CSE 847 (Spring 2022): Machine Learning-Homework 5

Maolin Gan

April 13, 2022

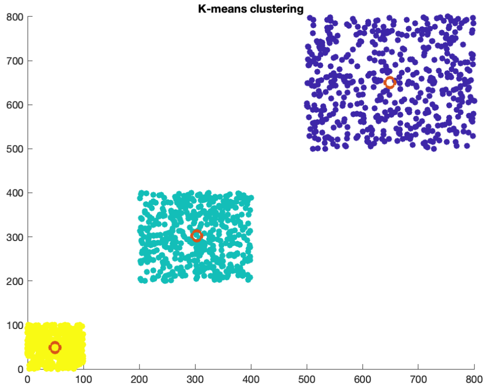
1 Github repository link

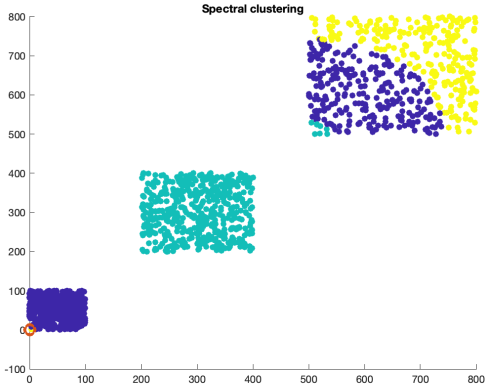
https://github.com/MaolinGanMSU/CSE847_homework5.git

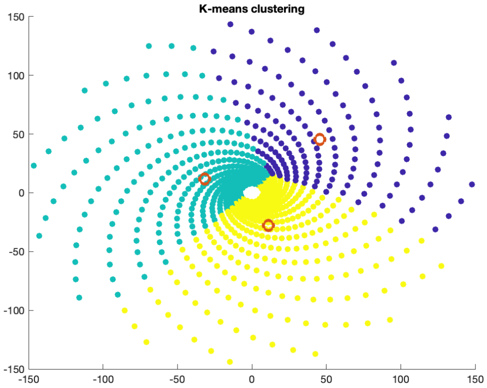
2 Clustering: K-means

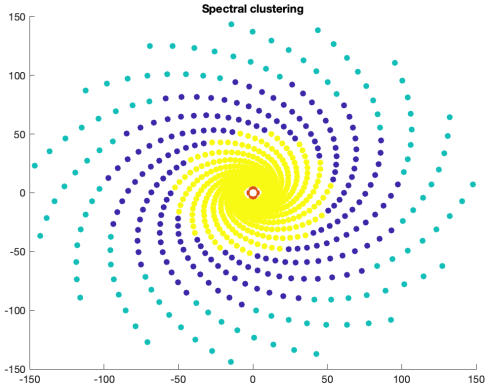
1. Elaborate the relationship between k-means and spectral relaxation of k-means. Is it possible that we obtain exact k-means solution using spectral relaxed k-means?
   Answer:
   K-Means divides the data into K-goups. Here, each group of data is close to each other and shares the same "mean" that represents that group. The K-Means algorithm aims to assign data points to the cluster with the smallest sum of squares of the distances to the nearest centroid. Spectral relaxation of k-means is to find the cluster in a graph form. Because it has connectivity, not compactness, you can find clusters of almost any shape, interwoven, spiral, and so on. The K-means algorithm generally assumes that the cluster is spherical or circular, that is, within the radius k from the cluster centroid. K-means requires multiple iterations to determine the cluster centroid. In spectral, clusters do not follow a fixed shape or pattern. Points that are distant but connected may belong to the same cluster, and points that are less distant may belong to different clusters if they are not connected. This means that spectral relaxation of k-means can be useful for data of various shapes and sizes.
   We can think that the spectral k-means is just k-means with a pre-processing step that involves transforming the data using the eigenvectors. If the converted result is in the original dataset, the k-means and spectral k-means results will be the same. Therefore, it is possible that we obtain exact k-means solution using spectral relaxed k-means.

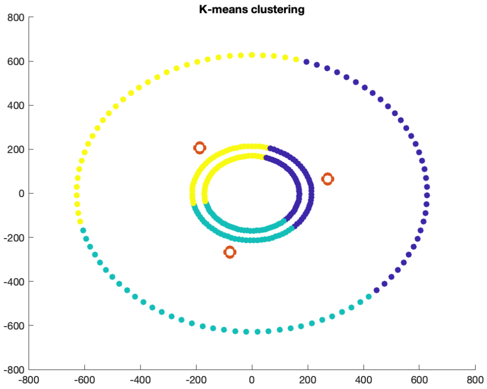
2. Implementation of k-means and spectral relaxation of k-means.
   Answer:
   We implement the k-means and spectral relaxation of k-means on three different random datasets. We notice that they have different effect. It seems k-means have better performance on linearly separable data. Spectral clustering have better performance on non-linearly separable data.

