Machine Learning Assignment3 - EM for GMM

2021021582 인공지능학과 양건준

1 Derive the equations for learning parameters of a Gaussian mixture model

Responsibility can be written as follows:

$$\gamma(z_{nk}) \equiv p(z_k = 1|x_n) = \frac{p(x_n|z_k = 1)p(z_k = 1)}{\sum_{j=1}^{K} p(x_n|z_i = 1)p(z_i = 1)}$$

In the context of the Gaussian mixture model, the log of the likelihood function is as follows:

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

As we want to maximize the log likelihood, take derivatives with respect to μ_k, Σ_k, π_k and set them to zero.

1.1 Take derivatives with respect to μ_k and setting to zero

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

$$= \sum_{n=1}^{N} \gamma(z_{nk}) \Sigma_k(x_n - \mu_k) \qquad \left(\frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} = \gamma(z_{nk})\right)$$

Multiply both side by Σ_k^{-1}

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

1.2 Take derivatives with respect to Σ_k and setting to zero

$$0 = \sum_{n=1}^{N} \frac{\pi_k N(x|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x|\mu_j, \Sigma_j)} \{\frac{1}{2} \sum_{k}^{-1} (x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{-1} - \frac{1}{2} \Sigma_k^{-1}\}$$

$$\leftrightarrow \sum_{n=1}^{N} \gamma(z_n k) \{\sum_{k}^{-1} (x_n - \mu_k)(x_n - \mu_k)^T - 1\} = 0$$

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

1.3 Take derivatives with respect to π_k and setting to zero

π_k can be obtained by Lagrangian multiplier because it is a constrained optimization

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

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

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

$$\leftrightarrow \sum_{k=1}^{K} \sum_{n=1}^{N} \gamma(z_{nk}) - \lambda = 0$$

$$\lambda = N$$

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

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

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