Aprendizagem 2024/25

Homework IV – Group 014

(ist106266, ist106583)

I. Pen-and-paper

1) Each iteration comprises the E-step, where the relative posteriors, γ_{ki}, are computed for each cluster-point pair according to the following formula:

$$N\left(x_i \middle|\ \mu_k,\ \Sigma_k\right) = \frac{\exp\left(-\frac{1}{2}\left(x_i-\mu_k\right)^T \Sigma_k^{-1}\left(x_i-\mu_k\right)\right)}{2\pi\sqrt{\left|\Sigma_k\right|}}$$

$$\gamma_{ki} = P\left(c_k \mid x_i\right) = \frac{P\left(c_k, x_i\right)}{P\left(x_i\right)} = \frac{\pi_k N\left(x_i \mid \mu_k, \Sigma_k\right)}{\Sigma_k^K\left(\pi_k N\left(x_i \mid \mu_k, \Sigma_k\right)\right)}$$

The E-step is then followed by the M-step, where the means, covariance matrices and priors are updated according to these rules:

$$N_k = \sum_{i=1}^{N} \gamma_{ki} \qquad \pi_k = \frac{N_k}{N} \qquad \mu_k = \frac{1}{N_k}\sum_{i=1}^{N} \gamma_{ki}\cdot x_i$$

$$\Sigma_k = \frac{1}{N_k}\sum_{i=1}^{N}\left(\gamma_{ki}\cdot\left(x_i-\mu_k\right)\left(x_i-\mu_k\right)^T\right)$$

With all of this out of the way, it is possible to start the first epoch's E-step:

$$\left|\Sigma_1\right| = \begin{vmatrix} 4 & 1 \\ 1 & 4 \end{vmatrix} = 15 \qquad \left|\Sigma_2\right| = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4$$

$$\Sigma_1^{-1} = \begin{bmatrix} 0.267 & -0.067 \\ -0.067 & 0.267 \end{bmatrix} \qquad \Sigma_2^{-1} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

$$N\left(x_1 \middle|\ \mu_1,\ \Sigma_1\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}\right)^T \begin{bmatrix} 0.267 & -0.067 \\ -0.067 & 0.267 \end{bmatrix}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}\right)\right)}{2\pi\sqrt{15}} = 0.029$$

$$N\left(x_1 \middle|\ \mu_2,\ \Sigma_2\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)^T \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)\right)}{2\pi\sqrt{4}} = 0.062$$

$$N\left(x_2 \middle|\ \mu_1,\ \Sigma_1\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}\right)^T \begin{bmatrix} 0.267 & -0.067 \\ -0.067 & 0.267 \end{bmatrix}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}\right)\right)}{2\pi\sqrt{15}} = 0.005$$

$$N\left(x_2 \middle|\ \mu_2,\ \Sigma_2\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)^T \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)\right)}{2\pi\sqrt{4}} = 0.048$$

$$N\left(x_3 \middle|\ \mu_1,\ \Sigma_1\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}\right)^T \begin{bmatrix} 0.267 & -0.067 \\ -0.067 & 0.267 \end{bmatrix}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}\right)\right)}{2\pi\sqrt{15}} = 0.036$$

$$N\left(x_3 \middle|\, \mu_2, \Sigma_2\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)^T \begin{bmatrix}0.5 & 0\\0 & 0.5\end{bmatrix}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)\right)}{2\pi\sqrt{4}} = 0.011$$

$$p\left(x_1, c_1\right) = 0.014 \quad p\left(x_1, c_2\right) = 0.031 \qquad \gamma_{11} = 0.311 \quad \gamma_{21} = 0.689$$

$$p\left(x_2, c_1\right) = 0.002 \quad p\left(x_2, c_2\right) = 0.024 \qquad \gamma_{12} = 0.077 \quad \gamma_{22} = 0.923$$

$$p\left(x_3, c_1\right) = 0.018 \quad p\left(x_3, c_2\right) = 0.006 \qquad \gamma_{13} = 0.750 \quad \gamma_{23} = 0.250$$

