Name - Rashmi Soni

Homework 5

CS 57300

1) Python Version used - 3.6

2) For the value of silhouette coefficient, B is calculated from the nearest neighbor cluster which is determined from the mean distance from all the points in the cluster.

3) For the value of NMI, the formula used is the one given in the class slides, i.e the range of NMI value will be from 0 to 0.5

4) Cdist function is used to calculate the Euclidean distance between two points.

6) If the graph for question 1 (exploration) needs to be plotted, change the value of variable 'ques' at line No.400 to 11 and 12 accordingly, (i.e. ques=1). And update the training file name accordingly at line No. 402 and 410 respectively.

7) If the graph for analysis questions and bonus needs to be plotted, change the value of variable ‘analysisQues’ at line No.455 according to each question. And update the training file name accordingly.