Finite Mixture Model

Consider a dataset $X_1, \ldots, X_n$ consisting of n p-dimensional observations. Let these observations be independent and identically distributed according to the probability distn given by,

$$g(\underline{x}; \psi) = \sum_{k=1}^{K} \tau_k f_k(\underline{x}, \theta_k) \quad ; \quad \sum_{k=1}^{K} \tau_k = 1$$

$g(\cdot; \psi)$ is known as a mixture model.

K is known as mixture order (# of components)

τ_k is called mixing proportion (prior probability that an obs. comes from k^{th} comp.)

$f_k(\underline{x}, \theta_k)$ is the density corresponding to k^{th} component.

It may be noted that the likelihood is given by,

$$L(\psi) = \prod_{i=1}^{n} \sum_{k=1}^{K} \tau_k f_k(\underline{x}_i, \theta_k)$$

$$\Rightarrow \log L(\psi) = \sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} \tau_k f_k(\underline{x}_i, \theta_k)\right)$$

This summation inside log is difficult to deal with. So we employ EM algorithm.

Suppose the membership labels of $\underline{x}_1, \ldots, \underline{x}_n$ are known (which is in fact missing) and they are given by $z_1, z_2, \ldots, z_n$. This gives us the complete data given by: $(\underline{x}_1, z_1), (\underline{x}_2, z_2), \ldots, (\underline{x}_n, z_n)$. Then the complete data likelihood is given by,

$$L_c(\psi) = \prod_{i=1}^{n} \prod_{k=1}^{K} \left(\tau_k \cdot f_k(\underline{x}_i, \theta_k)\right)^{I(z_i = k)}$$

$$\Rightarrow \log L_c(\psi) = \sum_{i=1}^{n} \sum_{k=1}^{K} I(z_i = k) \cdot \log\left(\tau_k \cdot f_k(\underline{x}_i, \theta_k)\right)$$

In the E-step of EM algorithm we take conditional expectation of complete log likelihood given $\underline{x}$. This is known as Q-function and given by,

$$Q(\psi \mid \psi^{(b-1)}, \underline{x}_1, \ldots, \underline{x}_n) = E\left[\log L_c(\psi) \mid \psi^{(b-1)}, \underline{x}_1, \ldots, \underline{x}_n\right]$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} E\left[I(z_i = k \mid \underline{x}_i)\right] \cdot \log\left(\tau_k f_k(\underline{x}_i, \theta_k)\right)$$

[Note: z_i is the only random component here and z_i only depends on $\underline{x}_i$ since $\underline{x}_i$'s are indep.]

$$E[I(z_i = k \mid \underline{x}_i)] = P(z_i = k \mid \underline{x}_i) = P(z_i = k, \underline{x}_i) / P(\underline{x}_i)$$

$$= \frac{P(z_i = k)\, P(\underline{x}_i \mid z_i = k)}{\sum_{k'=1}^{K} P(z_i = k')\, P(\underline{x}_i \mid z_i = k')} = \frac{\tau_k^{(b-1)} \cdot f_k(\underline{x}_i, \theta_k^{(b-1)})}{\sum_{k'=1}^{K} \tau_{k'}^{(b-1)} \cdot f_{k'}(\underline{x}_i, \theta_{k'}^{(b-1)})} = \pi_{ik}^{(b)}$$

$\pi_{ik}^{(b)}$ is known as the posterior probability that x_i belongs to k^{th} component.

In the M-step, we maximize the Q f. given by,

$$Q(\psi|\psi^{(b-1)}, x_1, \dots, x_n) = \sum_{i=1}^{n}\sum_{k=1}^{K} \pi_{ik}^{(b)} \cdot (\log \tau_k + \log f_k(x_i, \vartheta_k))$$

Let us first take derivative w.r.t. τ_k. The restriction $\sum_{k=1}^{K} \tau_k = 1$ would be taken care of by a Lagrange multiplier.

$$\frac{\partial Q}{\partial \tau_k} = \frac{\partial}{\partial \tau_k}\left[\sum_{i=1}^{n}\sum_{k=1}^{K} \pi_{ik}^{(b)} \cdot \log \tau_k - \lambda\left(\sum_{k=1}^{K} \tau_k - 1\right)\right] = \sum_{i=1}^{n} \pi_{ik}^{(b)} \cdot \frac{1}{\tau_k} - \lambda$$

