ENGR421

HW8

Doğa Demirtürk
68859

In this homework we implemented a spectral clustering algorithm in Python.
The data was the same as the data in the 7th homework, so I used the same code to import the data set. Then I started by constructing the B matrix (connectivity matrix) as described in the description of the homework and by setting the delta parameter to 1.25.

$$b_{ij} = \begin{cases} 1, & \|x_i - x_j\|_2 < \delta \\ 0, & \text{otherwise.} \end{cases}$$

$$b_{ii} = 0$$

Figure 1: Definition of connectivity matrix

After, I visualized the connectivity matrix by connecting each data point with its neighbors. The figure is the same as the figure given in the description file.
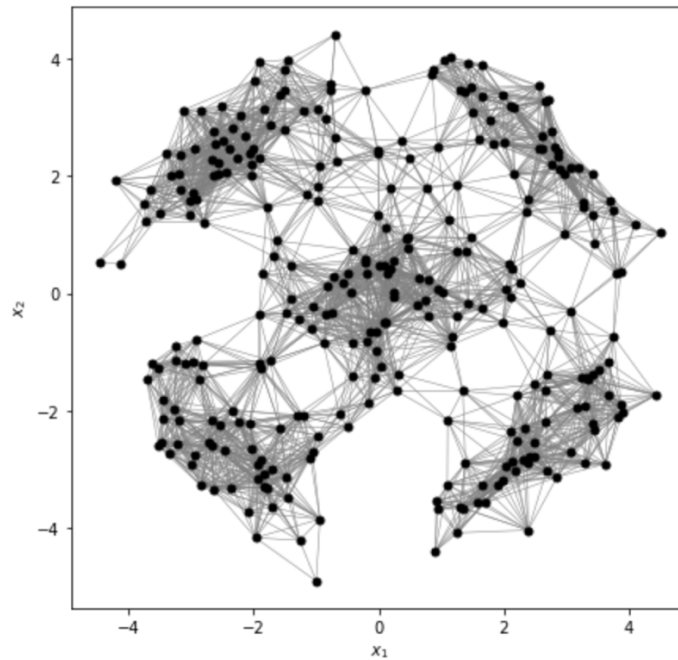

Figure 2: Visualiation of connectivity matrix

After that, I constructed D matrix by the definition given in the lecture notes. Then I calculated $L_{symmetric}$ matrix using the given formula.

$$d_{ii} = \sum_{j \neq i} b_{ij} \quad \forall i$$

$$\hookrightarrow \# \text{ of neighbors}$$
$$\text{of data point } i$$

$$d_{ij} = 0 \quad \forall (i, j \neq i)$$

Figure 3: Description of D matrix

$$\mathbf{L}_{symmetric} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{B}\mathbf{D}^{-1/2}$$

Figure 4: Formula for normalized Laplacian matrix

To continue with the spectral clustering algorithm, I found R eigenvectors corresponding to R smallest eigenvalues with R=5. I created Z matrix using those eigenvectors as columns of Z. Z is then an NxR matrix.

For k-means algorithm, I used the same code as in the lab 11. I assigned cluster number K as 5. For initialization of the centroids, I used the given rows of the Z matrix and I assigned memberships accordingly. After initialization step, I ran the k-means algorithm on Z. Finally, I calculated the means of the clusters according to the memberships of the X values since the centroids found were the means of the projected dots. I draw the figure as desired in the description file.

NOTE: I did this homework in a day, but the membership results were wrong. I tried a lot of different versions of the code and thought that there was a problem with my algorithm. I looked at the lecture notes, book, and internet, but it seemed like my algorithm was fine. Yet I tried a lot of different things for days. I finally figured out that I was starting with wrong rows of Z as initial centroids because in the pdf it says 29, 143,... but they should have been 28, 142,... due to indexing. It turned out the only problem was that. In conclusion, I think the pdf is confusing.
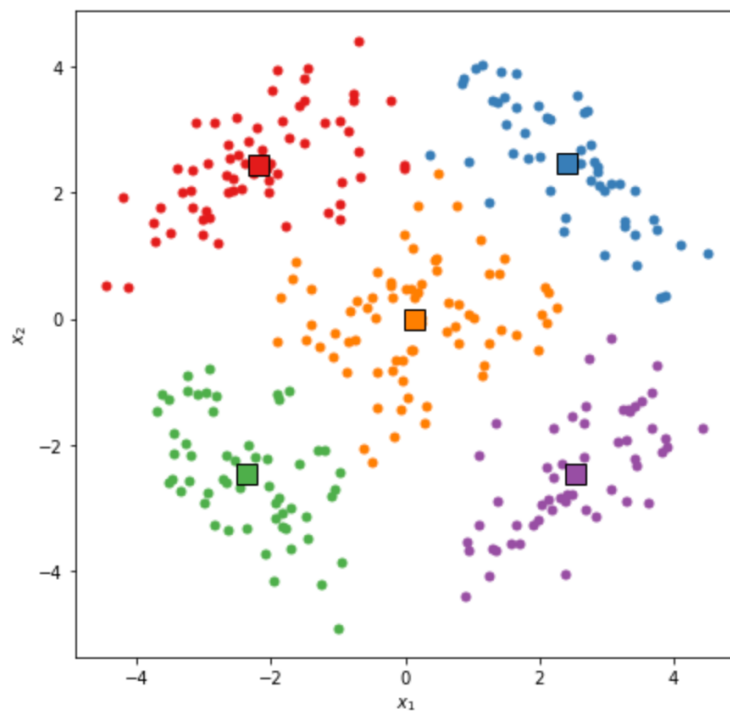
Figure 5: Visualization of the cluster assignments and mean of the clusters