Afterwards, the M-step is executed to update the parameters:

$$N_1 = 0.311 + 0.077 + 0.750 = 1.138 \qquad N_2 = 0.689 + 0.923 + 0.250 = 1.862$$

$$\mu_1^{\text{new}} = \frac{1}{1.138}\left(0.311\begin{bmatrix}1\\0\end{bmatrix} + 0.077\begin{bmatrix}0\\2\end{bmatrix} + 0.750\begin{bmatrix}3\\-1\end{bmatrix}\right) = \begin{bmatrix}2.250\\-0.524\end{bmatrix}$$

$$\mu_2^{\text{new}} = \frac{1}{1.862}\left(0.689\begin{bmatrix}1\\0\end{bmatrix} + 0.923\begin{bmatrix}0\\2\end{bmatrix} + 0.250\begin{bmatrix}3\\-1\end{bmatrix}\right) = \begin{bmatrix}0.773\\0.857\end{bmatrix}$$

$$\begin{aligned}\Sigma_1^{\text{new}} = \frac{1}{1.138}\Bigg(&0.311\begin{bmatrix}1-2.250\\0-(-0.524)\end{bmatrix}\begin{bmatrix}1-2.250 & 0-(-0.524)\end{bmatrix} + \\ &0.077\begin{bmatrix}0-2.250\\2-(-0.524)\end{bmatrix}\begin{bmatrix}0-2.250 & 2-(-0.524)\end{bmatrix} + \\ &0.750\begin{bmatrix}3-2.250\\-1-(-0.524)\end{bmatrix}\begin{bmatrix}3-2.250 & -1-(-0.524)\end{bmatrix}\Bigg) \\ = &\begin{bmatrix}1.140 & -0.799\\-0.799 & 0.655\end{bmatrix}\end{aligned}$$

$$\begin{aligned}\Sigma_2^{\text{new}} = \frac{1}{1.862}\Bigg(&0.689\begin{bmatrix}1-0.773\\0-0.857\end{bmatrix}\begin{bmatrix}1-0.773 & 0-0.857\end{bmatrix} + \\ &0.923\begin{bmatrix}0-0.773\\2-0.857\end{bmatrix}\begin{bmatrix}0-0.773 & 2-0.857\end{bmatrix} + \\ &0.250\begin{bmatrix}3-0.773\\-1-0.857\end{bmatrix}\begin{bmatrix}3-0.773 & -1-0.857\end{bmatrix}\Bigg) \\ = &\begin{bmatrix}0.981 & -1.065\\-1.065 & 1.382\end{bmatrix}\end{aligned}$$

$$\pi_1^{\text{new}} = \frac{1.138}{3} = 0.379 \qquad \pi_2^{\text{new}} = \frac{1.862}{3} = 0.621$$

The second epoch is then executed much like the first, just with the updated parameters instead of the original ones. Starting with the E-step once again:

$$|\Sigma_1| = 0.108 \qquad\qquad |\Sigma_2| = 0.222$$

$$\Sigma_1^{-1} = \begin{bmatrix} 6.065 & 7.398 \\ 7.398 & 10.556 \end{bmatrix} \qquad \Sigma_2^{-1} = \begin{bmatrix} 6.225 & 4.797 \\ 4.797 & 4.419 \end{bmatrix}$$

$$N\left(x_1 \middle|\, \mu_1,\, \Sigma_1\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.250\\-0.524\end{bmatrix}\right)^T \begin{bmatrix}6.065 & 7.398\\7.398 & 10.556\end{bmatrix}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.250\\-0.524\end{bmatrix}\right)\right)}{2\pi\sqrt{0.108}} = 0.127$$

$$N\left(x_2 \middle|\, \mu_1,\, \Sigma_1\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.250\\-0.524\end{bmatrix}\right)^T \begin{bmatrix}6.065 & 7.398\\7.398 & 10.556\end{bmatrix}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.250\\-0.524\end{bmatrix}\right)\right)}{2\pi\sqrt{0.108}} = 0$$

$$N\left(x_3 \middle|\, \mu_1,\, \Sigma_1\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2.250\\-0.524\end{bmatrix}\right)^T \begin{bmatrix}6.065 & 7.398\\7.398 & 10.556\end{bmatrix}\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2.250\\-0.524\end{bmatrix}\right)\right)}{2\pi\sqrt{0.108}} = 0.373$$

$$N\left(x_1 \middle|\, \mu_2,\, \Sigma_2\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.773\\0.857\end{bmatrix}\right)^T \begin{bmatrix}6.225 & 4.797\\4.797 & 4.419\end{bmatrix}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.773\\0.857\end{bmatrix}\right)\right)}{2\pi\sqrt{0.222}} = 0.144$$

