EM for GMMs

Wednesday, 22 March 2023 9:51 PM

The log likelihood for GMM given data X is given by

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

Derivating w.r.t μ_k and equating to 0 we get

$$0 = -\sum_{n=1}^{N} \underbrace{\frac{w_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_j w_j \mathcal{N}(x_n|\mu_j, \Sigma_j)}}_{\gamma(z_{nk})} \Sigma_k (x_n - \mu_k)$$

Multiplying by Σ_k^{-1}

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n$$

where $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$

$N_k \approx$ effective #. of points assigned to cluster k.

Mean μ_k of k^{th} Gaussian is obtained by taking weighted mean of all of points in the dataset.

Similarly

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T$$

Finally we need to maximize $\ln p(X|w, \mu, \Sigma)$ w.r.t to w_k

We consider

$$\ln p(X|w, \mu, \Sigma) + \lambda \left(\sum_{k=1}^{K} w_k - 1 \right)$$

which gives

$$0 = \sum_{n=1}^{N} \frac{\mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_j w_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} + \lambda$$

Multiply both sides by w_k and sum over k, we get $\lambda = -N$

Finally $w_k = \frac{N_k}{N}$

EM for guassian mixtures

1. Initialize μ_k, Σ_k and w_k and evaluate initial value of log likelihood

2. E-step

$$\gamma(z_{nk}) = \frac{w_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} w_j \mathcal{N}(x_n|\mu_j, \Sigma_j)}$$

3. M-step

$$\mu_k^{new} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{nk}) x_n$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

$$w_k^{new} = \frac{N_k}{N}$$

where $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$

4. Evaluate $\sum_{i=1}^{N} \ln \left\{ \sum_{k=1}^{K} w_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}$

Repeat 2-4 till convergence

EM – Alternate View

Goal of EM algorithm is to find M L solutions for models having latent variables.

Let Data – X

Latent Variable – Z

Parameter – θ

$$\ln p(X|\theta) = \ln \left\{ \sum_{Z} p(X, Z|\theta) \right\}$$

X – incomplete data.

(X, Z) – complete data

In practice we have only X & knowledge of Z in terms of posterior $p(Z|X, \theta)$

In E step :– Expected value of log likelihood under posterior of latent variable

M step :– Maximize the Expectation

General EM Algorithm

Given : $p(X, Z|\theta)$ over observed X and latent Z governed by θ

Goal : Maximize $p(X|\theta)$ w.r.t θ

Steps :

1. Initialize θ^{old}

2. E step:
Find value of $p(Z|X, \theta^{old})$

3. M step

$$\theta^{new} = \underset{\theta}{\operatorname{argmax}} \, Q(\theta, \theta^{old})$$

Here

$$Q(\theta, \theta^{old}) = \sum_{Z} p(Z|X, \theta^{old}) \ln p(X, Z|\theta)$$

$$\theta^{old} \leftarrow \theta^{new}$$

Repeat 2-3 till convergence.