Aprendizagem 2024/25

Homework 4 – Group 98

(João Pinho - ist1102765, Tiago Sousa - ist1110691)

Part I: Pen and paper

1 EM Clustering

The EM Clustering algorithm can be divided in 4 steps: Initialization, Expectation, Maximization and Evaluation of the log likelihood. For this exercise, we were provided the initialized Gaussian mixture parameters:

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Sigma_1 = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix} \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \pi_1 = 0.5, \pi_2 = 0.5$$

Let's now perform the first epoch of the EM algorithm. We first do the Expectation step (E-Step). In this step, we compute the weights that will be later used to update each cluster's components of the mixture. The weight of observation i for cluster k can be written (using Bayes' rule with normal distribution and defining the probability of cluster k to be $p(c_k) := \pi_k$) as:

$$\gamma_{ki} = p(c_k \mid \mathbf{x}_i) = \frac{p(\mathbf{x}_i \mid c_k)p(c_k)}{p(\mathbf{x}_i)} = \frac{N(\mathbf{x}_i \mid \mathbf{u}_k, \Sigma_k)\pi_k}{\sum_k N(\mathbf{x}_i \mid \mathbf{u}_k, \Sigma_k)\pi_k}$$

If we define $p(c_k, \mathbf{x}_i) := N(\mathbf{x}_i \mid \mathbf{u}_k, \Sigma_k)\pi_k$, we can rewrite the formula as:

$$\gamma_{ki} = \frac{p(c_k, \mathbf{x}_i)}{\sum_k p(c_k, \mathbf{x}_i)}$$

Let's start by writing the normal distributions of each cluster:

$$N(\mathbf{x}_i \mid \mathbf{u}_1, \Sigma_1) = \frac{e^{-\frac{1}{2}(\mathbf{x}_i - \mathbf{u}_1)^T \Sigma_1^{-1} (\mathbf{x}_i - \mathbf{u}_1)}}{(2\pi)^{\frac{2}{2}} \mid \Sigma_1 \mid^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}\left(\mathbf{x}_i - \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 4 & 1 \\ 1 & 4 \end{smallmatrix}\right]^{-1} \left(\mathbf{x}_i - \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]\right)}}{2\pi \left|\begin{smallmatrix} 4 & 1 \\ 1 & 4 \end{smallmatrix}\right|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}\left(\mathbf{x}_i - \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]\right)^T \frac{1}{15} \left[\begin{smallmatrix} 4 & -1 \\ -1 & 4 \end{smallmatrix}\right] \left(\mathbf{x}_i - \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]\right)}}{2\pi\sqrt{15}}$$

$$N(\mathbf{x}_i \mid \mathbf{u}_2, \Sigma_2) = \frac{e^{-\frac{1}{2}(\mathbf{x}_i - \mathbf{u}_2)^T \Sigma_2^{-1} (\mathbf{x}_i - \mathbf{u}_2)}}{(2\pi)^{\frac{2}{2}} \mid \Sigma_2 \mid^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}\left(\mathbf{x}_i - \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right]^{-1} \left(\mathbf{x}_i - \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]\right)}}{2\pi \left|\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}\left(\mathbf{x}_i - \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]\right)^T \frac{1}{4} \left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right] \left(\mathbf{x}_i - \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]\right)}}{2\pi\sqrt{4}}$$

Now, we may compute the $p(c_k, \mathbf{x}_i)$ probabilities:

$$p(c_1, \mathbf{x}_1) = N(\mathbf{x}_1 \mid \mathbf{u}_1, \Sigma_1)\pi_1 = \frac{e^{-\frac{1}{2}\left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]\right)^T \frac{1}{15} \left[\begin{smallmatrix} 4 & -1 \\ -1 & 4 \end{smallmatrix}\right] \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]\right)}}{2\pi\sqrt{15}} 0.5 = \frac{e^{-\frac{1}{30}\left[\begin{smallmatrix} -1 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 4 & -1 \\ -1 & 4 \end{smallmatrix}\right] \left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]}}{2\pi\sqrt{15}} 0.5 =$$

$$\frac{e^{-\frac{10}{30}}}{2\pi\sqrt{15}}0.5 \approx 0.015$$

$$p(c_1, \mathbf{x}_2) = N(\mathbf{x}_2 \mid \mathbf{u}_1, \Sigma_1)\pi_1 = \frac{e^{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}\right)^T \frac{1}{15}\begin{bmatrix}4&-1\\-1&4\end{bmatrix}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}\right)}}{2\pi\sqrt{15}}0.5 = \frac{e^{-\frac{1}{30}\begin{bmatrix}-2&3\end{bmatrix}\begin{bmatrix}4&-1\\-1&4\end{bmatrix}\begin{bmatrix}-2\\3\end{bmatrix}}}{2\pi\sqrt{15}}0.5 =$$

$$= \frac{e^{-\frac{64}{30}}}{2\pi\sqrt{15}}0.5 \approx 0.002$$

$$p(c_1, \mathbf{x}_3) = N(\mathbf{x}_3 \mid \mathbf{u}_1, \Sigma_1)\pi_1 = \frac{e^{-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}\right)^T \frac{1}{15}\begin{bmatrix}4&-1\\-1&4\end{bmatrix}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}\right)}}{2\pi\sqrt{15}}0.5 = \frac{e^{-\frac{1}{30}\begin{bmatrix}1&0\end{bmatrix}\begin{bmatrix}4&-1\\-1&4\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}}}{2\pi\sqrt{15}}0.5 =$$

$$= \frac{e^{-\frac{4}{30}}}{2\pi\sqrt{15}}0.5 \approx 0.018$$

$$p(c_2, \mathbf{x}_1) = N(\mathbf{x}_1 \mid \mathbf{u}_2, \Sigma_2)\pi_2 = \frac{e^{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)^T \frac{1}{4}\begin{bmatrix}2&0\\0&2\end{bmatrix}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)}}{2\pi\sqrt{4}}0.5 = \frac{e^{-\frac{1}{8}\begin{bmatrix}0&-1\end{bmatrix}\begin{bmatrix}2&0\\0&2\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}}}{4\pi}0.5 =$$