$$N\left(x_2 \middle|\, \mu_2,\, \Sigma_2\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.773\\0.857\end{bmatrix}\right)^T \begin{bmatrix}6.225 & 4.797\\4.797 & 4.419\end{bmatrix}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.773\\0.857\end{bmatrix}\right)\right)}{2\pi\sqrt{0.222}} = 0.203$$

$$N\left(x_1 \middle|\, \mu_2,\, \Sigma_2\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}0.773\\0.857\end{bmatrix}\right)^T \begin{bmatrix}6.225 & 4.797\\4.797 & 4.419\end{bmatrix}\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}0.773\\0.857\end{bmatrix}\right)\right)}{2\pi\sqrt{0.222}} = 0.014$$

$$\begin{array}{llll} p\left(x_1,\, c_1\right) = 0.048 & p\left(x_1,\, c_2\right) = 0.089 & \gamma_{11} = 0.350 & \gamma_{21} = 0.650 \\ p\left(x_2,\, c_1\right) = 0 & p\left(x_2,\, c_2\right) = 0.126 & \gamma_{12} = 0 & \gamma_{22} = 1 \\ p\left(x_3,\, c_1\right) = 0.141 & p\left(x_3,\, c_2\right) = 0.009 & \gamma_{13} = 0.940 & \gamma_{23} = 0.060 \end{array}$$

Finally, the M-step is executed one last time:

$$N_1 = 0.350 + 0 + 0.940 = 1.290 \qquad\qquad N_2 = 0.650 + 1 + 0.060 = 1.710$$

$$\mu_1^{new} = \frac{1}{1.290}\left(0.350\begin{bmatrix}1\\0\end{bmatrix} + 0\begin{bmatrix}0\\2\end{bmatrix} + 0.940\begin{bmatrix}3\\-1\end{bmatrix}\right) = \begin{bmatrix}2.457\\-0.729\end{bmatrix}$$

$$\mu_2^{new} = \frac{1}{1.710}\left(0.650\begin{bmatrix}1\\0\end{bmatrix} + 1\begin{bmatrix}0\\2\end{bmatrix} + 0.060\begin{bmatrix}3\\-1\end{bmatrix}\right) = \begin{bmatrix}0.485\\1.135\end{bmatrix}$$

$$\Sigma_1^{\text{new}} = \frac{1}{1.290}\left(0.350\begin{bmatrix}1-2.457\\0-(-0.729)\end{bmatrix}\begin{bmatrix}1-2.457 & 0-(-0.729)\end{bmatrix} + \right.$$

$$0\begin{bmatrix}0-2.457\\2-(-0.729)\end{bmatrix}\begin{bmatrix}0-2.457 & 2-(-0.729)\end{bmatrix} +$$

$$\left. 0.940\begin{bmatrix}3-2.457\\-1-(-0.729)\end{bmatrix}\begin{bmatrix}3-2.457 & -1-(-0.729)\end{bmatrix}\right)$$

$$= \begin{bmatrix}0.791 & -0.395\\-0.395 & 0.198\end{bmatrix}$$

$$\Sigma_2^{\text{new}} = \frac{1}{1.710}\left(0.650\begin{bmatrix}1-0.485\\0-1.135\end{bmatrix}\begin{bmatrix}1-0.485 & 0-1.135\end{bmatrix} + \right.$$

$$1\begin{bmatrix}0-0.485\\2-1.135\end{bmatrix}\begin{bmatrix}0-0.485 & 2-1.135\end{bmatrix} +$$

$$\left. 0.060\begin{bmatrix}3-0.485\\-1-1.135\end{bmatrix}\begin{bmatrix}3-0.485 & -1-1.135\end{bmatrix}\right)$$

$$= \begin{bmatrix}0.460 & -0.656\\-656 & 1.087\end{bmatrix}$$

$$\pi_2^{new} = \frac{1.710}{3} = 0.570 \qquad \pi_1^{new} = \frac{1.290}{3} = 0.430$$

