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

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

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

$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{\sqrt{15}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}-2 & 3\end{pmatrix} \cdot \begin{pmatrix}\frac{4}{15} & -\frac{1}{15}\\-\frac{1}{15} & \frac{4}{15}\end{pmatrix} \cdot \begin{pmatrix}-2\\3\end{pmatrix}\right) = \frac{e^{-\frac{32}{15}}}{2\pi \cdot \sqrt{15}} = \mathbf{0.004}$$

joint probability: $P(c = 1, x_2) = P(c = 1)p(x_2|c = 1) = \pi_1 \cdot \mathcal{N}(x_2|\mu_1, \sigma_1) = \mathbf{0.002}$

normalized posterior: $P(c = 1|x_2) = \dfrac{0.002}{0.024 + 0.002} = \mathbf{0.077}$

→ 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: $P(c = 2|x_2) = \frac{0.024}{0.024 + 0.002} = \mathbf{0.923}$

$\boxed{x_3}$

→ 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 $c = 2$:

$$\begin{aligned}
&\text{prior: } P(c=2) = \pi_2 = \mathbf{0.5}\\
&\text{likelihood: } p(x_3|c=2) = \mathcal{N}(x_3|\mu_2, \sigma_2)\\
&= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{det(\Sigma_2)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_3-\mu_2)^T \Sigma_2^{-1} \cdot (x_3-\mu_2)\right)\\
&= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{4}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix} - \begin{pmatrix}1\\1\end{pmatrix}\right)^T \cdot \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix} - \begin{pmatrix}1\\1\end{pmatrix}\right)\right)\\
&= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{4}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}2 & -2\end{pmatrix} \cdot \begin{pmatrix}\frac{1}{2} & 0\\0 & \frac{1}{2}\end{pmatrix} \cdot \begin{pmatrix}2\\-2\end{pmatrix}\right) = \frac{e^{-2}}{2\pi\cdot\sqrt{4}} = \mathbf{0.012}
\end{aligned}$$

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	0.077	0.923
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:

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

For the covariance matrices:

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

Where

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

→ Cluster $c = 1$:

$$\mu_1 = \frac{0.326 \cdot x_1 + 0.077 \cdot x_2 + 0.75 \cdot x_3}{0.326 + 0.077 + 0.75} = \frac{0.326 \cdot \binom{1}{0} + 0.077 \cdot \binom{0}{2} + 0.750 \cdot \binom{3}{-1}}{1.153} = \binom{2.234}{-0.517}$$

$$\Sigma_1 = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 1.146 & -0.797 \\ -0.797 & 0.650 \end{pmatrix}$$

$$\begin{aligned}\Sigma_{11} &= \frac{0.326 \cdot ((x_{11} - \mu_{11}) \cdot (x_{11} - \mu_{11})) + 0.077 \cdot ((x_{21} - \mu_{11}) \cdot (x_{21} - \mu_{11})) + 0.75 \cdot ((x_{31} - \mu_{11}) \cdot (x_{31} - \mu_{11}))}{1.153} \\ &= \frac{0.326 \cdot ((1 - 2.234)(1 - 2.234)) + 0.077 \cdot ((0 - 2.234)(0 - 2.234)) + 0.750 \cdot ((3 - 2.234)(3 - 2.234))}{1.153} \\ &= \mathbf{1.146}\end{aligned}$$

$$\begin{aligned}\Sigma_{21} &= \frac{0.326 \cdot ((x_{12} - \mu_{12}) \cdot (x_{11} - \mu_{11})) + 0.077 \cdot ((x_{22} - \mu_{12}) \cdot (x_{21} - \mu_{11})) + 0.75 \cdot ((x_{32} - \mu_{12}) \cdot (x_{31} - \mu_{11}))}{1.153} \\ &= \frac{0.326 \cdot ((0 - (-0.517))(1 - 2.234)) + 0.077 \cdot ((2 - (-0.517))(0 - 2.234)) + 0.75 \cdot ((-1 - (-0.517))(3 - 2.234))}{1.153} \\ &= -\mathbf{0.797}\end{aligned}$$

$$\Sigma_{12} = \Sigma_{21} = -\mathbf{0.797}$$

$$\begin{aligned}\Sigma_{22} &= \frac{0.326 \cdot ((x_{12} - \mu_{12}) \cdot (x_{12} - \mu_{12})) + 0.077 \cdot ((x_{22} - \mu_{12}) \cdot (x_{22} - \mu_{12})) + 0.75 \cdot ((x_{32} - \mu_{12}) \cdot (x_{32} - \mu_{12}))}{1.153} \\ &= \frac{0.326 \cdot ((0 - (-0.517))(0 - (-0.517))) + 0.077 \cdot ((2 - (-0.517))(2 - (-0.517))) + 0.75 \cdot ((-1 - (-0.517))^2)}{1.153} \\ &= \mathbf{0.650}\end{aligned}$$