$$\frac{e^{-\frac{2}{8}}}{4\pi}0.5 \approx 0.031$$

$$p(c_2, \mathbf{x}_2) = N(\mathbf{x}_2 \mid \mathbf{u}_2, \Sigma_2)\pi_2 = \frac{e^{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)^T \frac{1}{4}\begin{bmatrix}2&0\\0&2\end{bmatrix}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)}}{2\pi\sqrt{4}}0.5 = \frac{e^{-\frac{1}{8}\begin{bmatrix}-1&1\end{bmatrix}\begin{bmatrix}2&0\\0&2\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix}}}{4\pi}0.5 =$$

$$\frac{e^{-\frac{4}{8}}}{4\pi}0.5 \approx 0.024$$

$$p(c_2, \mathbf{x}_3) = N(\mathbf{x}_3 \mid \mathbf{u}_2, \Sigma_2)\pi_2 = \frac{e^{-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)^T \frac{1}{4}\begin{bmatrix}2&0\\0&2\end{bmatrix}\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}\right)}}{2\pi\sqrt{4}}0.5 = \frac{e^{-\frac{1}{8}\begin{bmatrix}2&-2\end{bmatrix}\begin{bmatrix}2&0\\0&2\end{bmatrix}\begin{bmatrix}2\\-2\end{bmatrix}}}{4\pi}0.5 =$$

$$\frac{e^{-\frac{16}{8}}}{4\pi}0.5 \approx 0.005$$

Hence:

$$\sum_{k=1}^{2} p(c_k, \mathbf{x}_1) \approx 0.015 + 0.031 = 0.046$$

$$\sum_{k=1}^{2} p(c_k, \mathbf{x}_2) \approx 0.002 + 0.024 = 0.026$$

$$\sum_{k=1}^{2} p(c_k, \mathbf{x}_3) \approx 0.018 + 0.005 = 0.023$$

We may now compute the gammas:

$$\gamma_{11} = \frac{p(c_1, \mathbf{x}_1)}{\sum_k p(c_k, \mathbf{x}_1)} \approx \frac{0.015}{0.046} \approx 0.326$$

$$\gamma_{21} = \frac{p(c_2, \mathbf{x}_1)}{\sum_k p(c_k, \mathbf{x}_1)} \approx \frac{0.031}{0.046} \approx 0.674$$

$$\gamma_{12} = \frac{p(c_1, \mathbf{x}_2)}{\sum_k p(c_k, \mathbf{x}_2)} \approx \frac{0.002}{0.026} \approx 0.077$$

$$\gamma_{22} = \frac{p(c_2, \mathbf{x}_2)}{\sum_k p(c_k, \mathbf{x}_2)} \approx \frac{0.024}{0.026} \approx 0.923$$

$$\gamma_{13} = \frac{p(c_1, \mathbf{x}_3)}{\sum_k p(c_k, \mathbf{x}_3)} \approx \frac{0.018}{0.023} \approx 0.783$$

$$\gamma_{23} = \frac{p(c_2, \mathbf{x}_3)}{\sum_k p(c_k, \mathbf{x}_3)} \approx \frac{0.005}{0.023} \approx 0.217$$

Now let-s perform the M-step (maximization step) of the first epoch of the EM algorithm. Let's first calculate the sum of the weights for each class:

$$N_1 = \sum_{i=1}^{3} \gamma_{1i} = 0.326 + 0.077 + 0.783 = 1.186$$

$$N_2 = \sum_{i=1}^{3} \gamma_{2i} = 0.674 + 0.923 + 0.217 = 1.814$$

We may now update the Gaussian mixture using the computed weights and sum of weights:

$$\mathbf{u}_{1_{new}} = \frac{1}{N_1} \sum_{i=1}^{3} \gamma_{1i} \mathbf{x}_i = \frac{0.326 \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] + 0.077 \left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] + 0.783 \left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right]}{1.186} = \begin{bmatrix} 2.255 \\ -0.530 \end{bmatrix}$$

$$\mathbf{u}_{2_{new}} = \frac{1}{N_2} \sum_{i=1}^{3} \gamma_{2i} \mathbf{x}_i = \frac{0.674 \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] + 0.923 \left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] + 0.217 \left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right]}{1.814} = \begin{bmatrix} 0.730 \\ 0.898 \end{bmatrix}$$

$$\Sigma_{1_{new}} = \frac{1}{N_1} \sum_{i=1}^{3} \gamma_{1i} (\mathbf{x}_i - \mathbf{u}_{1_{new}}) \bullet (\mathbf{x}_i - \mathbf{u}_{1_{new}})^T =$$

$$= \frac{1}{1.186} \Big(0.326 \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.255 \\ -0.530 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.255 \\ -0.530 \end{smallmatrix}\right] \right)^T + 0.077 \left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.255 \\ -0.530 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.255 \\ -0.530 \end{smallmatrix}\right] \right)^T +$$

$$+ 0.783 \left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.255 \\ -0.530 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.255 \\ -0.530 \end{smallmatrix}\right] \right)^T \Big) \approx \begin{bmatrix} 1.130 & -0.784 \\ -0.784 & 0.639 \end{bmatrix}$$

$$\Sigma_{2_{new}} = \frac{1}{N_2} \sum_{i=1}^{3} \gamma_{2i} (\mathbf{x}_i - \mathbf{u}_{2_{new}}) \bullet (\mathbf{x}_i - \mathbf{u}_{2_{new}})^T =$$

$$= \frac{1}{1.814} \Big(0.674 \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.730 \\ 0.898 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.730 \\ 0.898 \end{smallmatrix}\right] \right)^T + 0.923 \left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.730 \\ 0.898 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.730 \\ 0.898 \end{smallmatrix}\right] \right)^T +$$

$$+0.217\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}0.730\\0.898\end{bmatrix}\right)\bullet\left(\begin{bmatrix}3\\-1\end{bmatrix}-\begin{bmatrix}0.730\\0.898\end{bmatrix}\right)^T\approx\begin{bmatrix}0.915 & -1.015\\-1.015 & 1.348\end{bmatrix}$$

$$\pi_{1_{new}}=p(c_1)=\frac{N_1}{\sum_k N_k}=\frac{1.186}{1.186+1.814}\approx 0.395$$

$$\pi_{2_{new}}=p(c_2)=\frac{N_2}{\sum_k N_k}=\frac{1.814}{1.186+1.814}\approx 0.605$$