Then, $\partial Q/\partial \tau_k = 0 \Rightarrow \tau_k = \frac{1}{\lambda}\sum_{i=1}^{n} \pi_{ik}^{(b)}$

But we have, $\sum_{k=1}^{K} \tau_k = 1$. Hence, $\sum_{k=1}^{K}\frac{1}{\lambda}\sum_{i=1}^{n}\pi_{ik}^{(b)} = 1 \Rightarrow \lambda = \sum_{i=1}^{n}\sum_{k=1}^{K}\pi_{ik}^{(b)} = \sum_{i=1}^{n} 1 = n$

Thus, $\tau_k^{(b)} = \frac{1}{n}\sum_{i=1}^{n} \pi_{ik}^{(b)}$

Estimation of ϑ_k : This will be demonstrated on (1) univariate Gaussian component, (2) multivariate Gaussian component.

(1) <u>Univariate Gaussian Mixture</u>

$$Q(\psi|\psi^{(b-1)}, x_1, \dots, x_n) = \sum_{i=1}^{n}\sum_{k=1}^{K} \pi_{ik}^{(b)} \cdot [\log \tau_k + \log \phi(x_i; \mu_k, \sigma_k^2)]$$
$$= \sum_{i=1}^{n}\sum_{k=1}^{K} \pi_{ik}^{(b)} [\log \tau_k - \log(\sqrt{2\pi\sigma^2}) + \tfrac{1}{2} \cancel{\exp} (- \frac{(x_i-\mu_k)^2}{2\sigma_k^2})]$$

$$\frac{\partial Q}{\partial \mu_k} = \frac{\partial}{\partial \mu_k}\left[\sum_{i=1}^{n}\sum_{k=1}^{K} \pi_{ik}^{(b)} \cdot \left(-\frac{1}{2\sigma_k^2}(x_i-\mu_k)^2\right)\right] = \frac{1}{\sigma_k^2}\sum_{i=1}^{n} \pi_{ik}^{(b)} \cdot (x_i - \mu_k)$$

Then $\partial Q/\partial \mu_k = 0 \Rightarrow \mu_k^{(b)} = \sum_{i=1}^{n} \pi_{ik}^{(b)} \cdot x_i \Big/ \sum_{i=1}^{n} \pi_{ik}^{(b)}$

$$\frac{\partial Q}{\partial \sigma_k^2} = \frac{\partial}{\partial \sigma_k^2}\left[\sum_{i=1}^{n}\sum_{k=1}^{K} \pi_{ik}^{(b)}\left(-\frac{1}{2}\log \sigma_k^2 - \frac{1}{2\sigma_k^2}(x_i-\mu_k)^2\right)\right]$$
$$= -\frac{1}{2}\sum_{i=1}^{n} \pi_{ik}^{(b)}\left(\frac{1}{\sigma_k^2} - \frac{1}{\sigma_k^4}(x_i-\mu_k)^2\right)$$

Then $\partial Q/\partial \sigma_k^2 = 0 \Rightarrow \sigma_k^{2\,(b)} = \sum_{i=1}^{n} \pi_{ik}^{(b)} (x_i - \mu_k^{(b)})^2 \Big/ \sum_{i=1}^{n} \pi_{ik}^{(b)}$

② Multivariate Gaussian Mixture

$$Q(\psi | \psi^{(b-1)}, \underline{x}_1, \ldots, \underline{x}_n) = \sum_{i=1}^{n} \sum_{k=1}^{K} \pi_{ik}^{(b)} [\log \tau_k + \log \Phi(\underline{x}_i, \underline{\mu}_k, \Sigma_k)]$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} \pi_{ik}^{(b)} [\log \tau_k - \tfrac{p}{2} \log 2\pi - \tfrac{1}{2} \log |\Sigma_k| - \tfrac{1}{2} (\underline{x}_i - \underline{\mu}_k)^T \Sigma_k^{-1} (\underline{x}_i - \underline{\mu}_k)]$$

$$\frac{\partial Q}{\partial \underline{\mu}_k} = \frac{\partial}{\partial \underline{\mu}_k} [\sum_{i=1}^{n} \pi_{ik}^{(b)} (-\tfrac{1}{2} (\underline{x}_i - \underline{\mu}_k)^T \Sigma_k^{-1} (\underline{x}_i - \underline{\mu}_k))]$$