→ cluster $c = 2$

$$\mu_2 = \frac{0.674 \cdot x_1 + 0.923 \cdot x_2 + 0.250 \cdot x_3}{0.674 + 0.923 + 0.250} = \frac{0.674 \cdot \binom{1}{0} + 0.923 \cdot \binom{0}{2} + 0.250 \cdot \binom{3}{-1}}{1.847} = \binom{0.771}{0.864}$$

$$\Sigma_2 = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 0.989 & -1.072 \\ -1.072 & 1.388 \end{pmatrix}$$

$$\begin{aligned}\Sigma_{11} &= \frac{0.674 \cdot ((x_{11} - \mu_{21}) \cdot (x_{11} - \mu_{21})) + 0.923 \cdot ((x_{21} - \mu_{21}) \cdot (x_{21} - \mu_{21})) + 0.25 \cdot ((x_{31} - \mu_{21}) \cdot (x_{31} - \mu_{21}))}{1.847} \\ &= \frac{0.674 \cdot ((1 - 0.771)(1 - 0.771)) + 0.923 \cdot ((0 - 0.771)(0 - 0.771)) + 0.25 \cdot ((3 - 0.771)(3 - 0.771))}{1.847} \\ &= \mathbf{0.989}\end{aligned}$$

$$\begin{aligned}\Sigma_{21} &= \frac{0.674 \cdot ((x_{12} - \mu_{22}) \cdot (x_{11} - \mu_{21})) + 0.923 \cdot ((x_{22} - \mu_{22}) \cdot (x_{21} - \mu_{21})) + 0.25 \cdot ((x_{32} - \mu_{22}) \cdot (x_{31} - \mu_{21}))}{1.847} \\ &= \frac{0.674 \cdot ((0 - 0.864)(1 - 0.771)) + 0.923 \cdot ((2 - 0.864)(0 - 0.771)) + 0.25 \cdot ((-1 - 0.864)(3 - 0.771))}{1.847} \\ &= -\mathbf{1.072}\end{aligned}$$

$\Sigma_{12} = \Sigma_{21} = -\mathbf{1.072}$

$$\Sigma_{22} = \frac{0.674 \cdot ((x_{12} - \mu_{22}) \cdot (x_{12} - \mu_{22})) + 0.923 \cdot ((x_{22} - \mu_{22}) \cdot (x_{22} - \mu_{22})) + 0.25 \cdot ((x_{32} - \mu_{22}) \cdot (x_{32} - \mu_{22}))}{1.847}$$
$$= \frac{0.674 \cdot ((0 - 0.864)(0 - 0.864)) + 0.923 \cdot ((2 - 0.864)(2 - 0.864)) + 0.25 \cdot ((-1 - 0.864)^2)}{1.847}$$
$$= \mathbf{1.388}$$

Normalized priors:

$$P(c = 1) = \frac{0.326 + 0.077 + 0.75}{(0.326 + 0.077 + 0.75) + (0.674 + 0.923 + 0.25)} = 0.384$$
$$P(c = 2) = \frac{0.674 + 0.923 + 0.25}{(0.326 + 0.077 + 0.75) + (0.674 + 0.923 + 0.25)} = 0.616$$

SECOND EM EPOCH

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

$\boxed{x_1}$

→ Cluster $c = 1$:

prior: $P(c = 1) = \mathbf{0.384}$

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

$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{det(\Sigma_1)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (x_1 - \mu_1)^T \Sigma_1^{-1} \cdot (x_1 - \mu_1)\right)$$
$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{\sqrt{0.110}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}2.234\\-0.517\end{pmatrix}\right)^T \cdot \begin{pmatrix}1.146 & -0.797\\-0.797 & 0.650\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}2.234\\-0.517\end{pmatrix}\right)\right)$$
$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{\sqrt{0.110}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}-1.234 & 0.517\end{pmatrix} \cdot \begin{pmatrix}5.926 & 7.266\\7.266 & 10.448\end{pmatrix} \cdot \begin{pmatrix}-1.234\\0.517\end{pmatrix}\right) = \frac{e^{-1.273}}{2\pi \cdot \sqrt{0.110}}$$

$$= \mathbf{0.134}$$

joint probability: $P(c = 1, x_1) = P(c = 1)p(x_1|c = 1) = \pi_1 \cdot \mathcal{N}(x_1|\mu_1, \sigma_1) = \mathbf{0.051}$

normalized posterior: $P(c = 1|x_1) = \dfrac{0.051}{0.051 + 0.087} = \mathbf{0.370}$

