Engr 421 – HW #5 Report

Firstly, I defined four functions for updating the membership, plotting the current state, e step of the em algorithm, and m step of the em algorithm. Functions for updating the membership and plotting the current state are taken from the lab. I used the formulas which are discussed in the lecture below to implement the em algorithm.

EM Algorithm:

E-STEP: $E[L(\Phi|\mathcal{X},\mathcal{Z})|\mathcal{X},\Phi^{(t)}]$ iteration

M-STEP: $\Phi^{(t+1)} = \arg\max E[L(\Phi|\mathcal{X},\mathcal{Z})|\mathcal{X},\Phi^{(t)}]$

E-STEP: $h_{ik} = E[z_{ik}|\mathcal{X},\Phi^{(t)}] = \dfrac{p(x_i|C_k,\Phi^{(t)}) \cdot P(C_k)}{\sum_{c=1}^{K} p(x_i|C_c,\Phi^{(t)})P(C_c)}$

success probability

$h_{ik} \geq 0$

$\sum_{k=1}^{K} h_{ik} = 1$

M-STEP: $\hat{\mu}_k^{(t+1)} = \dfrac{\sum_{i=1}^{N} h_{ik} \cdot x_i}{\sum_{i=1}^{N} h_{ik}}$ $\quad \hat{\Sigma}_k^{(t+1)} = \dfrac{\sum_{i=1}^{N} h_{ik}(x_i-\hat{\mu}_k^{(t+1)})(x_i-\hat{\mu}_k^{(t+1)})^T}{\sum_{i=1}^{N} h_{ik}}$

$\hat{P}(C_k) = \dfrac{\sum_{i=1}^{N} h_{ik}}{N}$

multivariate Gaussians

I initialized centroids with the given data. Using the data set and centroids, I found memberships. I got the number of data points as N and set the iterations to 100. Then, I calculated the priors and covariances.

I iterated the em algorithm for 100 times and printed the mean vectors as shown below.

```
[[-2.04419197 -2.69776844]
 [ 2.6622246  -2.30911081]
 [ 2.48874351  2.67687075]
 [-2.67591954  2.44658905]
 [ 0.15535175  0.05773829]]
```

Lastly, I plotted the clustering result and drawed the gaussian densities where they equal to 0.05.