Aprendizagem 2024/25

Homework IV – Group 99

(ist1106059, ist1106329)

I. Pen-and-paper

1)

	y_1	y_2
x_1	1	0
x_2	0	2
x_3	3	-1

$u_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ $\quad \Sigma_1 = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$ $\quad \pi_1 = 0.5$

$u_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\quad \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ $\quad \pi_2 = 0.5$

Epoch 1:

E-step:

$$P(c_k|\, x_i) = \frac{posterior\ (c_k|\, x_i)}{\sum_{j=1}^{2} posterior(c_j|x_i)}$$

$$posterior\ (c_k|\, x_i) = P(x_i \mid c_k)\, P(c_k) = N(x_i \mid u_k, \Sigma_k)\pi_k$$

$$N(x_i \mid u_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{d}{2}} \mid \Sigma_k \mid^{\frac{1}{2}}} e^{-1/2(x-u_k)^T \Sigma_k^{-1}(x-u_k)}, \qquad com\ d = 2$$

$|\Sigma_1| = 15$

$|\Sigma_2| = 4$

- Para x_1:

$posterior(c = 1|x_1) = P(x_1|c = 1)\, P(c = 1) = N(x_1 \mid u_1, \Sigma_1)\, \pi_1 = 0.029 * 0.5 = 0.015$

$posterior(c = 2|x_1) = P(x_1|c = 2)\, P(c = 2) = N(x_1 \mid u_2, \Sigma_2)\, \pi_2 = 0.062 * 0.5 = 0.031$

$$P(c = 1|\, x_1) = \frac{posterior\ (c = 1|\, x_1)}{posterior\ (c = 1|\, x_1) + posterior\ (c = 2|\, x_1)} = \frac{0.015}{0.015 + 0.031} = 0.326$$

$$P(c=2|\,x_1) = \frac{posterior\ (c=2|\,x_1)}{posterior\ (c=1|\,x_1) + \ posterior\ (c=2|\,x_1)} = \frac{0.031}{0.015+0.031} = 0.674$$

- Para x_2:

$$posterior(c=1|x_2) = P(x_2|c=1)\ P(c=1) = N(x_2\mid u_1, \Sigma_1)\ \pi_1 = \ 0.005 * 0.5 = \ 0.003$$
$$posterior(c=2|x_2) = P(x_2|c=2)\ P(c=2) = N(x_2\mid u_2, \Sigma_2)\ \pi_2 = \ 0.048 * 0.5 = \ 0.024$$

$$P(c=1|\,x_2) = \frac{posterior\ (c=1|\,x_2)}{posterior\ (c=1|\,x_2) + \ posterior\ (c=2|\,x_2)} = \frac{0.003}{0.003+0.024} = 0.111$$
$$P(c=2|\,x_2) = \frac{posterior\ (c=2|\,x_2)}{posterior\ (c=1|\,x_2) + \ posterior\ (c=2|\,x_2)} = \frac{0.024}{0.003+0.024} = 0.889$$

- Para x_3:

$$posterior(c=1|x_3) = P(x_3|c=1)\ P(c=1) = N(x_3\mid u_1, \Sigma_1)\ \pi_1 = \ 0.036 * 0.5 = \ 0.018$$
$$posterior(c=2|x_3) = P(x_3|c=2)\ P(c=2) = N(x_3\mid u_2, \Sigma_2)\ \pi_2 = \ 0.011 * 0.5 = \ 0.006$$

$$P(c=1|\,x_3) = \frac{posterior\ (c=1|\,x_3)}{posterior\ (c=1|\,x_3) + \ posterior\ (c=2|\,x_3)} = \frac{0.018}{0.018+0.006} = 0.750$$
$$P(c=2|\,x_3) = \frac{posterior\ (c=2|\,x_3)}{posterior\ (c=1|\,x_3) + \ posterior\ (c=2|\,x_3)} = \frac{0.006}{0.018+0.006} = 0.250$$

x_i	x_1	x_2	x_3
$P(c=1\mid x_i)$	0.326	0.111	0.750
$P(c=2\mid x_i)$	0.674	0.889	0.250

