

## Part I: Pen and paper

### 1. Performing two epochs

We have three observations, represented as two-dimensional vectors:

$$x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

We assume a Gaussian Mixture Model with two components. The initial parameters are as follows:

$$\mu_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mu_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad \pi_1 = \pi_2 = 0.5$$

#### First Epoch:

$$\gamma_k = \frac{\pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k)}{\sum_k (\pi_k \cdot \mathcal{N}(x_i | \mu_k, \Sigma_k))}$$

$$\mathcal{N}(x | \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right)$$

#### E-Step

$$P(C_k | x_i) = \frac{\pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k)}{\sum_k (\pi_k \cdot \mathcal{N}(x_i | \mu_k, \Sigma_k))}$$

$$\mathcal{N}(x_1 | \mu_1, \Sigma_1) = \frac{1}{2\pi \cdot \sqrt{15}} \exp \left( -\frac{1}{2} \begin{bmatrix} -1 & 1 \end{bmatrix} \frac{1}{15} \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right) = \frac{1}{2\pi \cdot \sqrt{15}} \exp(-0.5 \times \frac{2}{3}) \approx 0.029$$

$$\mathcal{N}(x_2 | \mu_1, \Sigma_1) = \frac{1}{2\pi \cdot \sqrt{15}} \exp \left( -\frac{1}{2} \begin{bmatrix} -2 & 3 \end{bmatrix} \frac{1}{15} \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \end{bmatrix} \right) = \frac{1}{2\pi \cdot \sqrt{15}} \exp(-0.5 \times \frac{64}{15}) \approx 0.005$$

$$\mathcal{N}(x_3 | \mu_1, \Sigma_1) = \frac{1}{2\pi \cdot \sqrt{15}} \exp \left( -\frac{1}{2} \begin{bmatrix} 1 & 0 \end{bmatrix} \frac{1}{15} \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} \right) = \frac{1}{2\pi \cdot \sqrt{15}} \exp(-0.5 \times \frac{4}{15}) \approx 0.036$$

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

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

$$\mathcal{N}(x_3|\mu_2, \Sigma_2) = \frac{1}{2\pi \cdot \sqrt{4}} \exp\left(-\frac{1}{2} \begin{bmatrix} 2 & -2 \end{bmatrix} \frac{1}{4} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -2 \end{bmatrix}\right) = \frac{1}{2\pi \cdot \sqrt{4}} \exp(-0.5 \times 4) \approx 0.011$$

$$P(C_k, x_i) = P(x_i|C_k) \cdot P(C_k) = \mathcal{N}(x_i|\mu_k, \Sigma_k) \cdot \pi_k$$

$$\pi_1 = \pi_2 = 0.5$$

$$P(C_1|x_i) = 0.015 \quad P(C_2|x_i) = 0.031$$

$$P(C_1|x_2) = 0.003 \quad P(C_2|x_2) = 0.024$$

$$P(C_1|x_3) = 0.018 \quad P(C_2|x_3) = 0.006$$

$$\Rightarrow \gamma_{ik} = \frac{P(C_k|x_i)}{\sum_j P(C_j|x_i)}$$

$$\gamma_{1,1} = \frac{0.015}{0.015 + 0.031} = 0.326$$

$$\gamma_{2,1} = 1 - 0.326 = 0.674$$

$$\gamma_{1,2} = \frac{0.003}{0.003 + 0.024} = 0.111$$

$$\gamma_{2,2} = 1 - 0.111 = 0.889$$

$$\gamma_{1,3} = \frac{0.018}{0.018 + 0.006} = 0.750$$

$$\gamma_{2,3} = 1 - 0.750 = 0.250$$

## M-Step

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

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

$$N_1 = 0.326 + 0.111 + 0.750 = 1.187$$

$$N_2 = 0.674 + 0.869 + 0.250 = 1.813$$

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

$$\mu_2 = \frac{1}{1.813} \left( 0.674 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.869 \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} + 0.250 \cdot \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right) = \begin{bmatrix} 0.785 \\ 0.843 \end{bmatrix}$$

