DATA MINING – CSE 5334

PROJECT 1 - KNN

NAME: KARTHIK KUMARASUBRAMANIAN
STUDENT ID: 1001549999

1. The accuracy for knn classifier is given as,

   k = 1, 5, 10, 30

```
Enter the number of neighbors : 1
The accuracy for the data set is given as:  0.46

In [75]: runfile('C:/Users/Karthik Subramanian/Documents/
Data Mining/Project 1/KNN/q1')


Enter the number of neighbors : 5
The accuracy for the data set is given as:  0.49

In [76]: runfile('C:/Users/Karthik Subramanian/Documents/
Data Mining/Project 1/KNN/q1')


Enter the number of neighbors : 10
The accuracy for the data set is given as:  0.51

In [77]: runfile('C:/Users/Karthik Subramanian/Documents/
Data Mining/Project 1/KNN/q1')


Enter the number of neighbors : 30
The accuracy for the data set is given as:  0.46
```

OBSERVATION:

The model classifier has lower accuracy because of the underfitting problem. The test and train data provided is very low to design a classifier. The solution would be to use k-fold cross validation in-order to increase the number of data points.

2. The accuracy for knn classifier for problem 3 is given as,

```
Enter the number of neighbors : 1
The accuracy for the data set is given as:  0.45

In [83]: runfile('C:/Users/Karthik Subramanian/Documents/Classes/
Data Mining/Project 1/KNN/q3')


Enter the number of neighbors : 5
The accuracy for the data set is given as:  0.58

In [84]: runfile('C:/Users/Karthik Subramanian/Documents/Classes/
Data Mining/Project 1/KNN/q3')


Enter the number of neighbors : 10
The accuracy for the data set is given as:  0.55

In [85]: runfile('C:/Users/Karthik Subramanian/Documents/Classes/
Data Mining/Project 1/KNN/q3')


Enter the number of neighbors : 30
The accuracy for the data set is given as:  0.56
```

OBSERVATION:

The model classifier has lower accuracy because of the underfitting problem. The test and train data provided is very low to design a classifier. The solution would be to use k-fold cross validation in-order to increase the number of data points.