Part I: Pen and paper

Consider the bivariate observations

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

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

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

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

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

    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: $p(x_2|c = 2) = \mathcal{N}(x_2|\mu_2, \sigma_2)$

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

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

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

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: $p(x_3|c = 1) = \mathcal{N}(x_3|\mu_1, \sigma_1)$

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

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

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

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

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

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

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

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:

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

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 $c = 1$:

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

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

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

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

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

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

→ cluster $c = 2$

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

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

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

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

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

$$\Sigma_{22} = \frac{0.674 \cdot ((x_{12} - \mu_{22}) \cdot (x_{12} - \mu_{22})) + 0.889 \cdot ((x_{22} - \mu_{22}) \cdot (x_{22} - \mu_{22})) + 0.25 \cdot ((x_{32} - \mu_{22}) \cdot (x_{32} - \mu_{22}))}{1.813}$$

$$= \frac{0.674 \cdot ((0 - 0.843)(0 - 0.843)) + 0.889 \cdot ((2 - 0.843)(2 - 0.843)) + 0.25 \cdot ((-1 - 0.843)^2)}{1.813}$$

$$= \mathbf{1.389}$$

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

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

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

$$= \frac{1}{(2 \cdot \pi)} \cdot \frac{1}{\sqrt{0.147}} \cdot \exp\left(-\frac{1}{2} \cdot (-1.170 \quad 0.445) \cdot \begin{pmatrix} 5.507 & 6.339 \\ 6.339 & 8.533 \end{pmatrix} \cdot \begin{pmatrix} -1.170 \\ 0.445 \end{pmatrix}\right) = \frac{e^{-1.314}}{2\pi \cdot \sqrt{0.147}}$$

$$= \mathbf{0.112}$$

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

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

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

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

$$= \mathbf{0.144}$$

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

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

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

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

$$= \mathbf{0.003}$$

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: $p(x_2|c = 2) = \mathcal{N}(x_2|\mu_2, \sigma_2)$

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

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

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

$$= \mathbf{0.199}$$

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: $p(x_3|c = 1) = \mathcal{N}(x_3|\mu_1, \sigma_1)$

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

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

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

$$= \mathbf{0.310}$$

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

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

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

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

$$= \mathbf{0.015}$$

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 $c = 1$:

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

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

$$\Sigma_{11} = \frac{0.336 \cdot ((x_{11} - \mu_{11}) \cdot (x_{11} - \mu_{11})) + 0.008 \cdot ((x_{21} - \mu_{11}) \cdot (x_{21} - \mu_{11})) + 0.932 \cdot ((x_{31} - \mu_{11}) \cdot (x_{31} - \mu_{11}))}{1.276}$$

$$= \frac{0.336 \cdot ((1 - 2.455)(1 - 2.455)) + 0.008 \cdot ((0 - 2.455)(0 - 2.455)) + 0.932 \cdot ((3 - 2.455)(3 - 2.455))}{1.276}$$

$$= \mathbf{0.812}$$

$$\Sigma_{21} = \frac{0.336 \cdot ((x_{12} - \mu_{12}) \cdot (x_{11} - \mu_{11})) + 0.008 \cdot ((x_{22} - \mu_{12}) \cdot (x_{21} - \mu_{11})) + 0.932 \cdot ((x_{32} - \mu_{12}) \cdot (x_{31} - \mu_{11}))}{1.276}$$

$$= \frac{0.336 \cdot ((0 - (-0.718))(1 - 2.455)) + 0.008 \cdot ((2 - (-0.718))(0 - 2.455)) + 0.932 \cdot ((-1 - (-0.718))(3 - 2.455))}{1.276}$$

$$= \mathbf{-0.429}$$

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

$$\Sigma_{22} = \frac{0.336 \cdot ((x_{12} - \mu_{12}) \cdot (x_{12} - \mu_{12})) + 0.008 \cdot ((x_{22} - \mu_{12}) \cdot (x_{22} - \mu_{12})) + 0.932 \cdot ((x_{32} - \mu_{12}) \cdot (x_{32} - \mu_{12}))}{1.276}$$

$$= \frac{0.336 \cdot ((0 - (-0.718))(0 - (-0.718))) + 0.008 \cdot ((2 - (-0.718))(2 - (-0.718))) + 0.932 \cdot ((-1 - (-0.718))^2)}{1.276}$$

$$= \mathbf{0.240}$$

➡ cluster $c = 2$

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

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

$$\Sigma_{11} = \frac{0.664 \cdot ((x_{11} - \mu_{21}) \cdot (x_{11} - \mu_{21})) + 0.992 \cdot ((x_{21} - \mu_{21}) \cdot (x_{21} - \mu_{21})) + 0.068 \cdot ((x_{31} - \mu_{21}) \cdot (x_{31} - \mu_{21}))}{1.724}$$

$$= \frac{0.664 \cdot ((1 - 0.503)(1 - 0.503)) + 0.992 \cdot ((0 - 0.503)(0 - 0.503)) + 0.068 \cdot ((3 - 0.503)(3 - 0.503))}{1.724}$$

$$= \mathbf{0.487}$$

$$\Sigma_{21} = \frac{0.664 \cdot ((x_{12} - \mu_{22}) \cdot (x_{11} - \mu_{21})) + 0.992 \cdot ((x_{22} - \mu_{22}) \cdot (x_{21} - \mu_{21})) + 0.068 \cdot ((x_{32} - \mu_{22}) \cdot (x_{31} - \mu_{21}))}{1.724}$$

$$= \frac{0.664 \cdot ((0 - 1.111)(1 - 0.503)) + 0.992 \cdot ((2 - 1.111)(0 - 0.503)) + 0.068 \cdot ((-1 - 1.111)(3 - 0.503))}{1.724}$$

$$= \mathbf{-0.678}$$

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

$$\Sigma_{22} = \frac{0.664 \cdot ((x_{12} - \mu_{22}) \cdot (x_{12} - \mu_{22})) + 0.992 \cdot ((x_{22} - \mu_{22}) \cdot (x_{22} - \mu_{22})) + 0.068 \cdot ((x_{32} - \mu_{22}) \cdot (x_{32} - \mu_{22}))}{1.724}$$

$$= \frac{0.664 \cdot ((0 - 1.111)(0 - 1.111)) + 0.992 \cdot ((2 - 1.111)(2 - 1.111)) + 0.068 \cdot ((-1 - 1.111)^2)}{1.724}$$

$$= \mathbf{1.106}$$

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.**

   From the calculations above we have the following results:

   $$P(c=1|x_1) = 0.336 \quad P(c=2|x_1) = 0.664$$
   $$P(c=1|x_2) = 0.008 \quad P(c=2|x_2) = 0.992$$
   $$P(c=1|x_3) = 0.932 \quad P(c=2|x_3) = 0.068$$

   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.**

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

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

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

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.**