STAT 202A Final Project

Xiaohan Wang

Problem 1. Spectral Clustering

(b) For each \sigma, plot the associated similarity matrix to visualize the amount of "overlap" among groups. What is the effect of the scaling parameter \sigma?

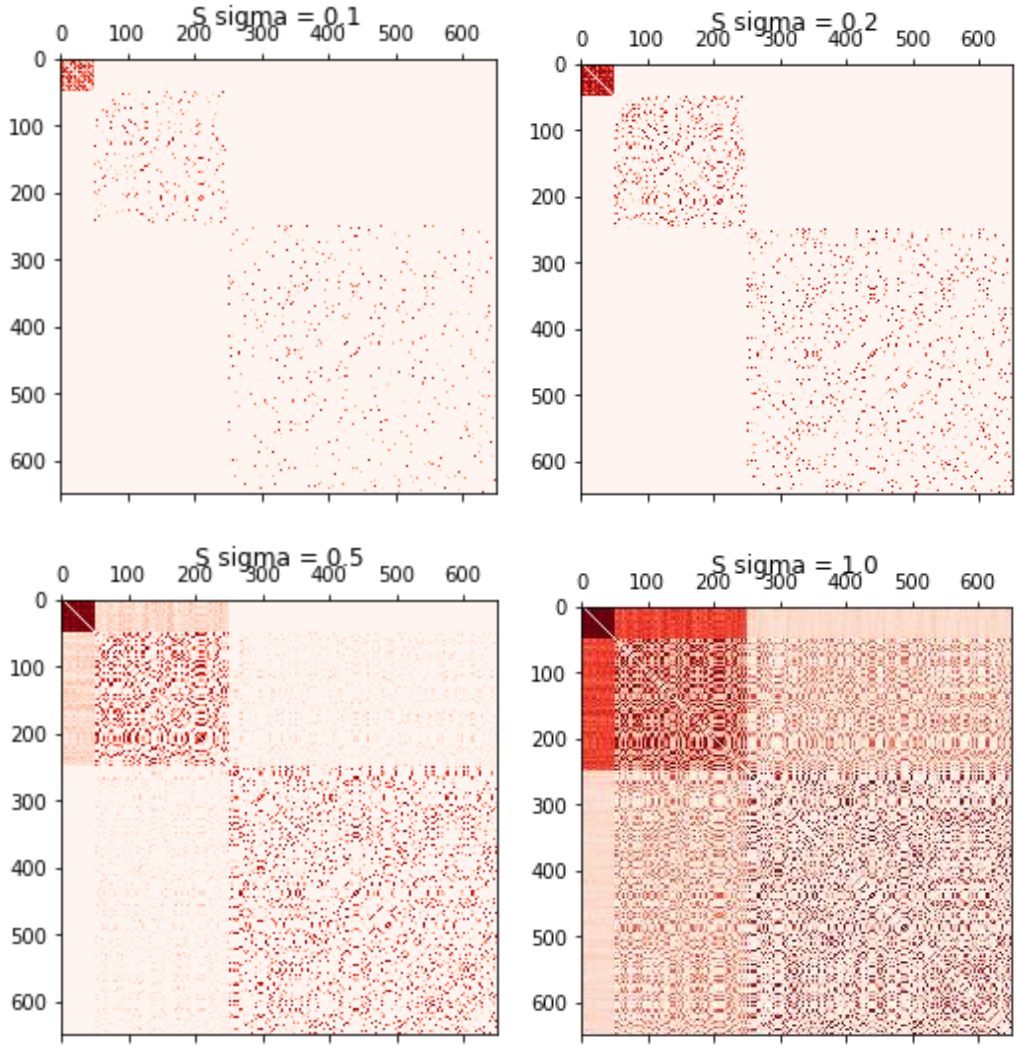
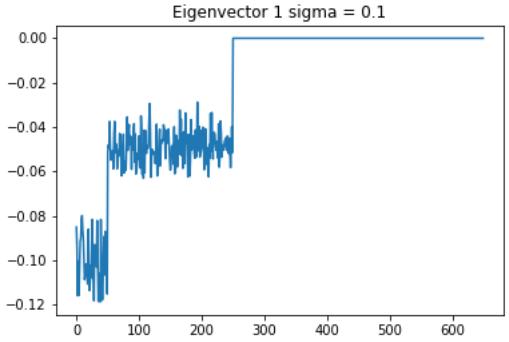