M-step:

$$N_k = \sum_{i=1}^{n} P(c_k \mid x_i)$$

$$u_k = \frac{1}{N_k}\sum_{i=1}^{n} P(c_k \mid x_i)\cdot x_i$$

$$\Sigma_k = \frac{1}{N_k}\sum_{i=1}^{n} P(c_k \mid x_i)\cdot (x_i - u_k)\cdot (x_i - u_k)^T$$

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

Aprendizagem 2024/25

Homework IV – Group 99

(ist1106059, ist1106329)

Cluster 1:

$N_1 = P(c = 1 \mid x_1) + P(c = 1 \mid x_2) + P(c = 1 \mid x_3) = 0.326 + 0.111 + 0.750 = 1.187$

$$u_1 = \frac{1}{N_1} * (P(c = 1 \mid x_1)x_1 + P(c = 1 \mid x_2)x_2 + P(c = 1 \mid x_3)x_3)$$
$$= \frac{1}{1.187} * (0.326\begin{bmatrix}1\\0\end{bmatrix} + 0.111\begin{bmatrix}0\\2\end{bmatrix} + 0.750\begin{bmatrix}3\\-1\end{bmatrix}) = \begin{bmatrix}2.170\\-0.445\end{bmatrix}$$

$\Sigma_1 = \frac{1}{N_1} * (P(c = 1 \mid x_1)(x_1 - u_1)(x_1 - u_1)^T + P(c = 1 \mid x_2)(x_2 - u_1)(x_2 - u_1)^T +$

$P(c = 1 \mid x_3)(x_3 - u_1)(x_3 - u_1)^T) = \frac{1}{1.187} * (0.326(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix})(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix})^T +$

$0.111(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix})(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix})^T + 0.750(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix})(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix})^T) =$

$= \begin{bmatrix}1.252 & -0.930\\-0.930 & 0.808\end{bmatrix}$

$$\pi_1 = P(c_1) = \frac{N_1}{3} = \frac{1.187}{3} = 0.396$$

Cluster 2:

$N_2 = P(c = 2 \mid x_1) + P(c = 2 \mid x_2) + P(c = 2 \mid x_3) = 0.674 + 0.889 + 0.250 = 1.813$

$$u_2 = \frac{1}{N_2} * (P(c = 2 \mid x_1)x_1 + P(c = 2 \mid x_2)x_2 + P(c = 2 \mid x_3)x_3)$$
$$= \frac{1}{1.813} * (0.674\begin{bmatrix}1\\0\end{bmatrix} + 0.889\begin{bmatrix}0\\2\end{bmatrix} + 0.250\begin{bmatrix}3\\-1\end{bmatrix}) = \begin{bmatrix}0.785\\0.843\end{bmatrix}$$

$\Sigma_2 = \frac{1}{N_2} * (P(c = 2 \mid x_1)(x_1 - u_2)(x_1 - u_2)^T + P(c = 2 \mid x_2)(x_2 - u_2)(x_2 - u_2)^T +$

$P(c = 2 \mid x_3)(x_3 - u_2)(x_3 - u_2)^T) = \frac{1}{1.813} * (0.674(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix})(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix})^T +$

$0.889(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix})(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix})^T + 0.250(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix})(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix})^T) =$

$\begin{bmatrix}0.996 & -1.076\\-1.076 & 1.389\end{bmatrix}$

$$\pi_2 = P(c_2) = \frac{N_2}{3} = \frac{1.813}{3} = 0.604$$

Aprendizagem 2024/25

Homework IV – Group 99

(ist1106059, ist1106329)

Epoch 2:

E-step:

$|\Sigma_1| = 0.147$ $|\Sigma_2| = 0.226$

- Para x_1:

$posterior(c = 1|x_1) = P(x_1|c = 1)\ P(c = 1) = N(x_1 \mid u_1, \Sigma_1)\ \pi_1 = 0.112 * 0.396 = 0.044$

$posterior(c = 2|x_1) = P(x_1|c = 2)\ P(c = 2) = N(x_1 \mid u_2, \Sigma_2)\ \pi_2 = 0.144 * 0.604 = 0.087$

$$P(c = 1|\ x_1) = \frac{posterior\ (c = 1|\ x_1)}{posterior\ (c = 1|\ x_1) + posterior\ (c = 2|\ x_1)} = \frac{0.044}{0.044 + 0.087} = 0.336$$

$$P(c = 2|\ x_1) = \frac{posterior\ (c = 2|\ x_1)}{posterior\ (c = 1|\ x_1) + posterior\ (c = 2|\ x_1)} = \frac{0.087}{0.044 + 0.087} = 0.664$$

