**CSE 5334 DM**

**Homework-1**

Name: Sagar Sharma

UTA ID: 1001626958

**Note: I will be using my one late day quota for this submission.**

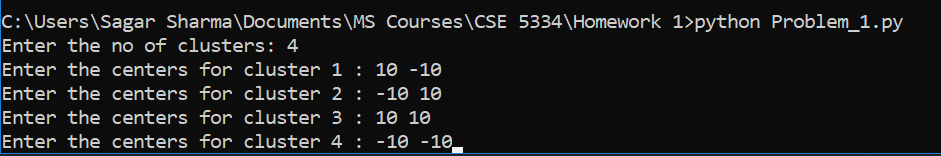
Programming Language Used: Python 3

Note: Might need to install packages like NumPy and matplotlib, if not already installed in order to successfully run the programs for both the problems.

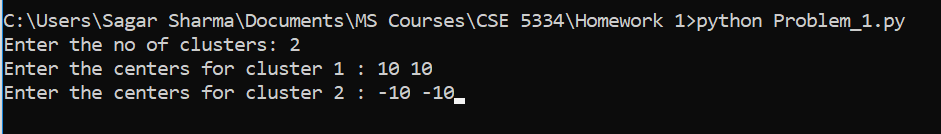
**Problem 1:**

*Part 1:* This zip file includes a python file named Problem1.py which consist the implementation for k-means algorithm according to the stopping conditions provided in the homework description. While running the program user needs to input the value for k which is the no of clusters we want to build. Also input the seed center values for the k clusters, when prompted to do so.

Snapshot demonstrating how to run the program from command line and the format in which the values for k and centers needs to be provided as well.



*Part 2:*

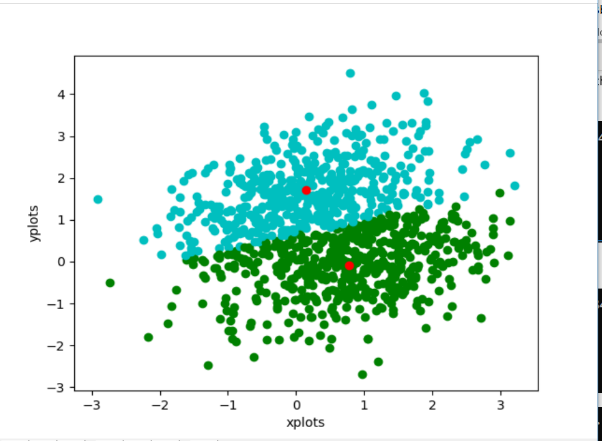


Centers found for both the clusters:

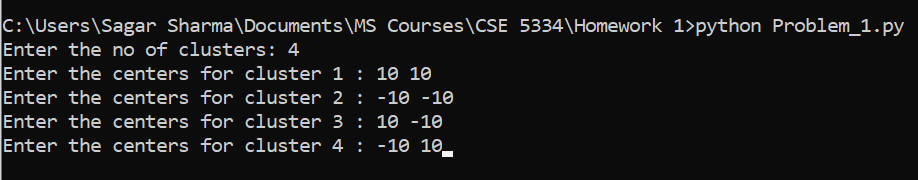


It took the algo **23 Iterations** to converge.

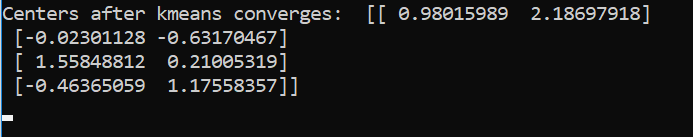
The snapshot of the scatterplot for the part 2 of the Problem 1 depicting the centers for the 2 clusters in red.



