Mixture Models and EM Algorithm

S. Sumitra
Department of Mathematics
Indian Institute of Space Science and Technology

Clustering problems could be solved by applying model-based approach, which consists in using certain models for clusters and attempting to optimize the fit between the data and the model. Each cluster (component) can be mathematically represented by a parametric distribution, for eg, Gaussian (continuous) or a poisson (discrete). The entire data set is therefore modelled by a mixture of these distributions. An individual distribution used to model a specific cluster is often referred to as a component distribution.

Let there be k clusters. Let the random variable C denote the component with values $1, ..k$. Here we are considering Gaussian mixture models. So $x_j/(C = i) \sim N(\mu_i, \sum_i)$ where μ_i and Σ_i are the mean and covariance matrix of the i^{th} class.

A data point is generated by first choosing a component and then generating a sample from that component. By total probability theorem,

$$p(x) = \sum_{i=1}^{k} p(C = i)p(x/C = i) \tag{1}$$

[$p(C = i)$ is analogous to $p(y = i)$ in Gaussian discriminant analysis.]

To determine in which cluster each x_j belongs, $p(C = i/x_j)$ has to be found. Now

$$p(C = i/x_j) = p_{ij} = \frac{p(C = i)p(x_j/C = i)}{p(x_j)}, i = 1, 2, \ldots k, j = 1, 2, \ldots N \tag{2}$$

Hence $\sum_{i=1}^{k} p_{ij} = 1$. Let $w_i = p(C = i), i = 1, 2. \ldots k$. Therefore the unknown parameters of a mixture of Gaussians are w_i, μ_i and Σ_i.

The EM algorithm can be applied to determine the unknown parameters. The EM algorithm has two main steps: E step & M step. In the E-step, it assumes the values of the model (that is w_i, μ_i and Σ_i) and find $P(C = i/x_j), i = 1, 2, \ldots k$, $j = 1, 2, \ldots N$. In the M-step, it updates the parameters of the model. The process iterates until convergence.

E step

In the E step, compute the probabilities $p_{ij}, i = 1, 2, \dots k,\ j = 1, 2, \dots N$.

M step

Compute the new mean, covariance and component weights as follows:

$$\mu_i = \frac{\sum_{j=1}^{N} p_{ij} x_j}{\sum_{j=1}^{N} p_{ij}}$$

[For sure event, $\mu_i = \dfrac{\sum_j 1\{x_j \in C = i\} x_j}{\sum_j 1.\{x_j \in C = i\}}$. Here, we don't know whether x_j is in component i. We only know $p(C = i/x_j)$.]

$$\Sigma_i = \frac{\sum_j p_{ij}(x_j - \mu_i)(x_j - \mu_i)^T}{\sum_{j=1}^{N} p_{ij}}$$

$$w_i = \frac{\sum_{j=1}^{N} p_{ij}}{N}$$

[*Compare these formulas with those of Gaussian discriminant analysis*]

The algorithm can be summarized as follows:

Algorithm 1 EM algorithm

Initialize μ_i, Σ_i, w_i, $i = 1, 2, \ldots k$

Iterate until covergence:

E Step

for $i = 1$ to k **do**

for $j = 1$ to N **do**

calculate $p(x_j/C = i) = \frac{1}{(2\pi)^{n/2}|\Sigma_i|^{1/2}} \exp -\frac{1}{2}(x_j - \mu_i)^T \Sigma_i^{-1}(x_j - \mu_i)$

calculate $p_{ij} = \frac{p(x_j/C = i)w_i}{\sum_{i=1}^{k} p(x_j/C = i)w_i}$

end for

$p_i = \sum_{j=1}^{N} p_{ij}$

end for

M Step

for $i = 1$ to k **do**

calculate $\mu_i = \frac{\sum_{j=1}^{N} p_{ij} x_j}{p_i}$

calculate $\Sigma_i = \frac{\sum_{j=1}^{N} p_{ij}(x_j - \mu_i)(x_j - \mu_i)^T}{p_i}$

set $w_i = \frac{p_i}{N}$

end for

end

References

(1) Artificial Intelligence by Stuart Russel and Peter Norwig
(2) Andrew Ng's Lecture Note