$$\begin{aligned} \Sigma_1 &= \frac{1}{1.187} \left( 0.326 \begin{bmatrix} -1.170 \\ 0.445 \end{bmatrix} \begin{bmatrix} -1.170 & 0.445 \end{bmatrix} \right. \\ &\quad \left. + 0.111 \begin{bmatrix} -2.170 \\ 2.445 \end{bmatrix} \begin{bmatrix} -2.170 & 2.445 \end{bmatrix} + 0.750 \begin{bmatrix} 0.830 \\ -0.555 \end{bmatrix} \begin{bmatrix} 0.830 & -0.555 \end{bmatrix} \right) \\ &= \begin{bmatrix} 1.252 & -0.930 \\ -0.930 & 0.808 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \Sigma_2 &= \frac{1}{1.813} \left( 0.674 \begin{bmatrix} 0.215 \\ -0.843 \end{bmatrix} \begin{bmatrix} 0.215 & -0.843 \end{bmatrix} \right. \\ &\quad \left. + 0.889 \begin{bmatrix} -0.785 \\ 1.157 \end{bmatrix} \begin{bmatrix} -0.785 & 1.157 \end{bmatrix} + 0.250 \begin{bmatrix} 2.215 \\ -1,843 \end{bmatrix} \begin{bmatrix} 2.215 & -1,843 \end{bmatrix} \right) \\ &= \begin{bmatrix} 0.996 & -1.076 \\ -1.076 & 1.389 \end{bmatrix} \end{aligned}$$

$$\pi_1 = \frac{N_1}{n} = \frac{1.187}{3} = 0.396$$

$$\pi_2 = \frac{N_2}{n} = \frac{1.513}{3} = 0.604$$

**Second Epoch:** Using the updated parameters from the first epoch, we repeat the E-step and M-step.

Observations:

$$x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

Parameters after 1st epoch:

$$\mu_1 = \begin{bmatrix} 2.170 \\ -0.445 \end{bmatrix}, \mu_2 = \begin{bmatrix} 0.785 \\ 0.843 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 1.252 & -0.930 \\ -0.930 & 0.808 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0.996 & -1.076 \\ -1.076 & 1.389 \end{bmatrix}, \quad \pi_1 = 0.396 \quad \pi_2 = 0.604$$

**E-Step**

$$P(C_k | \mathbf{x}_i) = \frac{\tau_k \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_k, \Sigma_k)}{\sum_k (\tau_k \cdot \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_k, \Sigma_k))}$$

$$\begin{aligned} \mathcal{N}(x_1 | \mu_1, \Sigma_1) &= \frac{1}{2\pi \cdot \sqrt{0.147}} \exp \left( -\frac{1}{2} \begin{bmatrix} -1.170 & 0.495 \end{bmatrix} \begin{bmatrix} 1.250 & -0.930 \\ -0.930 & 0.808 \end{bmatrix}^{-1} \begin{bmatrix} -1.17 \\ 0.495 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.147}} \exp(-0.5 \cdot 2.628) \approx 0.112 \end{aligned}$$

$$\begin{aligned} \mathcal{N}(x_2 | \mu_1, \Sigma_1) &= \frac{1}{2\pi \cdot \sqrt{0.147}} \exp \left( -\frac{1}{2} \begin{bmatrix} -2.170 & 2.448 \end{bmatrix} \begin{bmatrix} 1.250 & -0.930 \\ -0.930 & 0.808 \end{bmatrix}^{-1} \begin{bmatrix} -2.170 \\ 2.448 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.147}} \exp(-0.5 \cdot 9.684) \approx 0.003 \end{aligned}$$

$$\begin{aligned} \mathcal{N}(x_3 | \mu_1, \Sigma_1) &= \frac{1}{2\pi \cdot \sqrt{0.147}} \exp \left( -\frac{1}{2} \begin{bmatrix} 0.830 & -0.555 \end{bmatrix} \begin{bmatrix} 1.250 & -0.930 \\ -0.930 & 0.808 \end{bmatrix}^{-1} \begin{bmatrix} 0.830 \\ -0.555 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.147}} \exp(-0.5 \cdot 0.583) \approx 0.310 \end{aligned}$$

$$\begin{aligned} \mathcal{N}(x_1 | \mu_2, \Sigma_2) &= \frac{1}{2\pi \cdot \sqrt{0.226}} \exp \left( -\frac{1}{2} \begin{bmatrix} 0.215 & -0.843 \end{bmatrix} \begin{bmatrix} 0.996 & -1.076 \\ -1.076 & 1.389 \end{bmatrix}^{-1} \begin{bmatrix} 0.215 \\ -0.843 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.226}} \exp(-0.5 \cdot 1.693) \approx 0.144 \end{aligned}$$

