

# STAT 202A Statistics Programming

Chaojie Feng

Final Project: Spectral Cluster and tSNE

December 16, 2018

## 1 Problem 1: Spectral Clustering

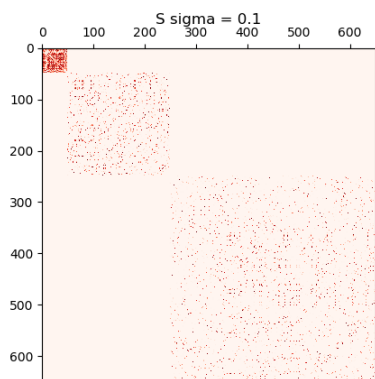
part(a): Similarity matrix with Gaussian Kernel is computed

part(b): From Figure 1, we can clearly see significant overlap as  $\sigma$  increases, suggesting with a higher scaling parameter, the relationship between data tends to extend farther.

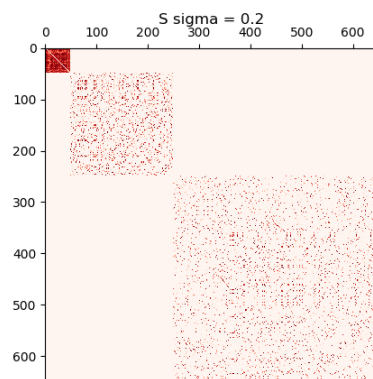
part(c): From Figure 2, we can determine that there are 3 groups of structure, by looking at small  $\sigma$ . Intuitively it makes sense because a small  $\sigma$  suggesting a closer relationship, and rejects all remote relationship.

part(d): Please refer to Figure 3 to Figure 6.

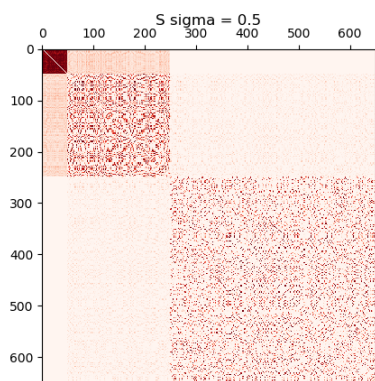
part(e): Although none of the choice of  $\sigma$  and  $k$  can truly represent the true clustering assignment, we can still see some patterns of implementing spectral clustering. For specific pair of  $\sigma$  and  $k$ . The best will be  $\sigma = 0.2$ ,  $k = 4$ .



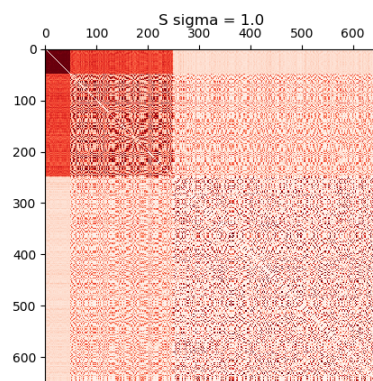
(a) Visualization of similarity matrix. Sigma = 0.1



(b) Visualization of similarity matrix. Sigma = 0.2

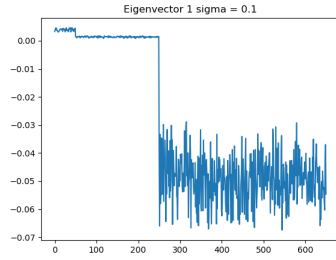


(c) Visualization of similarity matrix. Sigma = 0.5

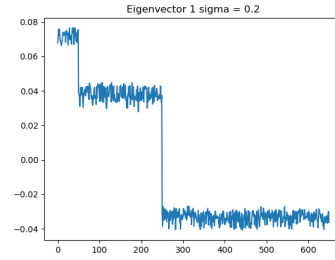


(d) Visualization of similarity matrix. Sigma = 1.0

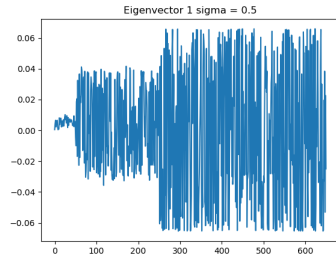
Figure 1: Comparison between similarity matrix using different Sigma



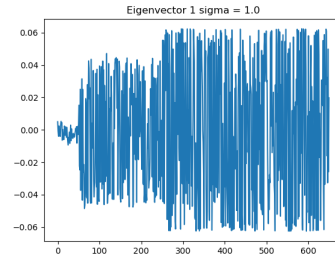
(a) Visualization of Eigenvector 1.  
Sigma = 0.1



