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

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

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

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

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: $P(c=1|x_1) = \dfrac{0.015}{0.015+0.007} = \mathbf{0.681}$

→ Cluster $c = 2$:

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

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

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

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

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{4} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}0 & -1\end{pmatrix} \cdot \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix} \cdot \begin{pmatrix}0\\-1\end{pmatrix}\right) = \frac{e^{-1}}{2\pi\cdot 4} = \mathbf{0.015}$$

joint probability: $P(c=2, x_1) = P(c=2)p(x_1|c=2) = \pi_2 \cdot \mathcal{N}(x_1|\mu_2, \sigma_2) = \mathbf{0.007}$

normalized posterior: $P(c=2|x_1) = \dfrac{0.007}{0.007+0.015} = \mathbf{0.318}$

$\boxed{x_2}$

→ Cluster $c = 1$:

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

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

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

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

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{15} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}-2 & 3\end{pmatrix} \cdot \begin{pmatrix}4 & 1\\1 & 4\end{pmatrix} \cdot \begin{pmatrix}-2\\3\end{pmatrix}\right) = \frac{e^{-20}}{2\pi\cdot 15} = \mathbf{0}$$

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: $P(c=1|x_2) = \dfrac{0}{0+0.003} = \mathbf{0}$

→ Cluster $c = 2$:

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

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

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

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{4} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}-1 & 1\end{pmatrix} \cdot \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix} \cdot \begin{pmatrix}-1\\1\end{pmatrix}\right) = \frac{e^{-2}}{2\pi\cdot 4} = \mathbf{0.005}$$

joint probability: $P(c=2, x_2) = P(c=2)p(x_2|c=2) = \pi_2 \cdot \mathcal{N}(x_2|\mu_2, \sigma_2) = \mathbf{0.003}$

normalized posterior: $P(c=2|x_2) = \frac{0.003}{0 + 0.003} = \mathbf{1}$

$\boxed{x_3}$

→ Cluster $c = 1$:

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

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

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

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{15} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}1 & 0\end{pmatrix} \cdot \begin{pmatrix}4 & 1\\1 & 4\end{pmatrix} \cdot \begin{pmatrix}1\\0\end{pmatrix}\right) = \frac{e^{-2}}{2\pi\cdot 15} = \mathbf{0.001}$$

joint probability: $P(c=1, x_3) = P(c=1)p(x_3|c=1) = \pi_1 \cdot \mathcal{N}(x_3|\mu_1, \sigma_1) = \mathbf{0.001}$

normalized posterior: $P(c=1|x_3) = \frac{0.001}{0 + 0.001} = \mathbf{1}$

➞ Cluster $c = 2$:

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

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

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

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

$$= \frac{1}{(2\cdot\pi)} \cdot \frac{1}{4} \cdot \exp\left(-\frac{1}{2} \cdot \begin{pmatrix}2 & -2\end{pmatrix} \cdot \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix} \cdot \begin{pmatrix}2\\-2\end{pmatrix}\right) = \frac{e^{-8}}{2\pi \cdot 4} = \mathbf{0}$$

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: $P(c=2|x_3) = \dfrac{0}{0 + 0.001} = \mathbf{0}$

| Observations | $c = 1$ | $c = 2$ |
|---|---|---|
| $x_1$ | 0.681 | 0.318 |
| $x_2$ | 0 | 1 |
| $x_3$ | 1 | 0 |

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

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

→ Cluster $c = 1$:

$$\mu_1 = \frac{0.681 \cdot x_1 + 0 \cdot x_2 + 1 \cdot x_3}{0.681 + 0 + 1} = \frac{0.681 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 \\ 2 \end{pmatrix} + 1 \cdot \begin{pmatrix} 3 \\ -1 \end{pmatrix}}{1.681} = \begin{pmatrix} 2.190 \\ -0.595 \end{pmatrix}$$

$$\Sigma_1 = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 0.964 & -0.482 \\ -0.482 & 0.241 \end{pmatrix}$$

$$\Sigma_{11} = \frac{0.681 \cdot ((x_{11} - \mu_{11}) \cdot (x_{11} - \mu_{11})) + 0 \cdot ((x_{21} - \mu_{11}) \cdot (x_{21} - \mu_{11})) + 1 \cdot ((x_{31} - \mu_{11}) \cdot (x_{31} - \mu_{11}))}{1.681}$$
$$= \frac{0.681 \cdot ((1 - 2.190) \cdot (1 - 2.190)) + 1 \cdot ((3 - 2.190) \cdot (3 - 2.190))}{1.681} = \frac{0.964 + 0.656}{1.681} = 0.964$$