$$\begin{aligned}\mathcal{N}(x_2|\mu_2, \Sigma_2) &= \frac{1}{2\pi \cdot \sqrt{0.226}} \exp \left( -\frac{1}{2} \begin{bmatrix} -0.785 & 1.157 \end{bmatrix} \begin{bmatrix} 0.996 & -1.076 \\ -1.076 & 1.389 \end{bmatrix}^{-1} \begin{bmatrix} -0.785 \\ 1.157 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.226}} \exp(-0.5 \cdot 1.040) \approx 0.199\end{aligned}$$

$$\begin{aligned}\mathcal{N}(x_3|\mu_2, \Sigma_2) &= \frac{1}{2\pi \cdot \sqrt{0.226}} \exp \left( -\frac{1}{2} \begin{bmatrix} 2.215 & -1.843 \end{bmatrix} \begin{bmatrix} 0.996 & -1.076 \\ -1.076 & 1.389 \end{bmatrix}^{-1} \begin{bmatrix} 2.215 \\ -1.843 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.226}} \exp(-0.5 \cdot 6.261) \approx 0.015\end{aligned}$$

$$P(C_k, x_i) = P(x_i|C_k) \cdot P(C_k) = \mathcal{N}(x_i|\mu_k, \Sigma_k) \cdot \pi_k$$

$$\pi_1 = 0.396 \quad \pi_2 = 0.604$$

$$P(C_1|x_i) = 0.044 \quad P(C_2|x_i) = 0.087$$

$$P(C_1|x_2) = 0.001 \quad P(C_2|x_2) = 0.120$$

$$P(C_1|x_3) = 0.123 \quad P(C_2|x_3) = 0.009$$

$$\Rightarrow \gamma_{ik} = \frac{P(C_k|x_i)}{\sum_j P(C_j|x_i)}$$

$$\gamma_{1,1} = \frac{0.044}{0.044 + 0.087} = 0.336$$

$$\gamma_{2,1} = 1 - 0.336 = 0.664$$

$$\gamma_{1,2} = \frac{0.001}{0.001 + 0.120} = 0.008$$

$$\gamma_{2,2} = 1 - 0.008 = 0.992$$

$$\gamma_{1,3} = \frac{0.0123}{0.123 + 0.009} = 0.932$$

$$\gamma_{2,3} = 1 - 0.932 = 0.068$$

## M-Step

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

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

$$N_1 = 0.336 + 0.008 + 0.932 = 1.276$$

$$N_2 = 0.664 + 0.992 + 0.068 = 1.724$$

$$\mu_1 = \frac{1}{1.276} \left( 0.336 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.008 \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} + 0.932 \cdot \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right) = \begin{bmatrix} 2.455 \\ -0.718 \end{bmatrix}$$

$$\mu_2 = \frac{1}{1.724} \left( 0.664 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.992 \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} + 0.068 \cdot \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right) = \begin{bmatrix} 0.503 \\ 1.111 \end{bmatrix}$$

$$\begin{aligned} \Sigma_1 &= \frac{1}{1.276} \left( 0.336 \begin{bmatrix} -1.455 \\ 0.718 \end{bmatrix} \begin{bmatrix} -1.455 & 0.718 \end{bmatrix} \right. \\ &\quad \left. + 0.008 \begin{bmatrix} -2.455 \\ 2.718 \end{bmatrix} \begin{bmatrix} -2.455 & 2.718 \end{bmatrix} + 0.932 \begin{bmatrix} 0.545 \\ -0.282 \end{bmatrix} \begin{bmatrix} 0.545 & -0.282 \end{bmatrix} \right) \\ &= \begin{bmatrix} 0.812 & -0.429 \\ -0.429 & 0.240 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \Sigma_2 &= \frac{1}{1.724} \left( 0.664 \begin{bmatrix} 0.497 \\ -1.111 \end{bmatrix} \begin{bmatrix} 0.497 & -1.111 \end{bmatrix} \right. \\ &\quad \left. + 0.992 \begin{bmatrix} -0.503 \\ 0.889 \end{bmatrix} \begin{bmatrix} -0.503 & 0.889 \end{bmatrix} + 0.068 \begin{bmatrix} 2.497 \\ -2.111 \end{bmatrix} \begin{bmatrix} 2.497 & -2.111 \end{bmatrix} \right) \\ &= \begin{bmatrix} 0.487 & -0.678 \\ -0.678 & 1.106 \end{bmatrix} \end{aligned}$$