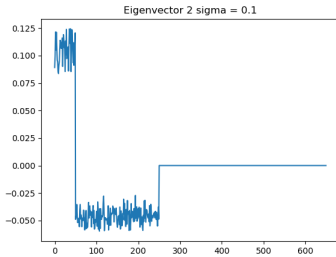
(b) Visualization of Eigenvector 1.  
Sigma = 0.2



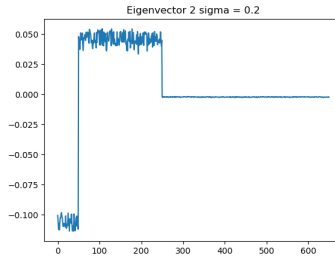
(c) Visualization of Eigenvector 1.  
Sigma = 0.5



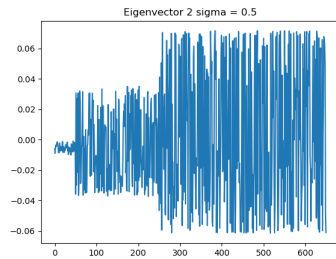
(d) Visualization of Eigenvector 1.  
Sigma = 1.0



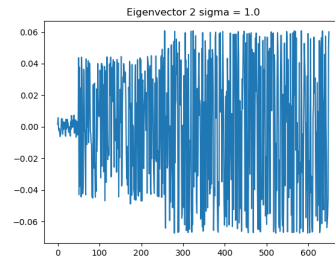
(e) Visualization of Eigenvector 2.  
Sigma = 0.1



(f) Visualization of Eigenvector 2.  
Sigma = 0.2

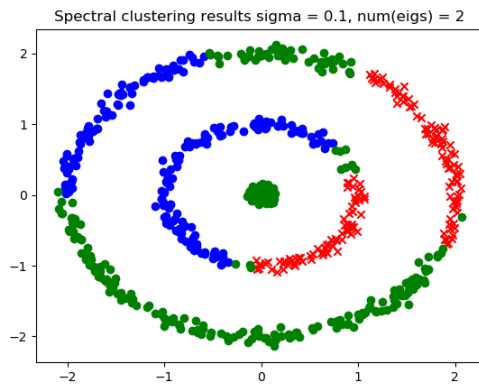


(g) Visualization of Eigenvector 2.  
Sigma = 0.5

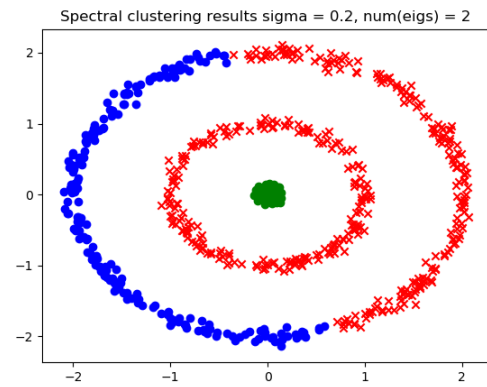


(h) Visualization of Eigenvector 2.  
Sigma = 1.0

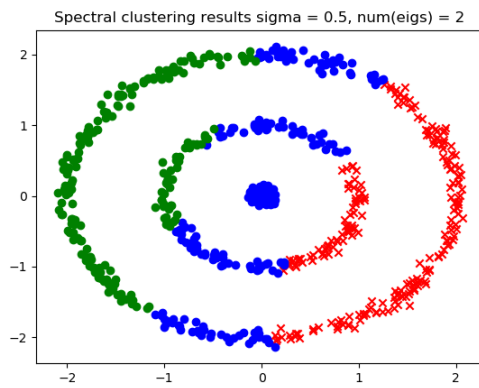
Figure 2: Eigenvectors with  $K = 2$



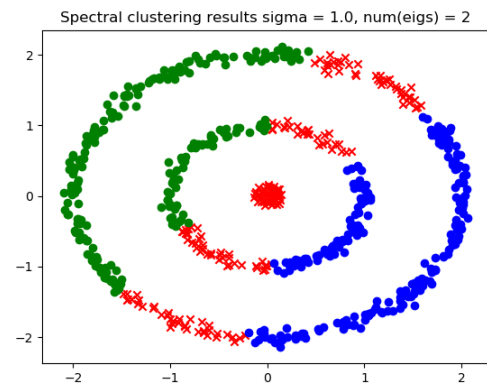
(a) Visualization of Clustering with Sigma = 0.1



(b) Visualization of Clustering with Sigma = 0.2

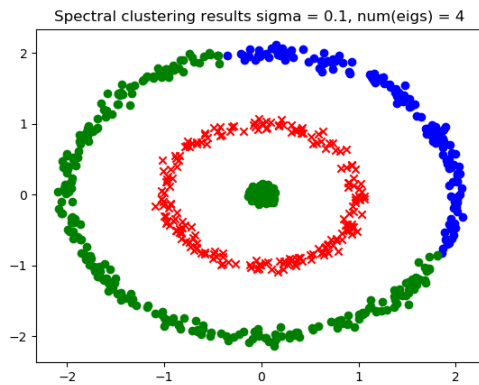


(c) Visualization of Clustering with Sigma = 0.5

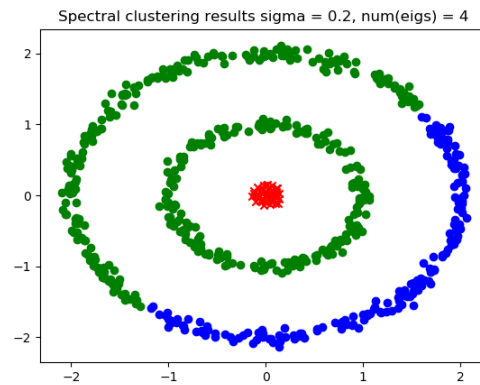


(d) Visualization of Clustering with Sigma = 1

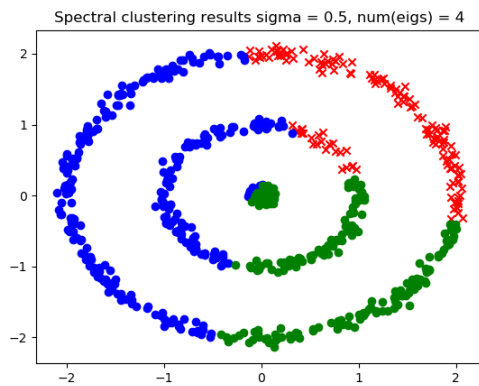
Figure 3: Visualization of Clustering with  $K = 1$



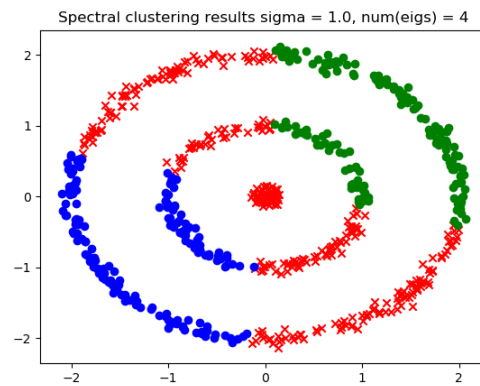
(a) Visualization of Clustering with Sigma = 0.1



(b) Visualization of Clustering with Sigma = 0.2

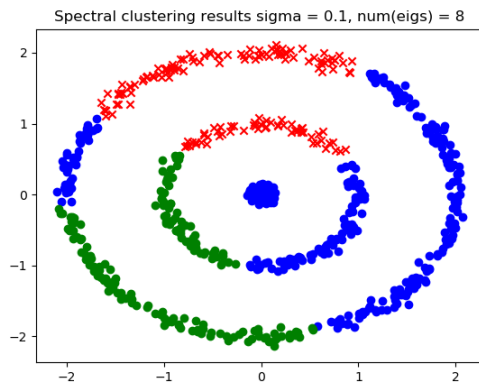


(c) Visualization of Clustering with Sigma = 0.5

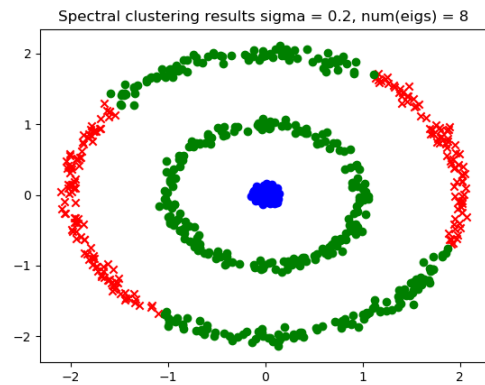


(d) Visualization of Clustering with Sigma = 1

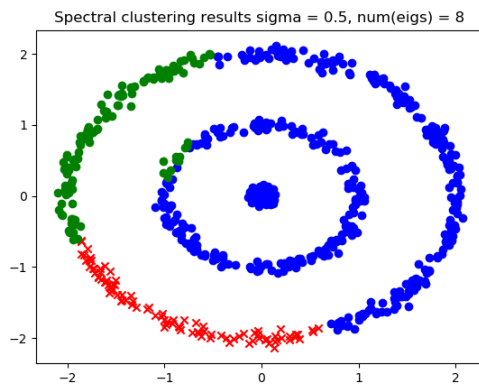
Figure 4: Visualization of Clustering with  $K = 2$



