Lecture 11: Mixture Model and EM

Course Teacher: Md. Shariful Islam Bhuyan

Gaussian Mixture Model

- $\mathbf{x}_i$ can be generated from any one of the K Gaussian
- Probability of selection of k^{th} Gaussian is w_k where $\sum_{k=1}^{K} w_k = 1$
- The probability of generation of $\mathbf{x}_i$ from k^{th} Gaussian is given as

$$P(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_k|}} e^{\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k)\right)}$$

- Log likelihood intractable, cannot be solved analytically

$$\ln p(\mathbf{d}|\boldsymbol{\theta}) = \ln p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w}) = \sum_{i=1}^{N} \ln p(\mathbf{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w}) = \sum_{i=1}^{N} \ln\left(\sum_{k=1}^{K} w_k N_k(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right)$$

Mixture Membership

$$P(\mathbf{x}_i) = \sum_{k=1}^{K} P(z_i = k) P(\mathbf{x}_i | z_i = k) = \sum_{k=1}^{K} w_k \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_k|}} e^{\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k)\right)}$$

$$p_{ik} = p(z_i = k | \mathbf{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w})$$

$$\begin{bmatrix} p_{11} & \cdots & p_{N1} \\ \vdots & \ddots & \vdots \\ p_{1K} & \cdots & p_{NK} \end{bmatrix} \qquad \sum_{k=1}^{K} p_{ik} = 1 \qquad \sum_{i=1}^{N} p_{ik} = N_k$$

1. **Initialize** the means $\boldsymbol{\mu}_k$, covariances $\boldsymbol{\Sigma}_k$ and mixing coefficients w_k, and evaluate the initial value of the log likelihood.

2. **E step:** Evaluate the conditional distribution of latent factors using the current parameter values

$$p_{ik} = p(z_i = k|\mathbf{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w}) = \frac{p(\mathbf{x}_i|z_i = k, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w})P(z_i = k|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w})}{p(\mathbf{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w})} = \frac{w_k N_k(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{k=1}^{K} w_k N_k(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}$$

3. **M step:** Re-estimate the parameters using the conditional distribution of latent factors

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^{N} p_{ik}\mathbf{x}_i}{\sum_{i=1}^{N} p_{ik}} \quad \boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^{N} p_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T}{\sum_{i=1}^{N} p_{ik}} \quad w_k = \frac{\sum_{i=1}^{N} p_{ik}}{N}$$

4. **Evaluate** the log likelihood and check for convergence of either the parameters or the log likelihood. If the convergence criterion is not satisfied return to step 2.

Clustering with EM