Mehmet Mert Bezirgan

Engr 421 – homework 6

**Problem**: Clustering

**Solution:**

For this homework we generated multivariate dataset with given means and covariances. At first step we apply k-means clustering algorithms 2 times to have basic idea of means covariances etc.

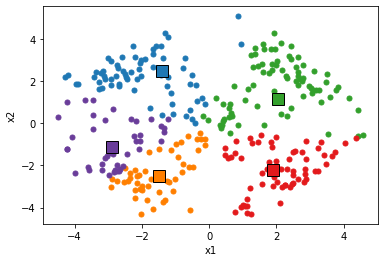


Figure 1 Dataset divided after kmeans

After K-means we need to run EM algorithm for 100 iteration. At the beginning of that loop we need to calculate probability matrix for our data which will give probabilities of a data point is a member of 1 cluster. After calculating this matrix we need to update means, covariances and prior probabilities and repeat this loop for 100 times.

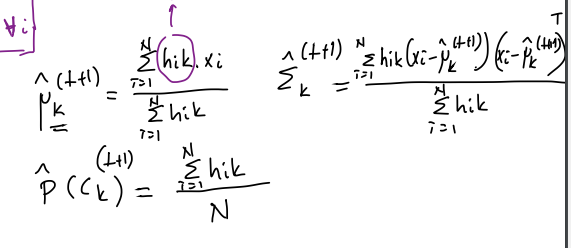


Figure 2 formulas for updating prior probabilities, covariances and means