2) a. To perform a hard assignment of the observations under a map assumption, it is necessary to compute $p(c_j|x_i) = p(x_i|c_j)p(c_j) = p(x_i,\ c_j)$ for each observation and assign them to the cluster that maximizes this value:

$$|\Sigma_1| = 0.001 \qquad |\Sigma_2| = 0.070$$

$$\Sigma_1^{-1} = \begin{bmatrix}198 & 395\\395 & 791\end{bmatrix} \qquad \Sigma_2^{-1} = \begin{bmatrix}15.529 & 9.371\\9.371 & 6.571\end{bmatrix}$$

$$N\left(x_1|\ \mu_1,\ \Sigma_1\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}2.457\\-0.729\end{bmatrix}\right)^T\begin{bmatrix}198 & 395\\395 & 791\end{bmatrix}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}2.457\\-0.729\end{bmatrix}\right)\right)}{2\pi\sqrt{0.001}} = 2.269$$

$$N\left(x_2|\ \mu_1,\ \Sigma_1\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}2.457\\-0.729\end{bmatrix}\right)^T\begin{bmatrix}198 & 395\\395 & 791\end{bmatrix}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}2.457\\-0.729\end{bmatrix}\right)\right)}{2\pi\sqrt{0.001}} = 0$$

$$N\left(x_3|\ \mu_1,\ \Sigma_1\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}2.457\\-0.729\end{bmatrix}\right)^T\begin{bmatrix}198 & 395\\395 & 791\end{bmatrix}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}2.457\\-0.729\end{bmatrix}\right)\right)}{2\pi\sqrt{0.001}} = 4.506$$

$$N\left(x_1|\ \mu_2,\ \Sigma_2\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}0.485\\1.135\end{bmatrix}\right)^T\begin{bmatrix}15.529 & 9.371\\9.371 & 6.571\end{bmatrix}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}0.485\\1.135\end{bmatrix}\right)\right)}{2\pi\sqrt{0.070}} = 0.266$$

$$N\left(x_2|\ \mu_2,\ \Sigma_2\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}0.485\\1.135\end{bmatrix}\right)^T\begin{bmatrix}15.529 & 9.371\\9.371 & 6.571\end{bmatrix}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}0.485\\1.135\end{bmatrix}\right)\right)}{2\pi\sqrt{0.070}} = 0.422$$

$$N\left(x_3|\,\mu_2,\,\Sigma_2\right) = \frac{\exp\left(-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}0.485\\1.135\end{bmatrix}\right)^T\begin{bmatrix}15.529 & 9.371\\9.371 & 6.571\end{bmatrix}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}0.485\\1.135\end{bmatrix}\right)\right)}{2\pi\sqrt{0.070}} = 0$$

$$p\left(x_1,\,c_1\right) = 0.976 > p\left(x_1,\,c_2\right) = 0.152 \Rightarrow x_1 \epsilon\, c_1$$

$$p\left(x_2,\,c_1\right) = 0 < p\left(x_2,\,c_2\right) = 0.241 \Rightarrow x_2 \epsilon\, c_2$$

$$p\left(x_3,\,c_1\right) = 1.938 > p\left(x_3,\,c_2\right) = 0 \Rightarrow x_3 \epsilon\, c_1$$

b. The silhouettes of observations and clusters take the following forms:

$$a(x_i) = \frac{1}{|c_k|-1}\sum_{x_j\in c_k,\, x_i\neq x_j} d(x_i,\,x_j),\ x_i \in c_k$$

$$b(x_i) = min\left\{\frac{1}{|c_j|}\sum_{y\in c_j} d(x_i,\,y)\right\},\ c_j \neq cluster(x_i)$$

$$S(X_i) = \begin{cases} 1 - \frac{a(X_i)}{b(X_i)} & \text{if } a\left(X_i\right) \leq b\left(X_i\right) \\ \frac{b(X_i)}{a(X_i)} - 1 & \text{if } a\left(X_i\right) > b\left(X_i\right) \end{cases}$$

$$S(c_i) = \frac{1}{|c_i|}\sum_{x\in c_i} S(x)$$

Therefore, calculating the silhouette of the larger cluster (c_1, since it has 2 observations while c_2 only has 1) goes like this:

$$a\left(x_1\right) = \left\|x_1 - x_3\right\|_2 = \sqrt{4+1} = \sqrt{5}$$

$$b\left(x_1\right) = \left\|x_1 - x_2\right\|_2 = \sqrt{4+1} = \sqrt{5}$$

$$a\left(x_3\right) = \left\|x_3 - x_1\right\|_2 = \sqrt{4+1} = \sqrt{5}$$

$$b\left(x_3\right) = \left\|x_3 - x_2\right\|_2 = \sqrt{9+9} = 3\sqrt{2}$$

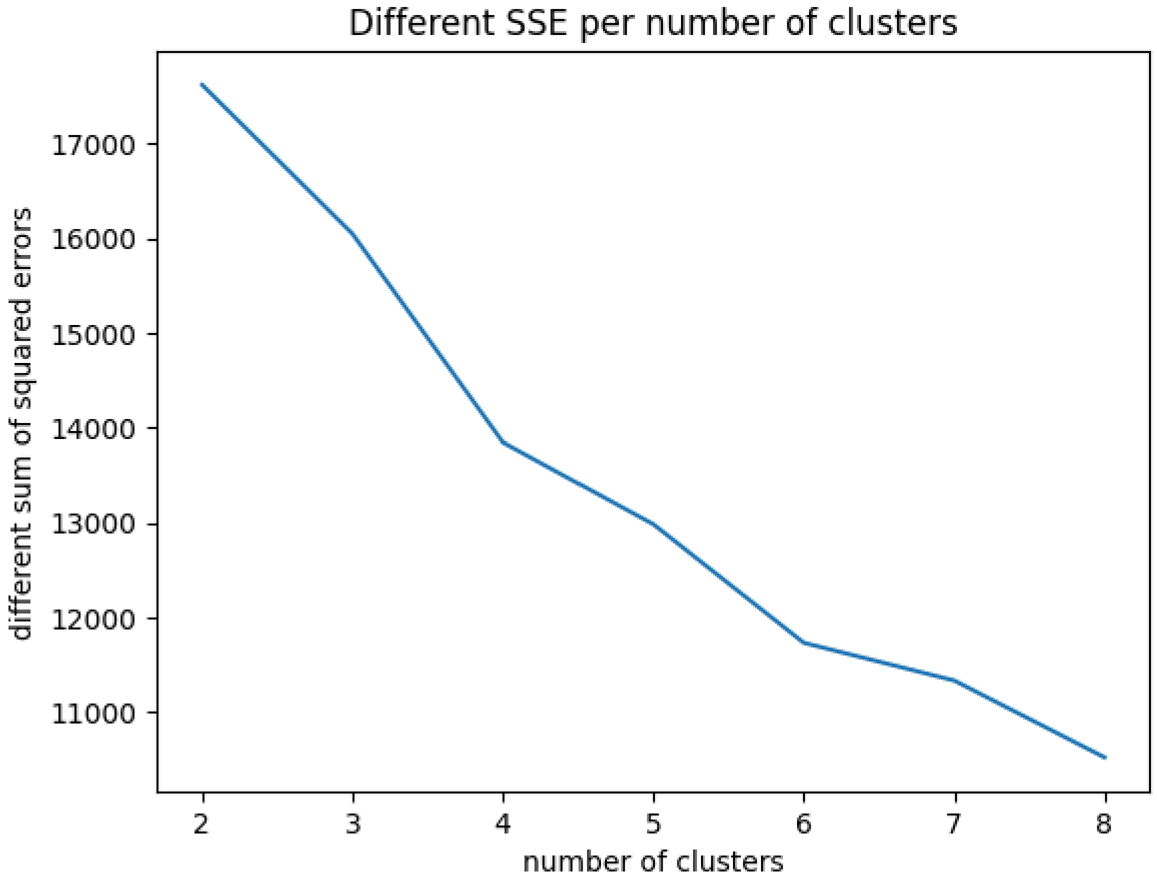
$$S\left(x_1\right) = 1 - \frac{a\left(X_1\right)}{b\left(X_1\right)} = 1 - 1 = 0$$

$$S\left(x_3\right) = 1 - \frac{a\left(X_3\right)}{b\left(X_3\right)} = 1 - \frac{\sqrt{5}}{3\sqrt{2}} = 0.473$$

$$S\left(c_1\right) = \frac{S\left(X_1\right)+S\left(X_3\right)}{2} = 0.237$$

II. Programming and critical analysis

1) a.