Therefore, in the end of the first epoch, we have:

$$\mathbf{u}_1=\begin{bmatrix}2.255\\-0.530\end{bmatrix},\mathbf{u}_2=\begin{bmatrix}0.730\\0.898\end{bmatrix}\Sigma_1=\begin{bmatrix}1.130 & -0.784\\-0.784 & 0.639\end{bmatrix}\Sigma_2=\begin{bmatrix}0.915 & -1.015\\-1.015 & 1.348\end{bmatrix},\pi_1=0.395,\pi_2=0.605$$

Let's now perform the second epoch of the EM algorithm, starting with the Expectation/E-step. We first compute the normal distributions:

$$N(\mathbf{x}_i\mid\mathbf{u}_1,\Sigma_1)=\frac{e^{-\frac{1}{2}(\mathbf{x}_i-\mathbf{u}_1)^T\Sigma_1^{-1}(\mathbf{x}_i-\mathbf{u}_1)}}{(2\pi)^{\frac{2}{2}}\mid\Sigma_1\mid^{\frac{1}{2}}}=\frac{e^{-\frac{1}{2}\left(\mathbf{x}_i-\begin{bmatrix}2.255\\-0.530\end{bmatrix}\right)^T\begin{bmatrix}1.130 & -0.784\\-0.784 & 0.639\end{bmatrix}^{-1}\left(\mathbf{x}_i-\begin{bmatrix}2.255\\-0.530\end{bmatrix}\right)}}{2\pi\begin{vmatrix}1.130 & -0.784\\-0.784 & 0.639\end{vmatrix}^{\frac{1}{2}}}\approx$$

$$\approx\frac{e^{-\frac{1}{2}\left(\mathbf{x}_i-\begin{bmatrix}2.255\\-0.530\end{bmatrix}\right)^T\begin{bmatrix}6.023 & 7.397\\7.397 & 10.652\end{bmatrix}\left(\mathbf{x}_i-\begin{bmatrix}2.255\\-0.530\end{bmatrix}\right)}}{2\pi\sqrt{0.106}}$$

$$N(\mathbf{x}_i\mid\mathbf{u}_2,\Sigma_2)=\frac{e^{-\frac{1}{2}(\mathbf{x}_i-\mathbf{u}_2)^T\Sigma_2^{-1}(\mathbf{x}_i-\mathbf{u}_2)}}{(2\pi)^{\frac{2}{2}}\mid\Sigma_2\mid^{\frac{1}{2}}}=\frac{e^{-\frac{1}{2}\left(\mathbf{x}_i-\begin{bmatrix}0.730\\0.898\end{bmatrix}\right)^T\begin{bmatrix}0.915 & -1.015\\-1.015 & 1.348\end{bmatrix}^{-1}\left(\mathbf{x}_i-\begin{bmatrix}0.730\\0.898\end{bmatrix}\right)}}{2\pi\begin{vmatrix}0.915 & -1.015\\-1.015 & 1.348\end{vmatrix}^{\frac{1}{2}}}\approx$$

$$\approx\frac{e^{-\frac{1}{2}\left(\mathbf{x}_i-\begin{bmatrix}0.730\\0.898\end{bmatrix}\right)^T\begin{bmatrix}6.625 & 4.986\\4.986 & 4.494\end{bmatrix}\left(\mathbf{x}_i-\begin{bmatrix}0.730\\0.898\end{bmatrix}\right)}}{2\pi\sqrt{0.204}}$$

Now we calculate the probabilities:

$$p(c_1,\mathbf{x}_1)=N(\mathbf{x}_1\mid\mathbf{u}_1,\Sigma_1)\pi_1=\frac{e^{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}2.255\\-0.530\end{bmatrix}\right)^T\begin{bmatrix}6.023 & 7.397\\7.397 & 10.652\end{bmatrix}\left(\begin{bmatrix}1\\0\end{bmatrix}-\begin{bmatrix}2.255\\-0.530\end{bmatrix}\right)}}{2\pi\sqrt{0.106}}0.395=$$

$$=\frac{e^{-\frac{1}{2}\begin{bmatrix}-1.255 & 0.530\end{bmatrix}\begin{bmatrix}6.023 & 7.397\\7.397 & 10.652\end{bmatrix}\begin{bmatrix}-1.255\\0.530\end{bmatrix}}}{2\pi\sqrt{0.106}}0.395\approx\frac{e^{\frac{-2.637}{2}}}{2\pi\sqrt{0.106}}0.395\approx 0.052$$

$$p(c_1,\mathbf{x}_2)=N(\mathbf{x}_2\mid\mathbf{u}_1,\Sigma_1)\pi_1=\frac{e^{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}2.255\\-0.530\end{bmatrix}\right)^T\begin{bmatrix}6.023 & 7.397\\7.397 & 10.652\end{bmatrix}\left(\begin{bmatrix}0\\2\end{bmatrix}-\begin{bmatrix}2.255\\-0.530\end{bmatrix}\right)}}{2\pi\sqrt{0.106}}0.395=$$

$$=\frac{e^{-\frac{1}{2}\begin{bmatrix}-2.255 & 2.530\end{bmatrix}\begin{bmatrix}6.023 & 7.397\\7.397 & 10.652\end{bmatrix}\begin{bmatrix}-2.255\\2.530\end{bmatrix}}}{2\pi\sqrt{0.106}}0.395\approx\frac{e^{\frac{-14.400}{2}}}{2\pi\sqrt{0.106}}0.395\approx 0.000$$

$$p(c_1, \mathbf{x}_3) = N(\mathbf{x}_3 \mid \mathbf{u}_1, \Sigma_1)\pi_1 = \frac{e^{-\frac{1}{2}\left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right]-\left[\begin{smallmatrix} 2.255 \\ -0.530 \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 6.023 & 7.397 \\ 7.397 & 10.652 \end{smallmatrix}\right]\left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right]-\left[\begin{smallmatrix} 2.255 \\ -0.530 \end{smallmatrix}\right]\right)}}{2\pi\sqrt{0.106}} 0.395 =$$

$$= \frac{e^{-\frac{1}{2}\left[\begin{smallmatrix} 0.745 \\ -0.470 \end{smallmatrix}\right]\left[\begin{smallmatrix} 6.023 & 7.397 \\ 7.397 & 10.652 \end{smallmatrix}\right]\left[\begin{smallmatrix} 0.745 \\ -0.470 \end{smallmatrix}\right]}}{2\pi\sqrt{0.106}} 0.395 \approx \frac{e^{\frac{-0.515}{2}}}{2\pi\sqrt{0.106}} 0.395 \approx 0.149$$