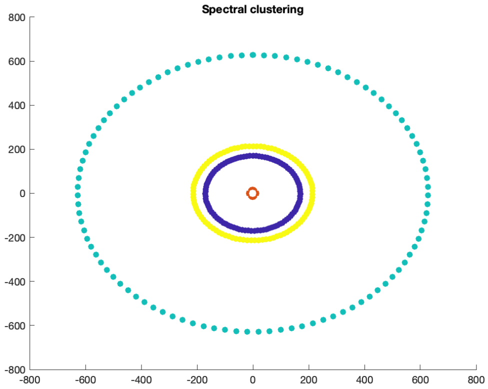






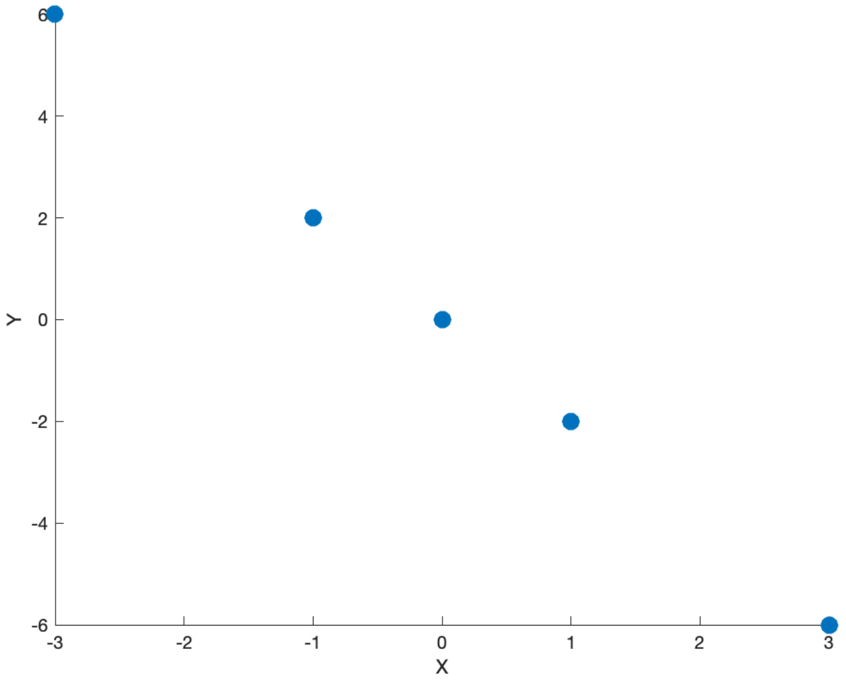






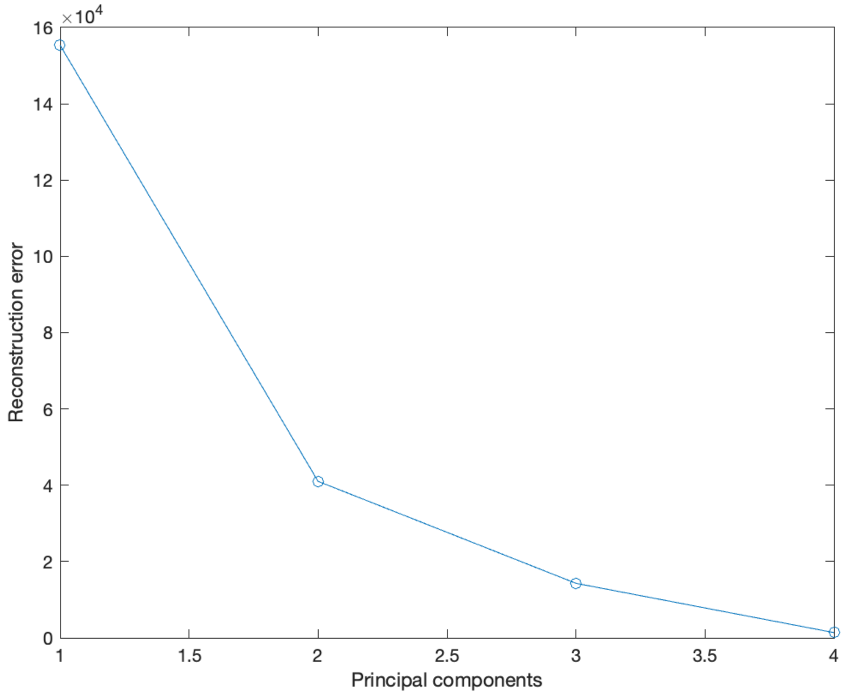
3 Principle Component Analysis

We draw the data points in 2d space as follows:

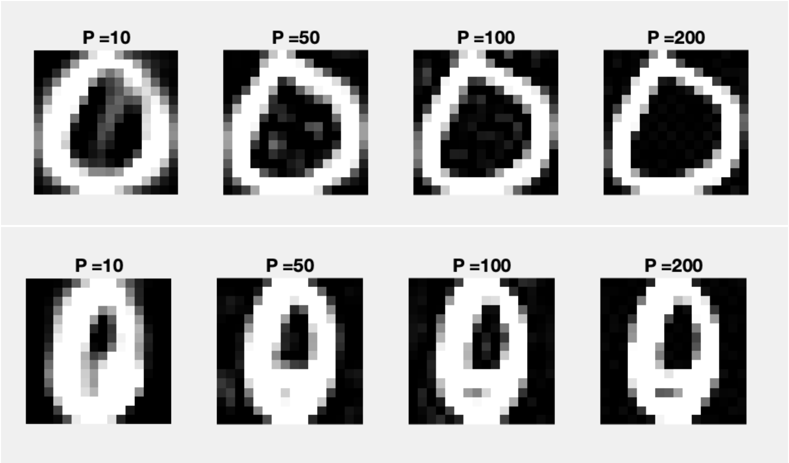


The principal components should be the unit vectors along which the data points are distributed. The first principal component is the line that best accounts for the shape of the point swarm. It is the line in the K-dimensional variable space that best approximates the data in the least squares sense and represents the maximum variance direction in the data. The second principal component is also represented by a line in the K-dimensional variable space, which is orthogonal to the first PC. This line also passes through the average point and reflects the second largest source of variation in the data.

The total reconstruction error for $p = 10, 50, 100, 200$:



A subset (the first two) of the reconstructed images for p = 10, 50, 100, 200:


P =10
P =50
P =100
P =200
P =10
P =50
P =100
P =200