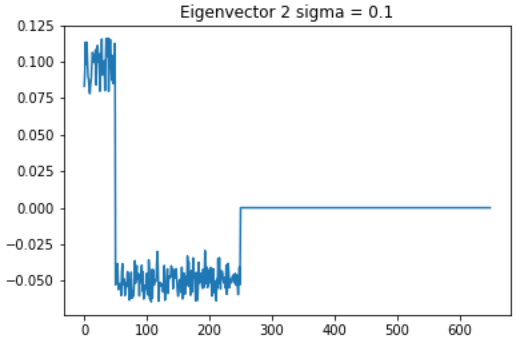
Figure 1. associated similarity matrices for sigmas

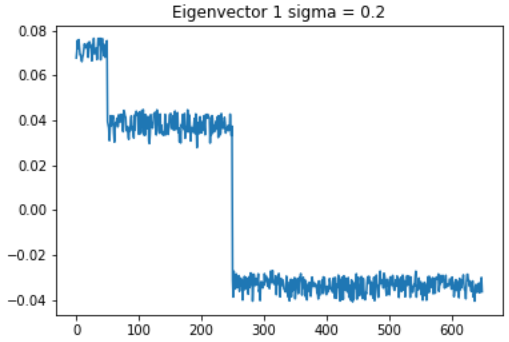
In the above figures, the deeper the color looks, the more similar X_i and X_j is. We can see a clear diagonal line in above figures, which fits with the formula that $A_{ii} = 0$. Compared these four figures, we find that as sigma increases, the color of the squares becomes deeper, and more overlapped with

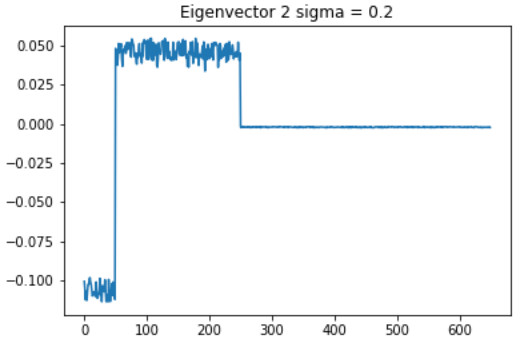
each other, because the number of "zeros" decreases. So by scaling parameter \sigma, we can increase the similarity between each pair of X_i and X_j.

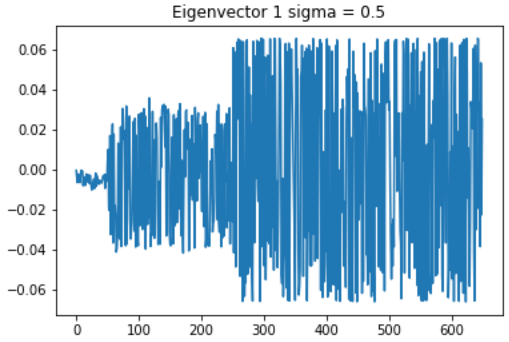
(c) Compute the degree matrix D and the Laplacian L. Find the k = 2 largest eigenvectors of L. Plot the components of the two largest eigenvectors for each \sigma = {0.1, 0.2, 0.5, 1}. Can you identify any group structure?