$$p(c_2, \mathbf{x}_1) = N(\mathbf{x}_1 \mid \mathbf{u}_2, \Sigma_2)\pi_2 = \frac{e^{-\frac{1}{2}\left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]-\left[\begin{smallmatrix} 0.730 \\ 0.898 \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 6.625 & 4.986 \\ 4.986 & 4.494 \end{smallmatrix}\right]\left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]-\left[\begin{smallmatrix} 0.730 \\ 0.898 \end{smallmatrix}\right]\right)}}{2\pi\sqrt{0.204}} 0.605 =$$

$$= \frac{e^{-\frac{1}{2}\left[\begin{smallmatrix} 0.270 & -0.898 \end{smallmatrix}\right]\left[\begin{smallmatrix} 6.625 & 4.986 \\ 4.986 & 4.494 \end{smallmatrix}\right]\left[\begin{smallmatrix} 0.270 \\ -0.898 \end{smallmatrix}\right]}}{2\pi\sqrt{0.204}} 0.605 \approx \frac{e^{\frac{-1.689}{2}}}{2\pi\sqrt{0.204}} 0.605 \approx 0.092$$

$$p(c_2, \mathbf{x}_2) = N(\mathbf{x}_2 \mid \mathbf{u}_2, \Sigma_2)\pi_2 = \frac{e^{-\frac{1}{2}\left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right]-\left[\begin{smallmatrix} 0.730 \\ 0.898 \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 6.625 & 4.986 \\ 4.986 & 4.494 \end{smallmatrix}\right]\left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right]-\left[\begin{smallmatrix} 0.730 \\ 0.898 \end{smallmatrix}\right]\right)}}{2\pi\sqrt{0.204}} 0.605 =$$

$$= \frac{e^{-\frac{1}{2}\left[\begin{smallmatrix} -0.730 & 1.102 \end{smallmatrix}\right]\left[\begin{smallmatrix} 6.625 & 4.986 \\ 4.986 & 4.494 \end{smallmatrix}\right]\left[\begin{smallmatrix} -0.730 \\ 1.102 \end{smallmatrix}\right]}}{2\pi\sqrt{0.204}} 0.605 \approx \frac{e^{\frac{-0.966}{2}}}{2\pi\sqrt{0.204}} 0.605 \approx 0.132$$

$$p(c_2, \mathbf{x}_3) = N(\mathbf{x}_3 \mid \mathbf{u}_2, \Sigma_2)\pi_2 = \frac{e^{-\frac{1}{2}\left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right]-\left[\begin{smallmatrix} 0.730 \\ 0.898 \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 6.625 & 4.986 \\ 4.986 & 4.494 \end{smallmatrix}\right]\left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right]-\left[\begin{smallmatrix} 0.730 \\ 0.898 \end{smallmatrix}\right]\right)}}{2\pi\sqrt{0.204}} 0.605 =$$

$$= \frac{e^{-\frac{1}{2}\left[\begin{smallmatrix} 2.270 & -1.898 \end{smallmatrix}\right]\left[\begin{smallmatrix} 6.625 & 4.986 \\ 4.986 & 4.494 \end{smallmatrix}\right]\left[\begin{smallmatrix} 2.270 \\ -1.898 \end{smallmatrix}\right]}}{2\pi\sqrt{0.204}} 0.605 \approx \frac{e^{\frac{-7.364}{2}}}{2\pi\sqrt{0.204}} 0.605 \approx 0.005$$

We may now compute the new gammas:

$$\sum_{k=1}^{2} p(c_k, \mathbf{x}_1) \approx 0.052 + 0.092 = 0.144$$

$$\sum_{k=1}^{2} p(c_k, \mathbf{x}_2) \approx 0.000 + 0.132 = 0.132$$

$$\sum_{k=1}^{2} p(c_k, \mathbf{x}_3) \approx 0.149 + 0.005 = 0.154$$

We may now compute the gammas:

$$\gamma_{11} = \frac{p(c_1, \mathbf{x}_1)}{\sum_k p(c_k, \mathbf{x}_1)} \approx \frac{0.052}{0.144} \approx 0.361$$

$$\gamma_{21} = \frac{p(c_2, \mathbf{x}_1)}{\sum_k p(c_k, \mathbf{x}_1)} \approx \frac{0.92}{0.144} \approx 0.639$$

$$\gamma_{12} = \frac{p(c_1, \mathbf{x}_2)}{\sum_k p(c_k, \mathbf{x}_2)} \approx \frac{0.000}{0.132} \approx 0.000$$

$$\gamma_{22} = \frac{p(c_2, \mathbf{x}_2)}{\sum_k p(c_k, \mathbf{x}_2)} \approx \frac{0.132}{0.132} \approx 1.000$$

$$\gamma_{13} = \frac{p(c_1, \mathbf{x}_3)}{\sum_k p(c_k, \mathbf{x}_3)} \approx \frac{0.149}{0.154} \approx 0.968$$

$$\gamma_{23} = \frac{p(c_2, \mathbf{x}_3)}{\sum_k p(c_k, \mathbf{x}_3)} \approx \frac{0.005}{0.154} \approx 0.032$$

Now we're ready to do the second maximization step:

$$N_1 = \sum_{i=1}^{3} \gamma_{1i} = 0.361 + 0.000 + 0.968 = 1.329$$

$$N_2 = \sum_{i=1}^{3} \gamma_{2i} = 0.639 + 1.000 + 0.032 = 1.671$$

This will be the second update of the Gaussian mixture:

$$\mathbf{u}_{1_{new}} = \frac{1}{N_1} \sum_{i=1}^{3} \gamma_{1i} \mathbf{x}_i = \frac{0.361 \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] + 0.000 \left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] + 0.968 \left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right]}{1.329} = \begin{bmatrix} 2.457 \\ -0.728 \end{bmatrix}$$

$$\mathbf{u}_{2_{new}} = \frac{1}{N_2} \sum_{i=1}^{3} \gamma_{2i} \mathbf{x}_i = \frac{0.639 \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] + 1.000 \left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] + 0.032 \left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right]}{1.814} = \begin{bmatrix} 0.440 \\ 1.178 \end{bmatrix}$$

$$\Sigma_{1_{new}} = \frac{1}{N_1} \sum_{i=1}^{3} \gamma_{1i} (\mathbf{x}_i - \mathbf{u}_{1_{new}}) \bullet (\mathbf{x}_i - \mathbf{u}_{1_{new}})^T =$$

$$= \frac{1}{1.329} \Big(0.361 \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right] \right)^T + 0.000 \left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right] \right)^T +$$

$$+ 0.968 \left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right] \right)^T \Big) \approx \begin{bmatrix} 0.793 & -0.394 \\ -0.394 & 0.199 \end{bmatrix}$$

$$\Sigma_{2_{new}} = \frac{1}{N_2} \sum_{i=1}^{3} \gamma_{2i} (\mathbf{x}_i - \mathbf{u}_{2_{new}}) \bullet (\mathbf{x}_i - \mathbf{u}_{2_{new}})^T =$$

$$= \frac{1}{1.671} \Big(0.639 \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.440 \\ 1.178 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.440 \\ 1.178 \end{smallmatrix}\right] \right)^T + 1.000 \left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.440 \\ 1.178 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.440 \\ 1.178 \end{smallmatrix}\right] \right)^T +$$

$$+ 0.032 \left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.440 \\ 1.178 \end{smallmatrix}\right] \right) \bullet \left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.440 \\ 1.178 \end{smallmatrix}\right] \right)^T \Big) \approx \begin{bmatrix} 0.600 & -0.960 \\ -0.960 & 1.714 \end{bmatrix}$$

$$\pi_{1_{new}} = p(c_1) = \frac{N_1}{\sum_k N_k} = \frac{1.329}{1.329 + 1.671} \approx 0.443$$

$$\pi_{2_{new}} = p(c_2) = \frac{N_2}{\sum_k N_k} = \frac{1.671}{1.329 + 1.671} \approx 0.557$$

Therefore, in the end of the second epoch, we have:

$$\mathbf{u}_1 = \begin{bmatrix} 2.457 \\ -0.728 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 0.440 \\ 1.178 \end{bmatrix} \Sigma_1 = \begin{bmatrix} 0.793 & -0.394 \\ -0.394 & 0.199 \end{bmatrix} \Sigma_2 = \begin{bmatrix} 0.600 & -0.960 \\ -0.960 & 1.714 \end{bmatrix}, \pi_1 = 0.443, \pi_2 = 0.557$$

2

2.1 Hard assignment of observations using MAP assumption

To assign each observation using a Maximum a Posterior assumption, we should calculate the posteriors for each observation and class. According to the Bayes' rule:

$$p(c_k|\mathbf{x}_i) = \frac{p(x_i \mid c_k)p(c_k)}{p(\mathbf{x}_i)} \approx p(\mathbf{x}_i \mid c_k)p(c_k)$$

We already calculated the priors in the previous question:

$$p(c_1) = \pi_1 = 0.443, p(c_2) = \pi_2 = 0.557$$

Now let's compute the likelihoods for each observation and class (The likelihood is a multivariate Gaussian distribution, because we used a Gaussian Mixture with the EM algorithm).

$$p(\mathbf{x}_i \mid c_k) \frac{e^{-\frac{1}{2}(\mathbf{x}_i - \mathbf{u}_k)^T \Sigma_k^{-1} (\mathbf{x}_i - \mathbf{u}_k)}}{(2\pi)^{\frac{2}{2}} \mid \Sigma_1 \mid^{\frac{1}{2}}}$$

$$p(\mathbf{x}_1 \mid c_1) = \frac{e^{-\frac{1}{2}\left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 0.793 & -0.394 \\ -0.394 & 0.199 \end{smallmatrix}\right]^{-1} \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right]\right)}}{(2\pi)^{\frac{2}{2}} \left| \begin{smallmatrix} 0.793 & -0.394 \\ -0.394 & 0.199 \end{smallmatrix} \right|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}\left[\begin{smallmatrix} -1.457 & 0.728 \end{smallmatrix}\right] \left[\begin{smallmatrix} 77.402 & 153.248 \\ 153.248 & 308.440 \end{smallmatrix}\right] \left[\begin{smallmatrix} -1.457 \\ 0.728 \end{smallmatrix}\right]}}{2\pi\sqrt{0.002571}} \approx$$
$$\approx 0.821$$

$$p(\mathbf{x}_2 \mid c_1) = \frac{e^{-\frac{1}{2}\left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 0.793 & -0.394 \\ -0.394 & 0.199 \end{smallmatrix}\right]^{-1} \left(\left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right]\right)}}{(2\pi)^{\frac{2}{2}} \left| \begin{smallmatrix} 0.793 & -0.394 \\ -0.394 & 0.199 \end{smallmatrix} \right|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}\left[\begin{smallmatrix} -2.457 & 2.728 \end{smallmatrix}\right] \left[\begin{smallmatrix} 77.402 & 153.248 \\ 153.248 & 308.440 \end{smallmatrix}\right] \left[\begin{smallmatrix} -2.457 \\ 2.728 \end{smallmatrix}\right]}}{2\pi\sqrt{0.002571}} \approx$$
$$\approx 0.000$$

$$p(\mathbf{x}_3 \mid c_1) = \frac{e^{-\frac{1}{2}\left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 0.793 & -0.394 \\ -0.394 & 0.199 \end{smallmatrix}\right]^{-1} \left(\left[\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2.457 \\ -0.728 \end{smallmatrix}\right]\right)}}{(2\pi)^{\frac{2}{2}} \left| \begin{smallmatrix} 0.793 & -0.394 \\ -0.394 & 0.199 \end{smallmatrix} \right|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}\left[\begin{smallmatrix} 0.543 & -0.272 \end{smallmatrix}\right] \left[\begin{smallmatrix} 77.402 & 153.248 \\ 153.248 & 308.440 \end{smallmatrix}\right] \left[\begin{smallmatrix} 0.543 \\ -0.272 \end{smallmatrix}\right]}}{2\pi\sqrt{0.002571}} \approx$$
$$\approx 2.605$$

$$p(\mathbf{x}_1 \mid c_2) = \frac{e^{-\frac{1}{2}\left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.440 \\ 1.178 \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 0.600 & -0.960 \\ -0.960 & 1.714 \end{smallmatrix}\right]^{-1} \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 0.440 \\ 1.178 \end{smallmatrix}\right]\right)}}{(2\pi)^{\frac{2}{2}} \left| \begin{smallmatrix} 0.600 & -0.960 \\ -0.960 & 1.714 \end{smallmatrix} \right|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}\left[\begin{smallmatrix} 0.560 & -1.178 \end{smallmatrix}\right] \left[\begin{smallmatrix} 16.049 & 8.989 \\ 8.989 & 5.61 \end{smallmatrix}\right] \left[\begin{smallmatrix} 0.560 \\ -1.178 \end{smallmatrix}\right]}}{2\pi\sqrt{0.107}} \approx$$

$$\approx 0.300$$

$$p(\mathbf{x}_2 \mid c_2) = \frac{e^{-\frac{1}{2}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.440 \\ 1.178 \end{bmatrix}\right)^T \begin{bmatrix} 0.600 & -0.960 \\ -0.960 & 1.714 \end{bmatrix}^{-1}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.440 \\ 1.178 \end{bmatrix}\right)}}{(2\pi)^{\frac{2}{2}} \begin{vmatrix} 0.600 & -0.960 \\ -0.960 & 1.714 \end{vmatrix}^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}[\,-0.440\ 0.822\,]\begin{bmatrix} 16.049 & 8.989 \\ 8.989 & 5.61 \end{bmatrix}[\,-0.440\ 0.822\,]}}{2\pi\sqrt{0.107}} \approx$$

$$\approx 0.398$$

$$p(\mathbf{x}_3 \mid c_2) = \frac{e^{-\frac{1}{2}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.440 \\ 1.178 \end{bmatrix}\right)^T \begin{bmatrix} 0.600 & -0.960 \\ -0.960 & 1.714 \end{bmatrix}^{-1}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.440 \\ 1.178 \end{bmatrix}\right)}}{(2\pi)^{\frac{2}{2}} \begin{vmatrix} 0.600 & -0.960 \\ -0.960 & 1.714 \end{vmatrix}^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}[\,2.560\ -2.178\,]\begin{bmatrix} 16.049 & 8.989 \\ 8.989 & 5.61 \end{bmatrix}\begin{bmatrix} 2.560 \\ -2.178 \end{bmatrix}}}{2\pi\sqrt{0.107}} \approx$$

$$\approx 0.000$$

We may now compute the posteriors to assign a class to each observation:

$$P(c_1|\mathbf{x}_1) \approx p(\mathbf{x}_1|c_1)p(c_1) = 0.821 \cdot 0.443 \approx 0.364$$

$$P(c_2|\mathbf{x}_1) \approx p(\mathbf{x}_1|c_2)p(c_2) = 0.300 \cdot 0.557 \approx 0.167$$

As $P(c_1|\mathbf{x}_1) > P(c_2|\mathbf{x}_1)$, we assign $\mathbf{x}_1$ to cluster c_1.

$$P(c_1|\mathbf{x}_2) \approx p(\mathbf{x}_2|c_1)p(c_1) = 0.000 \cdot 0.443 = 0.000$$

$$P(c_2|\mathbf{x}_2) \approx p(\mathbf{x}_2|c_2)p(c_2) = 0.398 \cdot 0.557 \approx 0.222$$

As $P(c_2|\mathbf{x}_2) > P(c_1|\mathbf{x}_2)$, we assign $\mathbf{x}_2$ to cluster c_2.

$$P(c_1|\mathbf{x}_3) \approx p(\mathbf{x}_3|c_1)p(c_1) = 2.605 \cdot 0.443 \approx 1.154$$

$$P(c_2|\mathbf{x}_3) \approx p(\mathbf{x}_3|c_2)p(c_2) = 0.000 \cdot 0.557 = 0.000$$

As $P(c_1|\mathbf{x}_3) > P(c_2|\mathbf{x}_3)$, we assign $\mathbf{x}_3$ to cluster c_1.

2.2 Silhouette of the larger cluster

The larger cluster as determined in the previous question is cluster c_1, with 2 observations. The silhouette of a cluster is the average of the silhouettes of each observation in that cluster, or:

$$s(c_k) = \frac{1}{|c_k|} \sum_{i\,:\,\mathbf{x}_i \text{ is in cluster } c_k} s(\mathbf{x}_i)$$

So we have:

$$s(c_1) = \frac{1}{2}(s(\mathbf{x}_1) + s(\mathbf{x}_3))$$

We'll use the Euclidean Distance to compute the silhouette of each observation, so let's first compute the Euclidean Distances for each pair of observations:

$$ED(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{\sum_i (u_i - v_i)^2}$$

ED	$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$
$\mathbf{x_1}$	0	$\sqrt{5} \approx 2.236$	$\sqrt{5} \approx 2.236$
$\mathbf{x_2}$	$\sqrt{5} \approx 2.236$	0	$\sqrt{18} \approx 4.243$
$\mathbf{x_3}$	$\sqrt{5} \approx 2.236$	$\sqrt{18} \approx 4.243$	0

Now, the silhouette of an observation in a cluster can be calculated as (if $a(\mathbf{x_i}) \leq b(\mathbf{x_i})$:

$$s(\mathbf{x_i}) = 1 - \frac{a(\mathbf{x_i})}{b(\mathbf{x_i})}$$

where $a(\mathbf{x_i})$ is the average distance of x_i to the other points in the same cluster and $b(\mathbf{x_i})$ is the minimum over the other clusters of the average distance between $\mathbf{x_i}$ and the points of that cluster. That is, if x_i is in cluster c_k:

$$a(\mathbf{x_i}) = \frac{1}{|c_k| - 1} \sum_{\mathbf{x_j} \in c_k, j \neq i} ED(\mathbf{x_i}, \mathbf{x_j})$$

$$b(\mathbf{x_i}) = \min_{l \neq k} \frac{1}{|c_l|} \sum_{\mathbf{x_j} \in c_l} ED(\mathbf{x_i}, \mathbf{x_j})$$

To compute $s(x_1)$, we may note that the only other point in the same cluster is x_3 and the only point in a different cluster is x_2. Therefore,

$$a(\mathbf{x_1}) = \frac{1}{2 - 1} ED(\mathbf{x_1}, \mathbf{x_3}) = 2.236$$

$$b(\mathbf{x_1}) = \frac{1}{1} ED(\mathbf{x_1}, \mathbf{x_2}) = 2.236$$

Hence, as $a(\mathbf{x_1}) \leq b(\mathbf{x_1})$, we get:

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

To compute $s(x_3)$, we may note as well that the only other point in the same cluster is x_1 and the only point in a different cluster is x_2. Therefore,

$$a(\mathbf{x_3}) = \frac{1}{2-1} ED(\mathbf{x_3}, \mathbf{x_1}) = 2.236$$

$$b(\mathbf{x_3}) = \frac{1}{1} ED(\mathbf{x_3}, \mathbf{x_2}) = 4.243$$

Hence, as $a(\mathbf{x_3}) \leq b(\mathbf{x_3})$, we get:

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

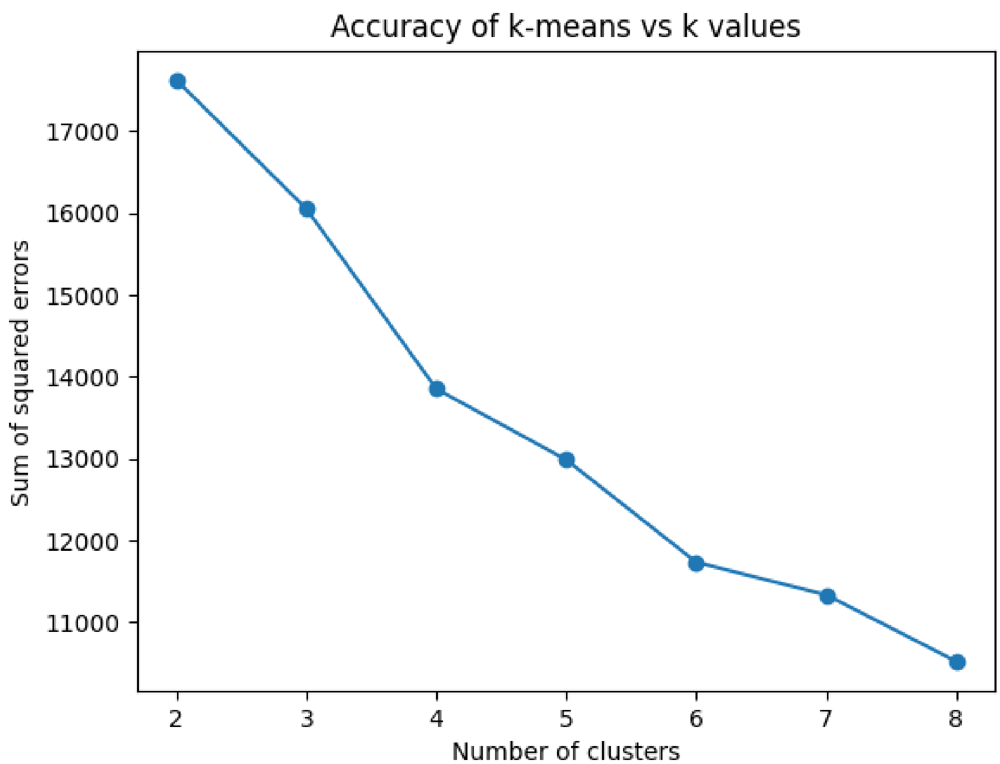
And finally we may compute the silhouette of cluster c_1:

$$s(c_1) = \frac{1}{2}(s(\mathbf{x}_1) + s(\mathbf{x}_3)) = \frac{0 + 0.473}{2} \approx 0.237$$

Part 2: Programming

1 Normalized data with MinMaxScaler

1.1 k-means clustering with different numbers of clusters



1.2 Discussing how many clusters there should be

By analyzing our plot, we can see that there are significant drops in the SSE from 2 to 3($\approx$1600) and 3 to 4($\approx$2200) clusters, but from then on the changes aren't as significant(4 to 5 has a difference of around 865). This can indicate that even though more clusters get can lower sum of squared errors, they are also costlier to compute, and the benefits aren't as significant.

Taking this into account, choosing 4 clusters seems to be a good compromise, as this simpler model should yield reasonably good clusters. More complex clustering, that could separate our population with less inertia, would also probably deal with relations that aren't obvious, and complicate our desired customer segments.

1.3 Would k-modes be a better clustering approach?

The 8 features selected were split into numerical: age and balance, categorical: marital status, education and job, and binary: housing, loan and default.

After applying pandas' get_dummies() function, we get a data frame with the same 2 numerical features and 19 binary features.

The algorithms mentioned, k-means and k-modes operate differently, and they handle some data types better than others. K-means calculates distances to the centroid of the clusters using Euclidean distances, which assumes a numerical feature space. Therefore, it may identify a continuous relation between binary features, instead of checking if they match or not.

