4.1 Problem 1[10 pts]

Read Section 9.1 in Bishop that discusses the K-means algorithm, and solve problem 9.1 which asks you to prove that it converges.
Consider the K-means algorithm discussed in Section 9.1. Show that as a consequence of there being a finite number of possible assignments for the set of discrete indicator variables r_{nk} and that for each such assignment there is a unique optimum for the $\{\mu_k\}$ the K-means algorithm must converge after a finite number of iterations.

1. Prove it converges. Using the derivative from Section 9.1, we can see that by assigning clusters the mean of all the data points that it contains, the loss function J can be reducing gradually. In this case, for each iteration, if the new clusters are different from the previous one, then the new one is supposed to have lower loss J than the previous clusters. And if the new one is just the same as the previous clusters, then the results will not change in the next following iterations as well, which means it has reached the local minimum. For the case that there is threshold for the difference in loss J, then the algorithm will finally reach the threshold point. Convergence problem proved.

$$\boldsymbol{\mu}_k = \frac{\sum_n r_{nk}\mathbf{x}_n}{\sum_n r_{nk}}. \qquad J = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$$

2. Assume that we have multiple assignments that are the same r_{nk}, whatever the assignments strategies we use. Instead of proving the uniqueness of the optimum for each assignment, we can prove that those identical assignments will finally converge to the same {uk}. In this case, since the assignment are identical, the initial clusters are the same as well, which leads to the same loss J. For each iteration/recomputation, the centroids of the clusters are assigned to the mean value position of the assignment points, which are the same as well. Because of that, the result clusters for every recomputation and reassignment are identical among those identical assignments. Since the convergence of K-means has been proved, every assignment will finally reach its optimum after finite iteration and the results are the same for those which have the identical assignments.

4.2 Problem 2[10 pts]

Read the beginning of Section 9.2 which describes Gaussian mixture models, and solve Problem 9.3.

Consider a Gaussian mixture model in which the marginal distribution $p(z)$ for the latent variable is given by (9.10) and the conditional distribution $p(x|z)$ for the observed variable is given by (9.11). Show that the marginal distribution $p(x)$ obtained by summing $p(z)p(x|z)$ over all possible values of z is a Gaussian mixture of the form (9.7).

given $P(Z) = \prod_{k=1}^{K} \pi_k^{z_k}$ $\quad p(x|Z) = \prod_{k=1}^{K} \mathcal{N}(x|\mu_k, \Sigma_k)^{z_k}$

marginal distribution

$$p(x) = \sum_{z} p(z) p(x|z) = \sum_{z} \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)\right)^{z_k}$$

$\because$ z : 1-of-K representation $\Big|$ $P(z_k=1) = \pi_k$ with K-dimension, $\sum_{k=1}^{K} z_k = 1$

$\therefore$ for every z, $\prod_{k=1}^{K} (\pi_k \mathcal{N}(x|\mu_k, \Sigma_k))^{z_k} = \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$ where $z_k = 1$

$$\therefore p(x) = \sum_{z} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) = \underbrace{\sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}_{\text{every class in } [1, K] \text{ happends once}}$$

4.3 Problem 3[10 pts]

Go through Section 12.1.2 which describes the Minimum-error formulation of PCA and perform omitted computations. Specifically, do all the derivations necessary to show that

0. Before (12.9)

$$\alpha_{nj} = \mathrm{x}_n^{\mathrm{T}} \mathrm{u}_j$$

1. (12.12)

$$z_{nj} = \mathrm{x}_n^{\mathrm{T}} \mathrm{u}_j$$

2. (12.13)

$$b_j = \overline{\mathrm{x}}^{\mathrm{T}} \mathrm{u}_j$$

3. In case of two-dimensional data space

$$\mathrm{S}\mathrm{u}_2 = \lambda_2 \mathrm{u}_2$$

$$J = \lambda_2$$

0. $\alpha_{nj} = x_n^T u_j \qquad x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{nD} \end{pmatrix} \quad u_i = \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iD} \end{pmatrix}$

$x_n = \sum_{i=1}^{D} \alpha_{ni} u_i$

$x_n^T = \sum_{i=1}^{D} u_i^T \alpha_{ni}^T$ inner product with u_j

$u_j x_n^T = \sum_{i=1}^{D} u_j u_i^T \alpha_{ni}^T = 1 \cdot \alpha_{nj}^T$

$\therefore \alpha_{nj} = x_n^T u_j$

1. $J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \tilde{x}_n \|^2 \qquad x_n = \sum_{i=1}^{D} (x_n^T u_i) u_i$

$\tilde{x}_n = \sum_{i=1}^{M} z_{ni} u_i + \sum_{i=M+1}^{D} b_i u_i$

$\therefore \frac{\partial J}{\partial z_{nj}} = -\frac{2}{N} \sum_n \left(\sum_i^D (x_n^T u_i) u_i - \sum_i^M (z_{ni} u_i) - \sum_{M+1}^{D} b_i u_i \right) \cdot u_j = 0$

$\sum_n (x_n^T u_j - z_{nj}) = 0, \quad \sum x_n^T u_j = \sum z_{nj}$

and u_j is complete orthonormal vector

$\therefore z_{nj} = x_n^T u_j$

2. $\frac{\partial J}{\partial b_j} = -\frac{2}{N} \sum_n \left(\sum_i^D (x_n^T u_i) u_i - \sum_{i=1}^{M} z_{ni} u_i - \sum_{M+1}^{D} b_i u_i \right) u_j = 0$

$\sum_n (x_n^T u_j - b_j) = 0$, b_j is constant

$b_j = \frac{\sum_n x_n^T u_j}{n} = \bar{x}^T u_j$

3. $J = \sum_{i=M+1}^{D} u_i^T S u_i = u_2^T S u_2$

$\tilde{J} = u_2^T S u_2 + \lambda_2 (1 - u_2^T u_2)$

$\frac{\partial \tilde{J}}{\partial u_2} = 2 S u_2 - 2 \lambda_2 u_2 = 0$

$S u_2 = \lambda_2 u_2$

$\cancel{J = \sum_{i=M+1}^{D} u_i^T S u_i = \sum_{i=M+1}^{D} \lambda_i u_i^T \cdot u_i = \sum_{i=M+1}^{D} \lambda_i}$

$J = u_2^T S u_2 = \lambda_2 u_2^T u_2 = \lambda_2$