Aprendizagem 2023
Homework I — Group 003
(ist1107028, ist1107137)

Part I: Pen and paper

Consider the bivariate observations

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

and the multivariate Gaussian mixture given by

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

Answer the following questions by presenting all intermediary steps, and use 3 decimal places in each.

1. **Perform two epochs of the EM clustering algorithm and determine the new parameters.**

FIRST EM EPOCH

(1) Expectation: **E-STEP**

$\boxed{x_1}$

→ Cluster $c = 1$:

prior: $P(c=1) = \pi_1 = \mathbf{0.5}$

likelihood: $p(x_1|c=1) = \mathcal{N}(x_1|\mu_1, \sigma_1)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_1)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\cdot(x_1-\mu_1)^T\Sigma_1^{-1}\cdot(x_1-\mu_1)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{15}} \cdot \exp\left(-\frac{1}{2}\cdot\left(\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}2\\-1\end{pmatrix}\right)^T \cdot \begin{pmatrix}4 & 1\\1 & 4\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}2\\-1\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{15}} \cdot \exp\left(-\frac{1}{2}\cdot\begin{pmatrix}-1 & 1\end{pmatrix} \cdot \begin{pmatrix}\frac{4}{15} & -\frac{1}{15}\\-\frac{1}{15} & \frac{4}{15}\end{pmatrix} \cdot \begin{pmatrix}-1\\1\end{pmatrix}\right) = \frac{e^{-\frac{1}{3}}}{2\pi\cdot\sqrt{15}} = \mathbf{0.029}$$

joint probability: $P(c=1, x_1) = P(c=1)p(x_1|c=1) = \pi_1 \cdot \mathcal{N}(x_1|\mu_1,\sigma_1) = \mathbf{0.015}$

normalized posterior: $P(c=1|x_1) = \dfrac{0.015}{0.031+0.015} = \mathbf{0.326}$

→ Cluster $c = 2$:

prior: $P(c=2) = \pi_2 = \mathbf{0.5}$

likelihood: $p(x_1|c=2) = \mathcal{N}(x_1|\mu_2, \sigma_2)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_2)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_1 - \mu_2)^T \Sigma_2^{-1} \cdot (x_1 - \mu_2)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{4}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}1\\1\end{pmatrix}\right)^T \cdot \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}1\\1\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{4}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}0 & -1\end{pmatrix} \cdot \begin{pmatrix}\frac{1}{2} & 0\\0 & \frac{1}{2}\end{pmatrix} \cdot \begin{pmatrix}0\\-1\end{pmatrix}\right) = \frac{e^{-\frac{1}{4}}}{2\pi\cdot\sqrt{4}} = \mathbf{0.062}$$

joint probability: $P(c=2, x_1) = P(c=2)p(x_1|c=2) = \pi_2 \cdot \mathcal{N}(x_1|\mu_2, \sigma_2) = \mathbf{0.031}$

normalized posterior: $P(c=2|x_1) = \dfrac{0.031}{0.031 + 0.015} = \mathbf{0.674}$

$\boxed{x_2}$

→ Cluster $c = 1$:

prior: $P(c=1) = \pi_1 = \mathbf{0.5}$

likelihood: $p(x_2|c=1) = \mathcal{N}(x_2|\mu_1, \sigma_1)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_1)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_2 - \mu_1)^T \Sigma_1^{-1} \cdot (x_2 - \mu_1)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{15}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}2\\-1\end{pmatrix}\right)^T \cdot \begin{pmatrix}4 & 1\\1 & 4\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}2\\-1\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{15}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}-2 & 3\end{pmatrix} \cdot \begin{pmatrix}\frac{4}{15} & -\frac{1}{15}\\-\frac{1}{15} & \frac{4}{15}\end{pmatrix} \cdot \begin{pmatrix}-2\\3\end{pmatrix}\right) = \frac{e^{-\frac{32}{15}}}{2\pi\cdot\sqrt{15}} = \mathbf{0.005}$$

joint probability: $P(c=1, x_2) = P(c=1)p(x_2|c=1) = \pi_1 \cdot \mathcal{N}(x_2|\mu_1, \sigma_1) = \mathbf{0.003}$

normalized posterior: $P(c=1|x_2) = \dfrac{0.003}{0.024 + 0.003} = \mathbf{0.111}$

→ Cluster $c = 2$:

prior: $P(c=2) = \pi_2 = \mathbf{0.5}$
likelihood: $p(x_2|c=2) = \mathcal{N}(x_2|\mu_2, \sigma_2)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_2)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_2 - \mu_2)^T \Sigma_2^{-1} \cdot (x_2 - \mu_2)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{4}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}1\\1\end{pmatrix}\right)^T \cdot \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}1\\1\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{4}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}-1 & 1\end{pmatrix} \cdot \begin{pmatrix}\frac{1}{2} & 0\\0 & \frac{1}{2}\end{pmatrix} \cdot \begin{pmatrix}-1\\1\end{pmatrix}\right) = \frac{e^{-\frac{1}{2}}}{2\pi\cdot\sqrt{4}} = \mathbf{0.048}$$

joint probability: $P(c=2, x_2) = P(c=2)p(x_2|c=2) = \pi_2 \cdot \mathcal{N}(x_2|\mu_2, \sigma_2) = \mathbf{0.024}$

normalized posterior: $P(c=2|x_2) = \frac{0.024}{0.024 + 0.003} = \mathbf{0.889}$

$\boxed{x_3}$

→ Cluster $c = 1$:

prior: $P(c=1) = \pi_1 = \mathbf{0.5}$
likelihood: $p(x_3|c=1) = \mathcal{N}(x_3|\mu_1, \sigma_1)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_1)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_3 - \mu_1)^T \Sigma_1^{-1} \cdot (x_3 - \mu_1)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{15}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix} - \begin{pmatrix}2\\-1\end{pmatrix}\right)^T \cdot \begin{pmatrix}4 & 1\\1 & 4\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix} - \begin{pmatrix}2\\-1\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{15}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}1 & 0\end{pmatrix} \cdot \begin{pmatrix}\frac{4}{15} & -\frac{1}{15}\\-\frac{1}{15} & \frac{4}{15}\end{pmatrix} \cdot \begin{pmatrix}1\\0\end{pmatrix}\right) = \frac{e^{-\frac{2}{15}}}{2\pi\cdot\sqrt{15}} = \mathbf{0.036}$$

joint probability: $P(c=1, x_3) = P(c=1)p(x_3|c=1) = \pi_1 \cdot \mathcal{N}(x_3|\mu_1, \sigma_1) = \mathbf{0.018}$

normalized posterior: $P(c=1|x_3) = \frac{0.018}{0.018 + 0.006} = \mathbf{0.750}$

