Exercise 3

Hou, Jue

014695647

1. Mixture model for binary data

Observed data log-likelihood:

Complete data log-likelihood:

Expected complete data log-likelihood:

E-step:

M-step:

After derivation with respect to and , we can get following two equation:

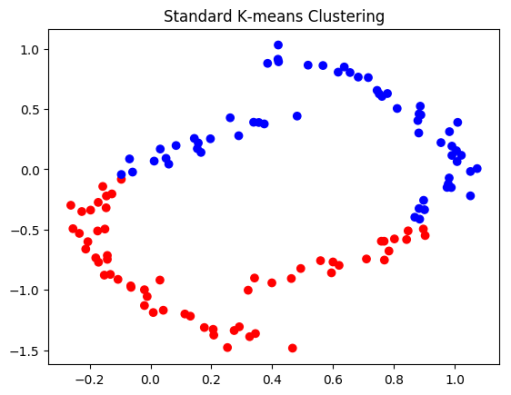
,

After derivation with respect to ,

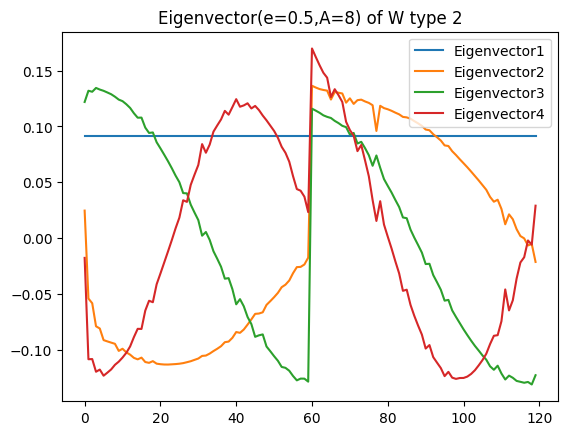
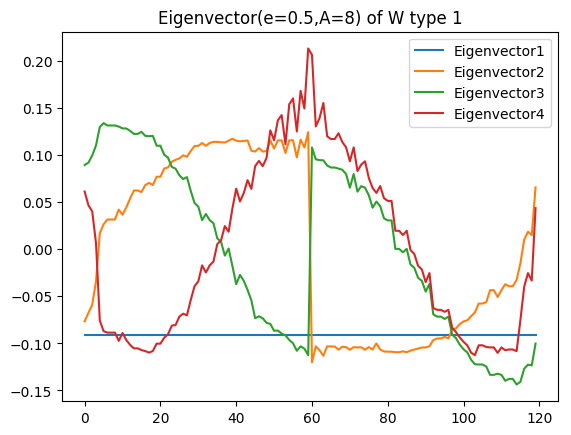
-

1. Spectral clustering

The following plot shows the result of standard K-means Clustering.

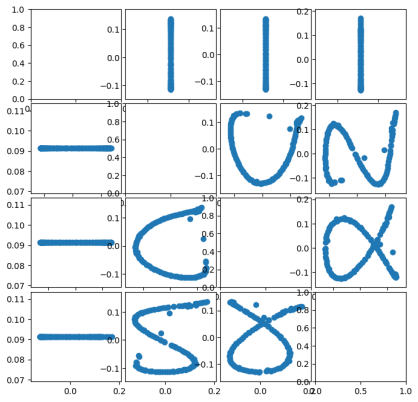
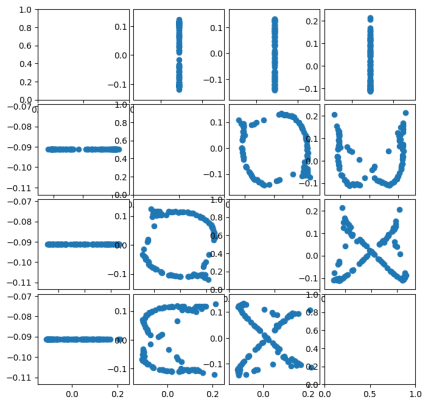


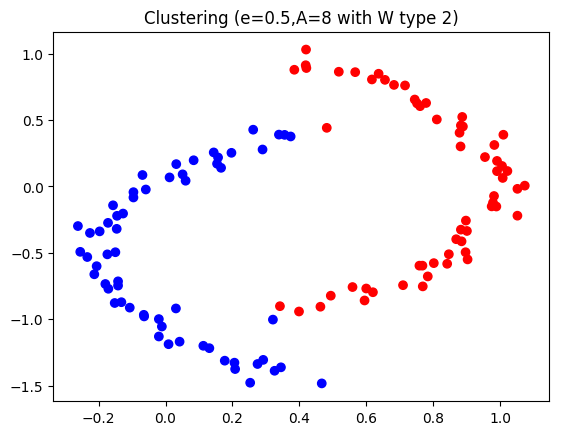
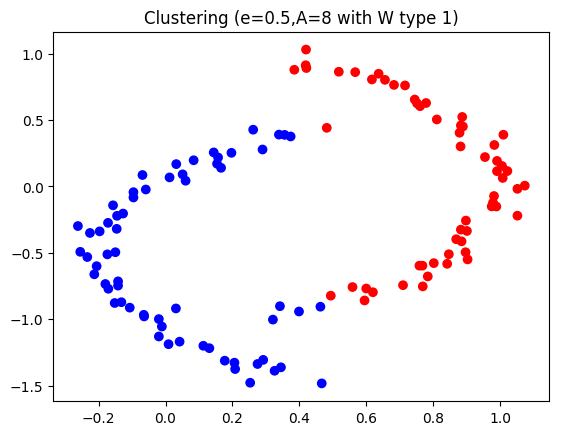
The following plot shows how the eigenvectors of both type of adjacency matrices as lines.



It is noticeable that both plot show a turn at point 60.

The scatter matrices are showed as follow. They present the scatters of eigenvectors of adjacency matrices I and II respectively.





These two plots above show clustering for adjacency matrices I and II respectively.

Compared to the one based on the original data, the one based on original data cannot cluster the half-circle data, while the other two can. These two clustering approaches obviously have a better performance.

If e is too small or too big, because of the limitation of K=2, most of data point will be clustered into one cluster. Only a small amount of data will be clustered into another cluster.

If A is too small or too big, just as the consequence of small e or big e, most of data point will be clustered into one cluster. Only several data points will be clustered into another cluster.

When M=2, the performance of both approaches is almost as good as it is when M=4.

If M is too big, the majority of data points are clustered into one group as well. The spectral clustering is based on the fact of only the eigenvectors that corresponds to first several smallest eigenvalues can reveal the cluster. Once there are too many eigenvectors is selected, the boundary between clusters becomes unreliable and the algorithm will be compromised. Therefore, an M of too big can lead to misclustering.

1. Deflationary orthogonalization

Therefore, W is orthonormal.

Estimated complete time :5 hours