→ Cluster $c = 2$:

prior: $P(c=2) = \mathbf{0.616}$

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

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

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.224}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}0.771\\0.864\end{pmatrix}\right)^T \cdot \begin{pmatrix}0.989 & -1.072\\-1.072 & 1.388\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}0.771\\0.864\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.224}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}0.229 & -0.864\end{pmatrix} \cdot \begin{pmatrix}6.209 & 4.795\\4.795 & 4.424\end{pmatrix} \cdot \begin{pmatrix}0.229\\-0.864\end{pmatrix}\right) = \frac{e^{-0.865}}{2\pi \cdot \sqrt{0.224}}$$

$$= \mathbf{0.142}$$

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: $P(c=2|x_1) = \dfrac{0.087}{0.051 + 0.087} = \mathbf{0.630}$

$\boxed{x_2}$

→ Cluster $c = 1$:

prior: $P(c=1) = \mathbf{0.384}$

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

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

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.110}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}2.234\\-0.517\end{pmatrix}\right)^T \cdot \begin{pmatrix}1.146 & -0.797\\-0.797 & 0.650\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}2.234\\-0.517\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.110}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}-2.234 & 2.517\end{pmatrix} \cdot \begin{pmatrix}5.926 & 7.266\\7.266 & 10.448\end{pmatrix} \cdot \begin{pmatrix}-2.234\\2.517\end{pmatrix}\right) = \frac{e^{-7.027}}{2\pi \cdot \sqrt{0.110}}$$

$$= \mathbf{0}$$

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: $P(c=1|x_2) = \dfrac{0}{0 + 0.126} = \mathbf{0}$

→ Cluster $c = 2$:

prior: $P(c = 2) = \mathbf{0.616}$
likelihood: $p(x_2|c = 2) = \mathcal{N}(x_2|\mu_2, \sigma_2)$

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

$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{\sqrt{0.224}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix} 0 \\ 2 \end{pmatrix} - \begin{pmatrix} 0.771 \\ 0.864 \end{pmatrix}\right)^T \cdot \begin{pmatrix} 0.989 & -1.072 \\ -1.072 & 1.388 \end{pmatrix}^{-1} \cdot \left(\begin{pmatrix} 0 \\ 2 \end{pmatrix} - \begin{pmatrix} 0.771 \\ 0.864 \end{pmatrix}\right)\right)$$

$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{\sqrt{0.224}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix} -0.771 & 1.136 \end{pmatrix} \cdot \begin{pmatrix} 6.209 & 4.795 \\ 4.795 & 4.424 \end{pmatrix} \cdot \begin{pmatrix} -0.771 \\ 1.136 \end{pmatrix}\right) = \frac{e^{-0.5}}{2\pi \cdot \sqrt{0.224}}$$

$$= \mathbf{0.204}$$

joint probability: $P(c = 2, x_2) = P(c = 2)p(x_2|c = 2) = \pi_2 \cdot \mathcal{N}(x_2|\mu_2, \sigma_2) = \mathbf{0.126}$

normalized posterior: $P(c = 2|x_2) = \dfrac{0.126}{0.126 + 0} = \mathbf{1}$

$\boxed{x_3}$

→ Cluster $c = 1$:

prior: $P(c = 1) = \mathbf{0.384}$
likelihood: $p(x_3|c = 1) = \mathcal{N}(x_3|\mu_1, \sigma_1)$

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

$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{\sqrt{0.110}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\begin{pmatrix} 3 \\ -1 \end{pmatrix} - \begin{pmatrix} 2.234 \\ -0.517 \end{pmatrix}\right)^T \cdot \begin{pmatrix} 1.146 & -0.797 \\ -0.797 & 0.650 \end{pmatrix}^{-1} \cdot \left(\begin{pmatrix} 3 \\ -1 \end{pmatrix} - \begin{pmatrix} 2.234 \\ -0.517 \end{pmatrix}\right)\right)$$

