Aprendizagem 2024/25

Homework IV – Group 028

(ist1106987, ist1107057)

I. Pen-and-paper

1)

$$\gamma_{ki} = p(c_k|\, x_i) = \frac{p(c_k, x_i)}{p(x_i)} = \frac{p(x_i|\, c_k)\, p(c_k)}{p(x_i)} = \frac{N(x_i|\mu_k, \Sigma_k)\, \pi_k}{\sum_{k=1}^{K} p(c_k, x_i)}$$

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

Epoch 1

E-step:

For observation x_1:

1. For cluster c_1:

$$\text{Prior: } p(c_1) = \pi_1 = 0.5$$

$$p(x_1|\, c_1) = N(x_1|\mu_1, \Sigma_1) = \frac{1}{2\pi \times \sqrt{15}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2\\-1\end{bmatrix}\right)^T \begin{bmatrix}4 & 1\\1 & 4\end{bmatrix}^{-1} \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2\\-1\end{bmatrix}\right)\right\} =$$
$$\frac{1}{2\pi \times \sqrt{15}} \exp\left\{-\frac{1}{2}\begin{bmatrix}-1 & 1\end{bmatrix} \begin{bmatrix}4/15 & -1/15\\-1/15 & 4/15\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix}\right\} = \frac{1}{2\pi \times \sqrt{15}} \exp\left\{-\frac{1}{3}\right\} \approx 0.029$$

$$p(c_1, x_1) = 0.029 \times 0.5 \approx 0.015$$

2. For cluster c_2:

$$\text{Prior: } p(c_2) = \pi_2 = 0.5$$

$$p(x_1|\, c_2) = N(x_1|\mu_2, \Sigma_2) = \frac{1}{2\pi \times \sqrt{4}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}\right)^T \begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}^{-1} \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}\right)\right\} =$$
$$\frac{1}{2\pi \times 2} \exp\left\{-\frac{1}{2}\begin{bmatrix}0 & -1\end{bmatrix} \begin{bmatrix}1/2 & 0\\0 & 1/2\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}\right\} = \frac{1}{2\pi \times 2} \exp\left\{-\frac{1}{4}\right\} \approx 0.062$$

$$p(c_2, x_1) = 0.062 \times 0.5 = 0.031$$

Therefore,

$$\gamma_{11} = \frac{0.015}{0.015 + 0.031} \approx 0.326$$
$$\gamma_{21} = \frac{0.031}{0.015 + 0.031} \approx 0.674$$

For observation x_2:

1. For cluster c_1:

$$\text{Prior: } p(c_1) = \pi_1 = 0.5$$

$$p(x_2|\, c_1) = N(x_2|\mu_1, \Sigma_1) = \frac{1}{2\pi \times \sqrt{15}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2\\-1\end{bmatrix}\right)^T \begin{bmatrix}4 & 1\\1 & 4\end{bmatrix}^{-1} \left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2\\-1\end{bmatrix}\right)\right\}$$
$$= \frac{1}{2\pi \times \sqrt{15}} \exp\left\{-\frac{1}{2}\begin{bmatrix}-2 & 3\end{bmatrix} \begin{bmatrix}4/15 & -1/15\\-1/15 & 4/15\end{bmatrix}\begin{bmatrix}-2\\3\end{bmatrix}\right\}$$
$$= \frac{1}{2\pi \times \sqrt{15}} \exp\left\{-\frac{32}{15}\right\} \approx 0.005$$

$$p(c_1, x_2) = 0.005 \times 0.5 \approx 0.003$$

2. For cluster c_2:

$$\text{Prior: } p(c_2) = \pi_2 = 0.5$$

$$p(x_2|\, c_2) = N(x_2|\mu_2, \Sigma_2) = \frac{1}{2\pi \times \sqrt{4}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}\right)^T \begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}^{-1} \left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}\right)\right\}$$

$$= \frac{1}{2\pi \times 2} \exp\left\{-\frac{1}{2}[-1 \quad 1]\begin{bmatrix}1/2 & 0\\0 & 1/2\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix}\right\} = \frac{1}{2\pi \times 2}\exp\left\{-\frac{1}{2}\right\} \approx 0.049$$

$$p(c_2, x_2) = 0.049 \times 0.5 = 0.024$$

Therefore,

$$\gamma_{12} = \frac{0.003}{0.003 + 0.024} \approx 0.111$$

$$\gamma_{22} = \frac{0.024}{0.003 + 0.024} \approx 0.889$$

For observation x_3:

1. For cluster c_1:

$$\text{Prior: } p(c_1) = \pi_1 = 0.5$$

$$p(x_3|\, c_1) = N(x_3|\mu_1, \Sigma_1) = \frac{1}{2\pi \times \sqrt{15}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2\\-1\end{bmatrix}\right)^T \begin{bmatrix}4 & 1\\1 & 4\end{bmatrix}^{-1} \left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2\\-1\end{bmatrix}\right)\right\}$$

$$= \frac{1}{2\pi \times \sqrt{15}} \exp\left\{-\frac{1}{2}[1 \quad 0]\begin{bmatrix}4/15 & -1/15\\-1/15 & 4/15\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\right\}$$

$$= \frac{1}{2\pi \times \sqrt{15}} \exp\left\{-\frac{2}{15}\right\} \approx 0.036$$

$$p(c_1, x_3) = 0.036 \times 0.5 \approx 0.018$$

2. For cluster c_2:

$$\text{Prior: } p(c_2) = \pi_2 = 0.5$$

$$p(x_3|\, c_2) = N(x_3|\mu_2, \Sigma_2) = \frac{1}{2\pi \times \sqrt{4}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}\right)^T \begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}^{-1} \left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}\right)\right\}$$

$$= \frac{1}{2\pi \times 2} \exp\left\{-\frac{1}{2}[2 \quad -2]\begin{bmatrix}1/2 & 0\\0 & 1/2\end{bmatrix}\begin{bmatrix}2\\-2\end{bmatrix}\right\} = \frac{1}{2\pi \times 2}\exp\{-2\} \approx 0.011$$

$$p(c_2, x_3) = 0.011 \times 0.5 \approx 0.006$$

Therefore,

$$\gamma_{13} = \frac{0.018}{0.018 + 0.006} \approx 0.750$$

$$\gamma_{23} = \frac{0.006}{0.018 + 0.006} \approx 0.250$$

M-Step:

For cluster c_1:

$$N_1 = \sum_{i=1}^{3} \gamma_{1i} = 0.326 + 0.111 + 0.750 = 1.187$$

$$\mu_1 = \frac{1}{N_1}\sum_{i=1}^{3}(\gamma_{1i} \cdot x_i) = \frac{1}{1.187}\left(0.326\begin{bmatrix}1\\0\end{bmatrix} + 0.111\begin{bmatrix}0\\2\end{bmatrix} + 0.750\begin{bmatrix}3\\-1\end{bmatrix}\right) = \begin{bmatrix}2.170\\-0.445\end{bmatrix}$$

$$\Sigma_1 = \frac{1}{N_1}\sum_{i=1}^{3}[\gamma_{1i} \cdot (x_i - \mu_1) \cdot (x_i - \mu_1)^T] =$$

$$= \frac{1}{1.187}\left\{0.326\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)^T + 0.111\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)^T + 0.750\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)^T\right\} =$$

$$= \frac{1}{1.187}\left\{0.326\begin{bmatrix}-1.170\\0.445\end{bmatrix}\begin{bmatrix}-1.170 & 0.445\end{bmatrix} + 0.111\begin{bmatrix}-2.170\\2.445\end{bmatrix}\begin{bmatrix}-2.170 & 2.445\end{bmatrix} + 0.750\begin{bmatrix}0.830\\-0.555\end{bmatrix}\begin{bmatrix}0.830 & -0.555\end{bmatrix}\right\} =$$

$$= \frac{1}{1.187}\left(\begin{bmatrix}0.446 & -0.170\\-0.170 & 0.065\end{bmatrix} + \begin{bmatrix}0.523 & -0.589\\-0.589 & 0.664\end{bmatrix} + \begin{bmatrix}0.517 & -0.345\\-0.345 & 0.231\end{bmatrix}\right) = \begin{bmatrix}1.252 & -0.930\\-0.930 & 0.809\end{bmatrix}$$

For cluster c_2:

$$N_2 = \sum_{i=1}^{3}\gamma_{2i} = 0.674 + 0.889 + 0.250 = 1.813$$

$$\mu_2 = \frac{1}{N_2}\sum_{i=1}^{3}(\gamma_{2i} \cdot x_i) = \frac{1}{1.813}\left(0.674\begin{bmatrix}1\\0\end{bmatrix} + 0.889\begin{bmatrix}0\\2\end{bmatrix} + 0.250\begin{bmatrix}3\\-1\end{bmatrix}\right) = \begin{bmatrix}0.785\\0.843\end{bmatrix}$$

$$\Sigma_2 = \frac{1}{N_2}\sum_{i=1}^{3}\gamma_{2i} \cdot (x_i - \mu_2) \cdot (x_i - \mu_2)^T =$$

$$= \frac{1}{1.813}\left\{0.674\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)^T + 0.889\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)^T + 0.250\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)^T\right\} =$$

$$= \frac{1}{1.813}\left\{0.674\begin{bmatrix}0.215\\-0.843\end{bmatrix}\begin{bmatrix}0.215 & -0.843\end{bmatrix} + 0.889\begin{bmatrix}-0.785\\1.157\end{bmatrix}\begin{bmatrix}-0.785 & 1.157\end{bmatrix} + 0.250\begin{bmatrix}2.215\\-1.843\end{bmatrix}\begin{bmatrix}2.215 & -1.843\end{bmatrix}\right\} =$$

$$= \frac{1}{1.813}\left(\begin{bmatrix}0.031 & -0.122\\-0.122 & 0.479\end{bmatrix} + \begin{bmatrix}0.548 & -0.807\\-0.807 & 1.190\end{bmatrix} + \begin{bmatrix}1.227 & -1.021\\-1.021 & 0.849\end{bmatrix}\right) = \begin{bmatrix}0.996 & -0.669\\-0.669 & 1.389\end{bmatrix}$$

$$\pi_1 = p(c_1) = \frac{N_1}{N} = \frac{1.187}{1.187 + 1.813} = 0.396$$

$$\pi_2 = p(c_2) = \frac{N_2}{N} = \frac{1.813}{1.187 + 1.813} = 0.604$$

Epoch 2

E-step:

For observation x_1:

3. For cluster c_1:

Prior: $p(c_1) = \pi_1 = 0.396$

$$p(x_1|\, c_1) = N(x_1|\mu_1, \Sigma_1) =$$

$$= \frac{1}{2\pi \times \sqrt{0{,}148}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)^T \begin{bmatrix}1.252 & -0.930\\-0.930 & 0.809\end{bmatrix}^{-1} \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)\right\}$$

$$= \frac{1}{2\pi \times \sqrt{0{,}148}} \exp\left\{-\frac{1}{2}\begin{bmatrix}-1.170 & 0.445\end{bmatrix}\begin{bmatrix}5.467 & 6.285\\6.285 & 8.461\end{bmatrix}\begin{bmatrix}-1.170\\0.445\end{bmatrix}\right\}$$

$$= \frac{1}{2\pi \times \sqrt{0{,}148}} \exp\{-1.307\} \approx 0.112$$

$$p(c_1, x_1) = 0.112 \times 0.396 \approx 0.044$$

4. For cluster c_2:

$$\text{Prior: } p(c_2) = \pi_2 = 0.604$$

$$p(x_1|\, c_2) = N(x_1|\mu_2, \Sigma_2)$$

$$= \frac{1}{2\pi \times \sqrt{0{,}936}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)^T \begin{bmatrix}0.996 & -0.669\\-0.669 & 1.389\end{bmatrix}^{-1} \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)\right\}$$

$$= \frac{1}{2\pi \times \sqrt{0{,}936}} \exp\left\{-\frac{1}{2}\begin{bmatrix}0.215 & -0.843\end{bmatrix}\begin{bmatrix}1.484 & 0.715\\0.715 & 1.064\end{bmatrix}\begin{bmatrix}0.215\\-0.843\end{bmatrix}\right\}$$

$$= \frac{1}{2\pi \times \sqrt{0{,}936}} \exp\{-0.283\} \approx 0.124$$

$$p(c_2, x_1) = 0.124 \times 0.604 \approx 0.075$$

Therefore,

$$\gamma_{11} = \frac{0.044}{0.044 + 0.075} \approx 0.370$$

$$\gamma_{21} = \frac{0.075}{0.044 + 0.075} \approx 0.630$$

For observation x_2:

3. For cluster c_1:

$$\text{Prior: } p(c_1) = \pi_1 = 0.396$$

$$p(x_2|\, c_1) = N(x_2|\mu_1, \Sigma_1) =$$

$$= \frac{1}{2\pi \times \sqrt{0{,}148}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)^T \begin{bmatrix}1.252 & -0.930\\-0.930 & 0.809\end{bmatrix}^{-1} \left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)\right\}$$

$$= \frac{1}{2\pi \times \sqrt{0{,}148}} \exp\left\{-\frac{1}{2}\begin{bmatrix}-2.170 & 2.445\end{bmatrix}\begin{bmatrix}5.467 & 6.285\\6.285 & 8.461\end{bmatrix}\begin{bmatrix}-2.170\\2.445\end{bmatrix}\right\}$$

$$= \frac{1}{2\pi \times \sqrt{0{,}148}} \exp\{-4.816\} \approx 0.003$$

$$p(c_1, x_2) = 0.003 \times 0.396 \approx 0.001$$

4. For cluster c_2:

$$\text{Prior: } p(c_2) = \pi_2 = 0.604$$

$$p(x_2|\ c_2) = \ N(x_2|\mu_2, \Sigma_2)$$

$$= \frac{1}{2\pi\sqrt{0{,}936}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)^T \begin{bmatrix}0.996 & -0.669\\-0.669 & 1.389\end{bmatrix}^{-1} \left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)\right\}$$

$$= \frac{1}{2\pi\sqrt{0{,}936}} \exp\left\{-\frac{1}{2}\begin{bmatrix}-0.785 & 1.157\end{bmatrix} \begin{bmatrix}1.484 & 0.715\\0.715 & 1.064\end{bmatrix} \begin{bmatrix}-0.785\\1.157\end{bmatrix}\right\}$$

$$= \frac{1}{2\pi\sqrt{0{,}936}} \exp\{-0.520\} \ \approx 0.098$$

$$p(c_2, x_2) = 0.098 \ \times 0.604 \ \approx 0.059$$

Therefore,

$$\gamma_{12} = \frac{0.001}{0.001 + 0.059} \approx 0.017$$

$$\gamma_{22} = \frac{0.059}{0.001 \ + 0.059} \approx 0.983$$

For observation x_3:

3. For cluster c_1:

$$\text{Prior: } p(c_1) = \ \pi_1 = 0.396$$

$$p(x_3|\ c_1) = \ N(x_3|\mu_1, \Sigma_1) =$$

$$= \frac{1}{2\pi \times \sqrt{0{,}148}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)^T \begin{bmatrix}1.252 & -0.930\\-0.930 & 0.809\end{bmatrix}^{-1} \left(\begin{bmatrix}3\\-1\end{bmatrix}\right.\right.$$

$$\left.\left. - \begin{bmatrix}2.170\\-0.445\end{bmatrix}\right)\right\}$$

$$= \frac{1}{2\pi \times \sqrt{0{,}148}} \exp\left\{-\frac{1}{2}\begin{bmatrix}0.830 & -0.555\end{bmatrix} \begin{bmatrix}5.467 & 6.285\\6.285 & 8.461\end{bmatrix} \begin{bmatrix}0.830\\-0.555\end{bmatrix}\right\}$$

$$= \frac{1}{2\pi \times \sqrt{0{,}148}} \exp\{-0.291\} \ \approx 0.309$$

$$p(c_1, x_3) = 0.309 \ \times 0.396 \ \approx 0.122$$

4. For cluster c_2:

$$\text{Prior: } p(c_2) = \ \pi_2 = 0.604$$

$$p(x_3|\ c_2) = \ N(x_3|\mu_2, \Sigma_2)$$

$$= \frac{1}{2\pi\sqrt{0{,}936}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)^T \begin{bmatrix}0.996 & -0.669\\-0.669 & 1.389\end{bmatrix}^{-1} \left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}0.785\\0.843\end{bmatrix}\right)\right\}$$

$$= \frac{1}{2\pi\sqrt{0{,}936}} \exp\left\{-\frac{1}{2}\begin{bmatrix}2.215 & -1.843\end{bmatrix} \begin{bmatrix}1.484 & 0.715\\0.715 & 1.064\end{bmatrix} \begin{bmatrix}2.215\\-1.843\end{bmatrix}\right\}$$

$$= \frac{1}{2\pi\sqrt{0{,}936}} \exp\{-2.529\} \ \approx 0.013$$

$$p(c_2, x_3) = 0.013 \times 0.604 \approx 0.008$$

Therefore,

$$\gamma_{13} = \frac{0.122}{0.122 + 0.008} \approx 0.938$$

$$\gamma_{23} = \frac{0.008}{0.122 + 0.008} \approx 0.062$$

M-Step:

For cluster c_1:

$$N_1 = \sum_{i=1}^{3} \gamma_{1i} = 0.370 + 0.017 + 0.938 = 1.325$$

$$\mu_1 = \frac{1}{N_1}\sum_{i=1}^{3}(\gamma_{1i} \cdot x_i) = \frac{1}{1.325}\left(0.370\begin{bmatrix}1\\0\end{bmatrix} + 0.017\begin{bmatrix}0\\2\end{bmatrix} + 0.938\begin{bmatrix}3\\-1\end{bmatrix}\right) = \begin{bmatrix}2.403\\-0.682\end{bmatrix}$$

$$\Sigma_1 = \frac{1}{N_1}\sum_{i=1}^{3}\gamma_{1i} \cdot (x_i - \mu_1) \cdot (x_i - \mu_1)^T =$$

$$= \frac{1}{1.325}\left\{0.370\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.403\\-0.682\end{bmatrix}\right)\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.403\\-0.682\end{bmatrix}\right)^T + 0.017\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.403\\-0.682\end{bmatrix}\right)\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.403\\-0.682\end{bmatrix}\right)^T + 0.938\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2.403\\-0.682\end{bmatrix}\right)\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}2.403\\-0.682\end{bmatrix}\right)^T\right\} =$$

$$= \frac{1}{1.325}\left\{0.370\begin{bmatrix}-1.403\\0.682\end{bmatrix}\begin{bmatrix}-1.403 & 0.682\end{bmatrix} + 0.017\begin{bmatrix}-2.403\\2.682\end{bmatrix}\begin{bmatrix}-2.403 & 2.682\end{bmatrix} + 0.938\begin{bmatrix}0.597\\-0.318\end{bmatrix}\begin{bmatrix}0.597 & -0.318\end{bmatrix}\right\} =$$

$$= \frac{1}{1.325}\left(\begin{bmatrix}0.728 & -0.354\\-0.354 & 0.172\end{bmatrix} + \begin{bmatrix}0.098 & -0.110\\-0.110 & 0.122\end{bmatrix} + \begin{bmatrix}0.334 & -0.178\\-0.178 & 0.095\end{bmatrix}\right) = \begin{bmatrix}0.875 & -0.485\\-0.485 & 0.294\end{bmatrix}$$