➞ Cluster $c = 2$:

prior: $P(c=2) = \pi_2 = \mathbf{0.5}$

likelihood: $p(x_3|c=2) = \mathcal{N}(x_3|\mu_2, \sigma_2)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_2)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\cdot(x_3-\mu_2)^T\Sigma_2^{-1}\cdot(x_3-\mu_2)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{4}} \cdot \exp\left(-\frac{1}{2}\cdot\left(\begin{pmatrix}3\\-1\end{pmatrix}-\begin{pmatrix}1\\1\end{pmatrix}\right)^T \cdot \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix}-\begin{pmatrix}1\\1\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{4}} \cdot \exp\left(-\frac{1}{2}\cdot\begin{pmatrix}2 & -2\end{pmatrix}\cdot\begin{pmatrix}\frac{1}{2} & 0\\0 & \frac{1}{2}\end{pmatrix}\cdot\begin{pmatrix}2\\-2\end{pmatrix}\right) = \frac{e^{-2}}{2\pi\cdot\sqrt{4}} = \mathbf{0.012}$$

joint probability: $P(c=1, x_3) = P(c=1)p(x_3|c=1) = \pi_1 \cdot \mathcal{N}(x_3|\mu_1, \sigma_1) = \mathbf{0.006}$

normalized posterior: $P(c=2|x_3) = \dfrac{0.006}{0.006+0.018} = \mathbf{0.250}$

Observations	$c=1$	$c=2$
x_1	0.326	0.674
x_2	0.111	0.889
x_3	0.750	0.250

Table 1: Normalized posteriors

(2) Maximization: **M-STEP**

For the recalculation we will use the following formulas (n represents the cluster):

For the means:

$$\mu_{\mathrm{n}} = \frac{P(c=n|x_1)\cdot x_1 + P(c=n|x_2)\cdot x_2 + P(c=n|x_3)\cdot x_3}{P(c=n|x_1)+P(c=n|x_2)+P(c=n|x_3)} \tag{1}$$

For the covariance matrices:

$$\Sigma_n = \begin{pmatrix}\Sigma_{11} & \Sigma_{21}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix} \tag{2}$$

Where

$$\Sigma_{ij} = \frac{P(c=n|x_1)((x_{1i}-\mu_{ni})(x_{1j}-\mu_{nj}))+P(c=n|x_2)((x_{2i}-\mu_{ni})(x_{2j}-\mu_{nj}))+P(c=n|x_3)((x_{3i}-\mu_{ni})(x_{3j}-\mu_{nj}))}{P(c=n|x_1)+P(c=n|x_2)+P(c=n|x_3)}$$

→ Cluster $c = 1$:

$$\mu_1 = \frac{0.326 \cdot x_1 + 0.111 \cdot x_2 + 0.75 \cdot x_3}{0.326 + 0.111 + 0.75} = \frac{0.326 \cdot \binom{1}{0} + 0.111 \cdot \binom{0}{2} + 0.750 \cdot \binom{3}{-1}}{1.187} = \binom{2.170}{-0.445}$$

$$\Sigma_1 = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 1.252 & -0.930 \\ -0.930 & 0.808 \end{pmatrix}$$

$$\begin{aligned}
\Sigma_{11} &= \frac{0.326 \cdot ((x_{11} - \mu_{11}) \cdot (x_{11} - \mu_{11})) + 0.111 \cdot ((x_{21} - \mu_{11}) \cdot (x_{21} - \mu_{11})) + 0.75 \cdot ((x_{31} - \mu_{11}) \cdot (x_{31} - \mu_{11}))}{1.187} \\
&= \frac{0.326 \cdot ((1 - 2.170)(1 - 2.170)) + 0.111 \cdot ((0 - 2.170)(0 - 2.170)) + 0.75 \cdot ((3 - 2.170)(3 - 2.170))}{1.187} \\
&= \mathbf{1.252}
\end{aligned}$$

$$\begin{aligned}
\Sigma_{21} &= \frac{0.326 \cdot ((x_{12} - \mu_{12}) \cdot (x_{11} - \mu_{11})) + 0.111 \cdot ((x_{22} - \mu_{12}) \cdot (x_{21} - \mu_{11})) + 0.75 \cdot ((x_{32} - \mu_{12}) \cdot (x_{31} - \mu_{11}))}{1.187} \\
&= \frac{0.326 \cdot ((0 - (-0.445))(1 - 2.170)) + 0.111 \cdot ((2 - (-0.445))(0 - 2.170)) + 0.75 \cdot ((-1 - (-0.445))(3 - 2.170))}{1.187} \\
&= -\mathbf{0.930}
\end{aligned}$$

$$\Sigma_{12} = \Sigma_{21} = -\mathbf{0.930}$$

$$\begin{aligned}
\Sigma_{22} &= \frac{0.326 \cdot ((x_{12} - \mu_{12}) \cdot (x_{12} - \mu_{12})) + 0.111 \cdot ((x_{22} - \mu_{12}) \cdot (x_{22} - \mu_{12})) + 0.75 \cdot ((x_{32} - \mu_{12}) \cdot (x_{32} - \mu_{12}))}{1.187} \\
&= \frac{0.326 \cdot ((0 - (-0.445))(0 - (-0.445))) + 0.111 \cdot ((2 - (-0.445))(2 - (-0.445))) + 0.75 \cdot ((-1 - (-0.445))^2)}{1.187} \\
&= \mathbf{0.808}
\end{aligned}$$

→ cluster $c = 2$

$$\mu_2 = \frac{0.674 \cdot x_1 + 0.889 \cdot x_2 + 0.250 \cdot x_3}{0.674 + 0.889 + 0.250} = \frac{0.674 \cdot \binom{1}{0} + 0.889 \cdot \binom{0}{2} + 0.250 \cdot \binom{3}{-1}}{1.813} = \binom{0.785}{0.843}$$

$$\Sigma_2 = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 0.996 & -1.076 \\ -1.076 & 1.389 \end{pmatrix}$$

$$\begin{aligned}
\Sigma_{11} &= \frac{0.674 \cdot ((x_{11} - \mu_{21}) \cdot (x_{11} - \mu_{21})) + 0.889 \cdot ((x_{21} - \mu_{21}) \cdot (x_{21} - \mu_{21})) + 0.25 \cdot ((x_{31} - \mu_{21}) \cdot (x_{31} - \mu_{21}))}{1.813} \\
&= \frac{0.674 \cdot ((1 - 0.785)(1 - 0.785)) + 0.889 \cdot ((0 - 0.785)(0 - 0.785)) + 0.25 \cdot ((3 - 0.785)(3 - 0.785))}{1.813} \\
&= \mathbf{0.996}
\end{aligned}$$

$$\begin{aligned}
\Sigma_{21} &= \frac{0.674 \cdot ((x_{12} - \mu_{22}) \cdot (x_{11} - \mu_{21})) + 0.889 \cdot ((x_{22} - \mu_{22}) \cdot (x_{21} - \mu_{21})) + 0.25 \cdot ((x_{32} - \mu_{22}) \cdot (x_{31} - \mu_{21}))}{1.813} \\
&= \frac{0.674 \cdot ((0 - 0.843)(1 - 0.785)) + 0.889 \cdot ((2 - 0.843)(0 - 0.785)) + 0.25 \cdot ((-1 - 0.843)(3 - 0.785))}{1.813} \\
&= -\mathbf{1.076}
\end{aligned}$$