$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{\sqrt{0.110}} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix} 0.766 & -0.483 \end{pmatrix} \cdot \begin{pmatrix} 5.926 & 7.266 \\ 7.266 & 10.448 \end{pmatrix} \cdot \begin{pmatrix} 0.766 \\ -0.483 \end{pmatrix}\right) = \frac{e^{-0.269}}{2\pi \cdot \sqrt{0.110}}$$

$$= \mathbf{0.367}$$

joint probability: $P(c = 1, x_3) = P(c = 1)p(x_3|c = 1) = \pi_1 \cdot \mathcal{N}(x_3|\mu_1, \sigma_1) = \mathbf{0.141}$

normalized posterior: $P(c = 1|x_3) = \dfrac{0.141}{0.141 + 0.009} = \mathbf{0.940}$

→ Cluster $c = 2$:

prior: $P(c=2) = \mathbf{0.616}$

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

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

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.224}} \cdot \exp\left(-\frac{1}{2}\cdot\left(\begin{pmatrix}3\\-1\end{pmatrix}-\begin{pmatrix}0.771\\0.864\end{pmatrix}\right)^T \cdot \begin{pmatrix}0.989 & -1.072\\-1.072 & 1.388\end{pmatrix}^{-1} \cdot \left(\begin{pmatrix}3\\-1\end{pmatrix}-\begin{pmatrix}0.771\\0.864\end{pmatrix}\right)\right)$$

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{\sqrt{0.224}} \cdot \exp\left(-\frac{1}{2}\cdot(2.229 \quad -1.864)\cdot\begin{pmatrix}6.209 & 4.795\\4.795 & 4.424\end{pmatrix}\cdot\begin{pmatrix}2.229\\-1.864\end{pmatrix}\right) = \frac{e^{-3.188}}{2\pi\cdot\sqrt{0.224}}$$

$= \mathbf{0.014}$

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: $P(c=2|x_3) = \frac{0.009}{0.009+0.141} = \mathbf{0.060}$

Observations	$c=1$	$c=2$
x_1	0.370	0.630
x_2	0	1
x_3	0.940	0.060

Table 2: Normalized posteriors

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

→ Cluster $c = 1$:

$$\mu_1 = \frac{0.370\cdot x_1 + 0\cdot x_2 + 0.94\cdot x_3}{0.370+0+0.94} = \frac{0.37\cdot\begin{pmatrix}1\\0\end{pmatrix} + 0\cdot\begin{pmatrix}0\\2\end{pmatrix} + 0.94\cdot\begin{pmatrix}3\\-1\end{pmatrix}}{1.31} = \begin{pmatrix}2.435\\-0.716\end{pmatrix}$$

$$\Sigma_1 = \begin{pmatrix}\Sigma_{11} & \Sigma_{21}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix} = \begin{pmatrix}0.811 & -0.405\\-0.405 & 0.203\end{pmatrix}$$

$$\Sigma_{11} = \frac{0.37 \cdot ((x_{11}-\mu_{11}) \cdot (x_{11}-\mu_{11})) + 0 \cdot ((x_{21}-\mu_{11}) \cdot (x_{21}-\mu_{11})) + 0.94 \cdot ((x_{31}-\mu_{11}) \cdot (x_{31}-\mu_{11}))}{1.31}$$
$$= \frac{0.37 \cdot ((1-2.435)(1-2.435)) + 0.94 \cdot ((3-2.435)(3-2.435))}{1.31}$$
$$= \mathbf{0.811}$$

$$\Sigma_{21} = \frac{0.37 \cdot ((x_{12}-\mu_{12}) \cdot (x_{11}-\mu_{11})) + 0 \cdot ((x_{22}-\mu_{12}) \cdot (x_{21}-\mu_{11})) + 0.94 \cdot ((x_{32}-\mu_{12}) \cdot (x_{31}-\mu_{11}))}{1.31}$$
$$= \frac{0.37 \cdot ((0-(-0.716))(1-2.435)) + 0.94 \cdot ((-1-(-0.716))(3-2.435))}{1.31}$$
$$= -\mathbf{0.405}$$

$$\Sigma_{12} = \Sigma_{21} = -\mathbf{0.405}$$

$$\Sigma_{22} = \frac{0.37 \cdot ((x_{12}-\mu_{12}) \cdot (x_{12}-\mu_{12})) + 0 \cdot ((x_{22}-\mu_{12}) \cdot (x_{22}-\mu_{12})) + 0.94 \cdot ((x_{32}-\mu_{12}) \cdot (x_{32}-\mu_{12}))}{1.31}$$
$$= \frac{0.37 \cdot ((0-(-0.716))(0-(-0.716))) + 0.94 \cdot ((-1-(-0.716))^2)}{1.31}$$
$$= \mathbf{0.203}$$

