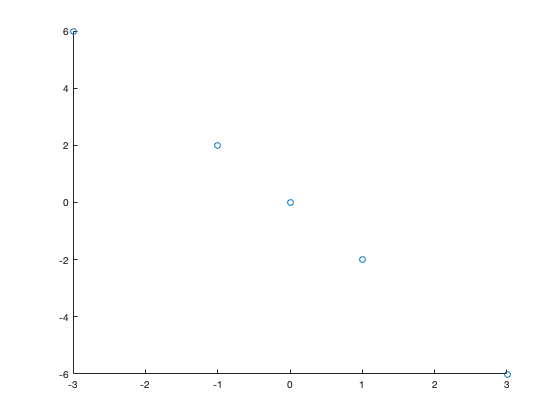
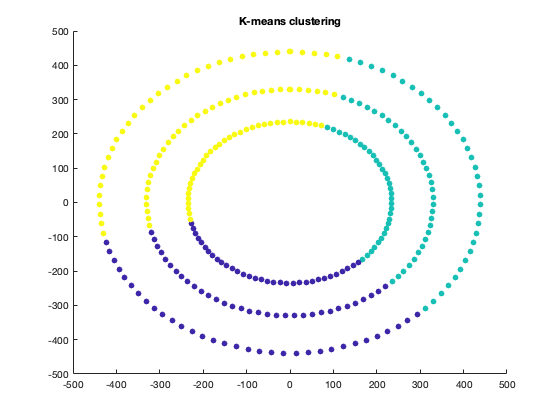
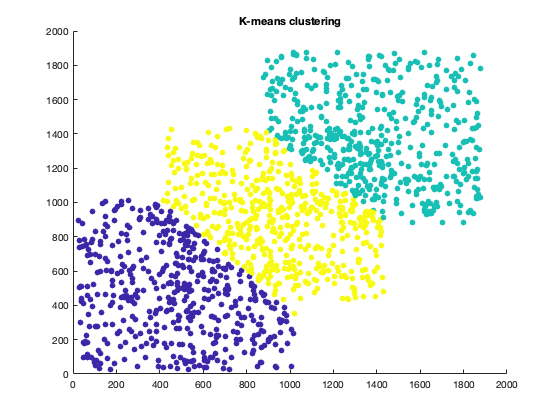
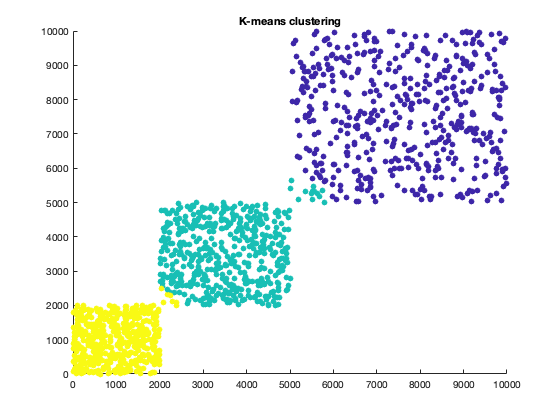
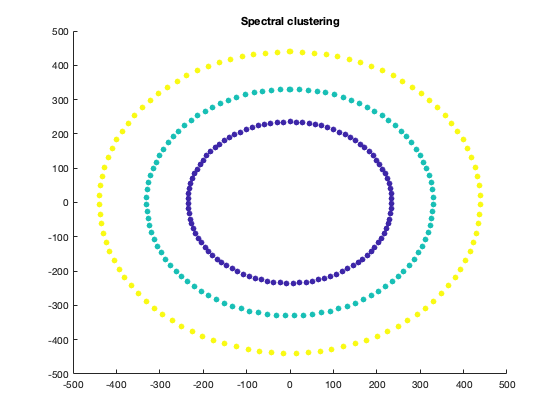
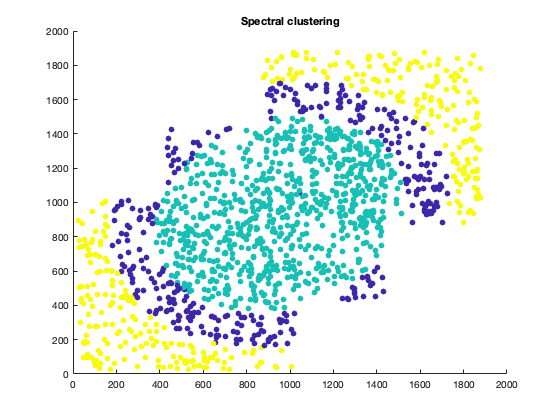
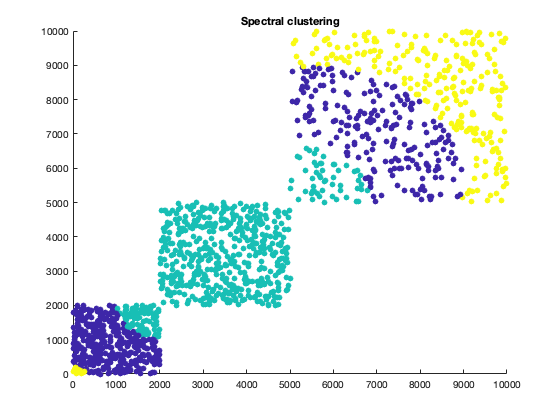
CSE 847 HW 5

Jonny Dowdall

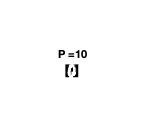
1. K-means determines centroids in the feature space of the original data. Points are assigned to clusters determined by the centroid they are closest to in the feature space. Spectral relaxation of k-means uses some distance metric to compute the pairwise distance between all data points. It then clusters the points according to the eigenspace of the distance between these points rather than the original feature space they exist in.
2. K-means seems to perform best on linearly separable data. Spectral clustering performs best on non-linearly separable data.

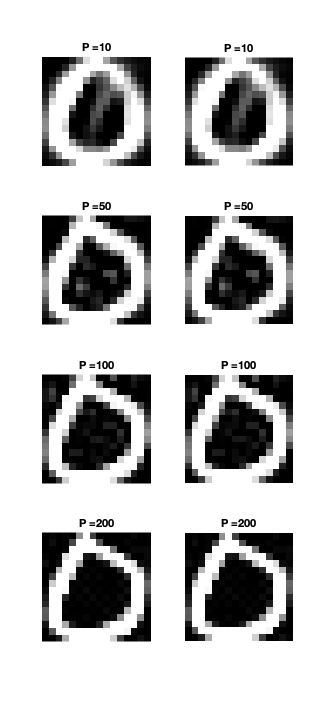


First PC

Second PC

2.

The first principal component points in the direction of the highest variance of the data. This is right along the line on which the data exists (-1 1). The second principal component points in the direction of the second highest variance of the data, orthogonal to the other principal component (1 1). There is no variance along this principal component since the data exists exclusively along X = -Y.



Reconstruction Error

6.9912

91.7083

16.4429

31.4649