

EM for Gaussian mixtures ($d = 1$)

October 23, 2025

We observe N realisations $X_1, \dots, X_N$ and K groups. We have introduced some latent variables (unobserved), denoted as $(z_1, \dots, z_N)$, such that

$$z_{ik} = \begin{cases} 1 & \text{if } X_i \text{ belongs to cluster } k \\ 0 & \text{otherwise} \end{cases}$$

The generative model is the following one:

$$\begin{cases} z_i \sim \mathcal{M}(1, \pi), \text{ with } \pi = (\pi_1, \dots, \pi_K) \text{ and } 0 < \pi_k < 1, \forall k \text{ and } \sum_{k=1}^K \pi_k = 1. \\ X_i | z_{ik} = 1 \sim \mathcal{N}(\mu_k, \sigma_k^2) \end{cases}$$

Note that we consider the case where we have only one variable ($d = 1$), but the general case can be easily derived ($d > 1$ when $X_i | z_{ik} = 1$ follows a multivariate Gaussian distribution).

The parameters are $\Theta = (\pi, \mu, \sigma)$ where $\pi = (\pi_1, \dots, \pi_K)$ are the probability of belonging to the clusters and $\mu = (\mu_1, \dots, \mu_K)$, $\sigma = (\sigma_1, \dots, \sigma_K)$ are the parameters of the data distribution.

EM algorithm Starting from an initialization $\Theta^{(0)}$, the EM algorithm consists of two steps proceeded iteratively:

- E-step: compute $Q(\Theta, \Theta^{(t)}) = \mathbb{E}[\log p(X, z | \Theta) | X, \Theta^{(t)}]$
- M-step: update the parameters with $\Theta^{(t+1)} \in \operatorname{argmax}_{\Theta} Q(\Theta, \Theta^{(t)})$

E-step

$$\begin{aligned} p(X, z | \Theta) &= \prod_{i=1}^N p(X_i, z_i; \Theta) \\ &= \prod_{i=1}^N \prod_{k=1}^K (\pi_k f_{\mathcal{N}(\mu_k, \sigma_k^2)}(X_i))^{z_{ik}} \end{aligned}$$

Therefore we have

$$\begin{aligned} \mathcal{L}_C(X, z | \Theta) &= \log p(X, z | \Theta) \\ &= \sum_{i=1}^N \sum_{k=1}^K z_{ik} \log(\pi_k f_{\mathcal{N}(\mu_k, \sigma_k^2)}(X_i)), \end{aligned}$$

where $f_{\mathcal{N}(\mu_k, \sigma_k^2)}(X_i) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp(-\frac{1}{2\sigma_k^2} (X_i - \mu_k)^2)$ is the probability density function of $\mathcal{N}(\mu_k, \sigma_k^2)$.

Let us compute the E-step; the only term which can not be out of the integral is z_{ik} .

$$\begin{aligned} Q(\Theta, \Theta^{(t)}) &= \mathbb{E}[\mathcal{L}_C(X, z|\Theta)|X, \Theta^{(t)}] \\ &= \sum_{i=1}^N \sum_{k=1}^K \mathbb{E}[z_{ik}|X, \Theta^{(t)}] \log(\pi_k f_{\mathcal{N}(\mu_k, \sigma_k^2)}(X_i)) \\ &= \sum_{i=1}^N \sum_{k=1}^K t_{ik}^{(t)} \left(-\log \sqrt{2\pi} - \frac{1}{2} \log \sigma_k^2 - \frac{1}{2\sigma_k^2} (X_i - \mu_k)^2 \right), \end{aligned}$$

with the notation $t_{ik}^{(t)} := \mathbb{E}[z_{ik}|X, \Theta^{(t)}]$

Finally, the E-step involves computing $t_{ik}^{(t)}$, which can be written as follows

$$\begin{aligned} t_{ik}^{(t)} &= \mathbb{E}[z_{ik}|X, \Theta^{(t)}] = \frac{p(X_i, z_i; \Theta^{(t)})}{p(X_i; \Theta^{(t)})} \\ &= \frac{\pi_k^{(t)} f_{\mathcal{N}(\mu_k^{(t)}, (\sigma_k^2)^{(t)})}(X_i)}{\sum_{\ell=1}^K \pi_{\ell}^{(t)} f_{\mathcal{N}(\mu_{\ell}^{(t)}, (\sigma_{\ell}^2)^{(t)})}(X_i)} \end{aligned}$$

M-step We have to solve $\Theta^{(t+1)} \in \operatorname{argmax}_{\Theta} Q(\Theta, \Theta^{(t)})$.

- The maximization over $\pi_k, \forall k \in \{1, \dots, K\}$ should be made under the constraint that $\sum_{k=1}^K \pi_k = 1$. We thus use Lagrange multiplier, and define the function $g(\pi_1, \dots, \pi_K, \lambda) = \sum_{k=1}^K t_{ik}^{(t)} \log \pi_k + \lambda(\sum_{k=1}^K \pi_k - 1)$

$$\begin{cases} \frac{\partial}{\partial \pi_k} g(\pi_1, \dots, \pi_K, \lambda) &= \frac{1}{\pi_k} \sum_{i=1}^N t_{ik}^{(t)} + \lambda \\ \frac{\partial}{\partial \lambda} g(\pi_1, \dots, \pi_K, \lambda) &= \sum_{k=1}^K \pi_k - 1 \end{cases}$$

$$\nabla g = 0 \Leftrightarrow (\dots) \Leftrightarrow \begin{cases} \pi_k^{(t+1)} = \frac{1}{N} \sum_{i=1}^N t_{ik}^{(t)} \\ \lambda = -N \end{cases}$$

- Maximization over $\mu_k, \forall k \in \{1, \dots, K\}$:

$$\begin{aligned} \frac{\partial}{\partial \mu_k} Q(\Theta, \Theta^{(t)}) &= \frac{1}{2} \sum_{i=1}^N t_{ik}^{(t)} \frac{1}{\sigma_k^2} 2(X_i - \mu_k) \\ &= \frac{1}{\sigma_k^2} \sum_{i=1}^N t_{ik}^{(t)} X_i - \frac{\mu_k}{\sigma_k^2} \sum_{i=1}^N t_{ik}^{(t)} \end{aligned}$$

$$\text{Thus, } \frac{\partial}{\partial \mu_k} Q(\Theta, \Theta^{(t)}) = 0 \Leftrightarrow \mu_k^{(t+1)} = \frac{\sum_{i=1}^N t_{ik}^{(t)} X_i}{\sum_{i=1}^N t_{ik}^{(t)}}$$

- Maximization over σ_k^2 :

$$\frac{\partial}{\partial \sigma_k^2} Q(\Theta, \Theta^{(t)}) = -\frac{1}{2\sigma_k^2} \sum_{i=1}^N t_{ik}^{(t)} + \frac{1}{2\sigma_k^4} \sum_{i=1}^N t_{ik}^{(t)} (X_i - \mu_k)^2$$

As this expression involves μ_k (unknown quantity), we can use plug-in with $\mu_k = \mu_k^{(t+1)}$. Thus, $\frac{\partial}{\partial \sigma_k^2} Q(\Theta, \Theta^{(t)}) = 0 \Leftrightarrow (\sigma_k^2)^{(t+1)} = \frac{\sum_{i=1}^N t_{ik}^{(t)} (X_i - \mu_k^{(t+1)})^2}{\sum_{i=1}^N t_{ik}^{(t)}}$