In this homework, I implemented an expectation-maximization (EM) clustering algorithm.

Note: I took Gaussian Ellipse code from someone’s github repository and I referenced it on the code.

**Part 1**

I read data from hw07\_data\_set which contains 300 data points generated randomly from five bivariate Gaussian densities.

**Part 2**

To initialize your EM algorithm, I took the centroids given in the file named hw07\_initial\_centroids.csv as the initial values for the mean vectors.

Graphical user interface, text, application

Description automatically generated

By assigning the data points to the nearest center, I estimated the initial covariance matrices and prior probabilities before applying EM algorithm.

**Part 3**

After the initialization step, I implemented EM algorithm.

E Step:

Diagram

Description automatically generated with low confidence

M Step:

Text, letter

Description automatically generated

For M Step I implemented three functions to update means, covariances and priors.

After defining functions, I run EM algorithm for 100 iterations. Then centroids which is mean vectors are found as follows:

Text

Description automatically generated

**Part 4**

I drew the clustering result obtained by your EM algorithm by coloring each cluster with a different color. I also drew the original Gaussian densities used to generate data points with dashed lines and the Gaussian densities EM algorithm with solid lines. I used gaussian ellipse code which is taken from github that is referenced above function.

Chart, diagram, bubble chart

Description automatically generated