

# Assignment 4

CMSC 726: Machine Learning  
October 30<sup>th</sup>, 2018

**Name:**

1. Problem 22.2 from the textbook.
2. Read Section 3 of the following notes: <http://www.cs.umd.edu/class/fall2018/cmsc726/materials/notes8.pdf>.

Derive the (M-step) update equation for  $\Sigma_j$  in the EM algorithm for Gaussian mixtures, i.e.,

$$\Sigma_j = \frac{\sum_{i=1}^m w_j^{(i)} (\mathbf{x}^{(i)} - \mu_j)(\mathbf{x}^{(i)} - \mu_j)^T}{\sum_{i=1}^m w_j^{(i)}}$$

3. (Programming Assignment) Implement the  $k$ -means algorithm over a dataset consisting of a mixture of four Gaussians. In particular, the means of the Gaussians are:  $\mu_1 = c[1, 1]^T$ ,  $\mu_2 = c[-1, 1]^T$ ,  $\mu_3 = c[-1, -1]^T$ ,  $\mu_4 = c[1, -1]^T$  where  $c = 2$ . Each Gaussian has an identity covariance.
  - (a) Generate a dataset with 250 samples from each Gaussian component. Overall, the dataset has 1,000 samples.
  - (b) Implement the  $k$ -means algorithm. How do the estimated cluster centroids compare to the mean vectors  $\mu_j$ 's?