(x_1) + S(x_3)}{2} = \underline{\underline{0,10557}}$$

Aprendizagem 2022/23

Homework IV – Group 041

II. Programming and critical analysis

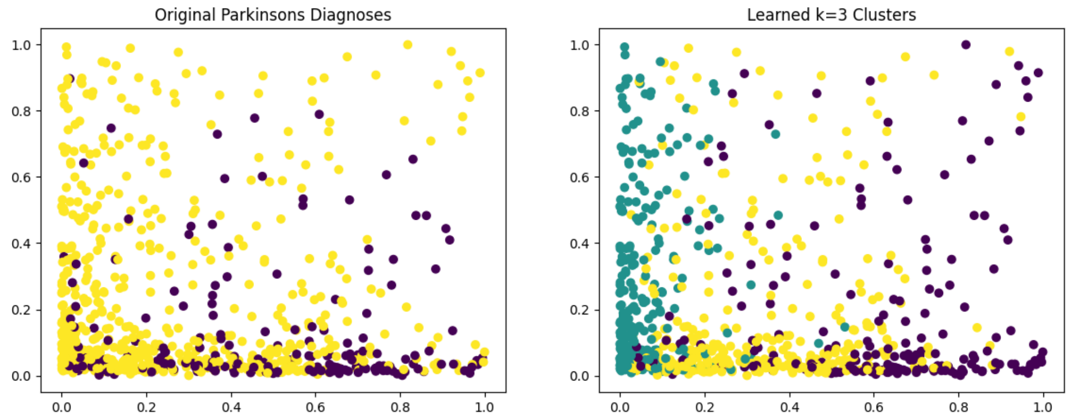
1)

```
[0] Silhouette (euclidian): 0.1136202757517943
[1] Silhouette (euclidian): 0.11403554201377072
[2] Silhouette (euclidian): 0.1136202757517943
[0] Purity: 0.7671957671957672
[1] Purity: 0.7632275132275133
[2] Purity: 0.7671957671957672
```

2)

O que está a causar o não determinismo é facto de estarmos inicialmente a considerar que os centroides das respetivas três clusters são completamente aleatórios e usando a distancia da euclidiana dos pontos a esses centroides associamos os pontos aos clusters dos respetivos centroides e caso o centroide altere quando os pontos estão associados aos seus respetivos clusters, então voltamos a recalcular os centroides até que o calculo do novo centroide resulte nos centroides usados no calculo dos mesmos.

3)



4)

É necessário 31 componentes principais para explicar mais de 80% de variabilidade

III. APPENDIX

```
import fractions
import pandas as pd
import numpy as np
```

```
#*##################################################################################
#
#* 1)
#*##################################################################################
#

#* import data
from scipy.io.arff import loadarff
data = loadarff("pd_speech.arff")
df = pd.DataFrame(data[0])
df['class'] = df['class'].str.decode('utf-8')
df['class'] = pd.to_numeric(df["class"])

#* aux variable
num_columns = df.shape[1]

#* pre-process data
from sklearn.preprocessing import MinMaxScaler
df_scaled = df.copy()
df_scaled.iloc[:, 0:num_columns-1] = MinMaxScaler().fit_transform(df.iloc[:,
0:num_columns-1])

#* partition data
X, y = df_scaled.iloc[:, 0:num_columns-1], df_scaled["class"]

#* parameterize clustering
from sklearn import cluster
kmeans_algo_0 = cluster.KMeans(n_clusters=3, random_state=0)
kmeans_algo_1 = cluster.KMeans(n_clusters=3, random_state=1)
kmeans_algo_2 = cluster.KMeans(n_clusters=3, random_state=2)

#* learn the model
kmeans_model_0 = kmeans_algo_0.fit(X)
kmeans_model_1 = kmeans_algo_1.fit(X)
kmeans_model_2 = kmeans_algo_2.fit(X)

#* produced clusters
y_pred_0 = kmeans_model_0.labels_
y_pred_1 = kmeans_model_1.labels_
y_pred_2 = kmeans_model_2.labels_

#* compute Silhouette
from sklearn import metrics
print("[0] Silhouette (euclidian):", metrics.silhouette_score(X, y_pred_0,
metric='euclidean'))
print("[1] Silhouette (euclidian):", metrics.silhouette_score(X, y_pred_1,
metric='euclidean'))
```

```
print("[2] Silhouette (euclidian):", metrics.silhouette_score(X, y_pred_2,
metric='euclidean'))

#/ compute Purity
import numpy as np
def purity_score(y_true, y_pred):
#/compute contingency/confusion matrix
    confusion_matrix = metrics.cluster.contingency_matrix(y_true, y_pred)
    return np.sum(np.amax(confusion_matrix, axis=0)) / np.sum(confusion_matrix)
y_true = y
print("[0] Purity:", purity_score(y_true, y_pred_0))
print("[1] Purity:", purity_score(y_true, y_pred_1))
print("[2] Purity:", purity_score(y_true, y_pred_2),"\n")

#*####################################################################################
#* 3)
#*####################################################################################
#

#* compute features' variances
from sklearn.feature_selection import VarianceThreshold
selection = VarianceThreshold().fit(X)

#* get second max variance
import heapq
max_three_variances = heapq.nlargest(3, selection.variances_)
third_max_variance = max_three_variances[2]

#* feature selection
X_new = VarianceThreshold(threshold=third_max_variance).fit_transform(X)

#* plot
import matplotlib.pyplot as plt
plt.figure(figsize=(14, 5))
plt.subplot(121)
plt.title(label="Original Parkinsons Diagnoses")
plt.scatter(X_new[:,0], X_new[:,1], c=y)
plt.subplot(122)
plt.title(label="Learned k=3 Clusters")
plt.scatter(X_new[:,0], X_new[:,1], c=y_pred_0)
plt.savefig("figures/plot.png")
plt.show()

#*####################################################################################
#
#* 4)
```

```
#*#####################################################################################
#

#* learn the transformation (components as linear combination of features)
from sklearn.decomposition import PCA
pca = PCA(n_components=X.shape[1])
pca.fit(X)
variability = 0
number_of_principal_components = 1
for f in pca.explained_variance_ratio_:
    variability += f
    if variability > 0.80:
        break
    number_of_principal_components += 1

print(number_of_principal_components)
```

END