Data Mining CSE-5334-001

Assignment 1- Report

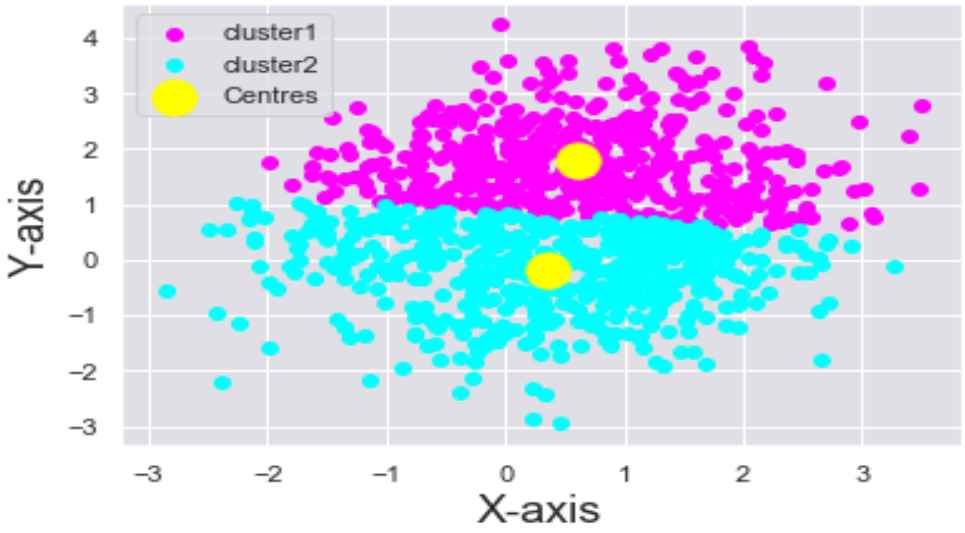
Name: Swati Sriram Mani

UTA Id: 1001648136

1) KMeans Algorithm:

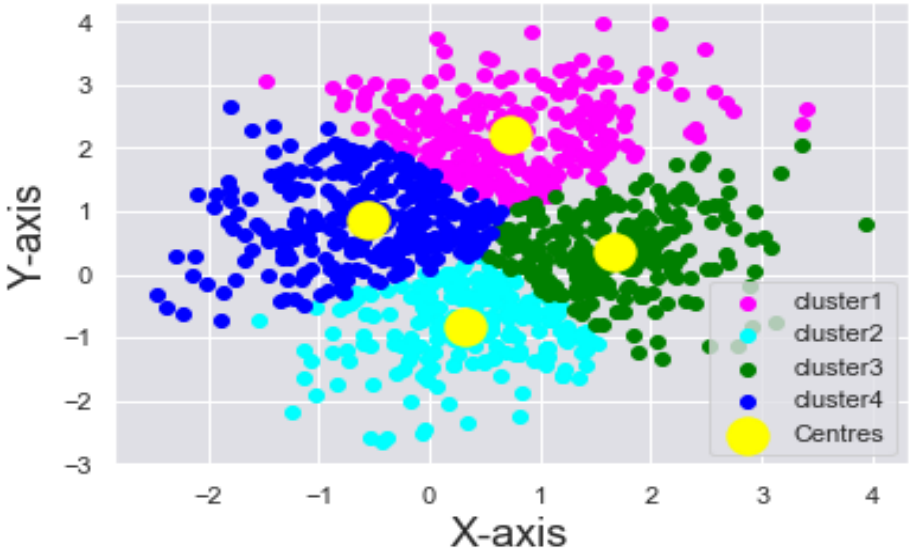
- I have generated the data randomly using the np.random.multivariate() and using the mean and the covariances provided in the question.
- The initial centers are taken from the user and stored into c.
- The data, number of Clusters(K) and centers are sent into the function mykmeans(X,K,c)
- The Euclidean Distance is calculated and then minimum of the values is stored into the C
- Another variable c_old is maintained, so that it is subtracted with the current centers and if the difference is less than 0.001 then exit from the loop.
- The number of iterations are counted using the count variable.
- Finally output is printed.