- Para x_2:

$posterior(c = 1|x_2) = P(x_2|c = 1)\ P(c = 1) = N(x_2 \mid u_1, \Sigma_1)\ \pi_1 = 0.003 * 0.396 = 0.001$

$posterior(c = 2|x_2) = P(x_2|c = 2)\ P(c = 2) = N(x_2 \mid u_2, \Sigma_2)\ \pi_2 = 0.199 * 0.604 = 0.120$

$$P(c = 1|\ x_2) = \frac{posterior\ (c = 1|\ x_2)}{posterior\ (c = 1|\ x_2) + posterior\ (c = 2|\ x_2)} = \frac{0.001}{0.001 + 0.120} = 0.008$$

$$P(c = 2|\ x_2) = \frac{posterior\ (c = 2|\ x_2)}{posterior\ (c = 1|\ x_2) + posterior\ (c = 2|\ x_2)} = \frac{0.120}{0.001 + 0.120} = 0.992$$

- Para x_3:

$posterior(c = 1|x_3) = P(x_3|c = 1)\ P(c = 1) = N(x_3 \mid u_1, \Sigma_1)\ \pi_1 = 0.310 * 0.396 = 0.123$

$posterior(c = 2|x_3) = P(x_3|c = 2)\ P(c = 2) = N(x_3 \mid u_2, \Sigma_2)\ \pi_2 = 0.015 * 0.604 = 0.009$

$$P(c = 1|\ x_3) = \frac{posterior\ (c = 1|\ x_3)}{posterior\ (c = 1|\ x_3) + posterior\ (c = 2|\ x_3)} = \frac{0.123}{0.123 + 0.009} = 0.932$$

$$P(c = 2|\ x_3) = \frac{posterior\ (c = 2|\ x_3)}{posterior\ (c = 1|\ x_3) + posterior\ (c = 2|\ x_3)} = \frac{0.009}{0.123 + 0.009} = 0.068$$

x_i	x_1	x_2	x_3
$P(c=1\|\ x_i)$	0.336	0.008	0.932
$P(c=2\|\ x_i)$	0.664	0.992	0.068

Homework IV – Group 99

(ist1106059, ist1106329)

M-step:

Cluster 1:

$N_1 = P(c = 1 \mid x_1) + P(c = 1 \mid x_2) + P(c = 1 \mid x_3) = 0.336 + 0.008 + 0.932 = 1.276$

$$u_1 = \frac{1}{N_1} * (P(c = 1 \mid x_1)x_1 + P(c = 1 \mid x_2)x_2 + P(c = 1 \mid x_3)x_3)$$
$$= \frac{1}{1.276} * (0.336 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.008 \begin{bmatrix} 0 \\ 2 \end{bmatrix} + 0.932 \begin{bmatrix} 3 \\ -1 \end{bmatrix}) = \begin{bmatrix} 2.455 \\ -0.718 \end{bmatrix}$$

$\Sigma_1 = \frac{1}{N_1} * (P(c = 1 \mid x_1)(x_1 - u_1)(x_1 - u_1)^T + P(c = 1 \mid x_2)(x_2 - u_1)(x_2 - u_1)^T +$

$P(c = 1 \mid x_3)(x_3 - u_1)(x_3 - u_1)^T) = \frac{1}{1.276} * (0.336(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.455 \\ -0.718 \end{bmatrix})(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.455 \\ -0.718 \end{bmatrix})^T +$

$0.008(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 2.455 \\ -0.718 \end{bmatrix})(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 2.455 \\ -0.718 \end{bmatrix})^T + 0.932(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2.455 \\ -0.718 \end{bmatrix})(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2.455 \\ -0.718 \end{bmatrix})^T) =$

$\begin{bmatrix} 0.812 & -0.429 \\ -0.429 & 0.240 \end{bmatrix}$

$$\pi_1 = P(c_1) = \frac{N_1}{3} = \frac{1.276}{3} = 0.425$$

Cluster 2:

$N_2 = P(c = 2 \mid x_1) + P(c = 2 \mid x_2) + P(c = 2 \mid x_3) = 0.664 + 0.992 + 0.068 = 1.724$

$$u_2 = \frac{1}{N_2} * (P(c = 2 \mid x_1)x_1 + P(c = 2 \mid x_2)x_2 + P(c = 2 \mid x_3)x_3)$$
$$= \frac{1}{1.724} * (0.664 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.992 \begin{bmatrix} 0 \\ 2 \end{bmatrix} + 0.068 \begin{bmatrix} 3 \\ -1 \end{bmatrix}) = \begin{bmatrix} 0.503 \\ 1.111 \end{bmatrix}$$

$\Sigma_2 = \frac{1}{N_2} * (P(c = 2 \mid x_1)(x_1 - u_2)(x_1 - u_2)^T + P(c = 2 \mid x_2)(x_2 - u_2)(x_2 - u_2)^T +$

$P(c = 2 \mid x_3)(x_3 - u_2)(x_3 - u_2)^T) = \frac{1}{1.724} * (0.664(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.503 \\ 1.111 \end{bmatrix})(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.503 \\ 1.111 \end{bmatrix})^T +$

$0.992(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.503 \\ 1.111 \end{bmatrix})(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.503 \\ 1.111 \end{bmatrix})^T + 0.068(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.503 \\ 1.111 \end{bmatrix})(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.503 \\ 1.111 \end{bmatrix})^T) =$

$\begin{bmatrix} 0.487 & -0.678 \\ -0.678 & 1.106 \end{bmatrix}$

$$\pi_2 = P(c_2) = \frac{N_2}{3} = \frac{1.724}{3} = 0.575$$

Aprendizagem 2024/25

Homework IV – Group 99

(ist1106059, ist1106329)

2)

a)

Parâmetros atualizados:

$u_1 = \begin{bmatrix} 2.455 \\ -0.718 \end{bmatrix}$ $\Sigma_1 = \begin{bmatrix} 0.812 & -0.429 \\ -0.429 & 0.240 \end{bmatrix}$ $\pi_1 = 0.425$

$u_2 = \begin{bmatrix} 0.503 \\ 1.111 \end{bmatrix}$ $\Sigma_2 = \begin{bmatrix} 0.487 & -0.678 \\ -0.678 & 1.106 \end{bmatrix}$ $\pi_2 = 0.575$

$|\Sigma_1| = 0.011$

$|\Sigma_2| = 0.079$

- Para x_1:

$posterior(c = 1|x_1) = P(x_1|c = 1)\, P(c = 1) = N(x_1 \mid u_1, \Sigma_1)\, \pi_1 = 0.374 * 0.425 = 0.159$

$posterior(c = 2|x_1) = P(x_1|c = 2)\, P(c = 2) = N(x_1 \mid u_2, \Sigma_2)\, \pi_2 = 0.256 * 0.575 = 0.147$

$argmax_{c=\{1,2\}}\, posterior(c|x_1) = argmax_{c=\{1,2\}}\, \{0.159, 0.147\} = c_1$

Ou seja, atribuímos a observação x_1 ao cluster c_1.

- Para x_2:

$posterior(c = 1|x_2) = P(x_2|c = 1)\, P(c = 1) = N(x_2 \mid u_1, \Sigma_1)\, \pi_1 = 0 * 0.425 = 0$

$posterior(c = 2|x_2) = P(x_2|c = 2)\, P(c = 2) = N(x_2 \mid u_2, \Sigma_2)\, \pi_2 = 0.391 * 0.575 = 0.225$

$argmax_{c=\{1,2\}}\, posterior(c|x_2) = argmax_{c=\{1,2\}}\, \{0, 0.225\} = c_2$

Ou seja, atribuímos a observação x_2 ao cluster c_2.

- Para x_3:

$posterior(c = 1|x_3) = P(x_3|c = 1)\, P(c = 1) = N(x_3 \mid u_1, \Sigma_1)\, \pi_1 = 1.262 * 0.425 = 0.536$

$posterior(c = 2|x_3) = P(x_3|c = 2)\, P(c = 2) = N(x_3 \mid u_2, \Sigma_2)\, \pi_2 = 0 * 0.575 = 0$

$argmax_{c=\{1,2\}}\, posterior(c|x_3) = argmax_{c=\{1,2\}}\, \{0.536, 0\} = c_1$

Ou seja, atribuímos a observação x_3 ao cluster c_1.

b)

O maior cluster é c_1 que contêm as observações x_1 *e* x_3.

