Goal : (1)

To derive the equations for learning parameters of a Gaussian mixture model

Gaussian Mixture Model (GMM)

For Given data x, GMM represents the probability of x occurring as the sum of several Gaussian probability density functions.

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \tag{1}$$

mixing coefficient, π_k represents the probability that the k-th Gaussian distribution will be selected. so π_k must satisfy follows:

$$0 \leq \pi_k \leq 1 \tag{2}$$

$$\sum_{k=1}^{K} \pi_k = 1 \tag{3}$$

1. Classification using GMM

Learning GMM means estimating appropriate π_k, μ_k, Σ_k for given $X = \{x_1, x_2, .., x_N\}$

Classification using GMM is to find out in which Gaussian distrubution a given data x_N was generated.

For this, responsibility is defined as follows.

$$\gamma(z_{nk}) = p(z_{nk} = 1)|x_n) \tag{4}$$

$z_{nk} \in \{0, 1\}$ is a binary variable with a value of 1 if the k-th Gaussian distribution of GMM is selected given xn or a value of 0 if not. That is, when z_{nk} is 1, it means that x_n is generated in the k-th Gaussian distribution.

Classification using GMM is to select the Gaussian distribution with the highest value by calculating the k number of $\gamma(z_{nk})$ given xn.

If the values of all parameters π, μ, Σ of GMM are determined through learning, $\gamma(z_{nk})$ can be calculated as follows using Bayes'theorem.

$$\gamma(z_{nk}) = p(z_{nk}=1|x_n) = \frac{p(z_{nk}=1)p(x_n|z_{nk}=1)}{\sum_{j=1}^{K} p(z_{nj}=1)p(x_n|z_{nj}=1)} = \frac{\pi_k N(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j,\Sigma_j)} \quad (5)$$

2. Learning GMM using Expectation-Maximization algorithm (EM algorithm)

For estimating parameters, log-likelihod $\mathcal{L}(X;\theta)$ is defined as

$$\mathcal{L}(X;\theta) = \ln p(X|\pi,\mu,\Sigma) = \ln\left\{\Pi_{n=1}^{N} p(x_n|\pi,\mu,\Sigma)\right\} = \sum_{n=1}^{N}\ln\left\{\sum_{k=1}^{K}\pi_k N(x_n|\mu_k,\Sigma_k)\right\} \quad (6)$$

Estimating π, μ, Σ when log-like hood is at its maximum has the same meaning as constructing GMM that best represents given data X.

For do this,
it partial differentiate $\mathcal{L}(X;\theta)$ for π_k, μ_k, Σ_k.

1) μ_k

$$\frac{\partial \mathcal{L}(X;\theta)}{\partial \mu_k} = \sum_{n=1}^{N}\frac{\pi_k N(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j,\Sigma_j)}\Sigma_k^{-1}(x_n-\mu_k) = 0$$

$$\Leftrightarrow \sum_{n=1}^{N}\gamma(z_{nk})(x_n-\mu_k) = 0$$

$$\therefore \mu_k = \frac{\sum_{n=1}^{N}\gamma(z_{nk})x_n}{\sum_{n=1}^{N}\gamma(z_{nk})} \quad (7)$$

.
.
.
.
.
.
.
.
.

2) Σ_k

$$\frac{\partial \mathcal{L}(X;\theta)}{\partial \Sigma_k} = \sum_{n=1}^{N}\frac{\pi_k N(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j,\Sigma_j)}\left\{\frac{1}{2}\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-1} - \frac{1}{2}\Sigma_k^{-1}\right\} = 0$$

$$\Leftrightarrow \sum_{n=1}^{N}\gamma(z_{nk})\left\{\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T - 1\right\} = 0$$

$$\therefore \Sigma_k = \frac{\sum_{n=1}^{N}\gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^{N}\gamma(z_{nk})} \quad (8)$$

.
.
.
.
.
.
.
.

3) π_k

π_k, the last parameter of GMM, must maximize log-likelihood while satisfying the condition of equ.3. so π_k is estimated by using Lagrange multiplier method, Lagrangian $J(X;\theta,\lambda)$ is as follows:

$$J(X;\theta,\lambda) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k N(x_n|\mu_k, \Sigma_k) + \lambda\left(1 - \sum_{k=1}^{K}\pi_k\right) \quad (9)$$

λ is found by partial differentiation of Lagrangean.

$$\frac{\partial J(X;\theta,\lambda)}{\partial \pi_k} = \sum_{n=1}^{N} \frac{N(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j,\Sigma_j)} - \lambda = 0$$

$$\Leftrightarrow \sum_{k=1}^{K}\sum_{n=1}^{N} \frac{\pi_k N(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j,\Sigma_j)} - \lambda\sum_{k=1}^{K}\pi_k = 0$$

$$\Leftrightarrow \sum_{k=1}^{K}\sum_{n=1}^{N} \gamma(z_{nk}) - \lambda = 0 \quad \left(\because \sum_{k=1}^{K}\pi_k = 1\right)$$

$$\therefore \lambda = N \quad \left(\because \sum_{k=1}^{K}\gamma(z_{nk}) = 1\right) \quad (10)$$

π_k can be estimated using the calculated λ.

$$\frac{\partial J(X;\theta,\lambda)}{\partial \pi_k} = \sum_{n=1}^{N} \frac{N(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j,\Sigma_j)} - N = 0$$

$$\Leftrightarrow \sum_{n=1}^{N} \frac{\pi_k N(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j,\Sigma_j)} - N\pi_k = 0$$

$$\therefore \pi_k = \frac{1}{N}\sum_{n=1}^{N}\gamma(z_{nk}) \quad (11)$$

For estimating parameters of GMM,

In **E-step** of EM algorithm,
$\gamma(z_{nk})$ is calculated by all datas and Gaussian distribution.

And then, In **M-step**of EM algorithm,
π, μ, Σ for all Gaussian distribution are estimated by using equ.7,8,11

These E-step and M-step are repeated until converging or a certain number of times.