CSE 847 Homework 5

By Aasiruddin Walajahi

Problem 1: Clustering

1.1.

K-Means clustering algorithm is very good at determine globular clusters.
Spectral relaxation algorithm allows clustering by determining the "shape" of the data. It does this by understanding the shape of the data by creating a similarity matrix.
Spectral relaxation is done with creating the eigenvectors of the laplacian matrix of the data. The top k eigenvectors are used as the data in the standard K-Means algorithm.
Yes, it is possible to obtain the same solution from both algorithms.

1.2.

We notice the effect of the different K-Means algorithm on two different datasets below:

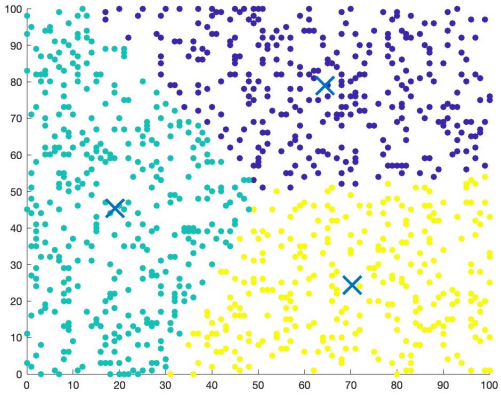

Fig 1. K-Means

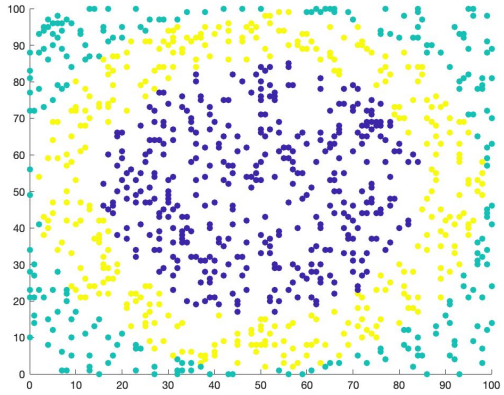

Fig 2. Spectral K-Means

In Fig 1. and Fig 2., we have an uniformly distribution data points. Upon applying the two clustering algorithms, we notice that they seperate the two dataset differently.

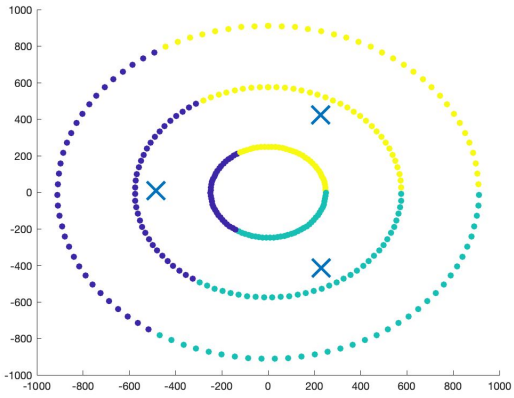

Fig 3. K-Means

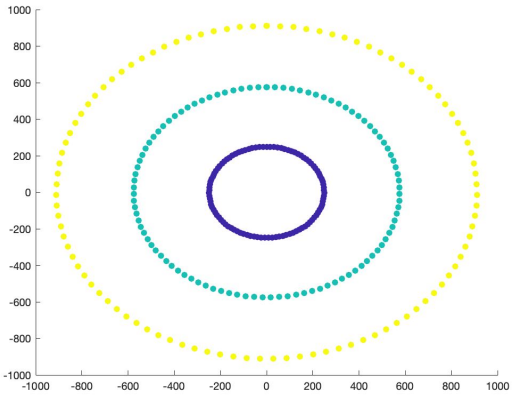

Fig 4. Spectral K-Means

We see that the above dataset, the spectral k-means dataset performs more accurate in determining the the correct clusters for the given dataset.

Problem 2: Principal Component Analysis

2.1.

