# Harun Sasmaz

# 59900

**Homework 6 Report**

1. **Generate data**

I used *mvrnorm* function in R with given class means, class covariances and class sizes, to generate bivariate Gaussian data points for five classes.

# Initialization with k-means

As stated in the pdf, I sampled k=5 datapoints randomly as initial centroids and calculated the assignments according to distances. According to new assignments (cluster memberships), I calculated new centroids. One more time, I calculated assignments from centroids and new centroids from new assignments. Last assignments I found will be used in EM clustering as initial means in the first iteration.

# Estimations before EM iteration

Before starting EM iterations, I calculated prior probabilities (number of assignments in class c divided by number of all datapoints) and covariance matrices according to initial clusters.

# Implementing EM algorithm and running it 100 iterations

To implement EM algorithm, firstly I defined posterior probability that *xi* is generated by *Gc,*

Text

Description automatically generated

After calculating *hit,* I updated priors, means and covariances respectively with following rules:

*Text, letter

Description automatically generated*

With these update rules, I calculated these equations 100 times and got the means after 100 iterations. Note that formulas are taken from course textbook.

# Plotting clustering results and drawing ellipses

With the last calculated values of means, I calculated distances and assignments for all datapoints and plotted them with different color. Since we have class covariances and class means provided in the homework description, I draw ellipses to show original Gaussian densities with dashed lines. Also, gaussian densities that my EM algorithm found is represented on the plot with solid lined ellipses. Assignments and densities, I found are very similar to expected output.