Introduction to Probabilistic Graphical Models **VU Thi Hai Yen**
Prof. Umut Simsekli haiyen96.hp@gmail.com

Practical Session 2

QUESTION 1:

In Gaussian Mixture Models, we often need to evaluate quantities that are defined as follows:

$$\gamma_i(x) = \frac{\pi_i \mathcal{N}(x; \mu_i, \Sigma_i)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x; \mu_j, \Sigma_j)}, \tag{1}$$

where $\pi_i \in [0;1]$ and $\mathcal{N}$ denotes the multivariate Gaussian distribution. A direct computation of these quantities might be problematic in practice since all the terms $(\pi_i \mathcal{N}(x; \mu_i, \Sigma_i))$ might be very small, and we might end up with $0/0$.
Derive mathematically (don't try to take the derivative!) and implement a function for numerically stable computation of $\{\gamma_i\}_{i=1}^{K}$.

SOLUTION:
Let $\ell_i = \log \pi_i \mathcal{N}(x; \mu_i, \Sigma_i)$, we first try to compute ℓ_i stably. We have:

$$\begin{aligned} \ell_i &= \log \pi_i \mathcal{N}(x; \mu_i, \Sigma_i) \\ &= \log \pi_i + \log \left[(2\pi)^{-d/2} \left(\det(\Sigma_i)\right)^{-1/2} \exp\left(-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right) \right] \\ &= \log \pi_i - \frac{d}{2} \log(2\pi) - \frac{1}{2} \log\left(\det(\Sigma_i)\right) - \frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i), \end{aligned} \tag{2}$$

where d is the dimension of x. To compute $\log\left(\det(\Sigma_i)\right)$ efficiently, we can utilize the function `numpy.linalg.slogdet`.

Now if we can compute $\ell_1, \ell_2, \ldots, \ell_K$, we can compute $\log \gamma_i(x)$ as follows:

$$\begin{aligned} \log \gamma_i(x) &= \log \pi_i \mathcal{N}(x; \mu_i, \Sigma_i) - \log \left(\sum_{j=1}^{K} \pi_j \mathcal{N}(x; \mu_j, \Sigma_j) \right) \\ &= \ell_i - \log \left(\sum_{j=1}^{K} \exp(\ell_j) \right). \end{aligned} \tag{3}$$

QUESTION 2:

Let us consider a Gaussian Mixture Model (GMM), given as follows:

$$p(x_n) = \sum_{i=1}^{K} \pi_i \mathcal{N}(x_n; \mu_i, \Sigma_i) \tag{4}$$

where $\{x_n\}_{n=1}^{N}$ is a set of observed data points. Derive the M-Step of the Expectation - Maximization algorithm for this model, to find $\pi_{1:K}^{(t+1)}, \mu_{1:K}^{(t+1)}, \Sigma_{1:K}^{(t+1)}$, where t denotes the iteration number.

SOLUTION:
We first recall the context of the Expectation-Maximization (EM) algorithm.

Given a set of observed data $\mathbf{X}$ which we suppose that depends on a set of hidden data $\mathbf{Z}$. The EM algorithm then try to model the data set with a set of parameters $\boldsymbol{\theta}$ by maximizing the marginal likelihood of the observed data:

$$L(\boldsymbol{\theta}; \mathbf{X}) = p(\mathbf{X} \mid \boldsymbol{\theta}) = \int p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})\, d\mathbf{Z}. \tag{5}$$

It iteratively performs the two following steps:

1. **E-step:** Compute the expectation of the log likelihood function:

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(t)}} \left[\log L(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Z})\right]. \tag{6}$$

2. **M-step:** Find the parameters that maximize the quantity:

$$\boldsymbol{\theta}^{(t+1)} = \underset{\boldsymbol{\theta}}{\arg\max}\, Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}). \tag{7}$$

For a Gaussian Mixture Model (GMM), we admit the following formula for the **E-step**:

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_k^{(t)}(x_n) \left[\log \pi_k - \frac{d}{2}\log(2\pi) - \frac{1}{2}\log\left(\det(\Sigma_k)\right) - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k)\right], \tag{8}$$

where $\gamma_k^{(t)}(x_n)$ is defined as in Equation (1) but for $\pi_k^{(t)}, \mu_k^{(t)}$ and $\Sigma_k^{(t)}$.

We perform the **M-step** by computing the parameters as follows:

$$\begin{aligned} \pi_k^{(t+1)} &= \arg\max_{\boldsymbol{\pi_k}} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_k^{(t)}(x_n) \log \pi_k \\ \mu_k^{(t+1)}, \Sigma_k^{(t+1)} &= \arg\max_{\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_k^{(t)}(x_n) \left[-\frac{1}{2} \log\left(\det(\Sigma_k)\right) - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) \right]. \end{aligned} \tag{9}$$

Taking into account that $\sum_{k=1}^{K} \gamma_k^{(t)}(x_n) = 1$, we derive the following formulas to perform the **M-step**:

$$\begin{aligned} \pi_k^{(t+1)} &= \frac{1}{N} \sum_{n=1}^{N} \gamma_k^{(t)}(x_n) \\ \mu_k^{(t+1)} &= \frac{\sum_{n=1}^{N} \gamma_k^{(t)}(x_n) x_n}{\sum_{n=1}^{N} \gamma_k^{(t)}(x_n)} \\ \Sigma_k^{(t+1)} &= \frac{\sum_{n=1}^{N} \gamma_k^{(t)}(x_n) \left(x_n - \mu_k^{(t+1)}\right) \left(x_n - \mu_k^{(t+1)}\right)^T}{\sum_{n=1}^{N} \gamma_k^{(t)}(x_n)}. \end{aligned} \tag{10}$$

For Question 3, the implementation is in the file `.ipynb`.