Centers for K=2(centers:((10,10),(-10 -10))



Iteration: 20

Centers for K=4: (centers: (10,10),(-10,-10),(10,-10),(-10,10))



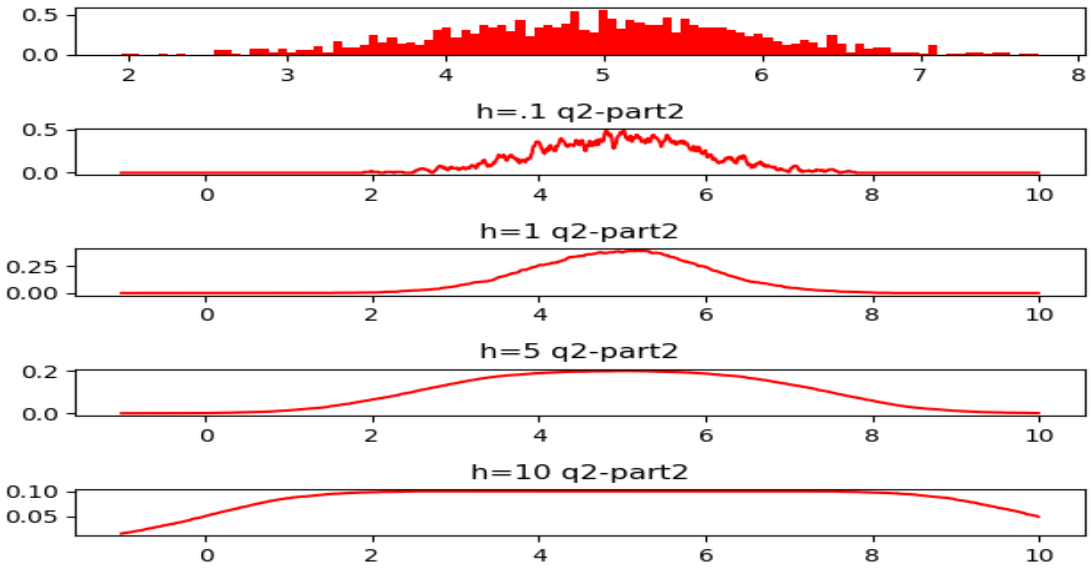
Iteration: 22

2.Kernel density Estimation Algorithm:

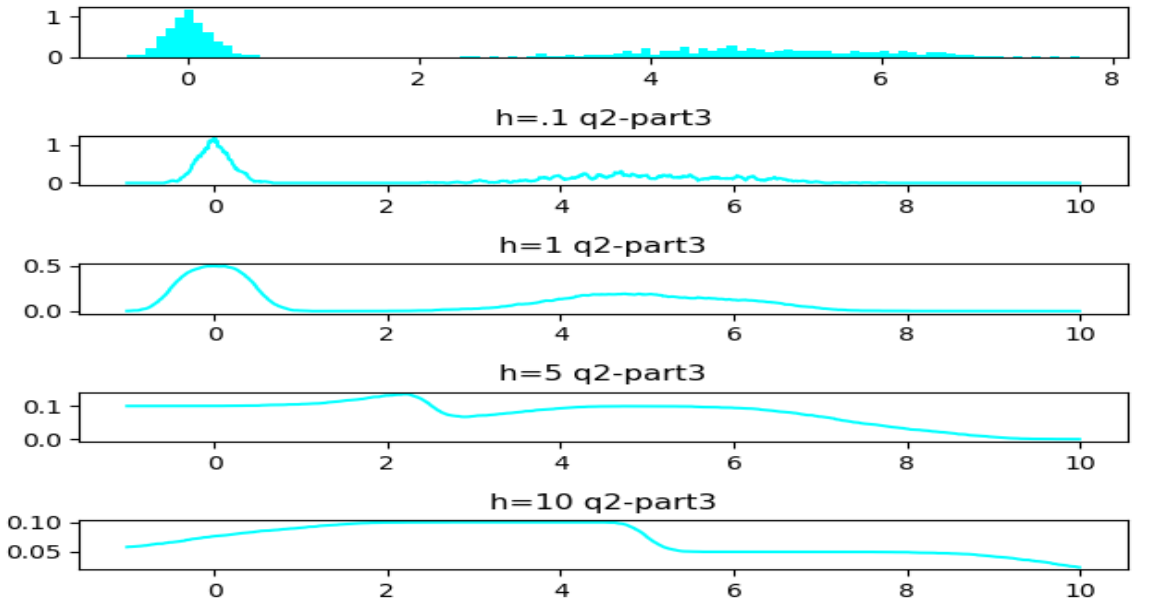
- I have generated the data using the np.random.normal() function
- I have implemented the formula for the kernel density estimation function using Parzen window function.
- For the 1-D, the value of d in the function is taken as 1, For 2-D, the value of d in the function is taken as 2 and further calculations are done.
- The values of probability are calculated for each and every point.