$$\Sigma_{21} = \frac{0.681 \cdot ((x_{12} - \mu_{12}) \cdot (x_{11} - \mu_{11})) + 0 \cdot ((x_{22} - \mu_{12}) \cdot (x_{21} - \mu_{11})) + 1 \cdot ((x_{32} - \mu_{12}) \cdot (x_{31} - \mu_{11}))}{1.681}$$
$$= \frac{0.681 \cdot ((0 - (-0.595)) \cdot (1 - 2.190)) + 1 \cdot ((-1 - (-0.595)) \cdot (3 - 2.190))}{1.681} = \frac{-0.482 + -0.328}{1.681}$$
$$= -0.482$$

$$\Sigma_{12} = \Sigma_{21} = -0.482$$

$$\Sigma_{22} = \frac{0.681 \cdot ((x_{12} - \mu_{12}) \cdot (x_{12} - \mu_{12})) + 0 \cdot ((x_{22} - \mu_{12}) \cdot (x_{22} - \mu_{12})) + 1 \cdot ((x_{32} - \mu_{12}) \cdot (x_{32} - \mu_{12}))}{1.681}$$
$$= \frac{0.681 \cdot ((0 - (-0.595)) \cdot (0 - (-0.595))) + 1 \cdot ((-1 - (-0.595)) \cdot (-1 - (-0.595)))}{1.681} = \frac{0.241 + 0.164}{1.681}$$
$$= 0.241$$

→ cluster $c = 2$

$$\mu_2 = \frac{0.318 \cdot x_1 + 1 \cdot x_2 + 0 \cdot x_3}{0.318 + 1 + 0} = \frac{0.318 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 \\ 2 \end{pmatrix} + 1 \cdot \begin{pmatrix} 3 \\ -1 \end{pmatrix}}{1.318} = \begin{pmatrix} 2.517 \\ -0.759 \end{pmatrix}$$

$$\Sigma_2 = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 5.362 & -5.546 \\ -5.546 & 7.795 \end{pmatrix}$$

$$\Sigma_{11} = \frac{0.318 \cdot ((x_{11} - \mu_{21}) \cdot (x_{11} - \mu_{21})) + 1 \cdot ((x_{21} - \mu_{21}) \cdot (x_{21} - \mu_{21})) + 0 \cdot ((x_{31} - \mu_{21}) \cdot (x_{31} - \mu_{21}))}{1.318}$$
$$= \frac{0.318 \cdot ((1 - 2.517) \cdot (1 - 2.517)) + 1 \cdot ((0 - 2.517) \cdot (0 - 2.517))}{1.318} = \frac{0.732 + 6.335}{1.318} = 5.362$$

$$\Sigma_{21} = \frac{0.318 \cdot ((x_{12} - \mu_{22}) \cdot (x_{11} - \mu_{21})) + 1 \cdot ((x_{22} - \mu_{22}) \cdot (x_{21} - \mu_{21})) + 0 \cdot ((x_{32} - \mu_{22}) \cdot (x_{31} - \mu_{21}))}{1.318}$$
$$= \frac{0.318 \cdot ((0 - (-0.759)) \cdot (1 - 2.517)) + 1 \cdot ((2 - (-0.759)) \cdot (0 - 2.517))}{1.318} = \frac{-0.366 + (-6.944)}{1.318}$$
$$= -5.546$$

$$\Sigma_{12} = \Sigma_{21} = -5.546$$

$$\Sigma_{22} = \frac{0.318 \cdot ((x_{12} - \mu_{22}) \cdot (x_{12} - \mu_{22})) + 1 \cdot ((x_{22} - \mu_{22}) \cdot (x_{22} - \mu_{22})) + 0 \cdot ((x_{32} - \mu_{22}) \cdot (x_{32} - \mu_{22}))}{1.318}$$
$$= \frac{0.318 \cdot ((0 - (-0.759)) \cdot (0 - (-0.759))) + 1 \cdot ((2 - (-0.759)) \cdot (2 - (-0.759)))}{1.318}$$
$$= \frac{0.183 + 7.612}{1.318} = 7.795$$

Normalized priors:

$$P(c=1) = \frac{0.681 + 0 + 1}{(0.681 + 0 + 1) + (0.318 + 1 + 0)} = 0.561$$
$$P(c=2) = \frac{0.318 + 1 + 0}{(0.681 + 0 + 1) + (0.318 + 1 + 0)} = 0.439$$

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

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

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

## **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.**