$$\pi_1 = \frac{N_1}{n} = \frac{1.276}{3} = 0.425$$

$$\pi_2 = \frac{N_2}{n} = \frac{1.724}{3} = 0.575$$

2. a) MAP assumption

$$\gamma_k = \frac{\pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k)}{\sum_k (\pi_k \cdot \mathcal{N}(x_i | \mu_k, \Sigma_k))}$$

$$\mathcal{N}(x | \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right)$$

$$P(C_k | x_i) = \frac{\pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k)}{\sum_k (\pi_k \cdot \mathcal{N}(x_i | \mu_k, \Sigma_k))}$$

**E Step**

$$\begin{aligned} \mathcal{N}(x_1 | \mu_1, \Sigma_1) &= \frac{1}{2\pi \cdot \sqrt{0.011}} \exp \left( -\frac{1}{2} \begin{bmatrix} -1.455 & 0.718 \end{bmatrix} \begin{bmatrix} 0.812 & -0.429 \\ -0.429 & 0.240 \end{bmatrix}^{-1} \begin{bmatrix} -1.455 \\ 0.718 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.011}} \exp (-0.5 \cdot 2.800) \approx 0.374 \end{aligned}$$

$$\begin{aligned} \mathcal{N}(x_2 | \mu_1, \Sigma_1) &= \frac{1}{2\pi \cdot \sqrt{0.011}} \exp \left( -\frac{1}{2} \begin{bmatrix} 2.455 & 2.718 \end{bmatrix} \begin{bmatrix} 0.812 & -0.429 \\ -0.429 & 0.240 \end{bmatrix}^{-1} \begin{bmatrix} 2.455 \\ 2.718 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.011}} \exp (-0.5 \cdot 158.690) \approx 0.000 \end{aligned}$$

$$\begin{aligned} \mathcal{N}(x_3 | \mu_1, \Sigma_1) &= \frac{1}{2\pi \cdot \sqrt{0.011}} \exp \left( -\frac{1}{2} \begin{bmatrix} 0.545 & -0.282 \end{bmatrix} \begin{bmatrix} 0.812 & -0.429 \\ -0.429 & 0.240 \end{bmatrix}^{-1} \begin{bmatrix} 0.545 \\ -0.282 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.011}} \exp (-0.5 \cdot 0.368) \approx 1.262 \end{aligned}$$

$$\begin{aligned} \mathcal{N}(x_1 | \mu_2, \Sigma_2) &= \frac{1}{2\pi \cdot \sqrt{0.079}} \exp \left( -\frac{1}{2} \begin{bmatrix} 0.497 & -1.111 \end{bmatrix} \begin{bmatrix} 0.996 & -1.076 \\ -1.076 & 1.389 \end{bmatrix}^{-1} \begin{bmatrix} 0.497 \\ -1.111 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.079}} \exp (-0.5 \cdot 1.590) \approx 0.256 \end{aligned}$$

$$\begin{aligned} \mathcal{N}(x_2 | \mu_2, \Sigma_2) &= \frac{1}{2\pi \cdot \sqrt{0.079}} \exp \left( -\frac{1}{2} \begin{bmatrix} -0.503 & 0.889 \end{bmatrix} \begin{bmatrix} 0.996 & -1.076 \\ -1.076 & 1.389 \end{bmatrix}^{-1} \begin{bmatrix} -0.503 \\ 0.889 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.079}} \exp (-0.5 \cdot 0.739) \approx 0.392 \end{aligned}$$

$$\begin{aligned} \mathcal{N}(x_3|\mu_2, \Sigma_2) &= \frac{1}{2\pi \cdot \sqrt{0.079}} \exp \left( -\frac{1}{2} \begin{bmatrix} 2.497 & -2.111 \end{bmatrix} \begin{bmatrix} 0.996 & -1.076 \\ -1.076 & 1.389 \end{bmatrix}^{-1} \begin{bmatrix} 2.497 \\ -2.111 \end{bmatrix} \right) \\ &= \frac{1}{2\pi \cdot \sqrt{0.079}} \exp(-0.5 \cdot 24.302) \approx 0.000 \end{aligned}$$

