# COMP 421 – HOMEWORK 06

## REPORT

Özgür Taylan Özcan

Initially, I generated the random data from the given Gaussian densities, using mvrnorm function for each class with size, mean and covariance parameters.

I plotted the randomly generated data on a graph and got the expected result.

For k-means clustering algorithm, I set K=5 and then initialized the centroids randomly by sampling from data points. I called the k-means function for 2 iterations. As a result, I received the updated centroids and cluster assignments for each data point.

Then, using the cluster assignments, I calculated the prior probabilities and initial covariance matrices to be able to use them as initial values for mean vectors in the EM algorithm.

After initialization, I implemented the EM algorithm, considering the related equations from the textbook. I firstly calculated h, using the following equation:

$$h_i^t = \frac{\pi_i |\mathbf{S}_i|^{-1/2} \exp[-(1/2)(\boldsymbol{x}^t - \boldsymbol{m}_i)^T \mathbf{S}_i^{-1} (\boldsymbol{x}^t - \boldsymbol{m}_i)]}{\sum_j \pi_j |\mathbf{S}_j|^{-1/2} \exp[-(1/2)(\boldsymbol{x}^t - \boldsymbol{m}_j)^T \mathbf{S}_j^{-1} (\boldsymbol{x}^t - \boldsymbol{m}_j)]}$$

Then using h, I updated the prior probabilities, mean vectors and covariances. I applied the following equations:

$$\pi_i = \frac{\sum_t h_i^t}{N}$$

$$\boldsymbol{m}_i^{l+1} = \frac{\sum_t h_i^t \boldsymbol{x}^t}{\sum_t h_i^t}$$

$$\mathbf{S}_i^{l+1} = \frac{\sum_t h_i^t (\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1})(\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1})^T}{\sum_t h_i^t}$$

At the end, I printed the mean vectors and plotted the clusters on a graph with gaussian density lines on them (original densities as dashed lines and my densities as solid lines).

I got similar results with the ones in the homework description, as expected.