Problem Set 3

Written Problems Solutions

Problem 1

9.1 Since both the E- and the M-step minimise the distortion measure (9.1), the algorithm will never change from a particular assignment of data points to prototypes, unless the new assignment has a lower value for (9.1).

Since there is a finite number of possible assignments, each with a corresponding unique minimum of (9.1) w.r.t. the prototypes, $\{\boldsymbol{\mu}_k\}$, the K-means algorithm will converge after a finite number of steps, when no re-assignment of data points to prototypes will result in a decrease of (9.1). When no-reassignment takes place, there also will not be any change in $\{\boldsymbol{\mu}_k\}$.

Problem 2

9.3 From (9.10) and (9.11), we have

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{x}|\mathbf{z})p(\mathbf{z}) = \sum_{\mathbf{z}} \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right)^{z_k}.$$

Exploiting the 1-of-K representation for $\mathbf{z}$, we can re-write the r.h.s. as

$$\sum_{j=1}^{K} \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right)^{I_{kj}} = \sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

where $I_{kj} = 1$ if $k = j$ and 0 otherwise.

Problem 3

Consider a binary classification dataset where the direction of maximum variance is parallel to the hyperplane separating the two classes. In this case, projecting the data onto the first principal component will make the two classes indistinguishable.