*Part 3:*

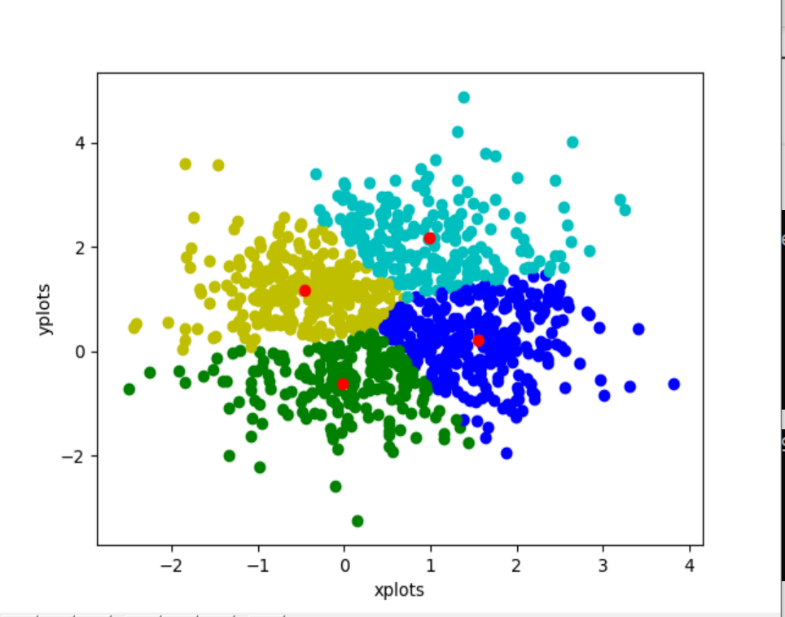
**

Centers found for all the 4 clusters.

**

*It took the algo* ***20 Iterations*** *to converge.*

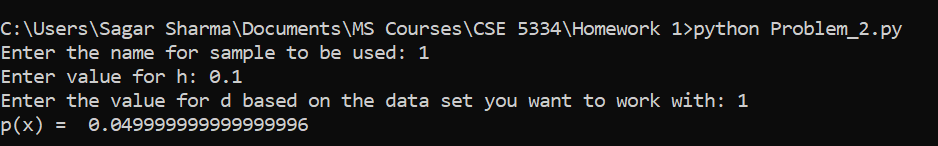
The snapshot of the scatterplot for the part 3 of the Problem 1 depicting the centers for the 4 clusters in red.

**

**Problem 2:**

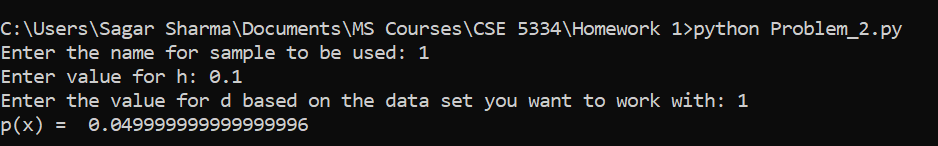
*Part 1:* This zip file includes a python file named Problem2.py which consist the implementation for KDE using Parzen window estimation based on the parameters asked in the homework description. While running the program user needs to input the value for data set which has options 1, 2 and 3 corresponding to the data sets generated for the part 2, part 3 and part 4 of this problem respectively. Also input the h values as desired (can use the values given in the different parts of this problem), also provide the d(dimension) value according to the data set used for the different parts of the problem when prompted to do so.

Snapshot demonstrating how to run the program from command line and the format in which the values for data set, h and d needs to be provided as well.

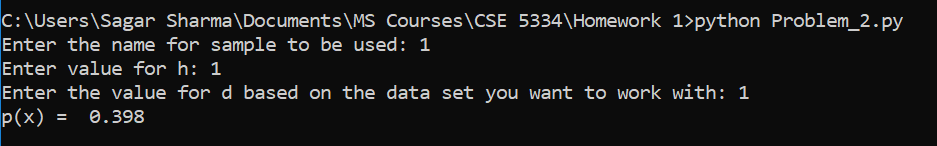


Part 1:

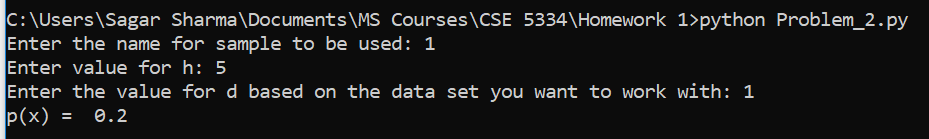
h = 0.1



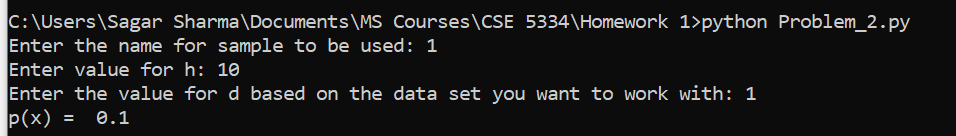
h = 1



h = 5

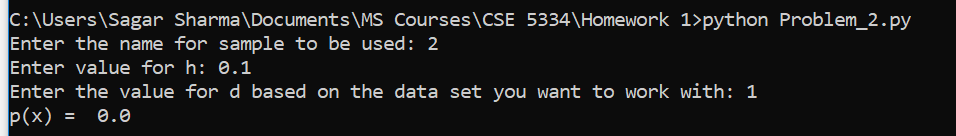


H = 10

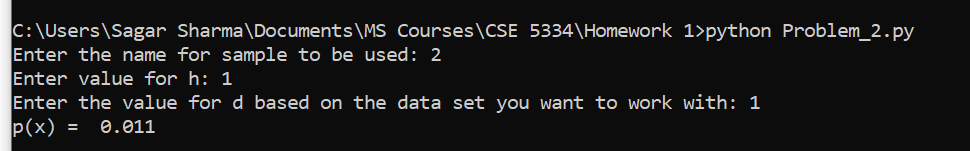


Part 3:

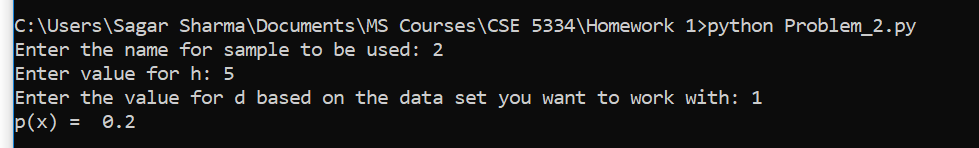
H = 0.1



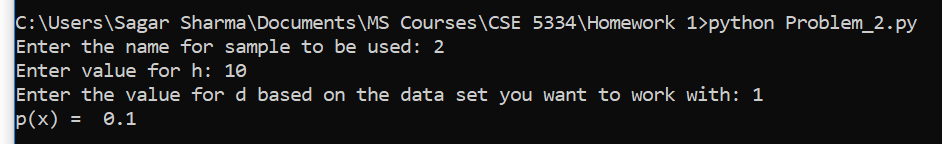
H = 1



H =5

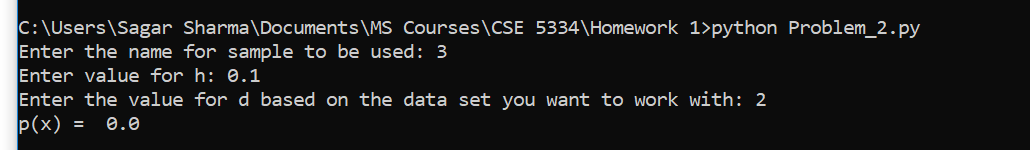


H = 10

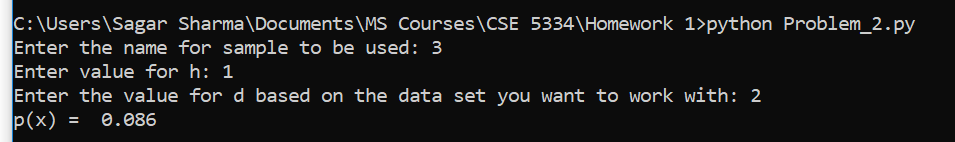


Part 4:

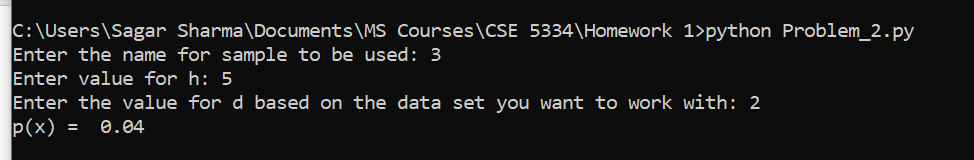
H = 0.1



H = 1



H = 5



H = 10

