Homework Set 6, CPSC 8420, Fall 2024

Last Name, First Name

Due 12/05/2024, 11:59PM EST

Problem 1

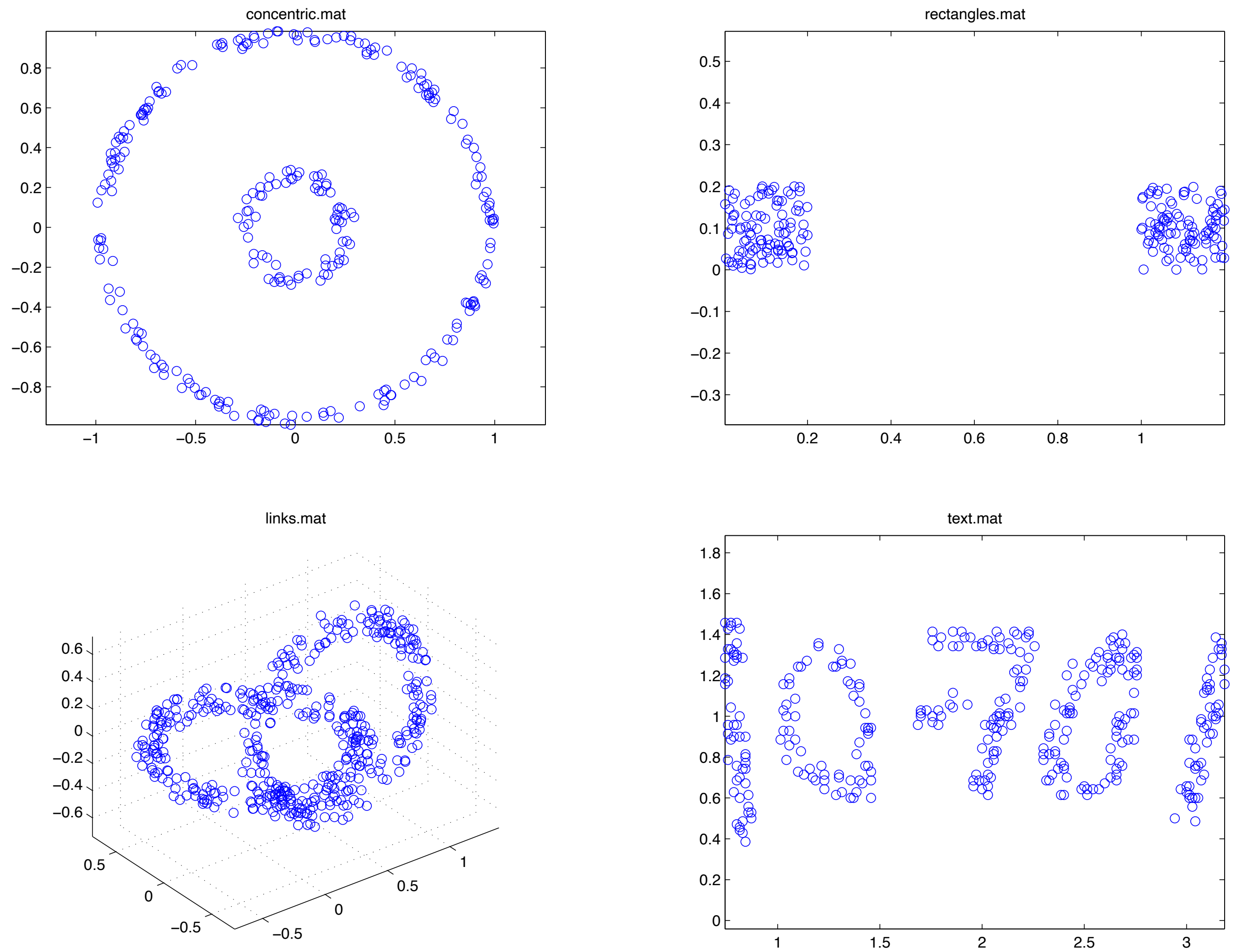
Frequently, the affinity matrix is constructed as:

$$A_{ij} = e^{-d(x_i,x_j)^2/\sigma} \tag{1}$$

where σ is some user-specified parameter. The best that we can hope for in practice is a near block-diagonal affinity matrix. It can be shown in this case, that after projecting to the space spanned by the top k eigenvectors, points which belong to the same block are close to each other in a euclidean sense. The steps are as follows:

- Construct an affinity matrix A using the above equation.
- Symmetrically 'normalize' the rows and columns of A to get a matrix N such that $N(i,j) = \frac{A(i,j)}{\sqrt{d(i)d(j)}}$, where $d(i) = \sum_k A(i,k)$.
- Construct a matrix Y whose columns are the first k eigenvectors of N.
- Normalize each row of Y such that it is of unit length.
- Cluster the dataset by running k-means on the set of embedded points, where each row of Y is a data-point.

1. Run k-means on the datasets provided in the .zip file. For text.mat, take $k = 6$. For all others use $k = 2$.
2. Implement the above spectral clustering algorithm and run it on the four provided datasets using the same k. Plot your clustering results using $\sigma = .025, .05, .2, .5$. Hints: You may find the MATLAB functions pdist and eig to be helpful. A function plotClusters.m has been provided to help visualize clustering results.
3. Plot the first 10 eigenvalues for the rectangles.mat and text.mat datasets when $\sigma = .05$. What do you notice?
4. How do k-means and spectral clustering compare?


concentric.mat
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−0.5
0
0.5
1
rectangles.mat
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
0.2
0.4
0.6
0.8
1
links.mat
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
0.5
0
−0.5
1
0.5
0
−0.5
text.mat
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1
1.5
2
2.5
3