



**KOÇ  
UNIVERSITY**



**DASC521: Introduction to  
Machine Learning**

Fall 2021

**Instructor:** Mehmet Gönen

**Homework 08-** Spectral Clustering

Gamze Keçibaş- 60211

03.01.2022

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. The provided data points are plotted in Figure 1:

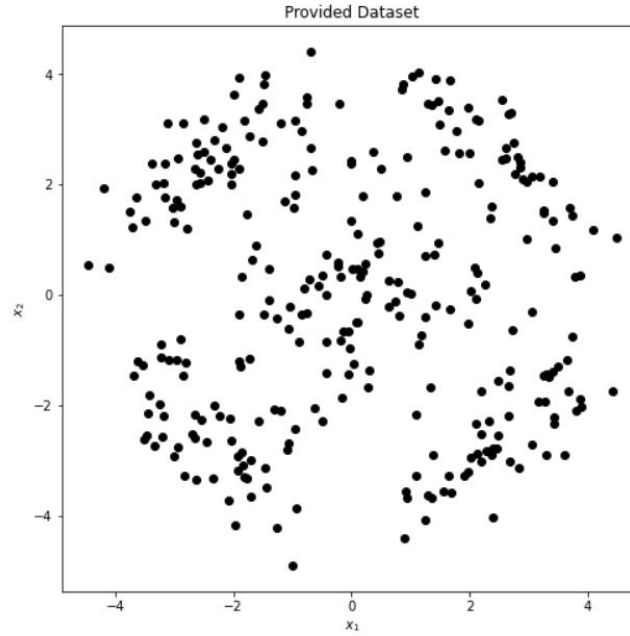


Figure 1: Provided data points in the homework

Euclidean distances are used to separate five clusters. Distance threshold that is called delta is given as 1.25. Distance pairs are calculated and visualized as below:

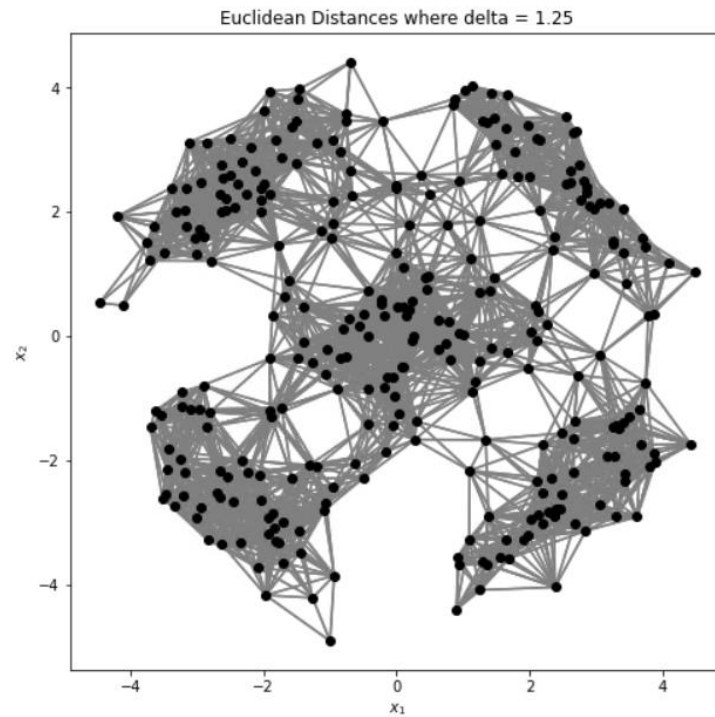


Figure 2: Euclidean Distances between data points

These distances are stored in B matrix in the code. After that D and L matrices are calculated by lecture notes and a function is created to get normalized Laplacian matrix. Finally, 300x5 Z matrix is built, final centroids are determined using provided initial centroids that are referred 29<sup>th</sup>, 143<sup>th</sup>, 204<sup>th</sup>, 271<sup>st</sup> and 277<sup>th</sup> rows of Z spectral clustering is completed. The final result is plotted as in Figure 3. Final centroids are also printed:

```
[array([ 0.18076212, -0.03097313]),
 array([2.24886322, 2.43861295]),
 array([-2.23674537, 2.40729986]),
 array([ 2.52795218, -2.44881535]),
 array([-2.29175988, -2.45402909])]
```

Figure 3: Final centroids of clusters

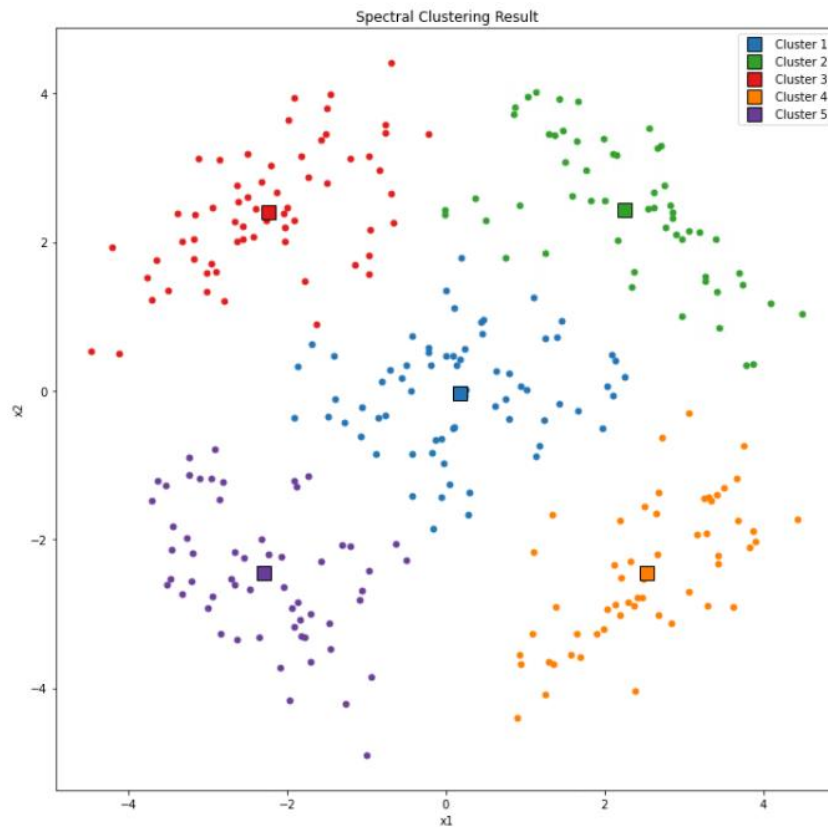


Figure 4: Result of Spectral Clustering