$\Sigma_{12} = \Sigma_{21} = \mathbf{-1.076}$

$$\Sigma_{22} = \frac{0.674 \cdot ((x_{12} - \mu_{22}) \cdot (x_{12} - \mu_{22})) + 0.889 \cdot ((x_{22} - \mu_{22}) \cdot (x_{22} - \mu_{22})) + 0.25 \cdot ((x_{32} - \mu_{22}) \cdot (x_{32} - \mu_{22}))}{1.813}$$
$$= \frac{0.674 \cdot ((0 - 0.843)(0 - 0.843)) + 0.889 \cdot ((2 - 0.843)(2 - 0.843)) + 0.25 \cdot ((-1 - 0.843)^2)}{1.813}$$
$$= \mathbf{1.389}$$

Normalized priors:

$$P(c = 1) = \frac{0.326 + 0.111 + 0.75}{(0.326 + 0.111 + 0.75) + (0.674 + 0.889 + 0.25)} = 0.396$$
$$P(c = 2) = \frac{0.674 + 0.889 + 0.25}{(0.326 + 0.111 + 0.75) + (0.674 + 0.889 + 0.25)} = 0.604$$

SECOND EM EPOCH

(1) Expectation: **E-STEP**

$\boxed{x_1}$

→ Cluster $c = 1$:

prior: $P(c = 1) = \mathbf{0.396}$

likelihood: $p(x_1|c = 1) = \mathcal{N}(x_1|\mu_1, \sigma_1)$

$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{det(\Sigma_1)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_1 - \mu_1)^T \Sigma_1^{-1} \cdot (x_1 - \mu_1)\right)$$
$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{\sqrt{0.147}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}2.170\\-0.445\end{pmatrix}\right)^T \cdot \begin{pmatrix}1.252 & -0.930\\-0.930 & 0.808\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}2.170\\-0.445\end{pmatrix}\right)\right)$$
$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{\sqrt{0.147}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}-1.170 & 0.445\end{pmatrix} \cdot \begin{pmatrix}5.507 & 6.339\\6.339 & 8.533\end{pmatrix} \cdot \begin{pmatrix}-1.170\\0.445\end{pmatrix}\right) = \frac{e^{-1.314}}{2\pi \cdot \sqrt{0.147}}$$
$$= \mathbf{0.112}$$

joint probability: $P(c = 1, x_1) = P(c = 1)p(x_1|c = 1) = \pi_1 \cdot \mathcal{N}(x_1|\mu_1, \sigma_1) = \mathbf{0.044}$

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

→ Cluster $c = 2$:

prior: $P(c=2) = \mathbf{0.604}$

likelihood: $p(x_1|c=2) = \mathcal{N}(x_1|\mu_2, \sigma_2)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_2)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_1 - \mu_2)^T \Sigma_2^{-1} \cdot (x_1 - \mu_2)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.226}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}0.785\\0.843\end{pmatrix}\right)^T \cdot \begin{pmatrix}0.996 & -1.076\\ -1.076 & 1.389\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}0.785\\0.843\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.226}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}0.215 & -0.843\end{pmatrix} \cdot \begin{pmatrix}6.155 & 4.768\\ 4.768 & 4.414\end{pmatrix} \cdot \begin{pmatrix}0.215\\ -0.843\end{pmatrix}\right) = \frac{e^{-0.846}}{2\pi \cdot \sqrt{0.226}}$$

$$= \mathbf{0.144}$$

joint probability: $P(c=2, x_1) = P(c=2)p(x_1|c=2) = \pi_2 \cdot \mathcal{N}(x_1|\mu_2, \sigma_2) = \mathbf{0.087}$

normalized posterior: $P(c=2|x_1) = \dfrac{0.087}{0.044 + 0.087} = \mathbf{0.664}$

$\boxed{x_2}$

→ Cluster $c = 1$:

prior: $P(c=1) = \mathbf{0.396}$

likelihood: $p(x_2|c=1) = \mathcal{N}(x_2|\mu_1, \sigma_1)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_1)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_2 - \mu_1)^T \Sigma_1^{-1} \cdot (x_2 - \mu_1)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.147}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}2.170\\-0.445\end{pmatrix}\right)^T \cdot \begin{pmatrix}1.252 & -0.930\\ -0.930 & 0.808\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}2.170\\-0.445\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.147}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}-2.17 & 2.445\end{pmatrix} \cdot \begin{pmatrix}5.507 & 6.339\\ 6.339 & 8.533\end{pmatrix} \cdot \begin{pmatrix}-2.17\\ 2.445\end{pmatrix}\right) = \frac{e^{-4.839}}{2\pi \cdot \sqrt{0.147}}$$

$$= \mathbf{0.003}$$

joint probability: $P(c=1, x_2) = P(c=1)p(x_2|c=1) = \pi_1 \cdot \mathcal{N}(x_2|\mu_1, \sigma_1) = \mathbf{0.001}$

normalized posterior: $P(c=1|x_2) = \dfrac{0.001}{0.001 + 0.120} = \mathbf{0.008}$

→ Cluster $c = 2$:

prior: $P(c=2) = \mathbf{0.604}$
likelihood: $p(x_2|c=2) = \mathcal{N}(x_2|\mu_2, \sigma_2)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_2)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_2 - \mu_2)^T \Sigma_2^{-1} \cdot (x_2 - \mu_2)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.226}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}0.785\\0.843\end{pmatrix}\right)^T \cdot \begin{pmatrix}0.996 & -1.076\\-1.076 & 1.389\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}0.785\\0.843\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.226}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}-0.785 & 1.157\end{pmatrix} \cdot \begin{pmatrix}6.155 & 4.768\\4.768 & 4.414\end{pmatrix} \cdot \begin{pmatrix}-0.785\\1.157\end{pmatrix}\right) = \frac{e^{-0.520}}{2\pi \cdot \sqrt{0.226}}$$

$$= \mathbf{0.199}$$

joint probability: $P(c=2, x_2) = P(c=2)p(x_2|c=2) = \pi_2 \cdot \mathcal{N}(x_2|\mu_2, \sigma_2) = \mathbf{0.120}$

normalized posterior: $P(c=2|x_2) = \dfrac{0.120}{0.120 + 0.001} = \mathbf{0.992}$

