EC500 - Introduction to Learning From Data

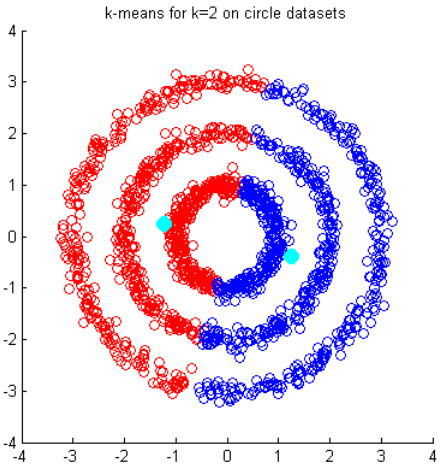
Matlab 4

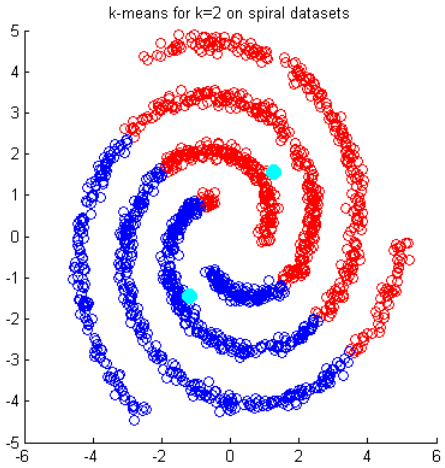
Mikhail Andreev

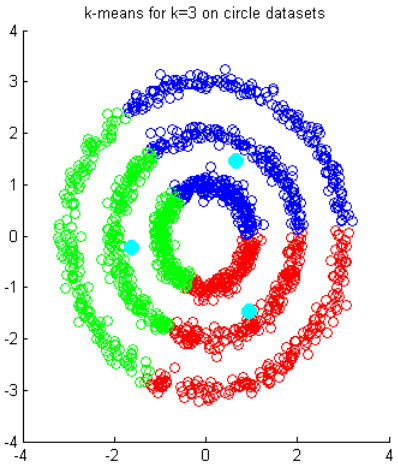
December 1, 2015

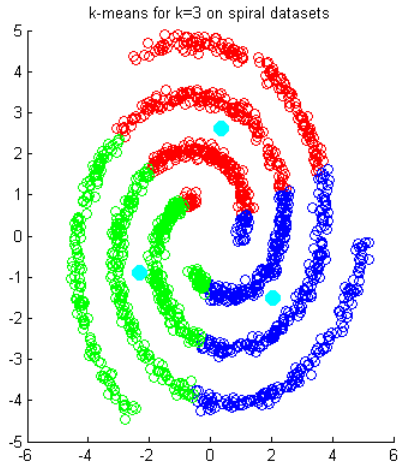
1 k-Means vs. Spectral Clustering

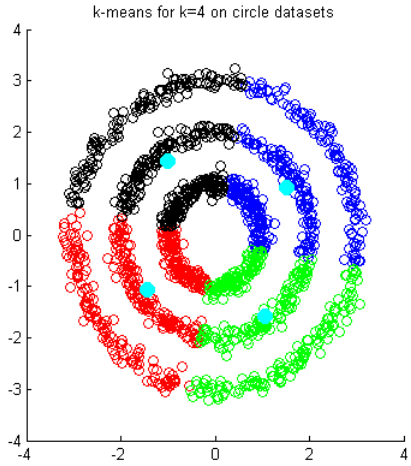
The k-means approach to the dataset yield the following three graphs for k=2,3,4 respectively:

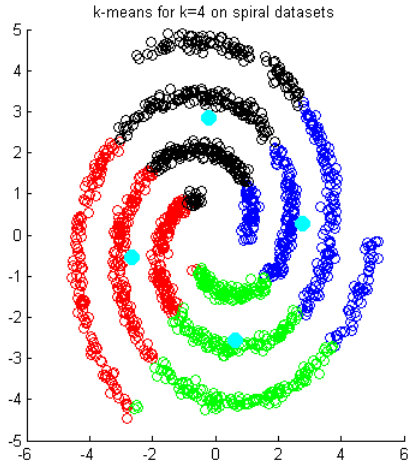








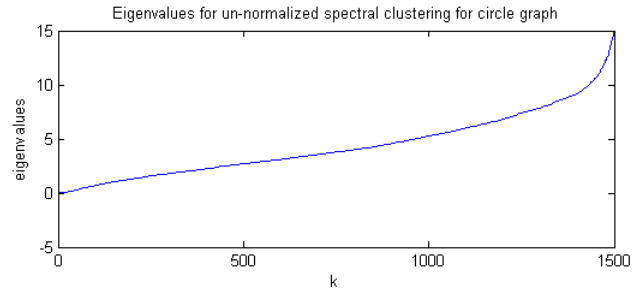


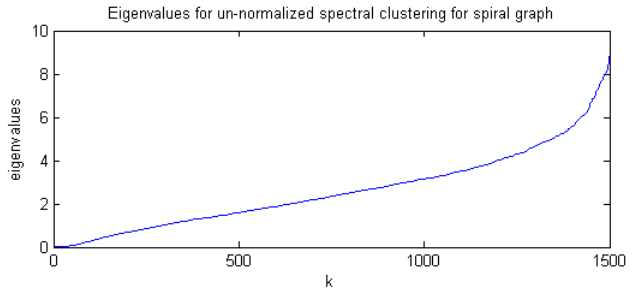


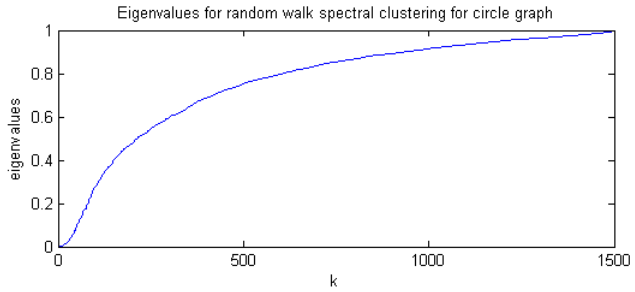
The cyan points in graphs indicate the centroids for each of the clusters. The overall sums of l2 distances between each point and the centroid for the clusters are:

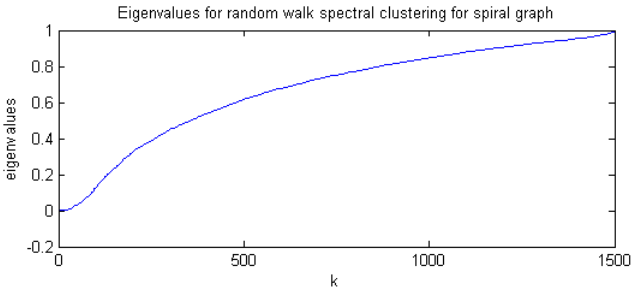
cluster 1 (red) - circle: 757.41, spiral: 1098.18
cluster 2 (blue) - circle: 721.62, spiral: 1088.39
cluster 3 (green) - circle: 522.01, spiral: 750.26
cluster 4 (black) - circle: 444.34, spiral: 609.27

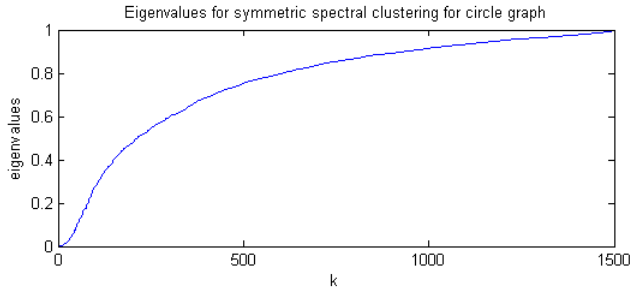
The eigenvalues of the Laplacian matrices can be seen here:

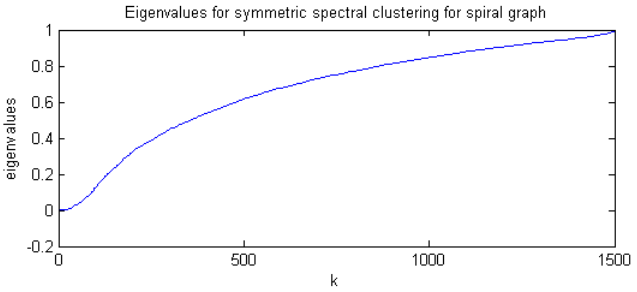




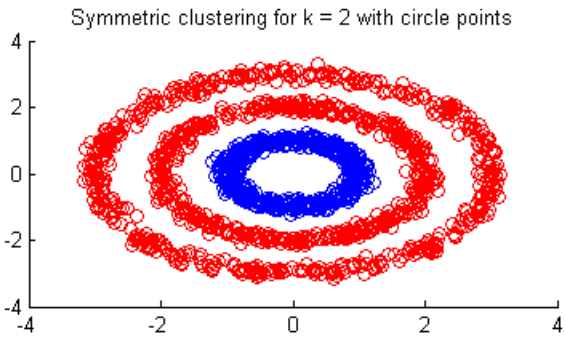


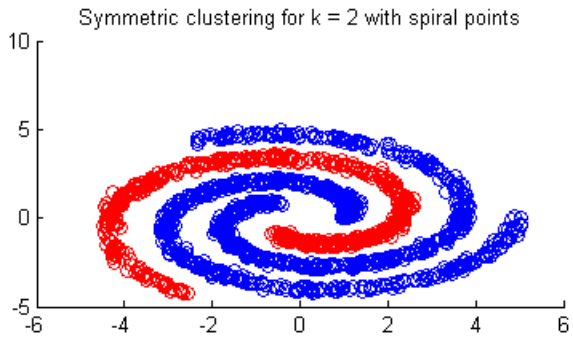


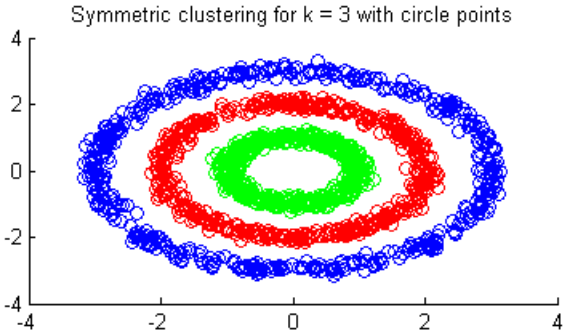


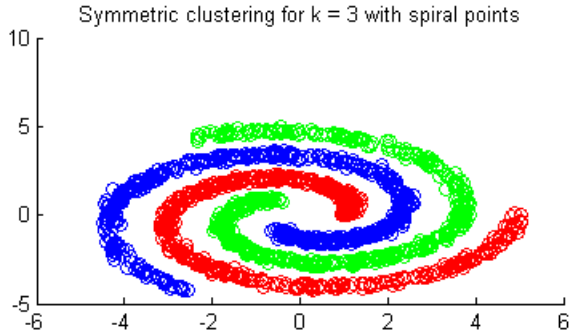


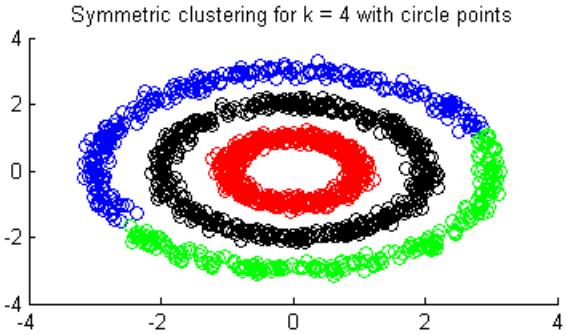
After applying spectral clustering we get the following output using symmetric normalization:

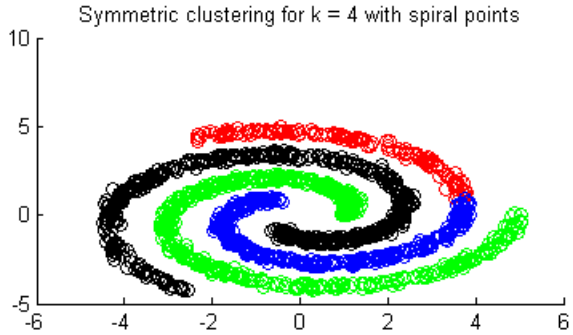




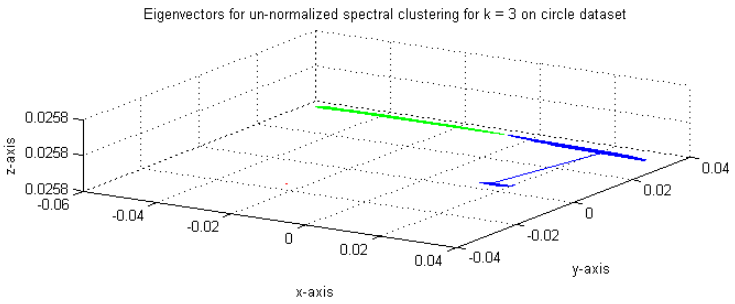


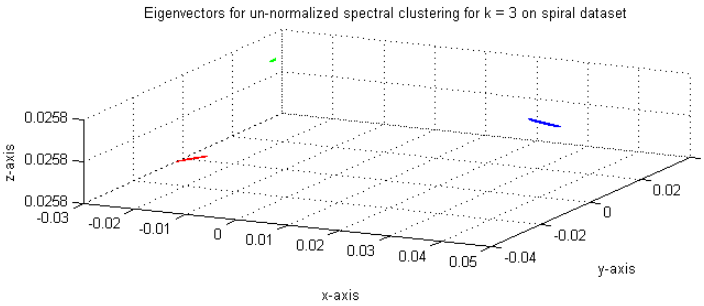


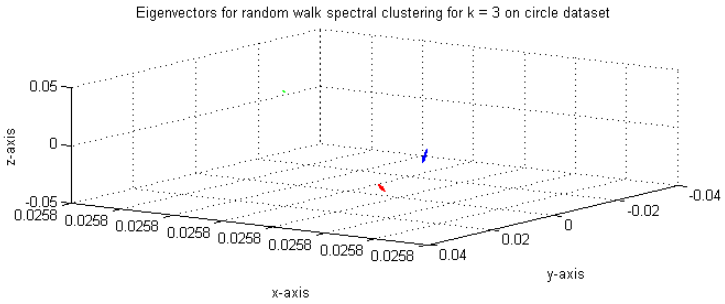


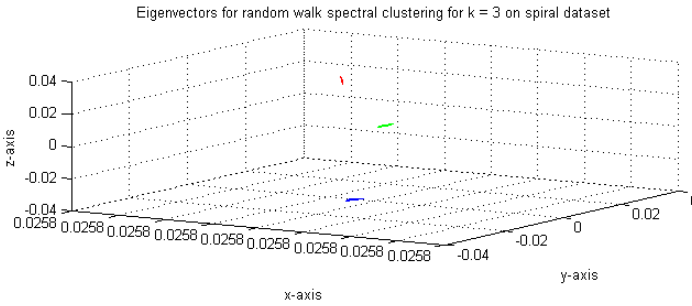


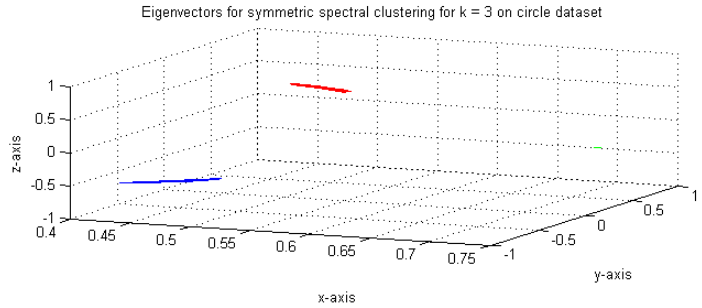
In the next figure we can see the 3D interpretation of the eigenvectors, showing that all three clusters are clearly separate in eigenspace:

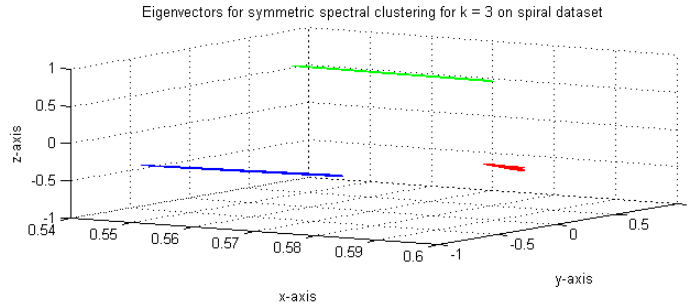




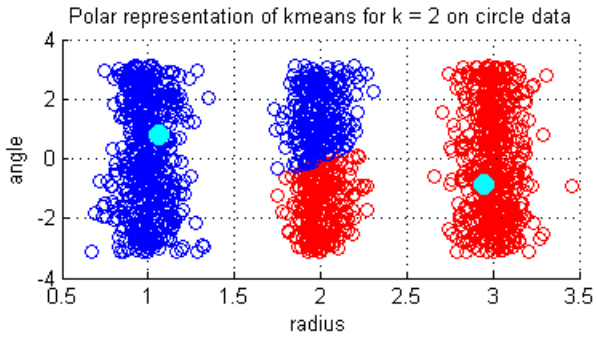


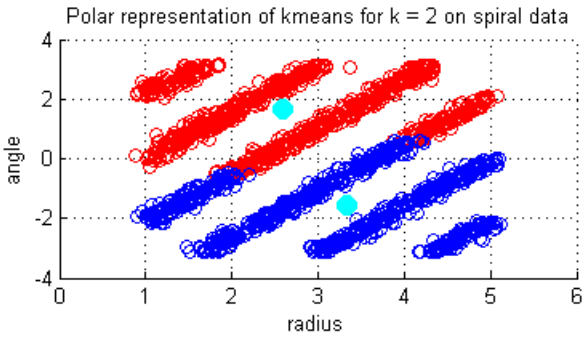


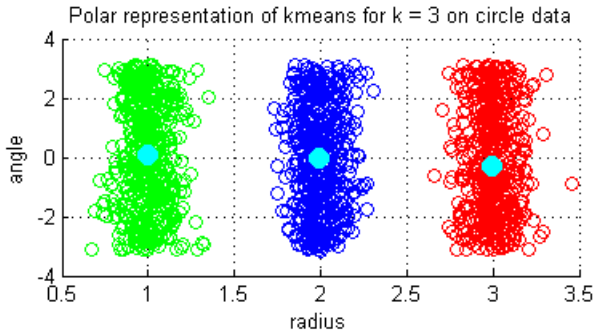


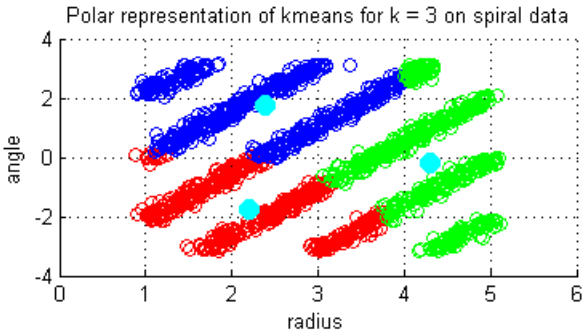


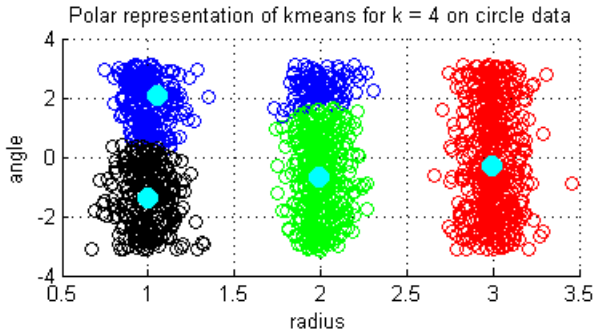
If we convert the data into polar coordinates, we see it is much easier for kmeans to classify the circle dataset, but still difficult to classify the spiral dataset.

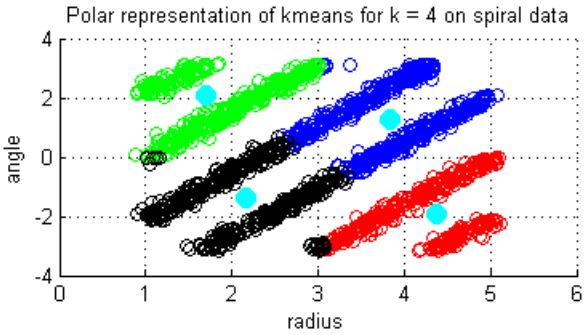












The cyan points in graphs indicate the centroids for each of the clusters. The overall sums of l2 distances between each point and the centroid for the clusters are:

cluster 1 (red) - circle: 904.57, spiral: 621.46
cluster 2 (blue) - circle: 742.63, spiral: 706.96
cluster 3 (green) - circle: 634.15, spiral: 539.10
cluster 4 (black) - circle: 236.26, spiral: 453.95

2 Spectral Clustering on Airbnb Data