Question 1: EM for Gaussian mixture model

Derive E-step and M-step formulas in a Gaussian mixture model

Given a set of data $X = \{x_1, \cdots, x_N\}$, there are k components

for each x_n, $p(x_n) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)$ $\quad \sum_{k=1}^{K} \pi_k = 1 \; (0 \le \pi_k \le 1)$

Thus $p(X) = \prod_{n=1}^{N} p(x_n) = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)$

Then the log likelihood that is gonna be maximized is:

$$\ln p(X \mid \pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln\left\{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)\right\}$$

$p(x_n) = \sum_{k=1}^{K} p(k) p(x_n \mid k)$ where $\pi_k = p(k)$ and $p(x_n \mid k) = \mathcal{N}(x_n \mid \mu_k, \Sigma_k)$

so $p(k \mid x_n) = \dfrac{p(k) p(x_n \mid k)}{\sum_{l=1}^{K} p(x_n \mid l) p(l)} = \dfrac{\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l \mathcal{N}(x_n \mid \mu_l, \Sigma_l)}$

set

$\dfrac{\partial \ln p(X \mid \pi, \mu, \Sigma)}{\partial \mu_k} = -\sum_{n=1}^{N} p(k \mid x_n) \Sigma_k (x_n - \mu_k)$ equals 0

so $\mu_k = \dfrac{1}{N_k} \sum_{n=1}^{N} p(k \mid x_n) x_n$ where $N_k = \sum_{n=1}^{N} p(k \mid x_n)$

Then set $\dfrac{\partial \ln p(X \mid \pi, \mu, \Sigma)}{\partial \Sigma_k} = 0$ so $\Sigma_k = \dfrac{1}{N_k} \sum_{n=1}^{N} p(k \mid x_n)(x_n - \mu_k)(x_n - \mu_k)^T$

Finally, maximize $\ln p(X \mid \pi, \mu, \Sigma)$ with respect to π_k with constraint $\sum_{k=1}^{K} \pi_k = 1$, $\dfrac{\partial\{\ln p(X \mid \pi, \mu, \Sigma) + \lambda(\sum_{k=1}^{K} \pi_k - 1)\}}{\partial \pi_k} = 0$

Thus, $\sum_{n=1}^{N} \dfrac{\mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l \mathcal{N}(x_n \mid \mu_l, \Sigma_l)} + \lambda = 0$ because $\sum_{k=1}^{K} \pi_k = 1$

$\sum_{n=1}^{N} \dfrac{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l \mathcal{N}(x_n \mid \mu_l, \Sigma_l)} + \sum_{k=1}^{K} \pi_k \lambda = 0$ so $\lambda = -N$

we can get $\pi_k = \dfrac{N_k}{N}$

Therefore

for μ, Σ and π initialized,

for E step: evaluate $N_k = \sum_{n=1}^{N} p(k|x_n) = \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{l=1}^{k} \pi_l \mathcal{N}(x_n|\mu_l, \Sigma_l)}$

for M step:

$$\begin{cases} \mu_k' = \frac{1}{N_k} \sum_{n=1}^{N} p(k|x_n) x_n \\ \Sigma_k' = \frac{1}{N_k} \sum_{n=1}^{N} p(k|x_n)(x_n - \mu_k')(x_n - \mu_k')^T \\ \pi_k' = \frac{N_k}{N} \end{cases}$$

Question 2: k-means and EM for Gaussian mixture model

for Gaussian mixture model,

$$p(k|x_n) = \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{l=1}^{k} \pi_l \mathcal{N}(x_n|\mu_l, \Sigma_l)} \quad \text{and} \quad \sum_{k=1}^{k} \pi_k = 1$$

for k-means,

distortion measurement $J = \sum_{n=1}^{N} \sum_{k=1}^{k} r_{nk} ||x_n - \mu_k||^2$

needs to minimize.

$$r_{nk} = \begin{cases} 1 & \text{if } x_n \text{ is assigned to cluster } k \\ 0 & \text{otherwise} \end{cases}$$

① fix μ, so $r_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j ||x_n - \mu_j||^2 \\ 0 & \text{otherwise} \end{cases}$

② fix r, $\frac{\partial J}{\partial \mu} = 0$ so $\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}}$

For each mixture component, $\Sigma_k = \varepsilon I$

when $\varepsilon \to 0$, EM = k-means.

in this case,

$$p(k|x_n) = \frac{\pi_k \mathcal{N}(x_n|\mu_k, \varepsilon I)}{\sum_{l=1}^{k} \pi_l \mathcal{N}(x_n|\mu_l, \varepsilon I)} = \begin{cases} 1 & x_n = \mu_k \\ 0 & x_n \neq \mu_k \end{cases}$$

so $\frac{\pi_k N(x_n \mid \mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l N(x_n \mid \mu_l, \Sigma_l)}$ could be represented by r_{nk}

Thus, in M step, $\mu_k' = \frac{1}{N_k} \sum_{n=1}^{N} P(k \mid x_n)\, x_n$

$$= \frac{\sum_{n=1}^{N} r_{nk} x_n}{\sum_{n=1}^{N} r_{nk}}$$ where $r_{nk} = 1$ or 0.