$$P(C_k, x_i) = P(x_i|C_k) \cdot P(C_k) = \mathcal{N}(x_i|\mu_k, \Sigma_k) \cdot \pi_k$$

$$\pi_1 = 0.425 \quad \pi_2 = 0.575$$

$$P(C_1|x_i) = 0.159 \quad P(C_2|x_i) = 0.147$$

$$P(C_1|x_2) = 0 \quad P(C_2|x_2) = 1$$

$$P(C_1|x_3) = 1 \quad P(C_2|x_3) = 0$$

$$\gamma_{1,1} = 0.520$$

$$\gamma_{2,1} = 0.480$$

$$\gamma_{1,2} = 0$$

$$\gamma_{2,2} = 1$$

$$\gamma_{1,3} = 1$$

$$\gamma_{2,3} = 0$$

Observation	Cluster 1	Cluster 2	Hard Assignment
$x_1$	0.520	0.480	$0.520 > 0.480 \Rightarrow x_1$ classified as Cluster 1
$x_2$	0	1	$1 > 0 \Rightarrow x_2$ classified as Cluster 2
$x_3$	1	0	$1 > 0 \Rightarrow x_3$ classified as Cluster 1

2. b) Silhouette of the larger cluster

$$-1 \leq S(x_i) = \frac{b(x_i) - a(x_i)}{\max(a(x_i), b(x_i))} \leq 1 \quad a(x_i) = \frac{1}{|C_i| - 1} \sum_{\substack{i \in C_i \\ i \neq j}} d(i, j) \quad b(x_i) = \min_{j \neq i} \frac{1}{|C_j|} \sum_{k \in C_j} d(i, j)$$

$$a_{x_1} = \sqrt{(1-3)^2 + (0+1)^2} = \sqrt{5} \quad b_{x_1} = \sqrt{(1-0)^2 + (0-2)^2} = \sqrt{5}$$

$$a_{x_3} = \sqrt{5} \quad b_{x_3} = \sqrt{(3-0)^2 + (-1-2)^2} = \sqrt{18}$$

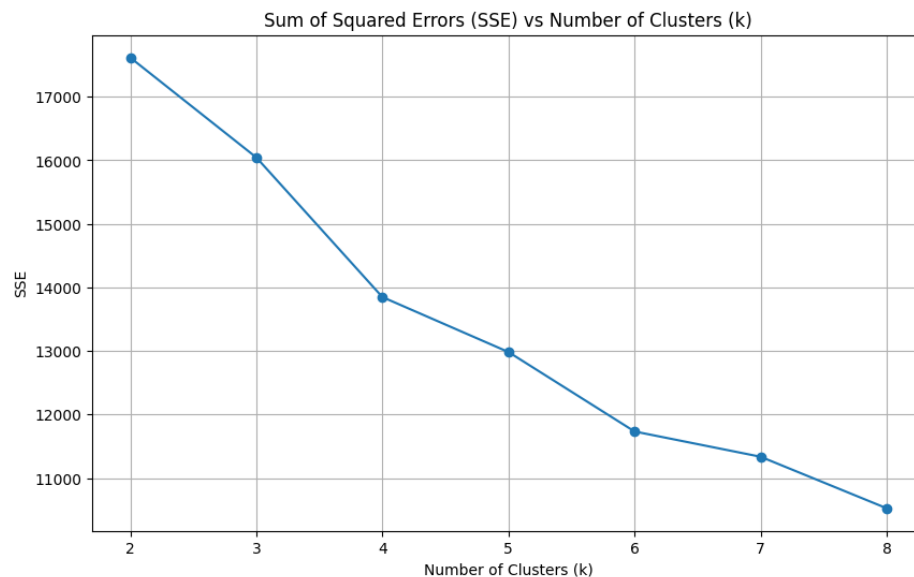
$$S(x_i) = 0 \quad S(x_3) = \frac{\sqrt{18} - \sqrt{5}}{\sqrt{18}} = 0.473$$