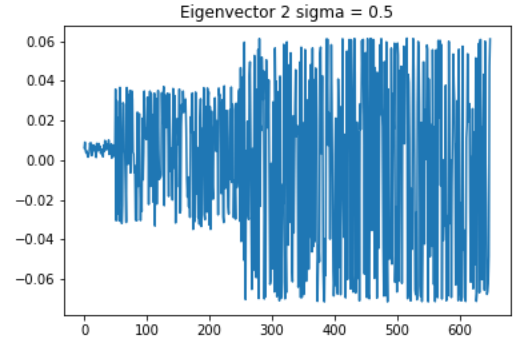












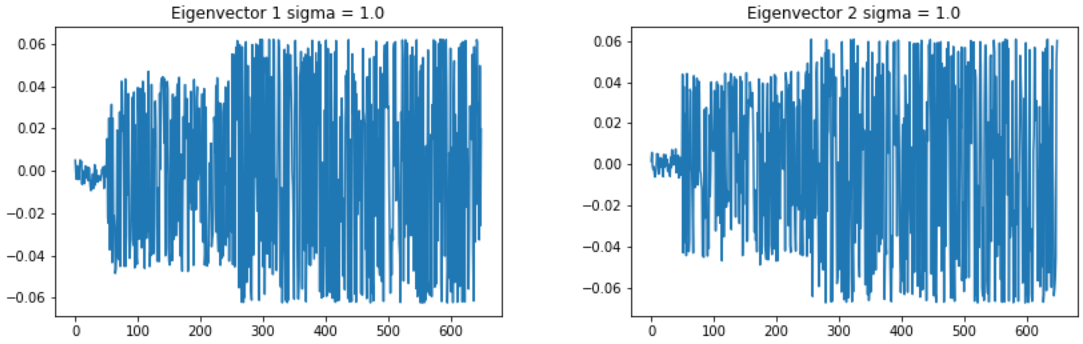
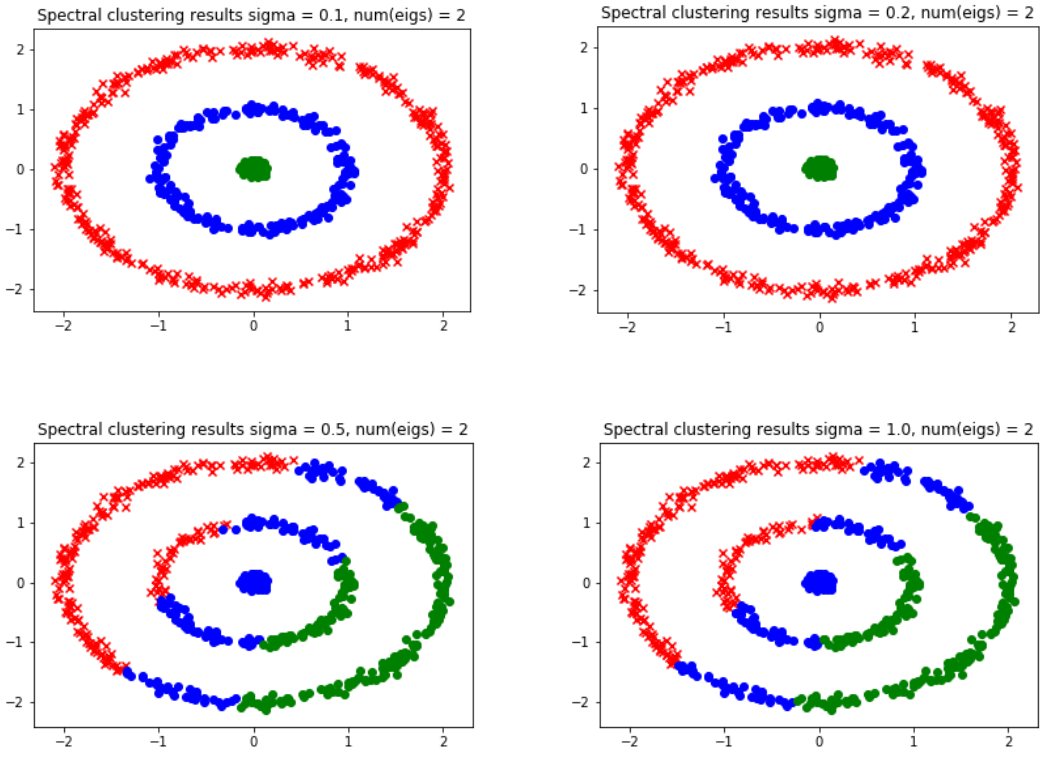
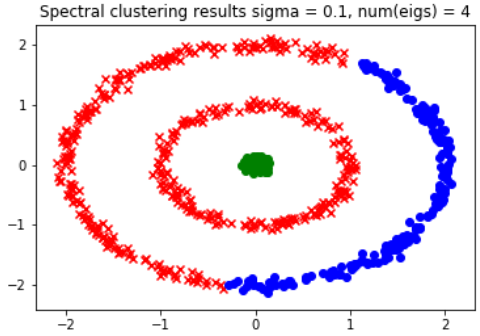


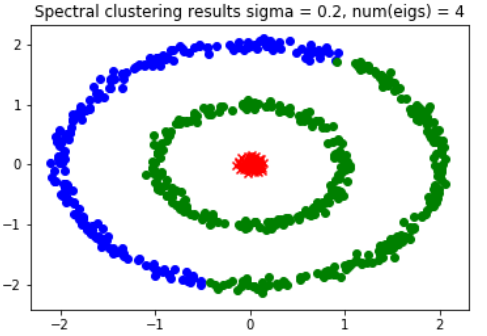
Figure 2. components of two largest eigenvectors for each sigma