$\boxed{x_3}$

→ Cluster $c = 1$:

prior: $P(c=1) = \mathbf{0.396}$
likelihood: $p(x_3|c=1) = \mathcal{N}(x_3|\mu_1, \sigma_1)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_1)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_3 - \mu_1)^T \Sigma_1^{-1} \cdot (x_3 - \mu_1)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.147}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix} - \begin{pmatrix}2.170\\-0.445\end{pmatrix}\right)^T \cdot \begin{pmatrix}1.252 & -0.930\\-0.930 & 0.808\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix} - \begin{pmatrix}2.170\\-0.445\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.147}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}0.830 & -0.555\end{pmatrix} \cdot \begin{pmatrix}5.507 & 6.339\\6.339 & 8.533\end{pmatrix} \cdot \begin{pmatrix}0.830\\-0.555\end{pmatrix}\right) = \frac{e^{-0.291}}{2\pi \cdot \sqrt{0.147}}$$

$$= \mathbf{0.310}$$

joint probability: $P(c=1, x_3) = P(c=1)p(x_3|c=1) = \pi_1 \cdot \mathcal{N}(x_3|\mu_1, \sigma_1) = \mathbf{0.123}$

normalized posterior: $P(c=1|x_3) = \dfrac{0.123}{0.123 + 0.009} = \mathbf{0.932}$

→ Cluster $c = 2$:

prior: $P(c=2) = \mathbf{0.604}$

likelihood: $p(x_3|c=2) = \mathcal{N}(x_3|\mu_2, \sigma_2)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_2)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\cdot(x_3-\mu_2)^T\Sigma_2^{-1}\cdot(x_3-\mu_2)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.226}} \cdot \exp\left(-\frac{1}{2}\cdot\left(\begin{pmatrix}3\\-1\end{pmatrix}-\begin{pmatrix}0.785\\0.843\end{pmatrix}\right)^T \cdot \begin{pmatrix}0.996 & -1.076\\-1.076 & 1.389\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix}-\begin{pmatrix}0.785\\0.843\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.226}} \cdot \exp\left(-\frac{1}{2}\cdot\begin{pmatrix}2.215 & -1.843\end{pmatrix}\cdot\begin{pmatrix}6.155 & 4.768\\4.768 & 4.414\end{pmatrix}\cdot\begin{pmatrix}2.215\\-1.843\end{pmatrix}\right) = \frac{e^{-3.131}}{2\pi\cdot\sqrt{0.226}}$$

$$= \mathbf{0.015}$$

joint probability: $P(c=1, x_3) = P(c=1)p(x_3|c=1) = \pi_1 \cdot \mathcal{N}(x_3|\mu_1, \sigma_1) = \mathbf{0.009}$

normalized posterior: $P(c=2|x_3) = \dfrac{0.009}{0.009+0.123} = \mathbf{0.068}$

Observations	$c=1$	$c=2$
x_1	0.336	0.664
x_2	0.008	0.992
x_3	0.932	0.068

Table 2: Normalized posteriors

(2) Maximization: **M-STEP**

→ Cluster $c = 1$:

$$\mu_1 = \frac{0.336\cdot x_1 + 0.008\cdot x_2 + 0.932\cdot x_3}{0.336+0.008+0.932} = \frac{0.336\cdot\begin{pmatrix}1\\0\end{pmatrix} + 0.008\cdot\begin{pmatrix}0\\2\end{pmatrix} + 0.932\cdot\begin{pmatrix}3\\-1\end{pmatrix}}{1.276} = \begin{pmatrix}2.455\\-0.718\end{pmatrix}$$

$$\Sigma_1 = \begin{pmatrix}\Sigma_{11} & \Sigma_{21}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix} = \begin{pmatrix}0.812 & -0.429\\-0.429 & 0.240\end{pmatrix}$$

$$\begin{aligned}
\Sigma_{11} &= \frac{0.336 \cdot ((x_{11} - \mu_{11}) \cdot (x_{11} - \mu_{11})) + 0.008 \cdot ((x_{21} - \mu_{11}) \cdot (x_{21} - \mu_{11})) + 0.932 \cdot ((x_{31} - \mu_{11}) \cdot (x_{31} - \mu_{11}))}{1.276} \\
&= \frac{0.336 \cdot ((1 - 2.455)(1 - 2.455)) + 0.008 \cdot ((0 - 2.455)(0 - 2.455)) + 0.932 \cdot ((3 - 2.455)(3 - 2.455))}{1.276} \\
&= \mathbf{0.812}
\end{aligned}$$

$$\begin{aligned}
\Sigma_{21} &= \frac{0.336 \cdot ((x_{12} - \mu_{12}) \cdot (x_{11} - \mu_{11})) + 0.008 \cdot ((x_{22} - \mu_{12}) \cdot (x_{21} - \mu_{11})) + 0.932 \cdot ((x_{32} - \mu_{12}) \cdot (x_{31} - \mu_{11}))}{1.276} \\
&= \frac{0.336 \cdot ((0 - (-0.718))(1 - 2.455)) + 0.008 \cdot ((2 - (-0.718))(0 - 2.455)) + 0.932 \cdot ((-1 - (-0.718))(3 - 2.455))}{1.276} \\
&= -\mathbf{0.429}
\end{aligned}$$

$$\Sigma_{12} = \Sigma_{21} = -\mathbf{0.429}$$

$$\begin{aligned}
\Sigma_{22} &= \frac{0.336 \cdot ((x_{12} - \mu_{12}) \cdot (x_{12} - \mu_{12})) + 0.008 \cdot ((x_{22} - \mu_{12}) \cdot (x_{22} - \mu_{12})) + 0.932 \cdot ((x_{32} - \mu_{12}) \cdot (x_{32} - \mu_{12}))}{1.276} \\
&= \frac{0.336 \cdot ((0 - (-0.718))(0 - (-0.718))) + 0.008 \cdot ((2 - (-0.718))(2 - (-0.718))) + 0.932 \cdot ((-1 - (-0.718))^2)}{1.276} \\
&= \mathbf{0.240}
\end{aligned}$$

