Machine Learning Homework 12

Problem 1

Let $x_n, \mu_k \in \mathcal{R}^D$ and $\Sigma_k = \Sigma = \sigma^2 I \in \mathcal{R}^{D \times D}$. Since Σ is fixed we do not reestimate it. The probability of observing a single samples x_n simplifies to

$$p(x_n|\mu_k, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}\sigma} e^{-\frac{1}{2\sigma^2}||x_n - \mu_k||^2} \tag{1}$$

It follows that the responsibilities r_{nk} are given by a softmax function, that is

$$r_{nk} = \frac{\pi_k e^{-\frac{1}{2\sigma^2}||x_n - \mu_k||^2}}{\sum_j \pi_j e^{-\frac{1}{2\sigma^2}||x_n - \mu_j||^2}} = \begin{cases} 1 & \text{if } k = \arg\min_j ||x_n - \mu_j|| \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

which results in hard assignments of samples to clusters. According to (9.40) in Bishop, the expected value of the complete-data log likelihood is given by

$$L = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \left(\log \pi_k + \log \mathcal{N}(x_n|\mu_k, \Sigma_k) \right) \tag{3}$$

In the limit, the Gaussian distribution simplifies to

$$\lim_{\sigma \to 0} \mathcal{N}(x_n|\mu_k, \Sigma_k) = \lim_{\sigma \to 0} -\log \sigma - \frac{1}{2\sigma^2}||x_n - \mu_k||^2 + \text{const.} \propto -||x_n - \mu_k||^2 \tag{4}$$

Finally, maximizing the the log-likelihood, given by the formula below, corresponds to the minimization of the K-Means error function.

$$L = \sum_{n=1}^{N} \sum_{k=1}^{K} -r_{nk}||x_n - \mu_k||^2 \tag{5}$$

Problem 2

$$p(x) = \sum_k \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{6}$$

Suppose we have two random variables of the same dimensionality: $X \sim \mathcal{N}(x|\mu_x, \Sigma_x)$ and $Y \sim \mathcal{N}(y|\mu_y, \Sigma_y)$. Let $z = \begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}(x|\mu_z, \Sigma_z)$ with $\mu_z = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$ and $\Sigma_z = \begin{pmatrix} \Sigma_x & 0 \\ 0 & \Sigma_y \end{pmatrix}$ We can construct a matrix A such that $Az = \pi_x x + \pi_y y$ It follows that $Az \sim \mathcal{N}(\mu_{xy}, \Sigma_{xy})$ with $\mu_{xy} = \pi_x \mu_x + \pi_y \mu_y$ and $\Sigma_{xy} = \pi_x^2 \Sigma_x + \pi_y^2 \Sigma_y$. This scheme generalizes to an arbitrary number of components. Specifically, a Gaussian Mixture Model with K components is characterized by the following values:

$$\begin{aligned} E[x] &= \sum_{k=1}^{K} \pi_k \mu_k \\ Cov(x) &= \sum_{k=1}^{K} \pi_k^2 \Sigma_k \end{aligned} \tag{7}$$