Paolo Furlanetto Ferrari

NetID: paolof2

Derivation of update rules

i. Probability update

The log-likelihood is

$$L = \sum_{i=1}^{n} \sum_{k=1}^{G} w_{ik} \log[p_k N(x_i, \mu_k, \Sigma)]$$

However, since the probability vector is constrained to have unit sum, the function that we need to optimize is

$$L' = \sum_{i=1}^{n} \sum_{k=1}^{G} w_{ik} \log[p_k N(x_i, \mu_k, \Sigma)] + \lambda(1 - \sum_{k=1}^{G} p_k)$$

Setting the partial derivative of L' with respect to p_k to zero, we have

$$0 = \sum_{i=1}^{n} \frac{w_{ik}}{p_k} + \lambda$$

Thus the update rule is

$$p_k = \sum_{i=1}^{n} \frac{w_{ik}}{\lambda}$$

Where λ is obtained by summing all equations above:

$$1 = \sum_{k=1}^{G} \sum_{i=1}^{n} \frac{w_{ik}}{\lambda}$$

$$\lambda = \sum_{k=1}^{G} \sum_{i=1}^{n} w_{ik} = \sum_{i=1}^{n} \sum_{k=1}^{G} w_{ik} = \sum_{i=1}^{n} 1 = n$$

ii. Mean update

Setting the partial derivative of L or L' with respect to μ_k to zero, we have:

$$0 = \sum_{i=1}^{n} \frac{w_{ik} \frac{\partial N(x_i, \mu_k, \Sigma)}{\partial \mu_k}}{p_k N(x_i, \mu_k, \Sigma)}$$

Since

$$\frac{\partial N(x_i, \mu_k, \Sigma)}{\partial \mu_k} = N(x_i, \mu_k, \Sigma) \frac{-1}{2} \frac{\partial((x_i - \mu_k)\Sigma^{-1}(x_i - \mu_k)^T))}{\partial \mu_k} = N(x_i, \mu_k, \Sigma).(x_i - \mu_k)$$

We have, after simplifying the multiplicative constant terms,

$$0 = \sum_{i=1}^{n} w_{ik}(x_i - \mu_k)$$

Thus

$$\mu_k = \sum_{i=1}^{n} w_{ik} x_i \Big/ \sum_{i=1}^{n} w_{ik}$$

iii. Covariance matrix update

Setting the partial derivative of L with respect to Σ to zero, we have:

$$0 = \sum_{i=1}^{n} \sum_{k=1}^{G} \frac{\partial \log[N(x_i, \mu_k, \Sigma)]}{\partial \Sigma} w_{ik}$$

Following the derivation of the derivative inside the sum from https://math.stackexchange.com/questions/1599966/derivative-of-multivariate-normal-distribution-wrt-mean-and-covariance, we have

$$\frac{\partial \log[p_k N(x_i, \mu_k, \Sigma)]}{\partial \Sigma} = \frac{1}{2}[\Sigma^{-T}(\mu_k - x_i)(\mu_k - x_i)^T \Sigma^T - \Sigma^{-T}]$$

Thus

$$0 = \sum_{i=1}^{n} \sum_{k=1}^{G} \frac{1}{2}[\Sigma^{-T}(\mu_k - x_i)(\mu_k - x_i)^T \Sigma^{-T} - \Sigma^{-T}]\, w_{ik}$$

Multiplying both sides on the right by Σ^T, we have

$$0 = \sum_{i=1}^{n} \sum_{k=1}^{G} [\Sigma^{-T}(\mu_k - x_i)(\mu_k - x_i)^T - I]\, w_{ik}$$

Thus

$$I \sum_{i=1}^{n} \sum_{k=1}^{G} w_{ik} = \Sigma^{-T} \sum_{i=1}^{n} \sum_{k=1}^{G} w_{ik}[(\mu_k - x_i)(\mu_k - x_i)^T]$$

Finally, multiplying both sides on the left by Σ^T, and since we get

$$\Sigma^T = \frac{\sum_{i=1}^{n}\sum_{k=1}^{G} w_{ik}[(\mu_k - x_i)(\mu_k - x_i)^T]}{\sum_{i=1}^{n}\sum_{k=1}^{GI} w_{ik}}$$

However, since $\Sigma^T = \Sigma$, we can use the above equation for updating Σ.