→ cluster $c = 2$

$$\mu_2 = \frac{0.664 \cdot x_1 + 0.992 \cdot x_2 + 0.068 \cdot x_3}{0.664 + 0.992 + 0.068} = \frac{0.664 \cdot \binom{1}{0} + 0.992 \cdot \binom{0}{2} + 0.068 \cdot \binom{3}{-1}}{1.724} = \binom{0.503}{1.111}$$

$$\Sigma_2 = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 0.487 & -0.678 \\ -0.678 & 1.106 \end{pmatrix}$$

$$\begin{aligned}
\Sigma_{11} &= \frac{0.664 \cdot ((x_{11} - \mu_{21}) \cdot (x_{11} - \mu_{21})) + 0.992 \cdot ((x_{21} - \mu_{21}) \cdot (x_{21} - \mu_{21})) + 0.068 \cdot ((x_{31} - \mu_{21}) \cdot (x_{31} - \mu_{21}))}{1.724} \\
&= \frac{0.664 \cdot ((1 - 0.503)(1 - 0.503)) + 0.992 \cdot ((0 - 0.503)(0 - 0.503)) + 0.068 \cdot ((3 - 0.503)(3 - 0.503))}{1.724} \\
&= \mathbf{0.487}
\end{aligned}$$

$$\begin{aligned}
\Sigma_{21} &= \frac{0.664 \cdot ((x_{12} - \mu_{22}) \cdot (x_{11} - \mu_{21})) + 0.992 \cdot ((x_{22} - \mu_{22}) \cdot (x_{21} - \mu_{21})) + 0.068 \cdot ((x_{32} - \mu_{22}) \cdot (x_{31} - \mu_{21}))}{1.724} \\
&= \frac{0.664 \cdot ((0 - 1.111)(1 - 0.503)) + 0.992 \cdot ((2 - 1.111)(0 - 0.503)) + 0.068 \cdot ((-1 - 1.111)(3 - 0.503))}{1.724} \\
&= -\mathbf{0.678}
\end{aligned}$$

$$\Sigma_{12} = \Sigma_{21} = -\mathbf{0.678}$$

$$\begin{aligned}
\Sigma_{22} &= \frac{0.664 \cdot ((x_{12} - \mu_{22}) \cdot (x_{12} - \mu_{22})) + 0.992 \cdot ((x_{22} - \mu_{22}) \cdot (x_{22} - \mu_{22})) + 0.068 \cdot ((x_{32} - \mu_{22}) \cdot (x_{32} - \mu_{22}))}{1.724} \\
&= \frac{0.664 \cdot ((0 - 1.111)(0 - 1.111)) + 0.992 \cdot ((2 - 1.111)(2 - 1.111)) + 0.068 \cdot ((-1 - 1.111)^2)}{1.724} \\
&= \mathbf{1.106}
\end{aligned}$$

Normalized priors:

$$P(c=1) = \frac{0.336+0.008+0.932}{(0.336+0.008+0.932)+(0.664+0.992+0.068)} = 0.425$$
$$P(c=2) = \frac{0.664+0.992+0.068}{(0.336+0.008+0.932)+(0.664+0.992+0.068)} = 0.575$$

2. **Using the final parameters computed in previous question:**

   a) **perform a hard assignment of observations to clusters under a MAP assumption.**

   Performing another E-STEP with the calculations above we obtain:

   $\boxed{x_1}$

   → Cluster $c=1$:

   prior: $P(c=1) = \pi_1 = \mathbf{0.425}$

   likelihood: $p(x_1|c=1) = \mathcal{N}(x_1|\mu_1, \sigma_1)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_1)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\cdot(x_1-\mu_1)^T\Sigma_1^{-1}\cdot(x_1-\mu_1)\right)$$
$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.011}} \cdot \exp\left(-\frac{1}{2}\cdot\left(\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}2.455\\-0.718\end{pmatrix}\right)^T\cdot\begin{pmatrix}0.812 & -0.429\\-0.429 & 0.240\end{pmatrix}^{-1}\cdot\left(\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}2.455\\-0.718\end{pmatrix}\right)\right)$$
$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.011}} \cdot \exp\left(-\frac{1}{2}\cdot\begin{pmatrix}-1.455 & 0.718\end{pmatrix}\cdot\begin{pmatrix}22.142 & 39.579\\39.579 & 74.915\end{pmatrix}\cdot\begin{pmatrix}-1.455\\0.718\end{pmatrix}\right) = \frac{e^{-1.4}}{2\pi\cdot\sqrt{0.011}} = \mathbf{0.374}$$

   joint probability: $P(c=1,x_1) = P(c=1)p(x_1|c=1) = \pi_1\cdot\mathcal{N}(x_1|\mu_1,\sigma_1) = \mathbf{0.159}$

   normalized posterior: $P(c=1|x_1) = \dfrac{0.159}{0.159+0.147} = \mathbf{0.52}$

   → Cluster $c=2$:

   prior: $P(c=2) = \pi_2 = \mathbf{0.575}$

   likelihood: $p(x_1|c=2) = \mathcal{N}(x_1|\mu_2, \sigma_2)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_2)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\cdot(x_1-\mu_2)^T\Sigma_2^{-1}\cdot(x_1-\mu_2)\right)$$
$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.079}} \cdot \exp\left(-\frac{1}{2}\cdot\left(\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}0.503\\1.111\end{pmatrix}\right)^T\cdot\begin{pmatrix}0.487 & -0.678\\-0.678 & 1.106\end{pmatrix}^{-1}\cdot\left(\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}0.503\\1.111\end{pmatrix}\right)\right)$$
$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.079}} \cdot \exp\left(-\frac{1}{2}\cdot\begin{pmatrix}0.497 & -1.111\end{pmatrix}\cdot\begin{pmatrix}14.011 & 8.589\\8.589 & 6.169\end{pmatrix}\cdot\begin{pmatrix}0.497\\-1.111\end{pmatrix}\right) = \frac{e^{-0.795}}{2\pi\cdot\sqrt{0.079}} = \mathbf{0.256}$$

   joint probability: $P(c=2,x_1) = P(c=2)p(x_1|c=2) = \pi_2\cdot\mathcal{N}(x_1|\mu_2,\sigma_2) = \mathbf{0.147}$

   normalized posterior: $P(c=2|x_1) = \dfrac{0.147}{0.147+0.159} = \mathbf{0.48}$

$\boxed{x_2}$

➞ Cluster $c = 1$:

prior: $P(c = 1) = \pi_1 = \mathbf{0.425}$
likelihood: $p(x_2|c = 1) = \mathcal{N}(x_2|\mu_1, \sigma_1)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_1)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\cdot(x_2-\mu_1)^T\Sigma_1^{-1}\cdot(x_2-\mu_1)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.011}} \cdot \exp\left(-\frac{1}{2}\cdot\left(\begin{pmatrix}0\\2\end{pmatrix}-\begin{pmatrix}2.455\\-0.718\end{pmatrix}\right)^T\cdot\begin{pmatrix}0.812 & -0.429\\-0.429 & 0.240\end{pmatrix}^{-1}\cdot\left(\begin{pmatrix}0\\2\end{pmatrix}-\begin{pmatrix}2.455\\-0.718\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.011}} \cdot \exp\left(-\frac{1}{2}\cdot\begin{pmatrix}-2.455 & 2.718\end{pmatrix}\cdot\begin{pmatrix}22.142 & 39.579\\39.579 & 74.915\end{pmatrix}\cdot\begin{pmatrix}-2.455\\2.718\end{pmatrix}\right) = \frac{e^{-79.345}}{2\pi\cdot\sqrt{0.011}} = \mathbf{0}$$

joint probability: $P(c = 1, x_2) = P(c = 1)p(x_2|c = 1) = \pi_1 \cdot \mathcal{N}(x_2|\mu_1, \sigma_1) = \mathbf{0}$

normalized posterior: $P(c = 1|x_2) = \dfrac{0}{0 + 0.} = \mathbf{0}$

➞ Cluster $c = 2$:

prior: $P(c = 2) = \pi_2 = \mathbf{0.575}$
likelihood: $p(x_2|c = 2) = \mathcal{N}(x_2|\mu_2, \sigma_2)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_2)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\cdot(x_2-\mu_2)^T\Sigma_2^{-1}\cdot(x_2-\mu_2)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.079}} \cdot \exp\left(-\frac{1}{2}\cdot\left(\begin{pmatrix}0\\2\end{pmatrix}-\begin{pmatrix}0.503\\1.111\end{pmatrix}\right)^T\cdot\begin{pmatrix}0.487 & -0.678\\-0.678 & 1.106\end{pmatrix}^{-1}\cdot\left(\begin{pmatrix}0\\2\end{pmatrix}-\begin{pmatrix}0.503\\1.111\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.079}} \cdot \exp\left(-\frac{1}{2}\cdot\begin{pmatrix}-0.503 & 0.889\end{pmatrix}\cdot\begin{pmatrix}14.011 & 8.589\\8.589 & 6.169\end{pmatrix}\cdot\begin{pmatrix}-0.503\\0.889\end{pmatrix}\right) = \frac{e^{-0.369}}{2\pi\cdot\sqrt{0.079}} = \mathbf{0.392}$$

joint probability: $P(c = 2, x_2) = P(c = 2)p(x_2|c = 2) = \pi_2 \cdot \mathcal{N}(x_2|\mu_2, \sigma_2) = \mathbf{0.225}$

normalized posterior: $P(c = 2|x_2) = \dfrac{0.225}{0.225 + 0} = \mathbf{1}$

$\boxed{x_3}$

→ Cluster $c = 1$:

prior: $P(c=1) = \pi_1 = \mathbf{0.425}$
likelihood: $p(x_3|c=1) = \mathcal{N}(x_3|\mu_1, \sigma_1)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_1)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\cdot(x_3-\mu_1)^T\Sigma_1^{-1}\cdot(x_3-\mu_1)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.011}} \cdot \exp\left(-\frac{1}{2}\cdot\left(\begin{pmatrix}3\\-1\end{pmatrix}-\begin{pmatrix}2.455\\-0.718\end{pmatrix}\right)^T \cdot \begin{pmatrix}0.812 & -0.429\\-0.429 & 0.240\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix}-\begin{pmatrix}2.455\\-0.718\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.011}} \cdot \exp\left(-\frac{1}{2}\cdot\begin{pmatrix}0.545 & -0.282\end{pmatrix}\cdot\begin{pmatrix}22.142 & 39.579\\39.579 & 74.915\end{pmatrix}\cdot\begin{pmatrix}0.545\\-0.282\end{pmatrix}\right) = \frac{e^{-0.184}}{2\pi\cdot\sqrt{0.011}} = \mathbf{1.262}$$

joint probability: $P(c=1, x_3) = P(c=1)p(x_3|c=1) = \pi_1 \cdot \mathcal{N}(x_3|\mu_1, \sigma_1) = \mathbf{0.536}$

normalized posterior: $P(c=1|x_3) = \dfrac{0.536}{0.536+0} = \mathbf{1}$

→ Cluster $c = 2$:

prior: $P(c=2) = \pi_2 = \mathbf{0.575}$
likelihood: $p(x_3|c=2) = \mathcal{N}(x_3|\mu_2, \sigma_2)$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_2)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\cdot(x_3-\mu_2)^T\Sigma_2^{-1}\cdot(x_3-\mu_2)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.079}} \cdot \exp\left(-\frac{1}{2}\cdot\left(\begin{pmatrix}3\\-1\end{pmatrix}-\begin{pmatrix}0.503\\1.111\end{pmatrix}\right)^T \cdot \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix}-\begin{pmatrix}0.503\\1.111\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.079}} \cdot \exp\left(-\frac{1}{2}\cdot\begin{pmatrix}2.497 & -2.111\end{pmatrix}\cdot\begin{pmatrix}14.011 & 8.589\\8.589 & 6.169\end{pmatrix}\cdot\begin{pmatrix}2.497\\-2.111\end{pmatrix}\right) = \frac{e^{-12.151}}{2\pi\cdot\sqrt{0.079}} = \mathbf{0}$$

joint probability: $P(c=1, x_3) = P(c=1)p(x_3|c=1) = \pi_1 \cdot \mathcal{N}(x_3|\mu_1, \sigma_1) = \mathbf{0}$

normalized posterior: $P(c=2|x_3) = \dfrac{0}{0.536+0} = \mathbf{0}$

Observations	$c=1$	$c=2$
x_1	0.52	0.48
x_2	0	1
x_3	1	0

Table 3: Normalized posteriors

Thus, we infer:

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

b) **compute the silhouette of the larger cluster (the one that has more observations assigned to it) using the Euclidean distance.**

$$s(x_1) = \frac{b(x_1)}{a(x_1)} - 1 = \frac{\|x_1 - x_2\|_2}{\|x_1 - x_3\|_2} - 1 = \frac{\sqrt{(1-0)^2 + (0-2)^2}}{\sqrt{(1-3)^2 + (0-(-1))^2}} - 1 = \frac{\sqrt{5}}{\sqrt{5}} - 1 = 0$$

$$s(x_3) = 1 - \frac{a(x_3)}{b(x_3)} = 1 - \frac{\|x_3 - x_1\|_2}{\|x_3 - x_2\|_2} = 1 - \frac{\sqrt{(3-1)^2 + (-1-0)^2}}{\sqrt{(3-0)^2 + (-1-2)^2}} = 1 - \frac{\sqrt{5}}{\sqrt{18}} = 0.473$$

$$s(c_1) = \frac{s(x_1) + s(x_3)}{2} = \frac{0 + (0.473)}{2} = 0.237$$

Part II: Programming

In the next exercise you will use the `accounts.csv` dataset. This dataset contains account details of bank clients, and the target variable y is binary ('has the client subscribed a term deposit?'). Select the first 8 features and remove duplicates and null values.

Hint: You can use `get_dummies()` to change the feature type from categorical to numerical (e.g. `pd.get_dummies(data, drop_first=True)`)

1. **Normalize the data using `MinMaxScaler`:**

   a) **Using `sklearn`, apply `k-means clustering` (without targets) on the normalized data with $k = \{2, 3, 4, 5, 6, 7, 8\}$. Apply k-means randomly initialized, using `max_iter = 500` and `random_state = 42`. Plot the different sum of squared errors (SSE) using the `_inertia` attribute of `k-means` according to the number of clusters.**

```python
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Load data
data = pd.read_csv('accounts.csv')
df = data.iloc[:, :8]
df = df.drop_duplicates().dropna()

# Convert categorical to numerical
data_dummies = pd.get_dummies(df, drop_first=True)

scaler = MinMaxScaler()
dn = scaler.fit_transform(data_dummies)

```

```python
SSE = []

for k in range(2, 9):
    kmeans = KMeans(n_clusters=k, max_iter=500, random_state=42)
    kmeans.fit(dn)
    SSE.append(kmeans.inertia_)

# Plot SSE vs number of clusters
plt.figure(figsize=(10, 6))
plt.plot(range(2, 9), SSE, marker='o')
plt.title('SSE vs Number of Clusters')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Sum of Squared Errors (SSE)')
plt.grid(True)
plt.show()
```

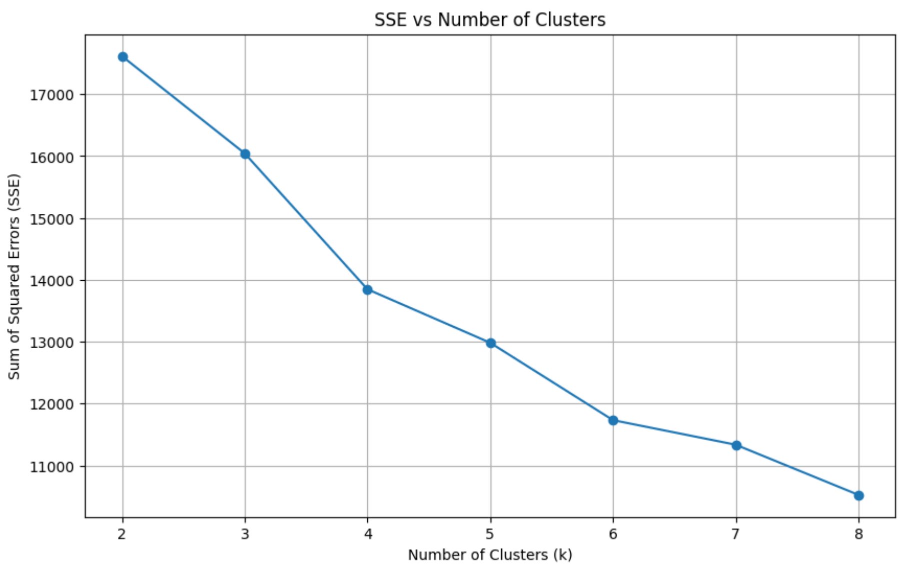

Figure 1: Sum of squared errors (SSE) according to the number of clusters

b) **According to the previous plot, how many underlying customer segments (clusters) should there be ? Explain based on the trade off between the clusters and inertia.**

As we can see in the previous plot, Sum of Squared Errors (SSE) decreases as the number of clusters (k) increases which means more clusters lead to better partitioning and lower variance (inertia). The highest decrease in SEE is in between $k = 2$ and $k = 4$. After that, it becomes more gradual, so there is a lower benefit for adding more clusters. Therefore, we suppose $k = 4$ is a reasonable number, as it balances the trade-off between cluster complexity and inertia.

c) **Would k-modes be a better clustering approach ? Explain why based on the dataset features.**

The dataset contains a mix of categorical features (job, marital, education, default, housing, loan, contact, month, poutcome, and deposit) and numerical features (age, balance, day, duration, campaign, pdays, previous). K-means clustering works by minimizing euclidean distance, which is effective for numerical data but unsuitable for categorical data and since this dataset has more categorical features, k-modes would be a better approach because instead of minimizing variance, it minimizes dissimilarity based on matching categories.

2. **Normalize the data using `StandardScaler`:**

   a) **Apply PCA to the data. How much variability is explained by the top 2 components?**

```python
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA

# Load the data
data = pd.read_csv('accounts.csv')
df = data.iloc[:, :8]
df = df.drop_duplicates().dropna()

# Convert categorical to numerical
data_dummies = pd.get_dummies(df, drop_first=True)

scaler = StandardScaler()

# Apply PCA
pca = PCA(n_components=2)  # Use 2 components
X_pca = pca.fit_transform(scaler.fit_transform(data_dummies))

print(f"Explained Variance Ratio for Top 2 PCs: {pca.
    explained_variance_ratio_}")
print(f"Total variability: {pca.explained_variance_ratio_.sum() *
    100:.2f}%")
```

   Variability explained by the top 2 components: 22.76%

   b) **Apply `k-means` clustering with $k = 3$ and `random_state=42` (all other arguments as default) and use the original 8 features. Next, provide a scatterplot according to the first 2 principal components. Can we clearly separate the clusters? Justify.**

```python
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Load the data
data = pd.read_csv('accounts.csv')
df = data.iloc[:, :8]
df = df.drop_duplicates().dropna()

```

```python
# Convert categorical to numerical
data_dummies = pd.get_dummies(df, drop_first=True)

scaler = StandardScaler()

# Apply PCA to reduce the data to 2 components
pca = PCA(n_components=2).fit_transform(scaler.fit_transform(
    data_dummies))

# Perform K-Means clustering
clusters = KMeans(n_clusters=3, random_state=42).fit_predict(scaler.
    fit_transform(data_dummies))

# Plot a scatterplot
plt.figure(figsize=(10, 6))
scatter = plt.scatter(pca[:, 0], pca[:, 1], c=clusters, cmap='plasma')
plt.title('PCA Scatterplot with K-Means Clustering (k=3)')
plt.legend(*scatter.legend_elements(), title='Clusters')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.grid(True)
plt.show()
```