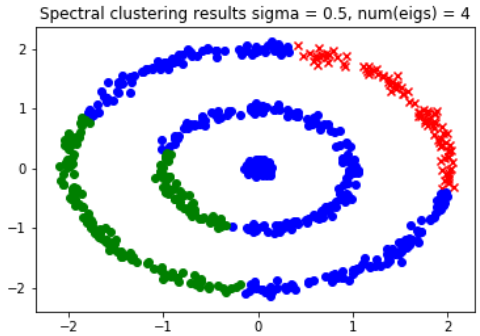
In each figure above, we can identify three groups, and components in the same group have close value to each other. The group structure is the same as the input X: group 1 – [1, 50], group 2 – [51, 250], group 3 – [251, 650].

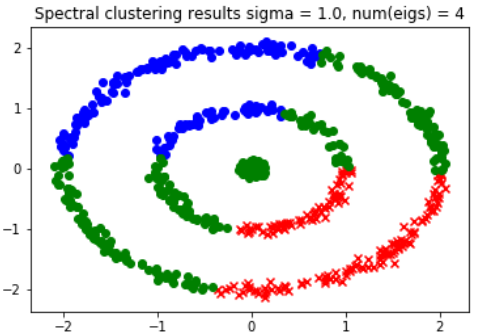
(d) For each Laplacian matrix L, run k-means (using random starting centroids) on the first k eigenvectors. Plot the cluster assignments.