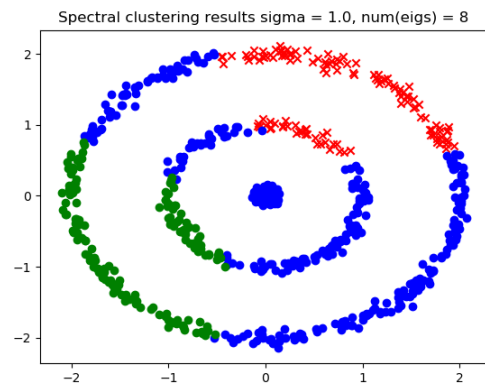
(a) Visualization of Clustering with Sigma = 0.1



(b) Visualization of Clustering with Sigma = 0.2

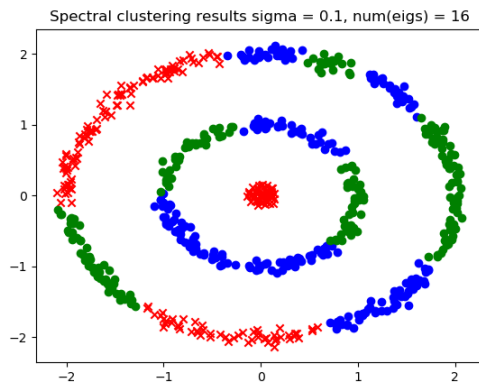


(c) Visualization of Clustering with Sigma = 0.5

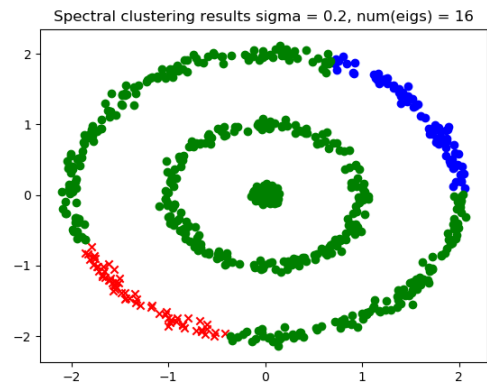


(d) Visualization of Clustering with Sigma = 1

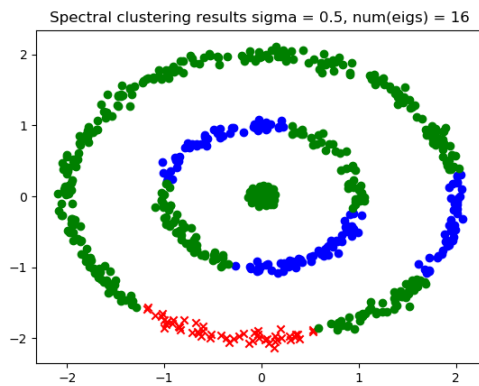
Figure 5: Visualization of Clustering with  $K = 4$



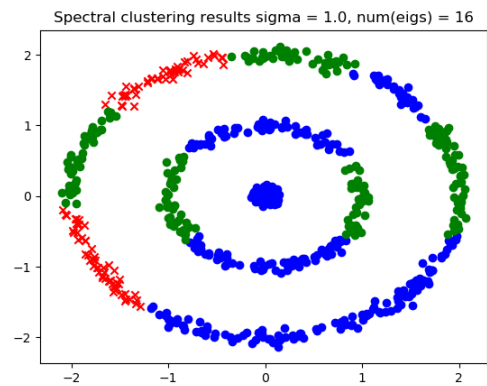
(a) Visualization of Clustering with Sigma = 0.1



(b) Visualization of Clustering with Sigma = 0.2



(c) Visualization of Clustering with Sigma = 0.5



(d) Visualization of Clustering with Sigma = 1

Figure 6: Visualization of Clustering with  $K = 8$

## 2 Problem 2: tSNE

part(a): Please refer to code.

part(b): The MNIST digits successfully clustered to different labels by using tSNE.

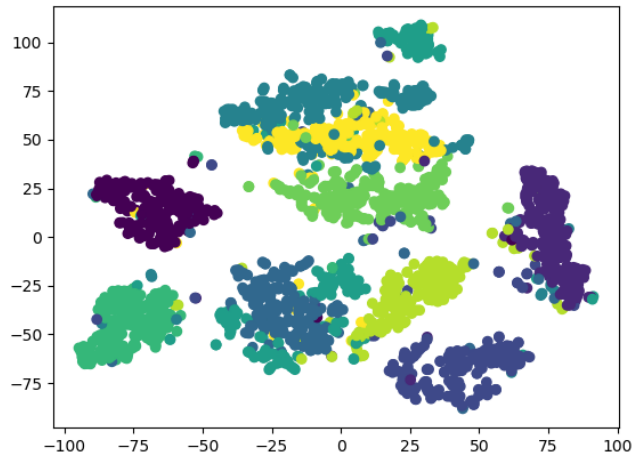


Figure 7: Clustering for MNIST digits