Homework: Expectation-Maximization

Tu T. Do, 21C29023

February 23, 2023

1 Gaussian Mixture Model

Given the latent variable z, random variable x follows Gaussian distribution

$$x_i|z_i = k; \theta \sim \mathcal{N}(\mu_k, \Sigma_k); \quad \theta = (\pi_{1:M}, \mu_{1:M}, \Sigma_{1:M}); \quad x_i \in \mathbb{R}^P \tag{1}$$

The probability density function (p.d.f) for x:

$$\begin{aligned} f(x_i|z_i = k, \theta) &= \mathcal{N}(x_i|\mu_k, \Sigma_k) \\ &= \prod_{k=1}^{M} \left[\mathcal{N}(x_i|\mu_k, \Sigma_k)\right]^{\mathbb{I}(z_i=k)} \end{aligned} \tag{2}$$

In which

$$\mathbb{I}(z_i = k) = \begin{cases} 1, & \text{If } z_i = k, \\ 0, & \text{If } z_i \neq k \end{cases}$$

Probability mass function (p.m.f) for z:

$$p(z_i = k|\theta) = \pi_k = \prod_{k=1}^{M} {\pi_k}^{\mathbb{I}(z_i=k)} \tag{3}$$

Joint probability of x, z

$$\begin{aligned} p(x_i, z_i|\theta) &= p(x_i|z_i, \theta)p(z_i|\theta) \\ &= \prod_{k=1}^{M} \left[\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)\right]^{\mathbb{I}(z_i=k)} \end{aligned} \tag{4}$$

$$\begin{aligned} P(X, Z|\theta) &= \prod_{i=1}^{N} p(x_i, z_i|\theta) \\ &= \prod_{i=1}^{N} \prod_{k=1}^{M} \left[\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)\right]^{I(z_i=k)} \end{aligned} \tag{5}$$

We need to maximize the quantity above, using **EM**-algorithm as follow

- The **E**-step
  - Consider the expectation of the complete log-likelihood

$$
\begin{aligned}
Q(\theta|\theta^t) &= \mathbb{E}_{Z|X,\theta^t}\left[\log P(X,Z|\theta)\right] \\
&= \mathbb{E}_{Z|X,\theta^t}\left[\log \prod_{i=1}^{N}\prod_{k=1}^{M}\left[\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)\right]^{\mathbb{I}(z_i=k)}\right] \\
&= \mathbb{E}_{Z|X,\theta^t}\left[\sum_{i=1}^{N}\sum_{k=1}^{M}\mathbb{I}(z_i=k)[\log \pi_k + \log \mathcal{N}(x_i|\mu_k,\Sigma_k)]\right] \\
&= \sum_{i=1}^{N}\sum_{k=1}^{M}\mathbb{E}_{z_i|x_i,\theta^t}[\mathbb{I}(z_i=k)][\log \pi_k + \log \mathcal{N}(x_i|\mu_k,\Sigma_k)]
\end{aligned} \tag{6}
$$

  - Because $\log \pi_k + \log \mathcal{N}(x_i|\mu_k,\Sigma_k)$ is constant with respect to z, we can take them out of the expectation. Consider the expectation of the indicator function

$$
\begin{aligned}
\mathbb{E}_{z_i|x_i,\theta^t}[\mathbb{I}(z_i=k)] &= \sum_{z_i=1}^{M} p(z_i|x_i,\theta^t)\mathbb{I}(z_i=k) \\
&= p(z_i=k|x_i,\theta^t) \\
&= \frac{p(x_i|z_i=k,\theta^t)p(z_i=k|\theta^t)}{\sum_{g=1}^{M} p(x_i|z_i=g,\theta^t)p(z_i=g|\theta^t)} \\
&= \frac{\pi_k^t \mathcal{N}(x_i|\mu_k^t,\Sigma_k^t)}{\sum_{g=1}^{M}\pi_g^t \mathcal{N}(x_i|\mu_g^t,\Sigma_g^t)} =: A_{ik}
\end{aligned} \tag{7}
$$

  - Subsititute eq.8 to eq.7 we have

$$
Q(\theta|\theta^t) = \sum_{i=1}^{N}\sum_{k=1}^{M} A_{ik}[\log \pi_k + \log \mathcal{N}(x_i|\mu_k,\Sigma_k)] \tag{8}
$$

- The **M**-step : We need to maximize $Q(\theta,\theta^t)$ that subject to constraint $\sum_{k=1}^{M}\pi_k - 1 = 0$
  Let

$$
K(\theta) = Q(\theta|\theta^t) - \lambda(\sum_{i=1}^{M}\pi_k - 1) \tag{9}
$$

– Solve for π_k

$$\begin{aligned}
&\frac{\partial K}{\partial \pi_k} = 0 \\
&\Longleftrightarrow \frac{\partial}{\partial \pi_k}\Big\{ \sum_{i=1}^{N}\sum_{k=1}^{M} A_{ik}[\log \pi_k + \log \mathcal{N}(x_i|\mu_k, \Sigma_k)] + \lambda(\sum_{k=1}^{M} -1)\Big\} = 0 \\
&\Longleftrightarrow \sum_{i=1}^{N} A_{ik}\frac{1}{\pi_k} - \lambda = 0 \\
&\Longleftrightarrow \pi_k = \frac{\sum_{i=1}^{N} A_{ik}}{\lambda}
\end{aligned} \tag{10}$$

Substitute π_k into the constrain

$$\begin{aligned}
&\sum_{k=1}^{M} \pi_k = 1 \\
\Longleftrightarrow &\sum_{k=1}^{M} \frac{\sum_{i=1}^{N} A_{ik}}{\lambda} = 1 \\
\Longleftrightarrow &\lambda = \sum_{k=1}^{M}\sum_{i=1}^{N} A_{ik} \\
&= N
\end{aligned} \tag{11}$$

