4.2

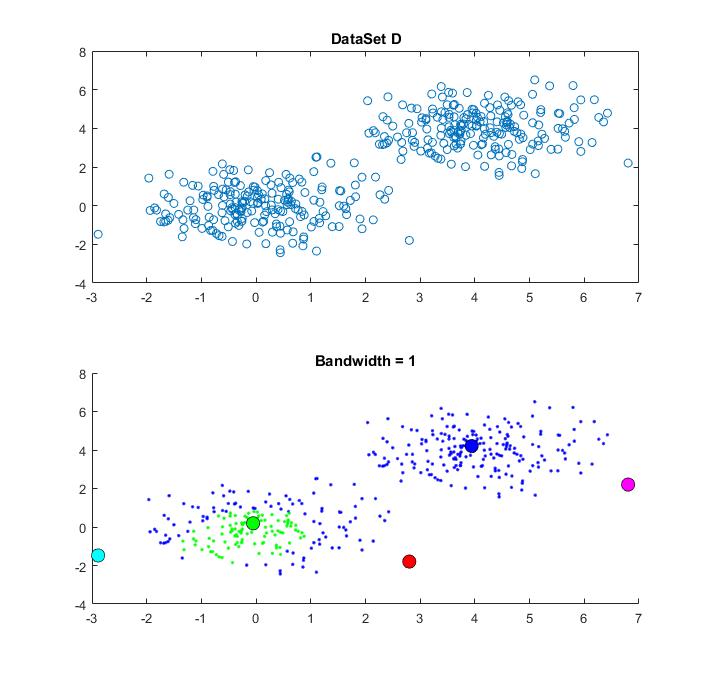
**Presetting the radius of flat kernel function to be radius(bandwidth) = 2**

Matlab Code:

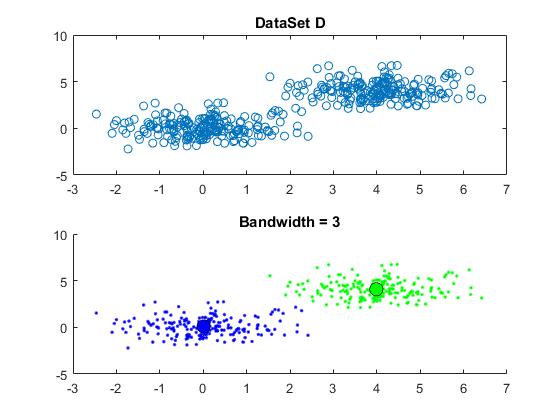




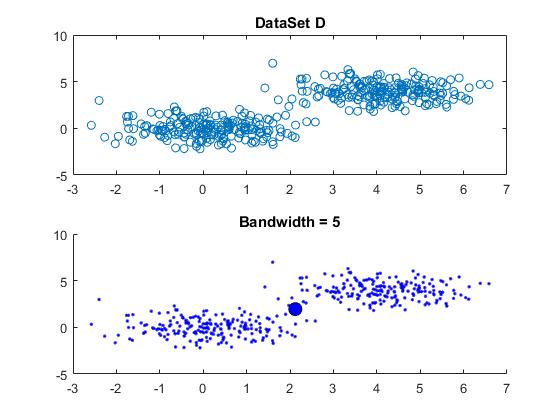
4.3

(1). Dataset D, choose radius(bandwidth) value r = 1, 3, 5, we get the result as

When bandwidth = 1, it can be seen that the number of clusters generated is 5, which is not we expected. The issue is that the dataset is consist of two multivariate Gaussian models with different mean value, which tends to be two clusters intuitively, however, the radius of kernel function is too small so that it cannot cover all the data points on the figure, which results in the extra points (especially edging point) will form a new cluster with very few points.

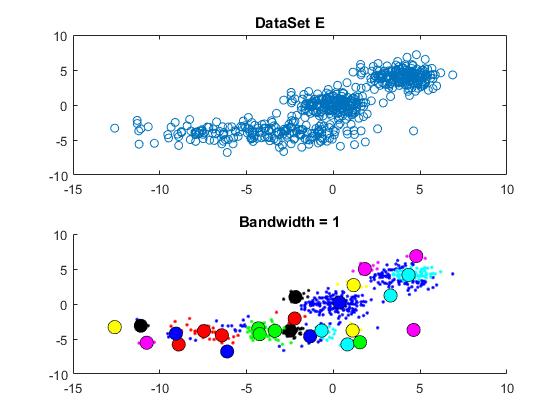


When bandwidth = 3, it can be seen that the number of clusters generated is 2. This model performs well under this bandwidth.

