



**KOÇ
UNIVERSITY**



**DASC521: Introduction to
Machine Learning**

Fall 2021

Instructor: Mehmet Gönen

Homework 07- Expectation-Maximization Clustering

Gamze Keçibaş- 60211

27.12.2021

In this homework, two-dimensional data set which includes 300 data points generated randomly from five bivariate Gaussian densities in a csv and its parameters (mean and covariance matrices) are provided. Initial mean vectors that are labelled as centroid are also provided. The provided data points are plotted in Figure 1:

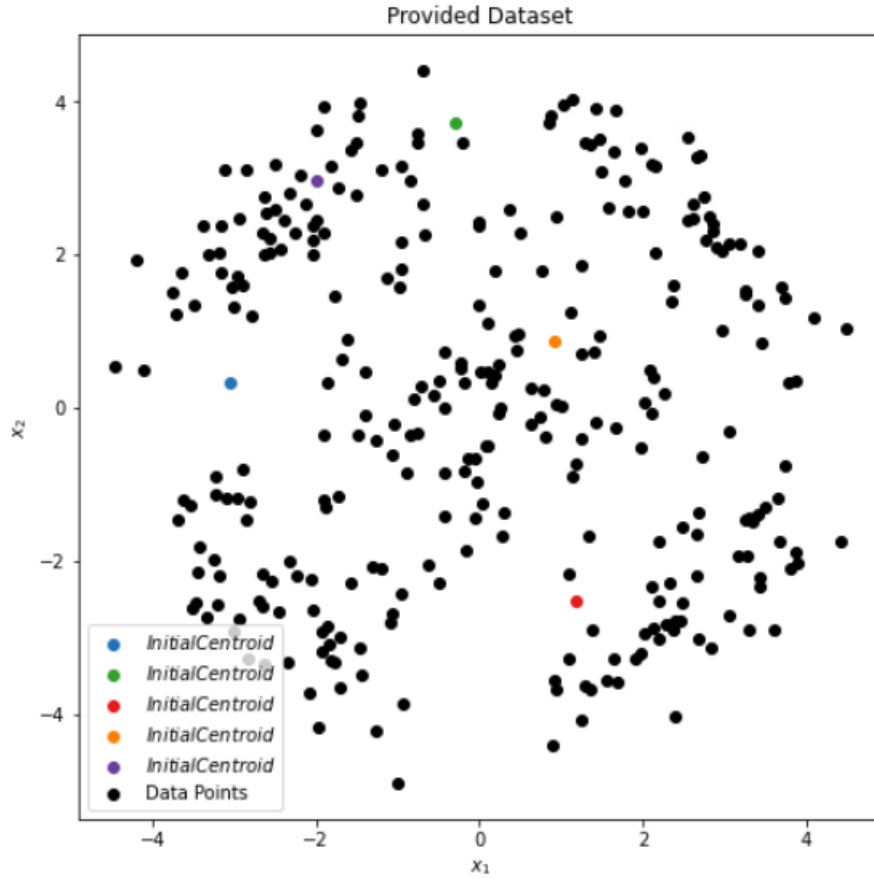


Figure 1: Provided data points in the homework

After importing the data, expectation-maximization (EM) algorithm is implemented as in related PS section. The algorithm runs 100 iterations and updated means (centroids) are below:

```
Initial centroids:
[[-3.0439416  0.32509753]
 [-0.30377397  3.73480678]
 [ 1.17587258 -2.50984601]
 [ 0.91037084  0.87608158]
 [-1.98588337  2.98466811]]

Final centroids:
[[ 2.52795218 -2.44881535]
 [-2.15654297  2.34986725]
 [-2.28528231 -2.4185532 ]
 [ 2.33410096  2.43982155]
 [ 0.27175327 -0.04157938]]
```

Figure 2: Initial and final centroids of clusters

In figure 2, initial centroids are also added to compare changing end of these iterations. After the 100 iterations, clusters seem in Figure 3:

Iteration#100:

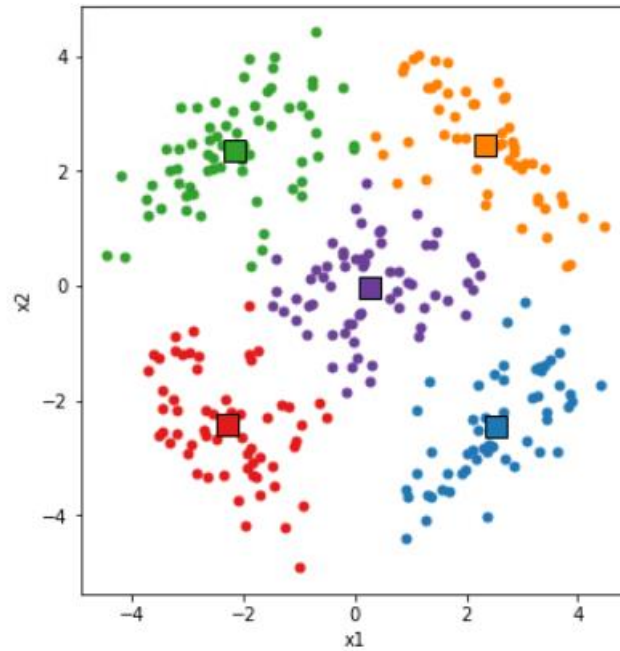


Figure 3: Updated clusters

In final step, confidence ellipses belonging to original Gaussian densities and Gaussian densities of EM algorithm where their values are equal to 0.05. Initial ellipses are represented dashed black lines and estimated ellipses are presented with appropriate color solid lines. The plot is shown in Figure 4:

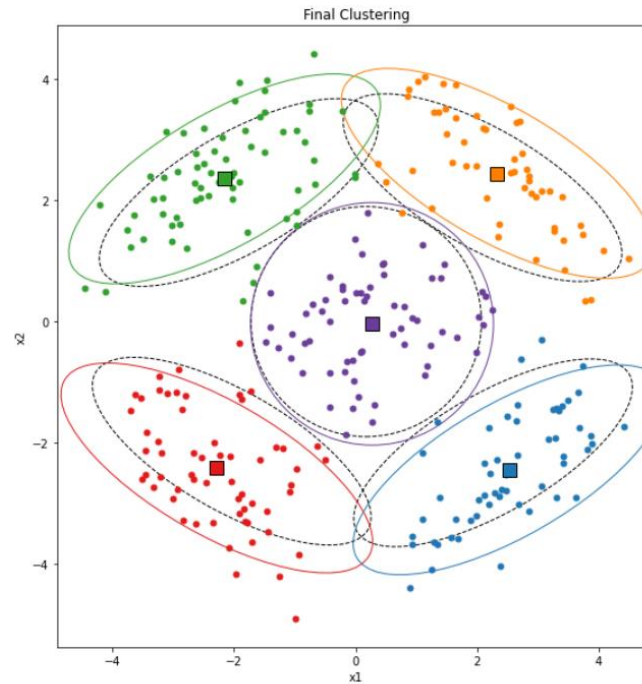


Figure 4: Final clustering and confidence ellipses