So that we have

$$\pi_k = \frac{\sum_{i=1}^{N} A_{ik}}{N} \tag{12}$$

– Solve for μ_k, recall the p.d.f of Multivariate Gaussian Distribution

$$\mathcal{N}(x_i|\mu_k, \Sigma_k) = (2\pi)^{\frac{-P}{2}} \det \Sigma_k{}^{-\frac{1}{2}} \exp\Big[-\frac{1}{2}(x_i - \mu_k)^\top \Sigma_k^{-1}(x_i - \mu_k)\Big]$$

$$\begin{aligned}
&\frac{\partial K}{\partial \mu_k} = 0 \\
&\Longleftrightarrow \frac{\partial}{\partial \mu_k}\sum_{i=1}^{N}\sum_{k=1}^{M} A_{ik}[-\frac{P}{2}\log 2\pi - \frac{1}{2}\log\det\Sigma_k - \frac{1}{2}(x_i - \mu_k)^\top \Sigma_k^{-1}(x_i - \mu_k)] = 0 \\
&\Longleftrightarrow \sum_{i=1}^{N} -A_{ik}\Sigma_k^{-1}(x_i - \mu_k) = 0 \\
&\Longleftrightarrow \mu_k = \frac{\sum_{i=1}^{N} A_{ik}x_i}{\sum_{i=1}^{N} A_{ik}}
\end{aligned} \tag{13}$$

– Solve for Σ_k

$$
\begin{aligned}
&\frac{\partial K}{\partial \Sigma_k} = 0 \\
&\Longleftrightarrow \frac{\partial}{\partial \Sigma_k} \sum_{i=1}^{N} \sum_{k=1}^{M} A_{ik} [-\frac{P}{2} \log 2\pi - \frac{1}{2} \log \det \Sigma_k - \frac{1}{2} (x_i - \mu_k)^\top \Sigma_k^{-1} (x_i - \mu_k)] = 0 \\
&\Longleftrightarrow \sum_{i=1}^{N} A_{ik} [-\frac{1}{2} (\Sigma_k^{-1})^\top + \frac{1}{2} (x_i - \mu_k)(x_i - \mu_k)^\top \Sigma_k^{-1} \mathbf{I} \Sigma_k^{-1}] = 0 \\
&\Longleftrightarrow \sum_{i=1}^{N} A_{ik} [(x_i - \mu_k)(x_i - \mu_k)^\top \Sigma_k^{-1}] = \sum_{i=1}^{N} A_{ik} \\
&\Longleftrightarrow \Sigma_k = \frac{\sum_{i=1}^{N} A_{ik} [(x_i - \mu_k)(x_i - \mu_k)^\top]}{\sum_{i=1}^{N} A_{ik}}
\end{aligned}
\tag{14}
$$

Eq. 13, 14, 15 give us the update rule for θ

2 Implementation

- To reproduce the result, see `readme.txt`.
- If there is too many data-points or each datapoint have too high dimension, floating-point underflow will happen.
- Initial location and scale used result from KMean.
- Git `https://github.com/young1906/EM`

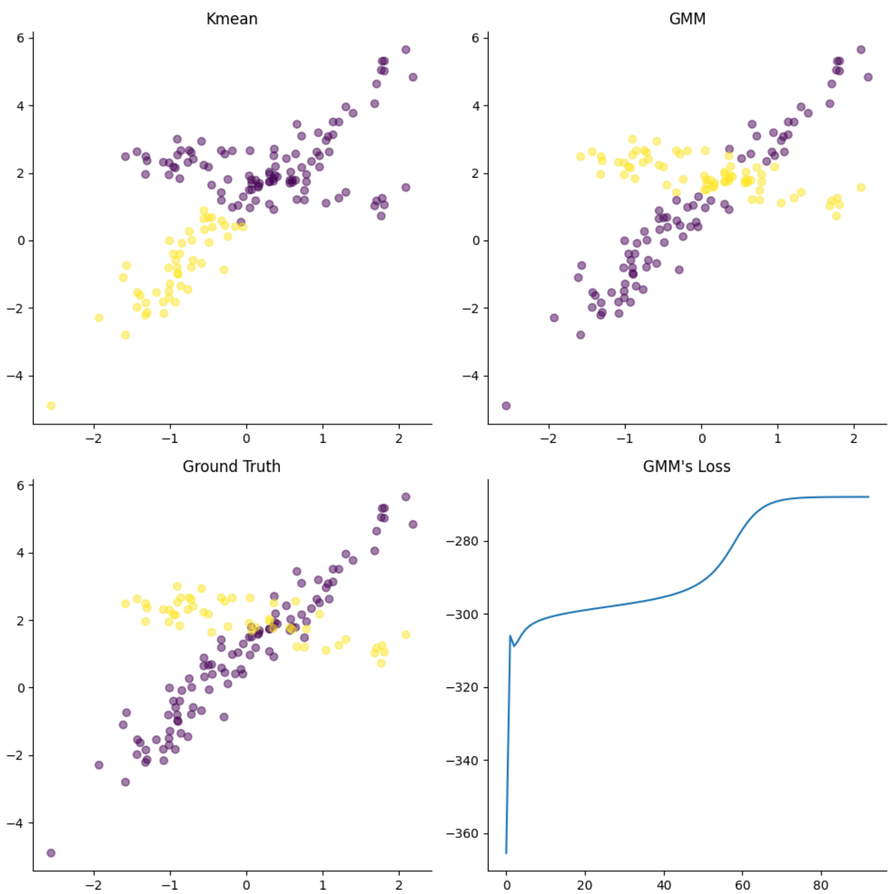

Figure 1: Comparison of KMean and GMM on a synthesis dataset