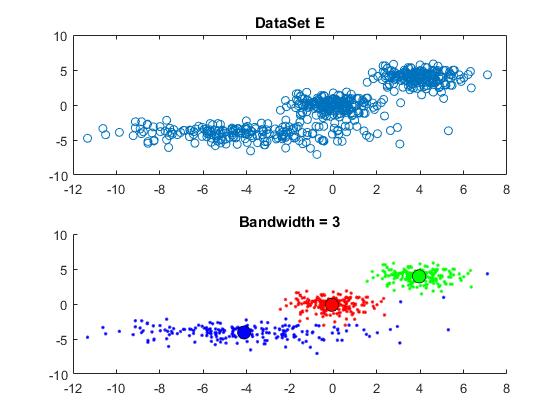


When bandwidth = 5, it can be seen that the number of clusters generated is 1, which is too few in this case. That’s because the bandwidth of the kernel function is too large, which makes the initial cluster cover all the data points so there are no points out of range of first cluster and the loop end.

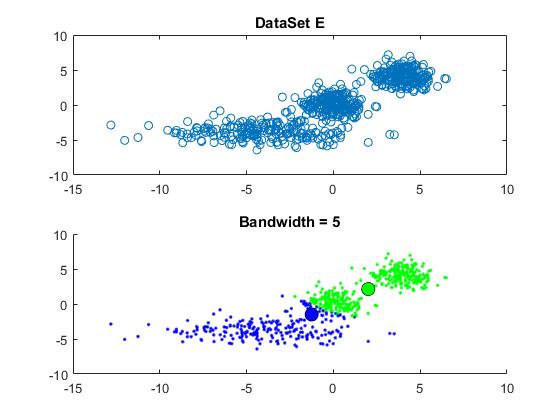
(2). Same as dataset d, now we try to modelling dataset e by the same radius (bandwidth) r = 1, 3, 5



Firstly, we can see the dataset E is consisting mainly from three regions, which expecting to see the number of clusters to be 3. When bandwidth = 1, it can be seen that the model generating a lots of clusters, and each one of them only representing few data points, which is due to the smaller value of bandwidth and it’s not we expected.



When bandwidth is equal to three, it can be seen that three clusters have been generated, and each of them have a clear centre which represent the most density part in that cluster. It is a good model to representing this dataset since we expected the clusters numbers to be three intuitively.



When bandwidth is equal to five, it can be seen there are only two clusters in this case, which is not we expected since it does not successfully distinguish the top two clusters, and the reason for that is the bandwidth of kernel function are too wide, same as dataset d.

Attached: mean shift algorithm test code



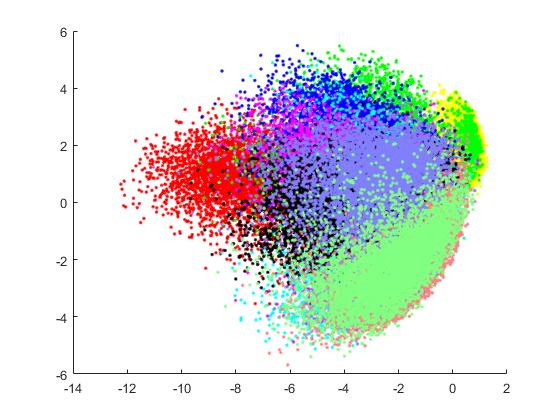
5.1

Principal Component Analysis with reduced dimension d = 2:

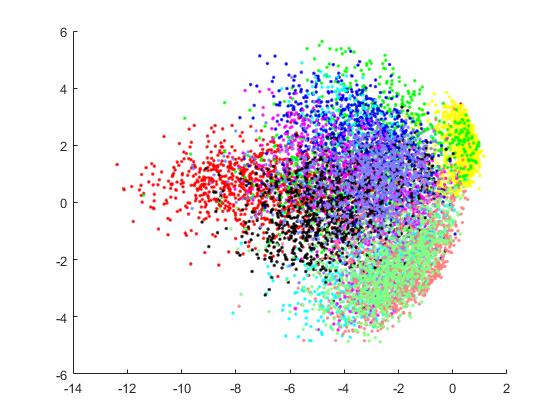


5.2

(1). The reduced data plotted in a two-dimension as shows (mnist\_train)



Testing the PCA with mnist\_test, we can get reduced data plotted as below.

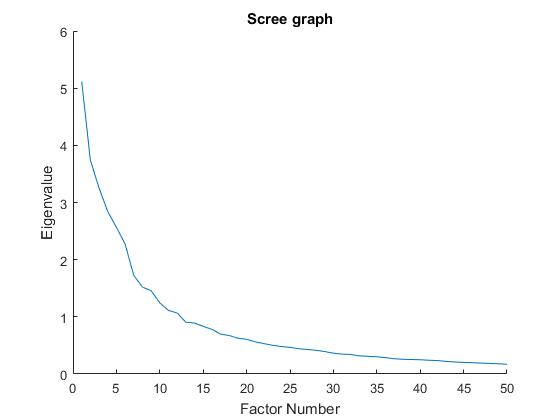


(2). Based on the output for the percentage of eigenvalue, the proportion of variance is calculated as 0.1680 (mnist\_train).

Testing under mnist\_test, the proportion of variance is 0.1759.

(3). Scree graph: Since after 50 factors, the eigenvalue will significantly tends to be 0, so in here it only shows the first maximum eigenvalue to the fiftieth maximum one.

Mnist\_train scree graph:



Mnist\_test scree graph:

