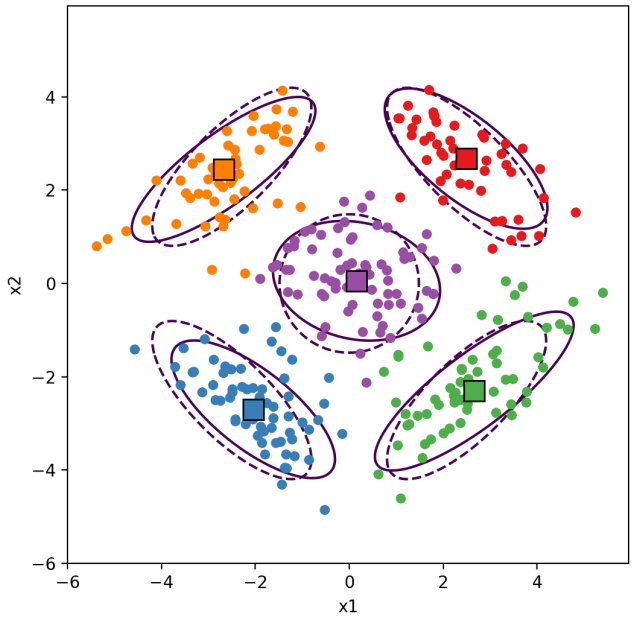
**ENGR 421 - HOMEWORK 5**
**Zeynep Öner**
**ID: 64912**

# Expectation-Maximization Clustering

This homework was an implementation of the expectation-maximization clustering algorithm. We were asked to group 5 different Gaussian densities.

To do so, we first need to run a k-means clustering algorithm and define the centroids along with the memberships. We already have covered this in the lab session. We update centroids and memberships until one of them stops changing.

We then need to run the Expectation-Maximization algorithm. The EM algorithm consists of 2 steps: E-step and M-step. In the E-step, the success probability function $h_{ik}$ is calculated. We multiple the multivariate Gaussian functions and the prior probabilities, and then divide it by the sum of the multiplications for all clusters. Using this success probability, we calculate the sample mean as sum of s.p. values times $x_i$ divided by the sum of s.p. values. The sample covariance is calculated by summing s.p. values and the distance between x and sample mean, along with the transpose of that last matrix. The prior probability is calculated by the sum of s.p. values divided by N. Using these values we update the memberships and plot the final clusters. We can draw the original and final Gaussian densities using the contour function of the matplotlib.pyplot library. The obtained picture is below:




The graph is not perfect, few data points are in different clusters but the final pdf looks the same. In addition, the mean vector generated is below:

```
[[-2.04376623 -2.69748647]
 [ 2.48903251  2.67603632]
 [ 2.6629993  -2.30718272]
 [ 0.15566804  0.05713617]
 [-2.67527591  2.44484063]]
```

The mean vector is exactly the same as the one presented in the homework description. Therefore, the algorithm worked successfully.