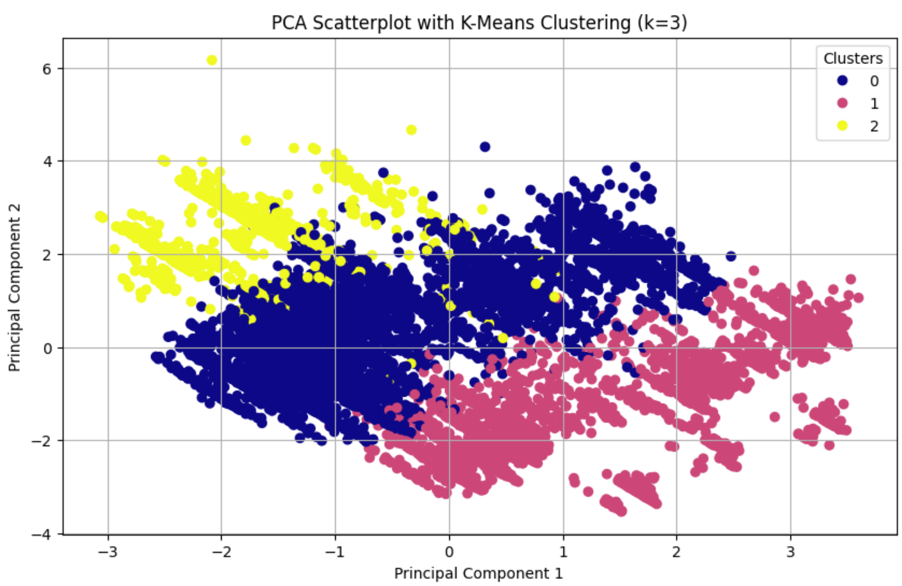

Figure 2: PCA Scatterplot

Some separation is visible for k-mean clustering with k = 3, but we can't say that they are clearly distinct since there is an overlap, which indicates that these clusters don't have distinct boundaries.

c) **Plot the cluster conditional features of the frequencies of 'job" and 'education" according to k-means, with** `multiple='dodge'`, `stat='density'`, `shrink=0.8`, `common_norm=False`. **Analyze the frequency plots using** `sns.displot`, **(see Data Exploration notebook). Describe the main differences between the clusters in no more than half page.**

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd


# Load the data
data = pd.read_csv('accounts.csv')
df = data.iloc[:, :8]
df = df.drop_duplicates().dropna()

df["clusters"] = clusters

plt.figure(figsize=(12, 6))

# Plot for "job"
sns.displot(data=df, x="job", hue ="clusters", multiple="dodge", stat=
    "density", shrink=0.8, common_norm=False)
plt.title('Frequency Distribution of Job')
plt.xticks(rotation=45)
plt.grid(True)
plt.show()

# Plot for "education"
sns.displot(data=df, x="education", hue ="clusters", multiple="dodge",
    stat="density", shrink=0.8, common_norm=False)
plt.title('Frequency Distribution of Education')
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```

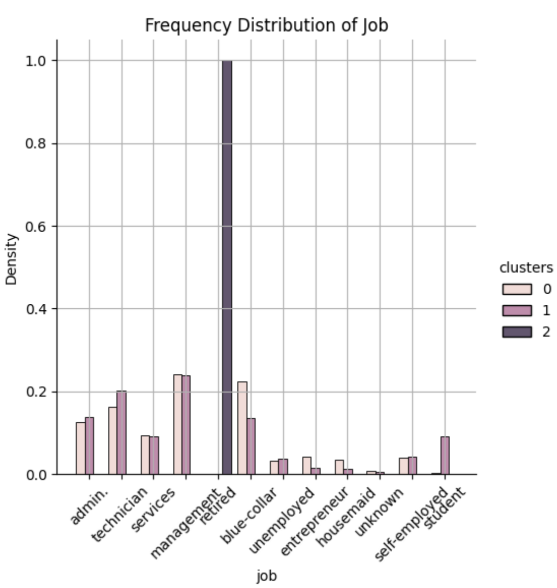

Figure 3: Frequency Distribution of Job

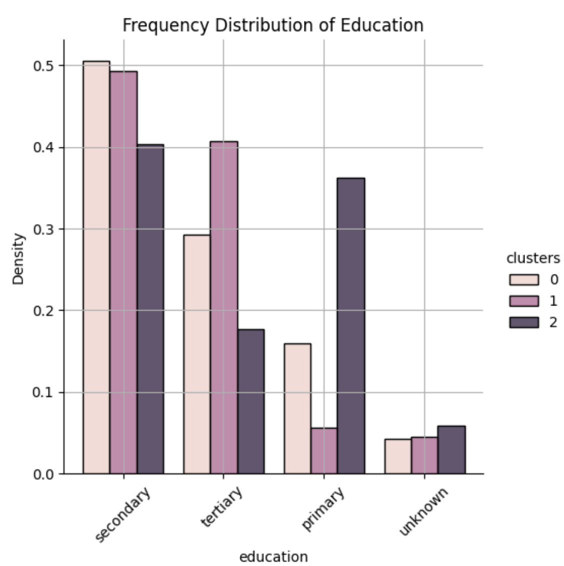

Figure 4: Frequency Distribution of Education

Figure 5: Cluster conditional features of the frequencies of 'job' and 'education'

For job frequency distribution:

- Cluster 0: Has relatively high density in **management** and **blue-collar**, which suggests that it is representing workers from a more traditional or stable sector.
- Cluster 1: Has very low density in **unemployment**, **entrepreneur**, **housemaid**, **self-employed** and practically zero for **unknown**. It is highly dense in **technician**, **management**, **blue-color** and **admin** which let us belive that represents worker from labor sector.
- Cluster 2: Since has density equal to zero for every job except for **retired**, it represents retired individuals.

For education frequency distribution:

- Cluster 0: The highest density value occurs in **secondary**, always decreasing for the remaining levels (in the order presented above), meaning it represents individuals with mid-level education.
- Cluster 1: Also has the highest density in **secondary**, but the value in **tertiary** is also pretty high. We can notice a large descent for the remaining levels (primary and unknown), these ones having almost the same value between each other. Thus, it represents people with higher-level education.
- Cluster 2: The highest values of density are found in **secondary** and **primary**, which suggests that it is refering to individuals with basic education.

In conclusion, cluster 0 seems to represent people with secondary education who work in blue-color and managment jobs, cluster 1 contains well-educated individuals working mostly on technician and management jobs and cluster 2 consists mainly of people with primary and secondary education, but also having the largest amount of individual in unknown education, who are retired.