- Para x_1:

  Como o cluster c_1 só tem 2 observações:

  $a(x_1) = \|x_1 - x_3\|_2^2 = \sqrt{(1-3)^2 + (0+1)^2} = 2.236$

  Como temos só 1 outro cluster e este só tem 1 observação:

  $b(x_1) = \|x_1 - x_2\|_2^2 = \sqrt{(1-0)^2 + (0-2)^2} = 2.236$

  Como $a(x_2) \geq b(x_2)$,

  $$s(x_1) = \frac{b(x_1)}{a(x_1)} - 1 = \frac{2.236}{2.236} - 1 = 0$$

- Para x_3:

  $a(x_3) = \|x_3 - x_1\|_2^2 = \sqrt{(3-1)^2 + (-1-0)^2} = 2.236$

  $b(x_3) = \|x_3 - x_2\|_2^2 = \sqrt{(3-0)^2 + (-1-2)^2} = 4.243$

  Como $b(x_3) < a(x_3)$,

  $$s(x_3) = 1 - \frac{a(x_3)}{b(x_3)} = 1 - \frac{2.236}{4.243} = 0.473$$

- Silhueta do cluster c_1 é $\frac{0+0.473}{2} = 0.237$

Aprendizagem 2024/25

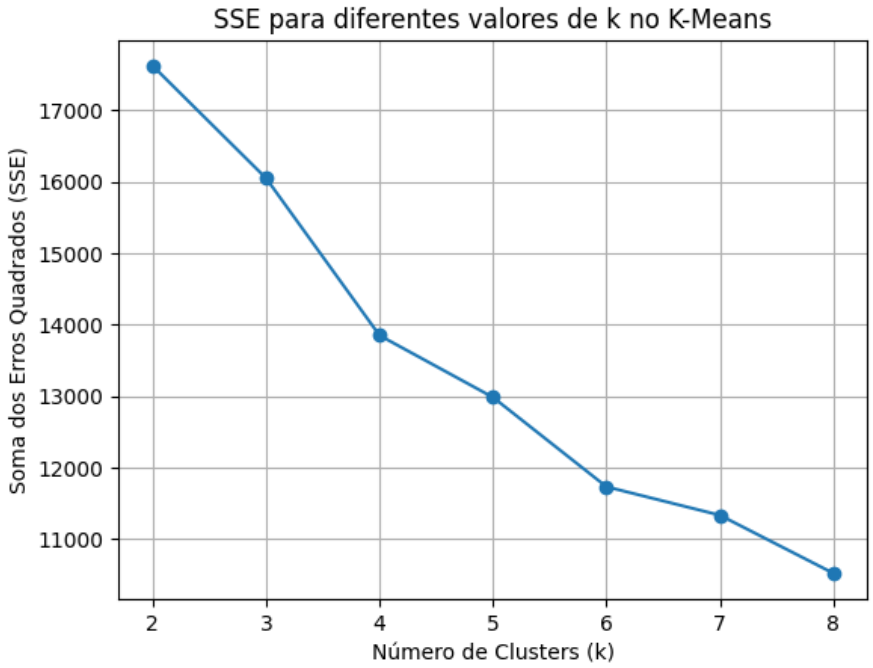
Homework IV – Group 99

(ist1106059, ist1106329)

II. Programming and critical analysis

1)

a)



b) Para responder à questão, podemos observar o gráfico obtido tendo em conta o conceito de "elbow finding".

Este conceito ajuda-nos a identificar o número ideal de clusters (k) de forma a haver uma boa divisão de dados sem criar clusters desnecessários.

Isto porque não basta olhar para o valor da SSE. Obviamente, quanto mais clusters tivermos, menor será o erro obtido porque os clusters vão dividir cada vez mais os dados e ajustar-se cada vez mais aos mesmos. No entanto, a partir de certo ponto, este aumento no número de clusters deixa de ser útil porque começamos a dividir dados com características similares e que deveriam pertencer ao mesmo cluster, o que dificulta a interpretação dos resultados, deixamos de conseguir captar padrões e de generalizar, e aumentamos a complexidade do modelo.