$$R : S(C_1) = \frac{0 + 0.473}{2} = 0.237$$

## Part II: Programming

### 1. 1a) Graphic for exercise 1



1b) R: In the SSE (Sum of Squared Errors) graph, the ideal number of clusters can be determined by the elbow method. The SSE decreases sharply from  $k = 2$  to  $k = 4$ , indicating that dividing the data into more clusters significantly reduces the error. From  $k = 4$  onward, the decrease in SSE begins to slow down, meaning that adding extra clusters does not provide a substantial improvement in fit. Therefore, based on the trade-off between the number of clusters and inertia,  $k = 4$  appears to be the ideal number of clusters. This is because it is the point where increasing the number of clusters stops providing significant improvements in compression (or reduction of SSE), balancing model simplicity and explained variability.

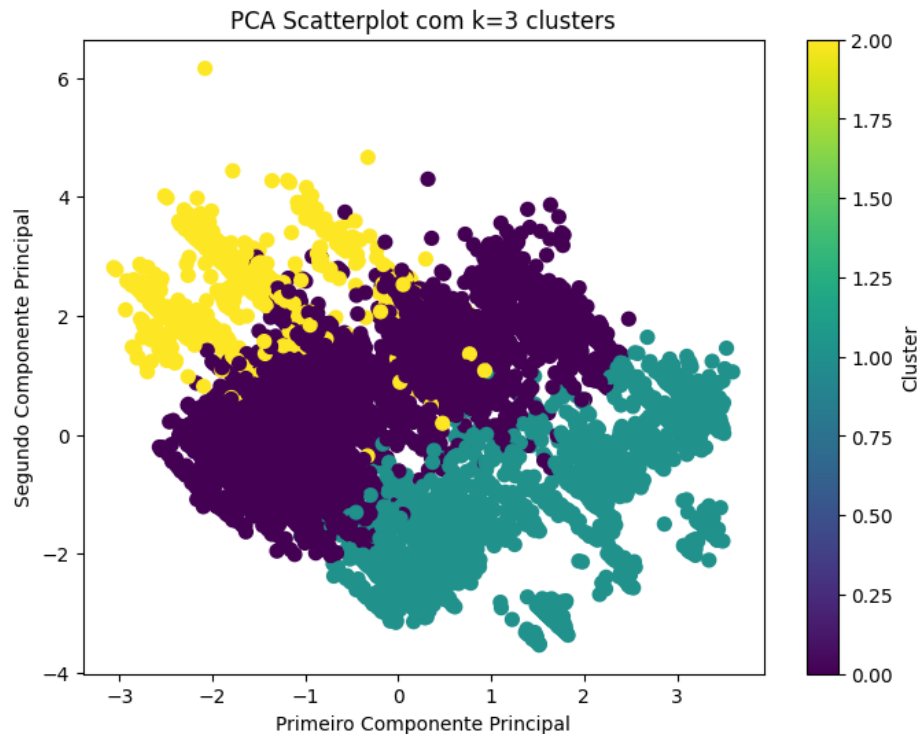
1c) R: K-modes is designed specifically for categorical data. It uses Hamming distance and can handle the mode (most frequent category) in place of the mean. K-means works well with numerical data as it minimizes the sum of squared distances (Euclidean distance). With this in mind, K-modes could be a better approach than traditional k-means since the majority of the features are categorical.

2. 2a) PCA

R: Explained variance by top 2 components: [0.11679024 0.11075988]