$$= \partial/\partial \underline{\mu}_k [\sum_{i=1}^{n} \pi_{ik}^{(b)} \cdot (-\tfrac{1}{2}) \cdot [\underline{x}_i^T \Sigma_k^{-1} \underline{x}_i - \underline{\mu}_k^T \Sigma_k^{-1} \underline{x}_i - \underline{x}_i^T \Sigma_k^{-1} \underline{\mu}_k + \underline{\mu}_k^T \Sigma_k^{-1} \underline{\mu}_k]]$$

$$= -\tfrac{1}{2} \sum_{i=1}^{n} \pi_{ik}^{(b)} \cdot \frac{\partial}{\partial \underline{\mu}_k} [\underline{\mu}_k^T \Sigma_k^{-1} \underline{\mu}_k - 2 \underline{\mu}_k^T \Sigma_k^{-1} \underline{x}_i]$$ (note: $\underline{\mu}_k^T \Sigma_k^{-1} \underline{x}_i = \underline{x}_i^T \Sigma_k^{-1} \underline{\mu}_k$ ∵ both are scalar)

$$= -\tfrac{1}{2} \sum_{i=1}^{n} \pi_{ik}^{(b)} \cdot (2 \Sigma_k^{-1} \underline{\mu}_k - 2 \Sigma_k^{-1} \underline{x}_i) = \sum_{i=1}^{n} \pi_{ik}^{(b)} (\Sigma_k^{-1} (\underline{x}_i - \underline{\mu}_k))$$

Then $\frac{\partial Q}{\partial \underline{\mu}_k} = 0 \Rightarrow \Sigma_k^{-1} \underline{\mu}_k \sum_{i=1}^{n} \pi_{ik}^{(b)} = \Sigma_k^{-1} \sum_{i=1}^{n} \pi_{ik}^{(b)} \underline{x}_i$ (multiply both sides by Σ_k)

$$\Rightarrow \underline{\mu}_k^{(b)} = \sum_{i=1}^{n} \pi_{ik}^{(b)} \underline{x}_i \Big/ \sum_{i=1}^{n} \pi_{ik}^{(b)}$$

$$\frac{\partial Q}{\partial \Sigma_k} = \frac{\partial}{\partial \Sigma_k} [\sum_{i=1}^{n} \pi_{ik}^{(b)} (-\tfrac{1}{2} \log |\Sigma_k| - \tfrac{1}{2} (\underline{x}_i - \underline{\mu}_k)^T \Sigma_k^{-1} (\underline{x}_i - \underline{\mu}_k))]$$

$$= -\tfrac{1}{2} \sum_{i=1}^{n} \pi_{ik}^{(b)} \cdot \frac{\partial}{\partial \Sigma_k} [\log |\Sigma_k| + (\underline{x}_i - \underline{\mu}_k)^T \cdot \Sigma_k^{-1} (\underline{x}_i - \underline{\mu}_k)]$$

$$= -\tfrac{1}{2} \sum_{i=1}^{n} \pi_{ik}^{(b)} [\Sigma_k^{-1} - \Sigma_k^{-1} (\underline{x}_i - \underline{\mu}_k)(\underline{x}_i - \underline{\mu}_k)^T \Sigma_k^{-1}]$$

$$= -\tfrac{1}{2} \Sigma_k^{-1} \cdot \sum_{i=1}^{n} \pi_{ik}^{(b)} [I - (\underline{x}_i - \underline{\mu}_k)(\underline{x}_i - \underline{\mu}_k)^T \cdot \Sigma_k^{-1}]$$

Then $\frac{\partial Q}{\partial \Sigma_k} = 0 \Rightarrow \sum_{i=1}^{n} \pi_{ik}^{(b)} [I - (\underline{x}_i - \underline{\mu}_k)(\underline{x}_i - \underline{\mu}_k)^T \Sigma_k^{-1}] \cdot \Sigma_k = 0$

$$\Rightarrow \Sigma_k^{(b)} = \sum_{i=1}^{n} \pi_{ik}^{(b)} \cdot (\underline{x}_i - \underline{\mu}_k^{(b)})(\underline{x}_i - \underline{\mu}_k^{(b)})^T \Big/ \sum_{i=1}^{n} \pi_{ik}^{(b)}$$