MA-INF 2201 Computer Vision Assignment-6

Pratika Kochar(3201097)
Tridivraj Bhattacharyya(3035538)

November 2018

1

1.1

We know, from the equation for multivariate Gaussian distribution,

$$\begin{aligned} r_{ik} &= \frac{\lambda_k Norm_{x_i}[\mu_k, \Sigma_k]}{\sum_{j=1}^{K} \lambda_j Norm_{x_i}[\mu_j, \Sigma_j]} \\ &= \frac{\lambda_k \frac{1}{\sqrt{2\pi}^d |I|} \exp(-\frac{1}{2}(x_i - \mu_k)^T I^{-1}(x_i - \mu_k))}{\sum_{j=1}^{K} \lambda_j \frac{1}{\sqrt{2\pi}^d |I|} \exp(-\frac{1}{2}(x_i - \mu_j)^T I^{-1}(x_i - \mu_j))} \quad [\text{d = dimension of } x_i \text{ vector}] \end{aligned}$$

1.2

We know λ_k has a constant probability value as it has an uniform distribution. Therefore,

$$\lambda_k = \frac{1}{b-a} \text{ where } a \leq \lambda_k \leq b$$

Therefore, removing the Identity matrix and bringing the constant terms outside with λ_k,

$$\begin{aligned} r_{ik} &= \frac{\lambda_k \frac{1}{\sqrt{2\pi}^d} \exp(-\frac{1}{2}(x_i - \mu_k)^2)}{\frac{1}{\sqrt{2\pi}^d} \sum_{j=1}^{K} \lambda_j \exp(-\frac{1}{2}(x_i - \mu_j)^2)} \\ &= \frac{C_i \exp(-\frac{1}{2}(x_i - \mu_k)^2)}{\sum_{j=1}^{K} \lambda_j \exp(-\frac{1}{2}(x_i - \mu_j)^2)} \text{ where } C_i = \lambda_k \end{aligned} \tag{1}$$

Applying logarithm to both sides,

$$\begin{aligned} \log(r_{ik}) &= \log(\frac{C_i \exp(-\frac{1}{2}(x_i - \mu_k)^2)}{\sum_{j=1}^{K} \lambda_j \exp(-\frac{1}{2}(x_i - \mu_j)^2)}) \\ &= -\frac{1}{2}(x_i - \mu_k)^2) - \log(\sum_{j=1}^{K} \exp(-\frac{1}{2}(x_i - \mu_j)^2)) \text{ [By cancelling the constants } C_i \text{ and } \sum_{j=1}^{K} \lambda_j = 1] \end{aligned} \tag{2}$$

1.3

The mean in the M-step of EM algorithm is,

$$\mu_k^{[t+1]} = \frac{\sum_{i=1}^{I} r_{ik}}{\sum_{j=1}^{K} \sum_{i=1}^{I} r_{ij}}$$

By applying the following, we assign each data point to its best fit Gaussian.

$$r_{ik} = \begin{cases} 1, & \text{if } k = \arg\max_k' \{r_{ik'}\} \\ 0, & \text{otherwise} \end{cases}$$

The modified mean updates using this become,

$$\mu_k^{[t+1]} = \begin{cases} \frac{\sum_{i=1}^{I} 1.x_i}{1}, & \text{if } k = \arg\max_k' \{r_{ik'}\} \\ 0, & \text{otherwise} \end{cases}$$

The covariance matrix being Identity, the shape of each cluster would resemble a circle. As per equations 1 and 2, the EM algorithm will then always assign each data point to a cluster greedily (in this case the gaussian), which might also be the local minima. This is similar to the K-means algorithm which assigns each point to the cluster with closest center.