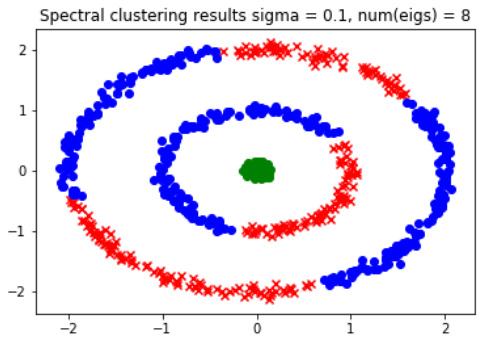


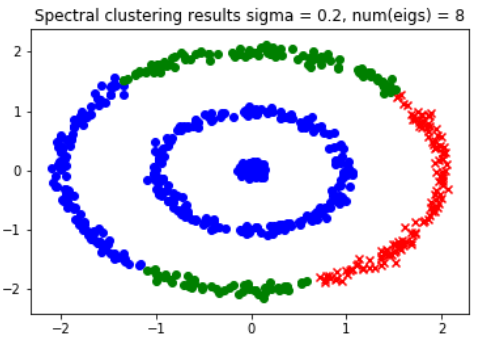

Spectral clustering results sigma = 0.1, num(eigs) = 4

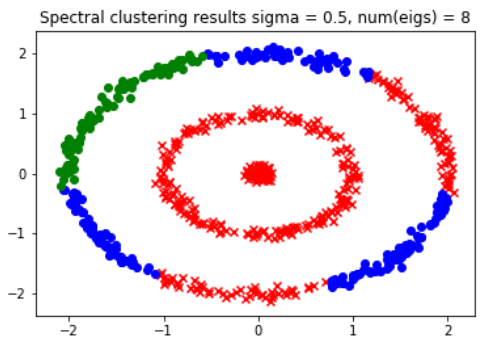

Spectral clustering results sigma = 0.2, num(eigs) = 4

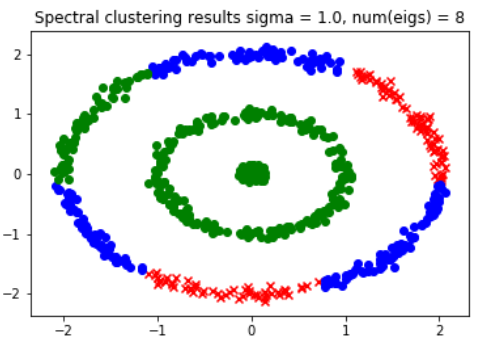

Spectral clustering results sigma = 0.5, num(eigs) = 4


Spectral clustering results sigma = 1.0, num(eigs) = 4


Spectral clustering results sigma = 0.1, num(eigs) = 8


Spectral clustering results sigma = 0.2, num(eigs) = 8


Spectral clustering results sigma = 0.5, num(eigs) = 8


Spectral clustering results sigma = 1.0, num(eigs) = 8

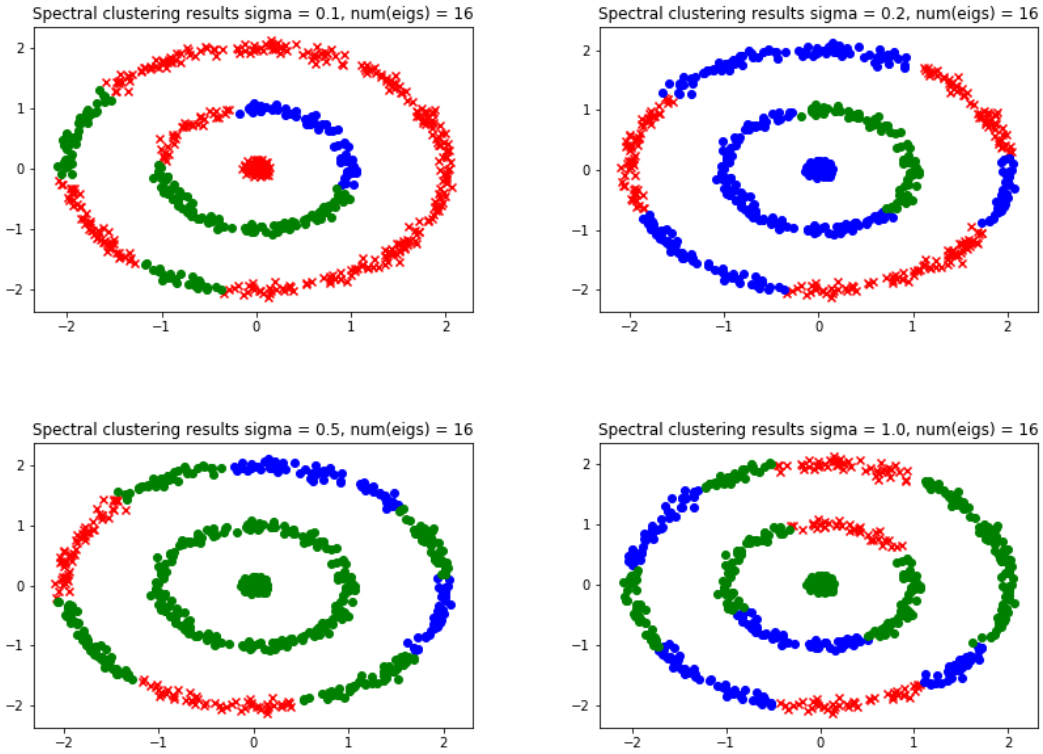


Figure 3. spectral clustering results for each sigma and number of eigens

(e) Report which choice of scaling \sigma and k eigenvectors lead to desirable clustering.
Compared results in figure 3, the best parameter combination is $\sigma = 0.1, \text{number of eigens} = 2$.

Problem 2. t-SNE

(b) Report a scatter-plot of the t-SNE result where color indicates the MNIST digit.

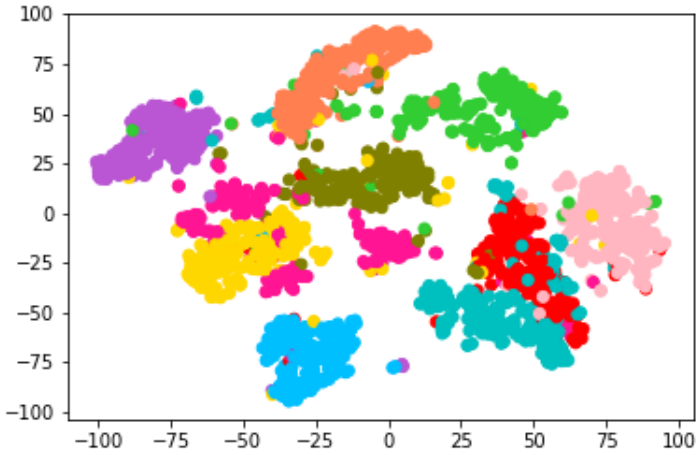


Figure 4. t-SNE result of MINIST dataset