EM with GMM

Md. Hasanul Islam
ID:1205096

1 Reason behind using GMM

Sometimes Data might follow a mixture model that looks like multimodal, i.e. there is more than one "peak" in the distribution of data. Trying to fit a multimodal distribution with a unimodal (one "peak") model will generally give a poor fit, as shown in the figure 1 . Since many simple distribution is multimodal, an obvious way to model a multimodal distribution would be to assume that it is generated by multiple unimodal distributions. The most commonly used distribution in modeling real world unimodal data is the Gaussian distribution. Thus, modeling multimodal data as a mixture of many unimodal Gaussian distibutions makes intuitive sense.

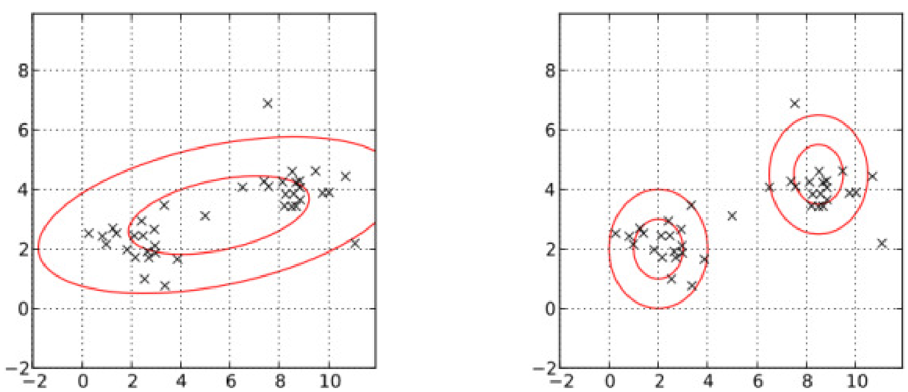

Figure 1: (Left)Fit with unimodal Gaussian. (Right) Fit with Gaussian mixture model with two mode.

2 Model of this data for GMM

Model of the given daa for GMM is given in figure 2

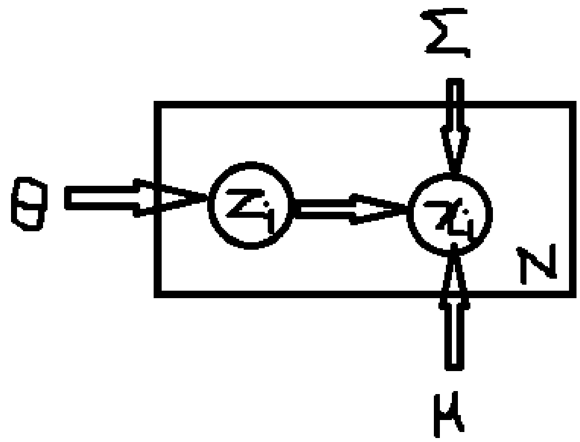

Figure 2: Model for this EM algorithm

3 Update Equations in M step

Complete Log likelihood:

$$
\begin{aligned}
L_c &= \log p(X, Z|\theta) \\
&= \log \prod_{i}^{N} p(x_i; z_i|\theta) \\
&= \log \prod_{i}^{N} \prod_{k}^{K} [p(x_i|z_i = k, \theta) p(z_i = k|\theta)]^{\pi_{ik}} \\
&= \sum_{i}^{N} \sum_{k}^{K} \pi_{ik} \log p(x_i|z_i = k, \theta) + \pi_{ik} \log \theta_k
\end{aligned}
$$

Update Eqauation for Mixing Co-efficient:

$$
\begin{aligned}
\sum_{k=1}^{K} \theta_k = 1 \\
L_1 = \sum_{i}^{N} \sum_{k}^{K} \pi_{ik} \log \theta_k + \lambda(1 - \sum_{k=1}^{K}) \\
\frac{\partial L_1}{\partial \theta_k} = \frac{\sum_{i}^{N} \pi_{ik}}{\theta_k} - \lambda = 0 \\
\sum_{i}^{N} \pi_{ik} - \lambda \theta_k = 0 \\
\sum_{i}^{N} \sum_{k=1}^{K} \pi_{ik} - \lambda \theta_k \sum_{k=1}^{K} = 0 \\
\sum_{i}^{N} \sum_{k=1}^{K} \pi_{ik} = N \\
\theta_k = \frac{\sum_{i=1}^{N} \pi_{ik}}{N}
\end{aligned}
\tag{1}
$$

Update Equation for Mean:

$$
\begin{aligned}
\frac{\partial L}{\partial \mu_k} = 0 \\
\frac{\partial}{\partial \mu_k} \sum_{i=1}^{N} \pi_{ik}(-\frac{1}{2})(x_i - \mu_k)^T \sigma_k^{-1}(x_i - \mu_k) = 0 \\
\sum_{i}^{N} \pi_{ik}(x_i - \mu_k)^T \sigma_k^{-1} = 0 \\
\mu_k = \frac{\sum_{i}^{N} \pi_{ik} x_i}{\sum_{i}^{N} \pi_{ik}}
\end{aligned}
$$

Update Equation for Variance:

$$\frac{\partial L}{\partial \sigma_k^{-1}} = 0$$

$$\frac{\partial}{\partial \sigma_k^{-1}} \sum_{i=1}^{N} \pi_{ik} \left\{ \frac{1}{2} \log|\sigma_k^{-1}| - \frac{1}{2}(x_i - \mu_k)^T \sigma_k^{-1} (x_i - \mu_k) \right\} = 0$$

$$\sum_{i}^{N} \pi_{ik} \left\{ \frac{1}{2}\sigma_k - \frac{1}{2}(x_i - \mu_k)^T (x_i - \mu_k) \right\} = 0$$

$$\sigma_k = \frac{\sum_i^N \pi_{ik}(x_i - \mu_k)^T (x_i - \mu_k)}{\sum_i^N \pi_{ik}}$$