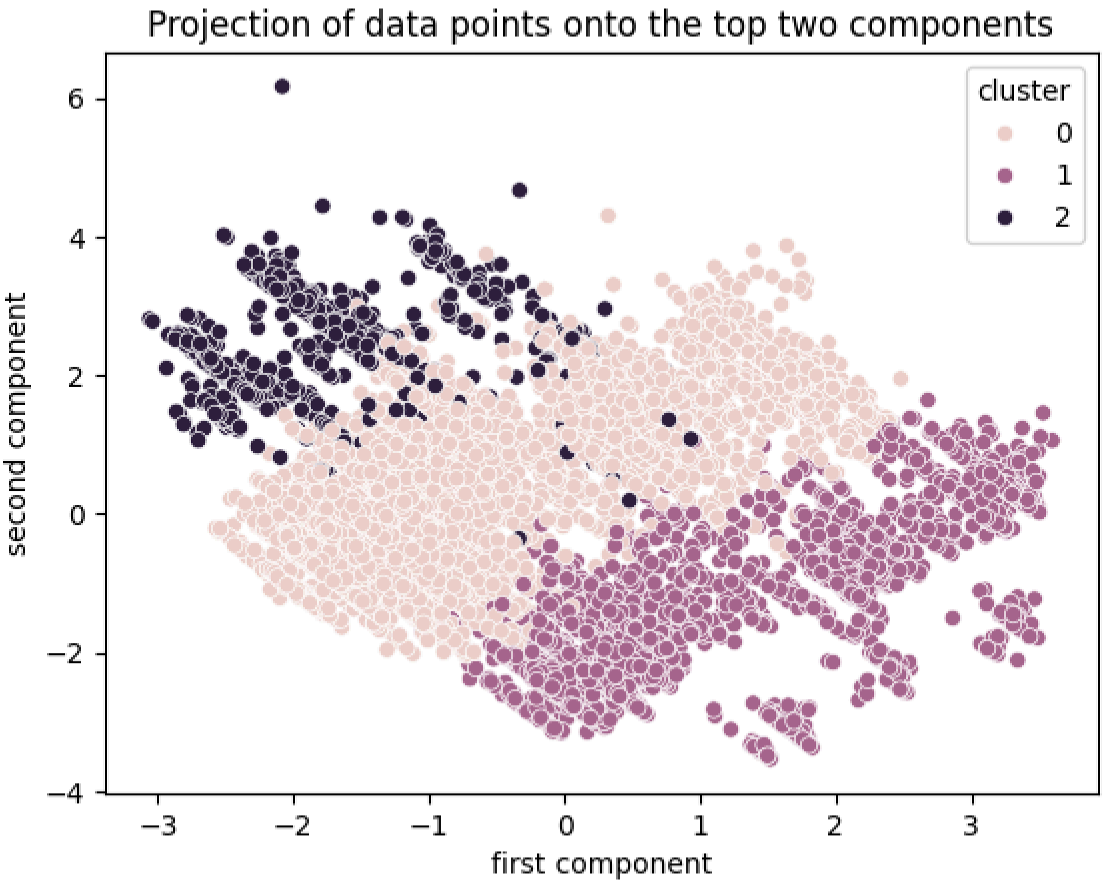


b. The graph shows a steep descent in inertia (different sum of squared errors) when increasing the number of clusters. This effect, however, seems to stop after more than six clusters are used, since the difference between the inertia of six, seven and eight clusters is quite small, at least when compared to the drops experienced before that point. Therefore, according to the elbow rule, there should be around six underlying customer segments.

c. When considering the features of the dataset, out of the eight features used, only two of them, age and balance, are numerical whereas the remaining six are categorical. However, these six categorical features cannot be fully captured by k-means since it relies on euclidean distance, which isn't suitable for them. Therefore, in this case, k-modes is likely a better clustering approach because it uses metrics specifically tailored to categorical features and modes instead of means, making it a better match for the dataset's features.

2) a. The first principal component explains a variability of 2.453, which represents 11.679% of all variation. The second explains a variation of 2.326, which amounts to 11.076% of the total variability. Together, the two justify 22.755% of all variation.

b. After training, the clusters according to the top two principal components can be displayed as follows:



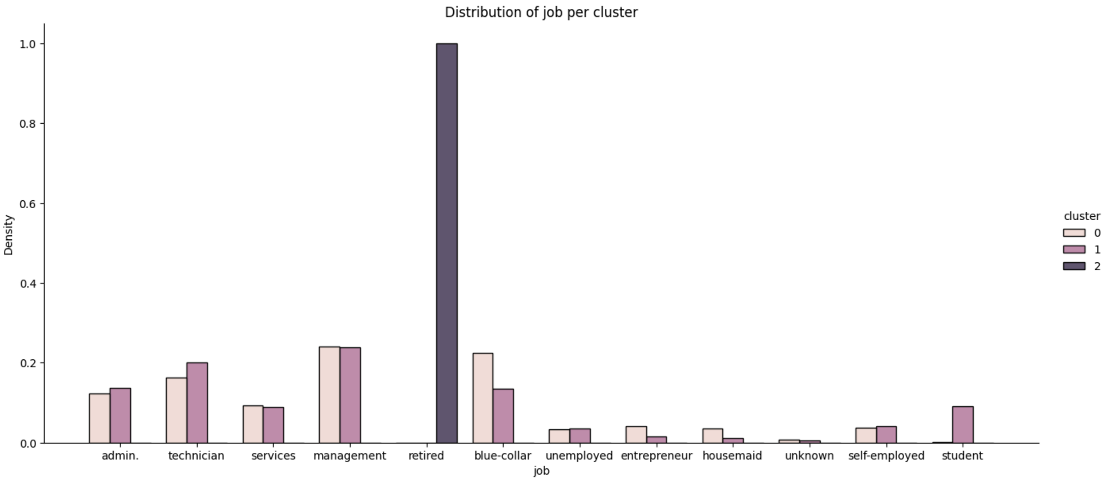
It is quite difficult to clearly separate the clusters since there are a lot of overlapping data points, from different clusters. This is due to the fact that the two components explain only 22.755% of the total variation of the data, which means a lot of information is lost when projecting the data onto these two components, making it hard to differentiate between clusters.

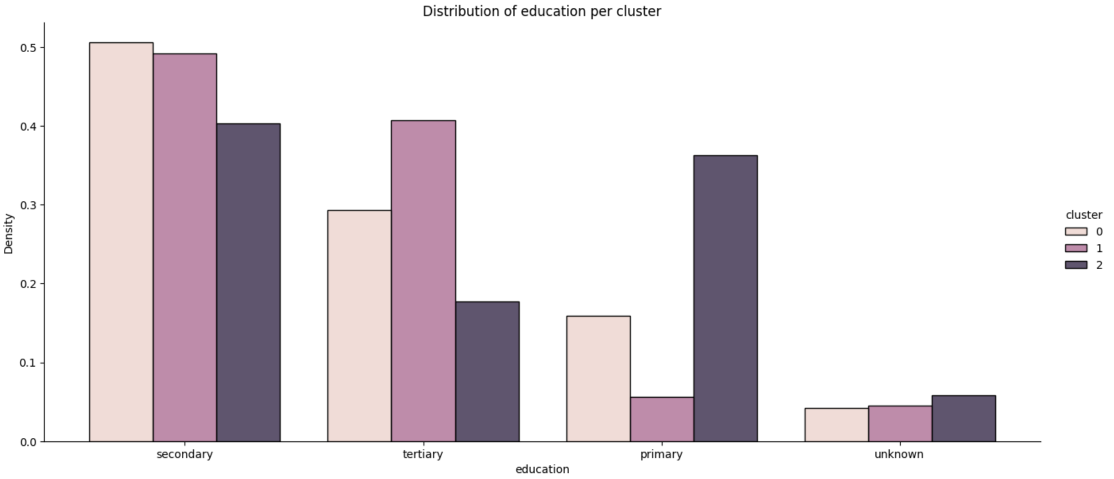
c. As the graphs below show, cluster 2 is composed entirely of retirees, which no other cluster has. In terms of education, this cluster mostly comprises people with primary and secondary education, with only around 20% having completed tertiary education.

Regarding cluster 1, proportionally, it is the cluster with the biggest share of people with tertiary education (40% of the cluster) and the smallest percentage of people with only primary education (less than 10%). Additionally, it is also the only cluster to have students.

Although cluster 0 has about the same proportion of people with secondary education as cluster 1 (around 50% of the cluster), it has less people with tertiary education (30% of the cluster), but more with primary education (about 15% of the cluster). In terms of jobs, even though it doesn't have students like cluster 1, the two clusters are very closely distributed, with cluster 1 having a slightly bigger share of technicians and administration workers and cluster 0 twice the proportion of blue-collar workers, entrepreneurs and handmaids.

All clusters have about the same proportion of unknown jobs and education levels.





END