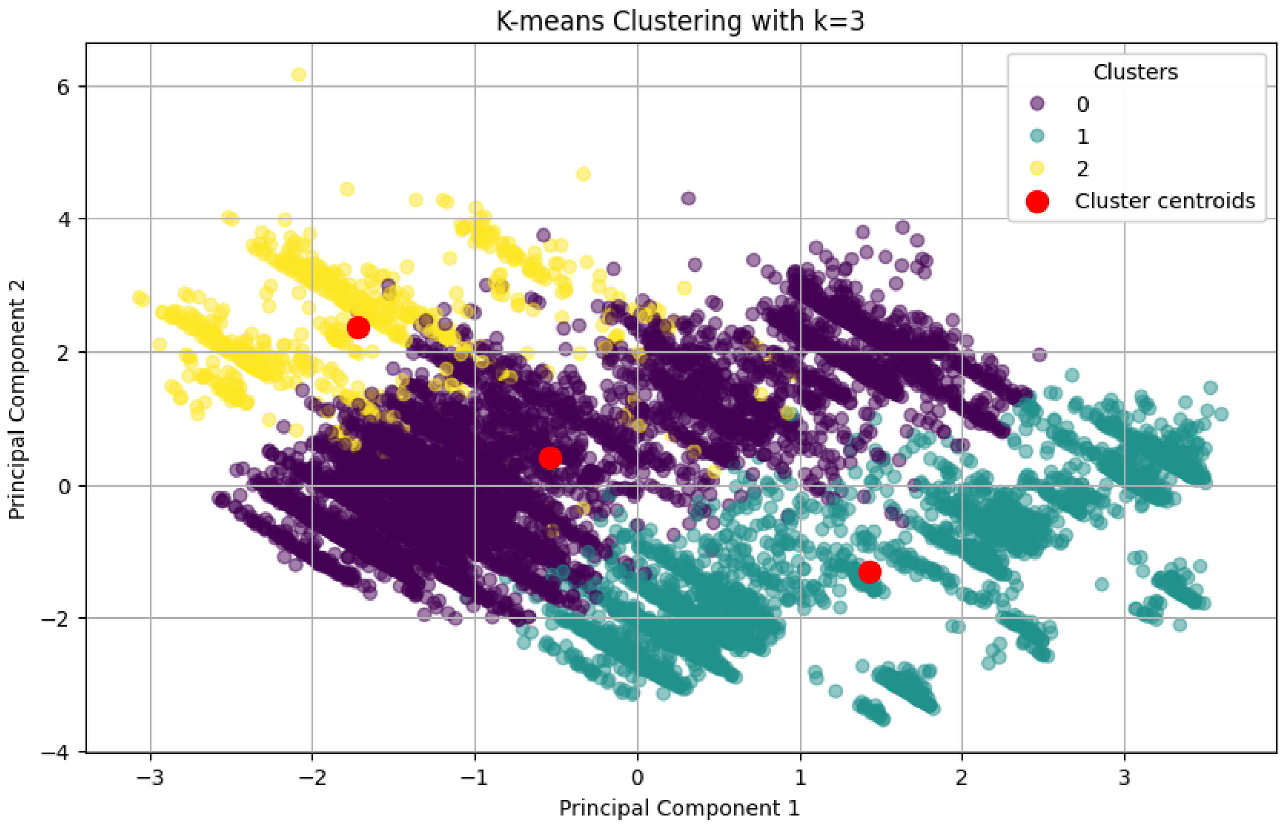
However, the K-modes approach is better for categorical and binary data, as it calculates distances using the number of mismatches, similarly to Hamming distance. This method should yield more significant clusters, because it doesn't presume binary features to be continuous, so it doesn't misinterpret the distance between points, unlike the k-means approach. Based on the predominance of binary features in our data, a k-modes approach would likely result in more meaningful clusters, which respect the nature of our binary data.

2 Normalized data with StandardScaler

2.1 Applying PCA to the data

The amount of variability explained by the top two components after applying PCA to the data is $\approx 22.75\%$. This shows that these features capture a significant amount of the dataset's variability, making them effective for dimensionality reduction and preserving its information.

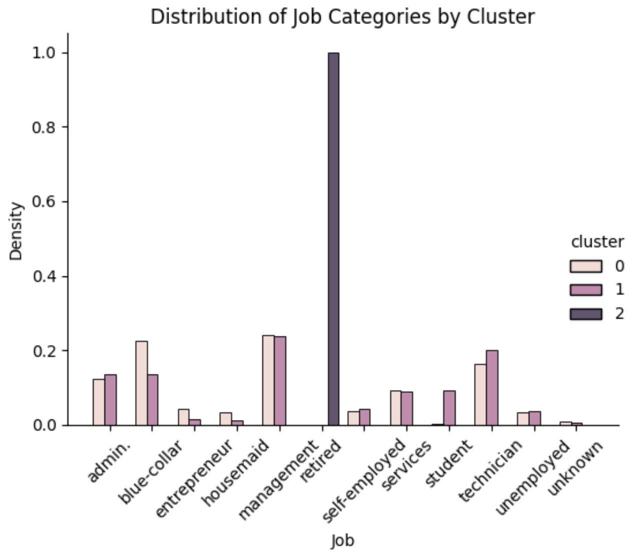
2.2 K-means clustering with k=3 and scatter plot according to PCA

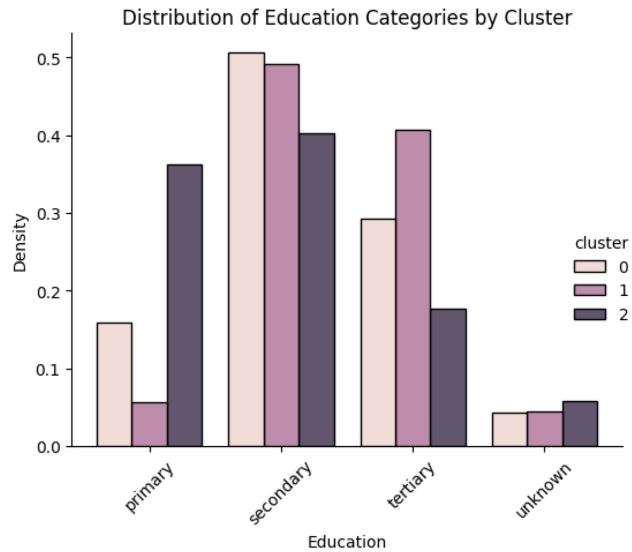


The scatter plot above shows clearly separated clusters, with little overlapping, as we can see 3 well defined regions, colored yellow, purple and blue. This means that the kmeans algorithm identified patterns in the data that represent 3 different customer segments, which can be analyzed separately.

However, the three regions are close to each other and there are some outlier points, which show that our approach was not perfect. This might mean that our data is not perfectly separable or that three clusters is not the optimal choice.

2.3 Cluster conditional features





Visually analyzing the plots above, we can draw some observations:

Cluster 0 has mostly individuals in blue-collar jobs, management or technicians, and mostly with secondary or tertiary education. Note that there are practically no students in this cluster, indicating that it represents individuals with mid to high level job positions and good educations.

Cluster 1 is more evenly distributed among jobs, with a slight focus on management and technicians too. Regarding education levels, they are similar to the previous cluster, mostly secondary and tertiary. We can assume that this cluster also represents individuals with jobs and education like those of cluster 0.

Cluster 2 contains all retired individuals, without any other job representation. The education plot for this cluster represents the education levels of retirees, which are mostly primary and secondary, and about 20 percent tertiary. This cluster seems to be entirely dedicated to retirees.

Both cluster 0 and 1 seem to represent the general population, composed mostly of mid to high level jobs, with mostly secondary or tertiary education, and the features that distinguish them apart are probably not the job or education level, but rather some other feature in the dataset. Cluster 2 represents retired people, which are group of the population with generally lower levels of education. This cluster is likely mostly defined by only the job as it represents a perfect split of the population.