Logo, é necessário encontrar um equilíbrio entre diminuir a inercia e obter um valor de k que evite as desvantagens mencionadas. Para isso, procuramos no gráfico o "elbow", ou seja, o ponto em que a redução da SSE fica menos acentuada e, por isso, não há um ganho tão significativo com a redução do SSE coo nos valores de k iniciais.

No plot obtido, a taxa de diminuição da SSE não varia muito e, por isso, este "elbow" não é muito óbvio. No entanto, é possível identificar que o valor ideal do número de clusters é k = 4 já que, de k=2 a k=4 a redução da SSE é mais acentuada que de valores de k superiores a 4 (em que a redução da SSE fica mais constante).

c) O k-modes é uma adaptação do k-mean para variáveis categóricas. Utiliza a distância de Hamming (contagem de valores diferentes entre observações) em vez da distância Euclidiana para calcular a distância entre as observações e os centróides dos clusters. No entanto, tal como o k-mean não apresenta resultados ótimos para variáveis categóricas, que não têm nenhuma relação de proximidade numérica, o k-modes não é ideal para variáveis numéricas porque iria apresentar o mesmo valor para duas observações com atributos numéricos quer muito próximos, quer muito distantes (desde que sejam diferentes).

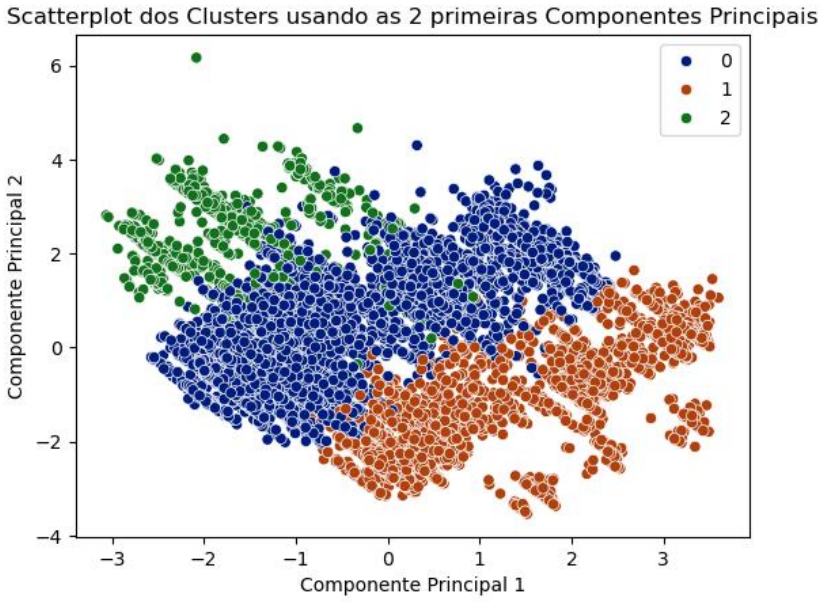
No que toca ao dataset em questão, temos tanto variáveis numéricas (como age e balance) como variáveis categóricas (como job, marital, education...). No entanto, como este contém mais variáveis categóricas que numéricas, o k-modes poderia ser uma melhor abordagem e poderia gerar clusters mais representativos. É de notar que esta escolha teria sempre prós e contras e, para além da proporção existente de cada tipo de variável, também é preciso ter em conta o significado e relevância de cada uma em relação aos resultados pretendidos.

Existe ainda outro aspeto que é importante mencionar. Neste dataset, foi aplicado uma normalização e conversão das variáveis categóricas para valores binários (com o get_dummies()), de forma que o k-means processe variáveis categóricas como se fossem numéricas, melhorando assim os resultados obtidos. No entanto, esta abordagem aumenta consideravelmente a dimensionalidade do dataset e, consequentemente, a complexidade do modelo e dificuldade na interpretação dos resultados, podendo fazer com que o k-modes ainda fosse uma abordagem melhor.

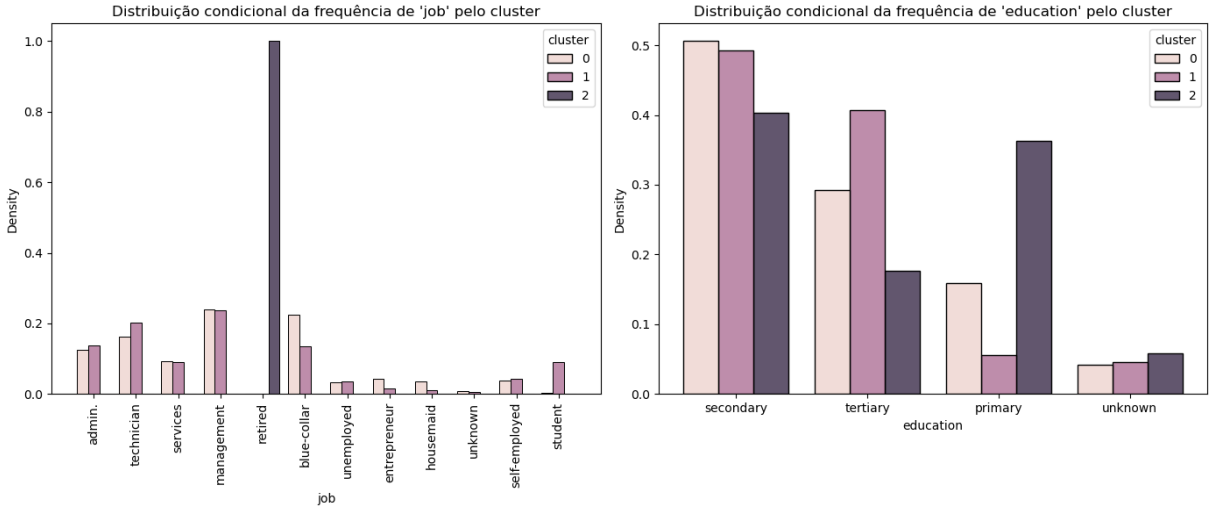
2)

a) Variância explicada pelos dois primeiros componentes principais: 22.76%

b) Ao observar o scatterplot, é possível distinguir 3 grupos distintos e, ainda que haja alguns pontos sobrepostos, não parece ser em quantidade significativa que impeça de formar estes 3 grupos. Logo, podemos afirmar que conseguimos separar os clusters tendo em conta as 2 componentes principais.



c)



Ao analisar os gráficos da distribuição condicional das variáveis "job" e "education" por cluster, conseguimos identificar algumas características distintas para cada grupo.

O Cluster 0 apresenta uma clara maior frequência nas categorias "blue-collar", "entrepreneur" e "housemaid" na feature "job". Para a feature "education", a categoria predominante deste cluster é "secondary". Desta forma, o cluster 0 aparenta representar um grupo de ocupações e escolaridade de nível médio.

Para o Cluster 1 conseguimos identificar maior probabilidade de incluir observações das categorias "admin", "student" e "technician" para a feature "job". Já no gráfico de "education" identificamos a categoria mais comum sendo "tertiary". Estes fatores indicam que este cluster tende a agrupar clientes com níveis educacionais mais elevados e ocupações profissionais de nível administrativo ou técnico.

Finalmente o Cluster 2 é caracterizado pela categoria "retired" na feature "job", estando todas as observações para essa categoria concentradas nesse cluster. No gráfico de "education", identificamos a categoria "primary" como associada a este cluster, que indica que este grupo de clientes apresenta menor escolaridade e são reformados.

Para além disso, na feature "job" ainda temos mais algumas categorias nas quais não conseguimos identificar um cluster predominante. Nas categorias "services", "management", "unemployed", "unknown" e "self-employed" a probabilidade divide-se entre pertencerem ao cluster 0 ou ao cluster 1. No gráfico da feature "education", a categoria "unknown" não tem um claro cluster predominante sendo que as probabilidades desta categoria pertencer a um determinado cluster decrescem do cluster 2 para o cluster 0.

Podemos então concluir que há uma boa distinção entre clusters e que cada um possui características demográficas diferentes, relacionando níveis de educação mais elevados com áreas de trabalho mais técnicas e níveis de educação mais baixos com áreas de emprego que exigem menos aptidões académicas ou mesmo desemprego.

END