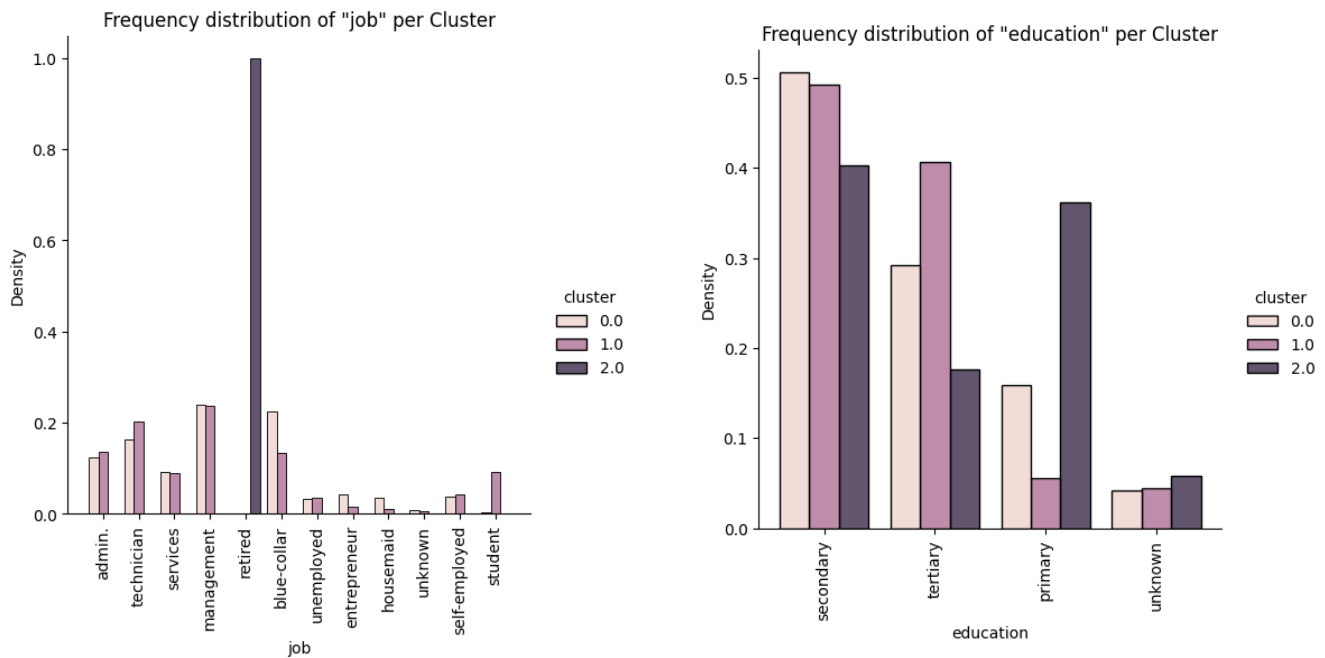
Total variance explained by the top 2 components: 22.76%

2b) Graphic for exercise 2b



R: There is significant overlap between the clusters, especially in the central region of the plot. The boundary between the purple and teal clusters is not well-defined, indicating that the data points in these clusters are not well-separated. The yellow cluster appears to be more distinguishable from the other two clusters, but there is still some mixing in the lower areas. The fact that there is substantial overlap between clusters suggests that clear separation is not possible in this 2D representation. This indicates that the features (or components) captured by the first two principal components do not fully explain the variance needed to separate the clusters (22.76%).

## 2c) Graphics for exercise 2c



R:Based on the provided graphs, here are the main differences between the clusters:

### Job Distribution Differences:

**Cluster 0:** This cluster has a higher representation of blue-collar, management, and technician roles. There is a more evenly distributed presence of other professions, like admin., services, and retired, with no single job type dominating overwhelmingly.

**Cluster 1:** This cluster has a noticeable presence in management and technician jobs, similar to Cluster 0, but it is also the only cluster that shows a slightly increased presence in student jobs. It targets a broad spectrum of job categories with no extreme outliers.

**Cluster 2:** This cluster is heavily dominated by the retired category, showing an extreme concentration here compared to other clusters. Other job categories are underrepresented, making it clear that Cluster 2 primarily groups individuals who are retired.

### Education Distribution Differences:

**Cluster 0:** This cluster has a predominant representation of individuals with secondary education. There is also a reasonable proportion of individuals with tertiary education, and a smaller, but notable presence of those with primary education.

**Cluster 1:** This cluster has a strong presence of individuals with secondary and tertiary education, with tertiary education being relatively more common here than in Cluster 0. This suggests that this cluster might be more associated with higher educational levels.

**Cluster 2:** There is a higher proportion of individuals with primary education in this cluster compared to others. It also has a considerable portion of individuals with secondary education but a much lower representation of tertiary education, indicating lower educational levels overall.

Having this in mind, Cluster 2 stands out for its dominance in the retired category and a focus on primary education, indicating an older demographic with lower educational attainment. Clusters 0 and 1 have more diverse job categories, with Cluster 0 leaning slightly more towards manual or technical jobs and Cluster 1 towards professional or academic roles. When it comes to education, Cluster 1 has a higher level of educational attainment, especially in tertiary education, compared to Cluster 0.