For cluster c_2:

$$N_2 = \sum_{i=1}^{3} \gamma_{2i} = 0.630 + 0.983 + 0.062 = 1.675$$

$$\mu_2 = \frac{1}{N_2}\sum_{i=1}^{3}(\gamma_{2i} \cdot x_i) = \frac{1}{1.675}\left(0.630\begin{bmatrix}1\\0\end{bmatrix} + 0.983\begin{bmatrix}0\\2\end{bmatrix} + 0.062\begin{bmatrix}3\\-1\end{bmatrix}\right) = \begin{bmatrix}0.487\\1.137\end{bmatrix}$$

$$\Sigma_2 = \frac{1}{N_2}\sum_{i=1}^{3}\gamma_{2i} \cdot (x_i - \mu_2) \cdot (x_i - \mu_2)^T =$$

$$= \frac{1}{1.675}\left\{0.630\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.487\\1.137\end{bmatrix}\right)\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.487\\1.137\end{bmatrix}\right)^T + 0.983\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.487\\1.137\end{bmatrix}\right)\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.487\\1.137\end{bmatrix}\right)^T + 0.062\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}0.487\\1.137\end{bmatrix}\right)\left(\begin{bmatrix}3\\-1\end{bmatrix} - \begin{bmatrix}0.487\\1.137\end{bmatrix}\right)^T\right\} =$$

$$= \frac{1}{1.675}\left\{0.630\begin{bmatrix}0.513\\-1.137\end{bmatrix}\begin{bmatrix}0.513 & -1.137\end{bmatrix} + 0.983\begin{bmatrix}-0.487\\0.863\end{bmatrix}\begin{bmatrix}-0.487 & 0.863\end{bmatrix} + 0.062\begin{bmatrix}2.513\\-2.137\end{bmatrix}\begin{bmatrix}2.513 & -2.137\end{bmatrix}\right\} =$$

$$= \frac{1}{1.675}\left(\begin{bmatrix}0.166 & -0.367\\-0.367 & 0.814\end{bmatrix} + \begin{bmatrix}0.233 & -0.413\\-0.413 & 0.732\end{bmatrix} + \begin{bmatrix}0.392 & -0.333\\-0.333 & 0.283\end{bmatrix}\right) = \begin{bmatrix}0.472 & -0.664\\-0.664 & 1.092\end{bmatrix}$$

$$\pi_1 = p(c_1) = \frac{N_1}{N} = \frac{1.325}{1.325 + 1.675} \approx 0.442$$

$$\pi_2 = p(c_2) = \frac{N_2}{N} = \frac{1.675}{1.325 + 1.675} \approx 0.558$$

2)

a)

For observation x_1:

1. For cluster c_1:

$$\text{Prior: } p(c_1) = \pi_1 = 0.442$$

$$p(x_1 | c_1) = N(x_1|\mu_1, \Sigma_1) = \frac{1}{2\pi \times \sqrt{0.022}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.403\\-0.682\end{bmatrix}\right)^T \begin{bmatrix}0.875 & -0.485\\-0.485 & 0.294\end{bmatrix}^{-1} \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2.403\\-0.682\end{bmatrix}\right)\right\} = \frac{1}{2\pi \times \sqrt{0.022}} \exp\left\{-\frac{1}{2}\begin{bmatrix}-1.403 & 0.682\end{bmatrix}\begin{bmatrix}13.348 & 22.020\\22.020 & 39.728\end{bmatrix}\begin{bmatrix}-1.403\\0.682\end{bmatrix}\right\} = \frac{1}{2\pi \times \sqrt{0.022}} \exp\{-1.307\} \approx 0.290$$

$$p(c_1, x_1) = 0.290 \times 0.442 \approx 0.128$$

2. For cluster c_2:

$$\text{Prior: } p(c_2) = \pi_2 = 0.558$$

$$p(x_1 | c_2) = N(x_1|\mu_2, \Sigma_2) = \frac{1}{2\pi \times \sqrt{0.075}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.487\\1.137\end{bmatrix}\right)^T \begin{bmatrix}0.472 & -0.664\\-0.664 & 1.092\end{bmatrix}^{-1} \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.487\\1.137\end{bmatrix}\right)\right\} = \frac{1}{2\pi \times \sqrt{0.075}} \exp\left\{-\frac{1}{2}\begin{bmatrix}0.513 & -1.137\end{bmatrix}\begin{bmatrix}14.652 & 8.909\\8.909 & 6.333\end{bmatrix}\begin{bmatrix}0.513\\-1.137\end{bmatrix}\right\} = \frac{1}{2\pi \times \sqrt{0.075}} \exp\{-0.821\} \approx 0.256$$

$$p(c_2, x_1) = 0.256 \times 0.585 = 0.150$$

Therefore,

$$\gamma_{11} = \frac{0.128}{0.128 + 0.150} \approx 0.460$$

$$\gamma_{21} = \frac{0.031}{0.128 + 0.150} \approx 0.540$$

For observation x_2:

1. For cluster c_1:

$$\text{Prior: } p(c_1) = \pi_1 = 0.442$$

$$p(x_1 | c_1) = N(x_1|\mu_1, \Sigma_1) = \frac{1}{2\pi \times \sqrt{0.022}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.403\\-0.682\end{bmatrix}\right)^T \begin{bmatrix}0.875 & -0.485\\-0.485 & 0.294\end{bmatrix}^{-1} \left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}2.403\\-0.682\end{bmatrix}\right)\right\} = \frac{1}{2\pi \times \sqrt{0.022}} \exp\left\{-\frac{1}{2}\begin{bmatrix}-2.403 & 2.682\end{bmatrix}\begin{bmatrix}13.348 & 22.020\\22.020 & 39.728\end{bmatrix}\begin{bmatrix}-2.403\\2.682\end{bmatrix}\right\} = \frac{1}{2\pi \times \sqrt{0.022}} \exp\{-39.507\} \approx 0.000$$

$$p(c_1, x_1) = 0.000 \times 0.442 = 0$$

2. For cluster c_2:

$$\text{Prior: } p(c_2) = \pi_2 = 0.558$$

$$p(x_1 | c_2) = N(x_1|\mu_2, \Sigma_2) = \frac{1}{2\pi \times \sqrt{0.075}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.487\\1.137\end{bmatrix}\right)^T \begin{bmatrix}0.472 & -0.664\\-0.664 & 1.092\end{bmatrix}^{-1} \left(\begin{bmatrix}0\\2\end{bmatrix} - \begin{bmatrix}0.487\\1.137\end{bmatrix}\right)\right\} = \frac{1}{2\pi \times \sqrt{0.075}} \exp\left\{-\frac{1}{2}\begin{bmatrix}-0.487 & 0.863\end{bmatrix}\begin{bmatrix}14.652 & 8.909\\8.909 & 6.333\end{bmatrix}\begin{bmatrix}-0.487\\0.683\end{bmatrix}\right\} = \frac{1}{2\pi \times \sqrt{0.075}} \exp\{-0.250\} \approx 0.453$$

$$p(c_2, x_1) = 0.453 \times 0.585 = 0.265$$

Therefore,

$$\gamma_{12} = \frac{0}{0 + 0.265} = 0$$

$$\gamma_{22} = \frac{0.265}{0 + 0.265} = 1$$

For observation x_3:

1. For cluster c_1:

$$\text{Prior: } p(c_1) = \pi_1 = 0.442$$

$$p(x_1|\, c_1) = N(x_1|\mu_1, \Sigma_1) = \frac{1}{2\pi \times \sqrt{0.022}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2.403 \\ -0.682 \end{bmatrix}\right)^T \begin{bmatrix} 0.875 & -0.485 \\ -0.485 & 0.294 \end{bmatrix}^{-1} \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 2.403 \\ -0.682 \end{bmatrix}\right)\right\} = \frac{1}{2\pi \times \sqrt{0.022}} \exp\left\{-\frac{1}{2}\begin{bmatrix} 0.597 & -0.318 \end{bmatrix} \begin{bmatrix} 13.348 & 22.020 \\ 22.020 & 39.728 \end{bmatrix} \begin{bmatrix} 0.597 \\ -0.318 \end{bmatrix}\right\} = \frac{1}{2\pi \times \sqrt{0.022}} \exp\{-0.207\} \approx 0.872$$

$$p(c_1, x_1) = 0.872 \times 0.442 \approx 0.385$$

2. For cluster c_2:

$$\text{Prior: } p(c_2) = \pi_2 = 0.558$$

$$p(x_1|\, c_2) = N(x_1|\mu_2, \Sigma_2) = \frac{1}{2\pi \times \sqrt{0.075}} \exp\left\{-\frac{1}{2}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.487 \\ 1.137 \end{bmatrix}\right)^T \begin{bmatrix} 0.472 & -0.664 \\ -0.664 & 1.092 \end{bmatrix}^{-1} \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.487 \\ 1.137 \end{bmatrix}\right)\right\} = \frac{1}{2\pi \times \sqrt{0.075}} \exp\left\{-\frac{1}{2}\begin{bmatrix} 2.513 & -2.137 \end{bmatrix} \begin{bmatrix} 14.652 & 8.909 \\ 8.909 & 6.333 \end{bmatrix} \begin{bmatrix} 2.513 \\ -2.137 \end{bmatrix}\right\} = \frac{1}{2\pi \times \sqrt{0.075}} \exp\{-12.881\} \approx 0.000$$

$$p(c_2, x_1) = 0.000 \times 0.585 = 0$$

Therefore,

$$\gamma_{13} = \frac{0.385}{0.385 + 0} = 1$$

$$\gamma_{23} = \frac{0}{0.385 + 0} = 0$$

We conclude that, under a MAP assumption,

$$c_1 = \{x_3\},\ c_2 = \{x_1, x_2\}$$

b)

$$a(x_1) = \|x_1 - x_2\| = \sqrt{(1-0)^2 + (0-2)^2} = \sqrt{5}$$

$$b(x_1) = \|x_1 - x_3\| = \sqrt{(1-3)^2 + (0+1)^2} = \sqrt{5}$$

$$s(x_1) = \frac{b(x_1)}{a(x_1)} - 1 = \frac{\sqrt{5}}{\sqrt{5}} - 1 = 0$$

$$a(x_2) = \|x_2 - x_1\| = \sqrt{(0-1)^2 + (2-0)^2} = \sqrt{5}$$

$$b(x_2) = \|x_2 - x_3\| = \sqrt{(0-3)^2 + (2+1)^2} = \sqrt{18}$$

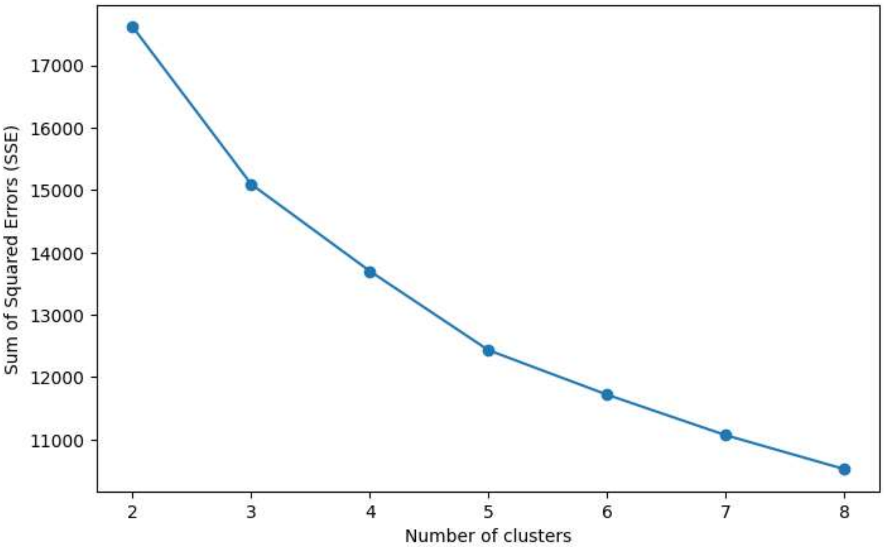
$$s(x_2) = 1 - \frac{a(x_2)}{b(x_2)} = 1 - \frac{\sqrt{5}}{\sqrt{18}} = 0.473$$

$$s(c_2) = \frac{s(x_1) + s(x_2)}{2} = \frac{0 + 0.473}{2} = 0.237$$

II. Programming and critical analysis

1)

a)



b) Adding too many clusters, aside from causing more processing time and complexity, could cause overfitting, where the model may capture unmeaningful values, making it difficult to generalize to new data. In this case, instead of identifying meaningful customer segments, the model might divide similar customers into separate, random groups.

Analyzing the graph, it's clear that the SSE decreases as the number of clusters increases, but after 5 clusters, the decrease is not as significant as before. One way to find the optimal number of clusters is using the elbow method. In this case the elbow occurs at 5 clusters, this number of clusters should provide a reasonable balance between SSE(inertia) values and maintaining a reasonable model without overfitting.

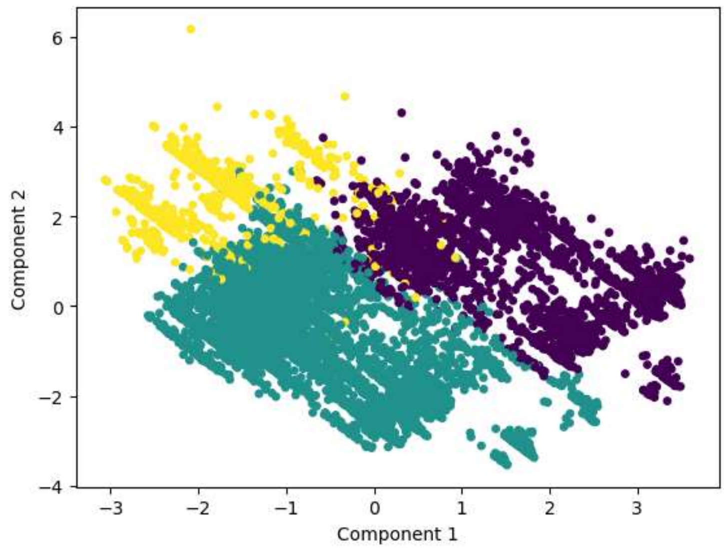
c) Analyzing the contents of this dataset, we see that it contains several categorical features, and k-means works by using squared Euclidean distances to minimize within-cluster variances. The problem here is exactly the way that k-means works with categorical data, since calculating squared Euclidean distances between categories like "married" and "single", even if possible, won't have any real meaning in the real world and could lead to misleading results.

On the other hand, k-modes is specifically made to deal with categorical data, grouping similar data points into clusters based on their categorical attributes. Given this, we think that k-modes would result in more adequate clusters, by treating categories as they really are and avoiding the distortions in the dataset caused by the conversion of categorical data into numeric format (pd.get_dummies()) that is needed in k-means.

2)

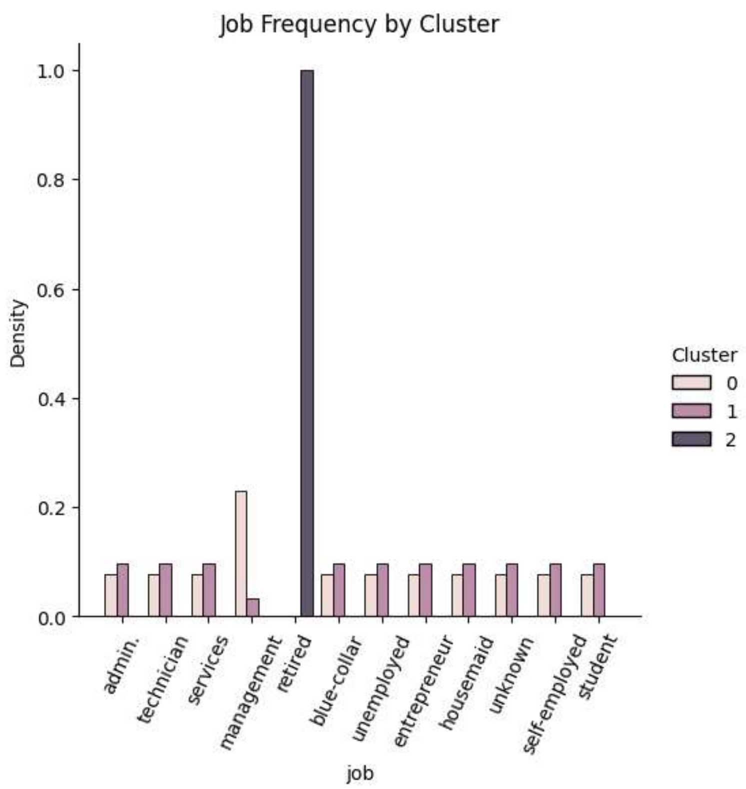
a) 22.76%

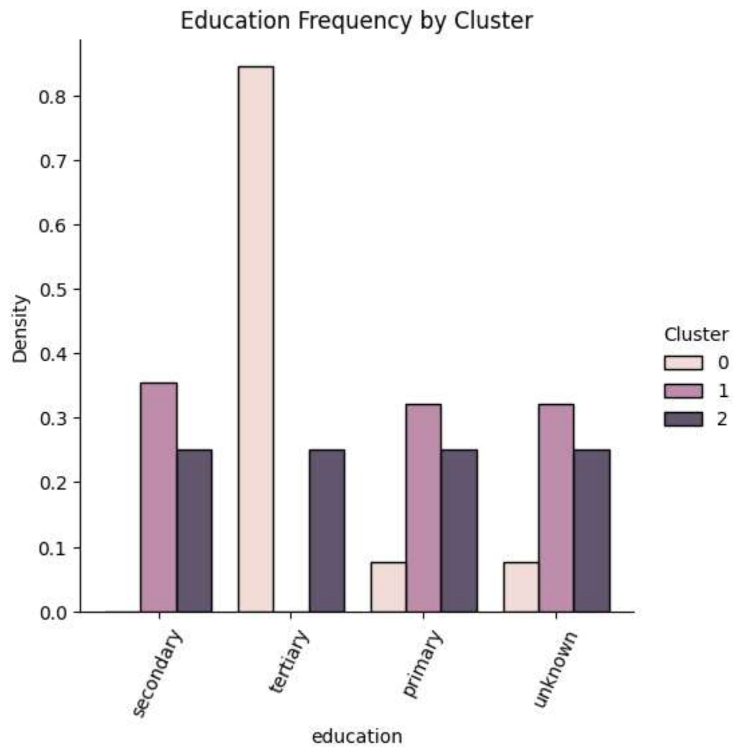
b)



The graph is basically divided into 3 zones, each one corresponding to a different cluster, except for a small number of points that seem misplaced, making the clusters not clearly separated. However, these few points only cause some overlaps between each zone, meaning that the PCA-reduced dimensions captured some meaningful differences between clusters.

c)





By observing the graph of the job frequency, we can see that all of Cluster 2's density is concentrated in the "retired" population, and in the education graph we can see that it is evenly destributed. This indicates that this cluster represents only this subgroup of clients.

Cluster 0 has a really high density in "tertiary" education, and in the job graph, this cluster shows an even distribution across various occupations, though it has a slightly higher density in "management" roles. This suggests that Cluster 0 represents clients with higher education but with diverse career backgrounds.

Cluster 1 is evenly distributed in both the job and education graphs. However, this cluster has a slightly higher presence than Cluster 0 in most jobs, except for "management". This variety of different education levels and job roles suggests that Cluster 1 represents a broader population with moderate education.

END