→ cluster $c = 2$

$$\mu_2 = \frac{0.630 \cdot x_1 + 1 \cdot x_2 + 0.06 \cdot x_3}{0.63 + 1 + 0.06} = \frac{0.63 \cdot \binom{1}{0} + 1 \cdot \binom{0}{2} + 0.06 \cdot \binom{3}{-1}}{1.69} = \binom{0.479}{1.148}$$

$$\Sigma_2 = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 0.463 & -0.657 \\ -0.657 & 1.085 \end{pmatrix}$$

$$\Sigma_{11} = \frac{0.63 \cdot ((x_{11}-\mu_{21}) \cdot (x_{11}-\mu_{21})) + 1 \cdot ((x_{21}-\mu_{21}) \cdot (x_{21}-\mu_{21})) + 0.06 \cdot ((x_{31}-\mu_{21}) \cdot (x_{31}-\mu_{21}))}{1.69}$$
$$= \frac{0.63 \cdot ((1-0.479)(1-0.479)) + 1 \cdot ((0-0.479)(0-0.479)) + 0.06 \cdot ((3-0.479)(3-0.479))}{1.69}$$
$$= \mathbf{0.463}$$

$$\Sigma_{21} = \frac{0.63 \cdot ((x_{12}-\mu_{22}) \cdot (x_{11}-\mu_{21})) + 1 \cdot ((x_{22}-\mu_{22}) \cdot (x_{21}-\mu_{21})) + 0.06 \cdot ((x_{32}-\mu_{22}) \cdot (x_{31}-\mu_{21}))}{1.69}$$
$$= \frac{0.63 \cdot ((0-1.148)(1-0.479)) + 1 \cdot ((2-1.148)(0-0.479)) + 0.06 \cdot ((-1-1.148)(3-0.479))}{1.69}$$
$$= -\mathbf{0.657}$$

$$\Sigma_{12} = \Sigma_{21} = -\mathbf{0.657}$$

$$\Sigma_{22} = \frac{0.63 \cdot ((x_{12}-\mu_{22}) \cdot (x_{12}-\mu_{22})) + 1 \cdot ((x_{22}-\mu_{22}) \cdot (x_{22}-\mu_{22})) + 0.06 \cdot ((x_{32}-\mu_{22}) \cdot (x_{32}-\mu_{22}))}{1.69}$$
$$= \frac{0.63 \cdot ((0-1.148)(0-0-1.148)) + 1 \cdot ((2-1.148)(2-1.148)) + 0.06 \cdot ((-1-1.148)^2)}{1.69}$$
$$= \mathbf{1.085}$$

Normalized priors:

$$P(c=1) = \frac{0.37+0+0.94}{(0.37+0+0.94)+(0.63+1+0.06)} = 0.643$$
$$P(c=2) = \frac{0.63+1+0.06}{(0.37+0+0.94)+(0.63+1+0.06)} = 0.829$$

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

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

   From the calculations above we have the following results:

$$\begin{array}{ll} P(c=1|x_1) = 0.37 & P(c=2|x_1) = 0.63 \\ P(c=1|x_2) = 0.00 & P(c=2|x_2) = 1.00 \\ P(c=1|x_3) = 0.94 & P(c=2|x_3) = 0.06 \end{array}$$

   Thus, we infer:

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

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

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

$$s(x_2) = \frac{b(x_2)}{a(x_2)} - 1 = \frac{\|x_2 - x_1\|_2}{\|x_2 - x_3\|_2} - 1 = \frac{\sqrt{5}}{\sqrt{18}} - 1 = \frac{\sqrt{10}}{6} = 0.527$$

$$s(c_2) = \frac{s(x_1) + s(x_2)}{2} = 0.264$$

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?').

1. **Select the first 8 features and remove duplicates and null values. Normalize the data using `MinMaxScaler`. 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.

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

2. **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.**

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

4. **Apply PCA to the data :**

   a) **Use `StandardScaler` to scale the data before you apply `fit_transform`. How much variability is explained by the top 2 components ?**

   b) **Provide a scatterplot according to the first 2 principal components and color the points according to $k = 3$ clusters. Can we clearly separate the clusters ? Justify.**

5. **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.**