Results:

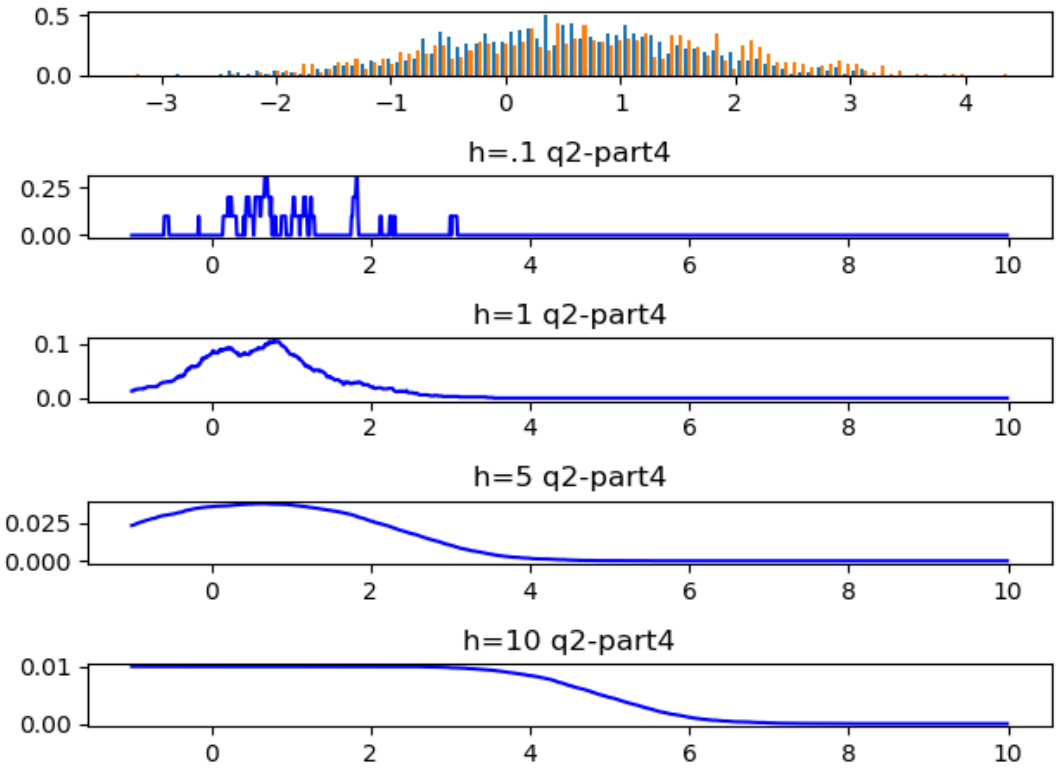
Part 2:



Part 3:



Part 4:


0.5
0.0
−3
−2
−1
0
1
2
3
4
h=.1 q2-part4
0.25
0.00
0
2
4
6
8
10
h=1 q2-part4
0.1
0.0
0
2
4
6
8
10
h=5 q2-part4
0.025
0.000
0
2
4
6
8
10
h=10 q2-part4
0.01
0.00
0
2
4
6
8
10