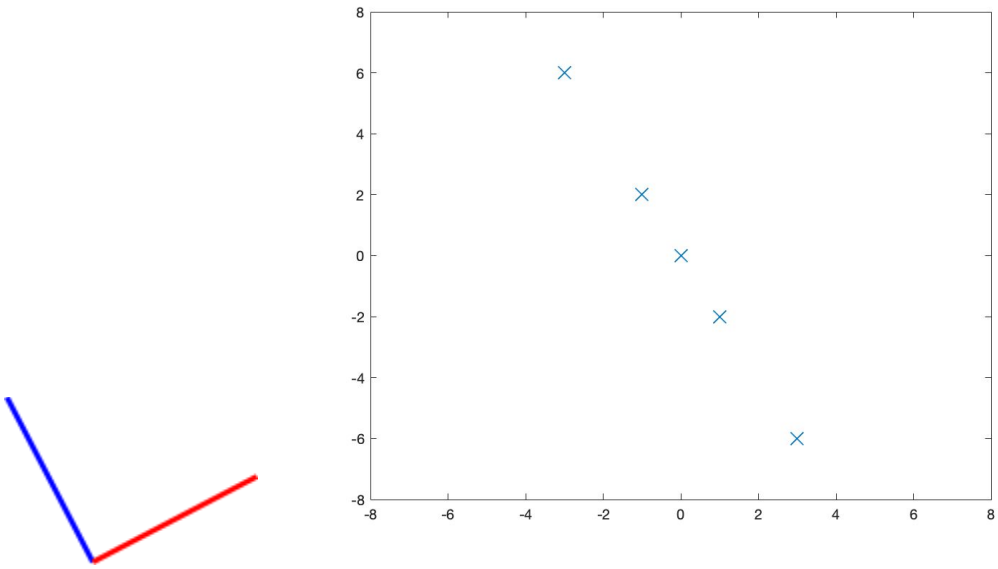

Fig 5. Given points plotted in 2D space

The first principal component is along the direction of all the points as this direction provides us with the most information regarding the dataset. In this case, it is the blue line depicted on the left figure above.
The second principal component is the orthogonal component to the first component in the second dimension. This is shown by the red line depicted in the figure of the left above.

2.2.

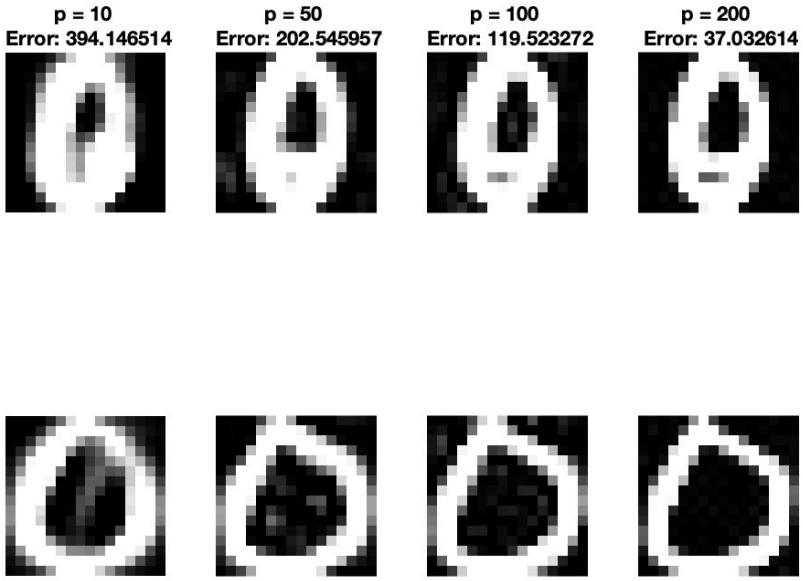

Fig 6. Reconstructed Images with their corresponding errors

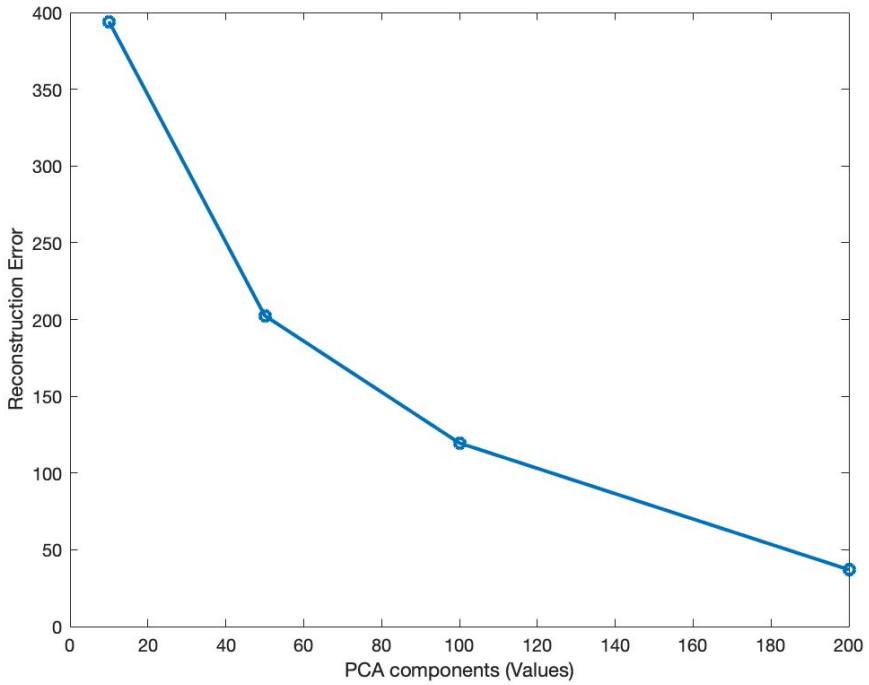

Fig 7. Reconstructed Errors with their corresponding p values

Source Code Location:

https://github.com/Aasir/CSE847/tree/master/HW5