AI Assignment – 2 Report

Anurag Gupta – 201601006

Garvit Kataria – 201601028

OUTLINE:

Task_1: Find k to be used in k-nearest neighbor classifier (k-NNC) using a 3-fold cross validation. Give the results in your report. You can draw a plot and with k value thus found employ k-NNC and find the error rate.

Task_2: Employ the Naive Bayes classifier with the given data set. You can use log of the posterior (to overcome the small numbers problem).

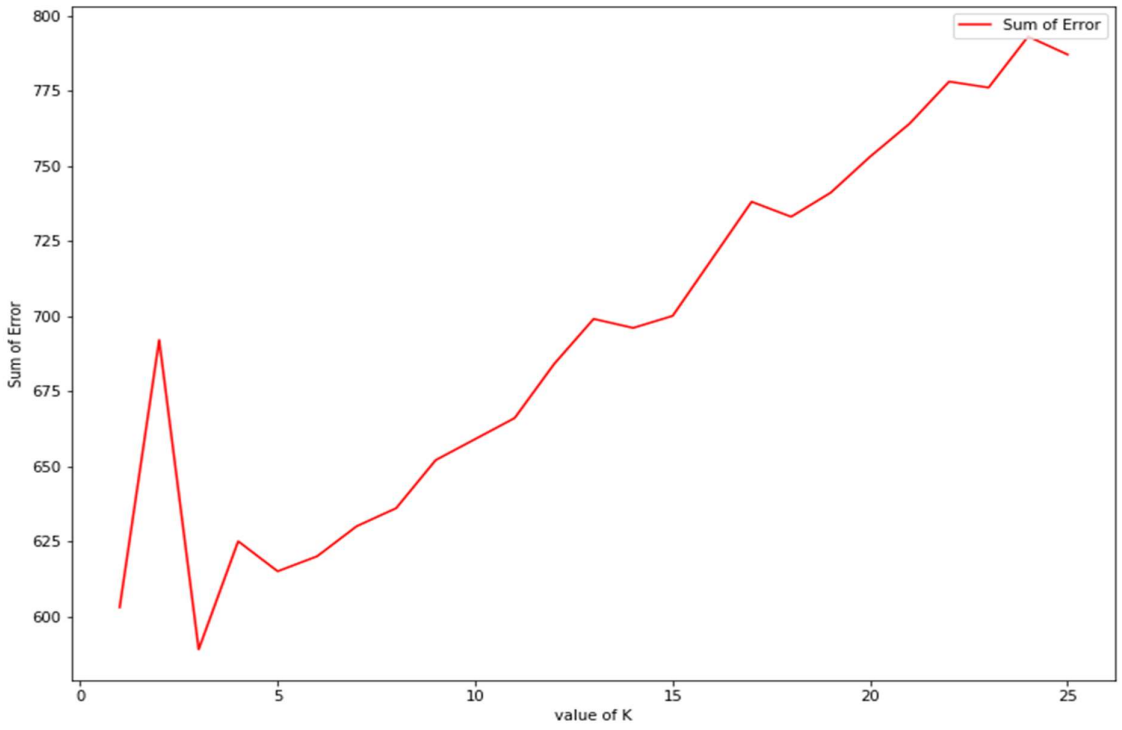
ANALYSIS:

For task 1(KNN Classifier):

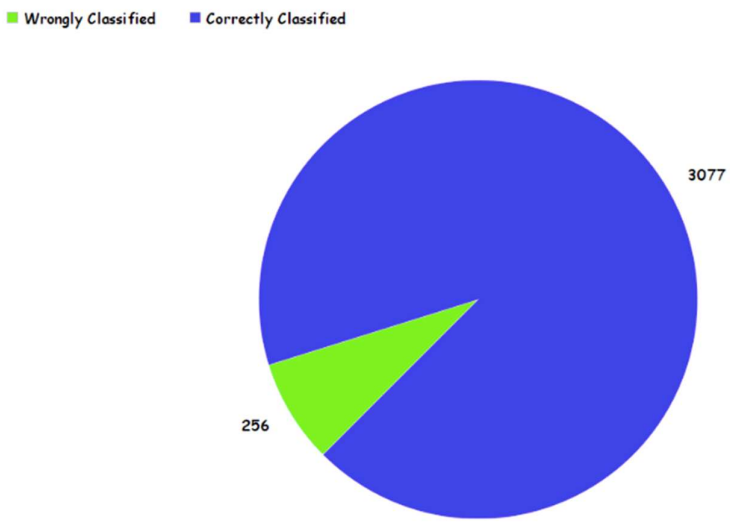
Intuition:

1. Given an unclassified example, we can assign it a class-group by observing what its
nearest neighbours belong to.
2. Out of k nearest neighbours some might be assigned to class A and some others might be
assigned to class B.
3. So we choose the class with most probability and assign it to our unclassified example.

We used 3-fold cross entropy as mentioned in problem. We ranged the value of k from 1 to 25 and found that the k for which the error was minimum is 13.
The error for the different values of k is shown in graph.



The following pie chart shows the analysis of wrong and correct classes.



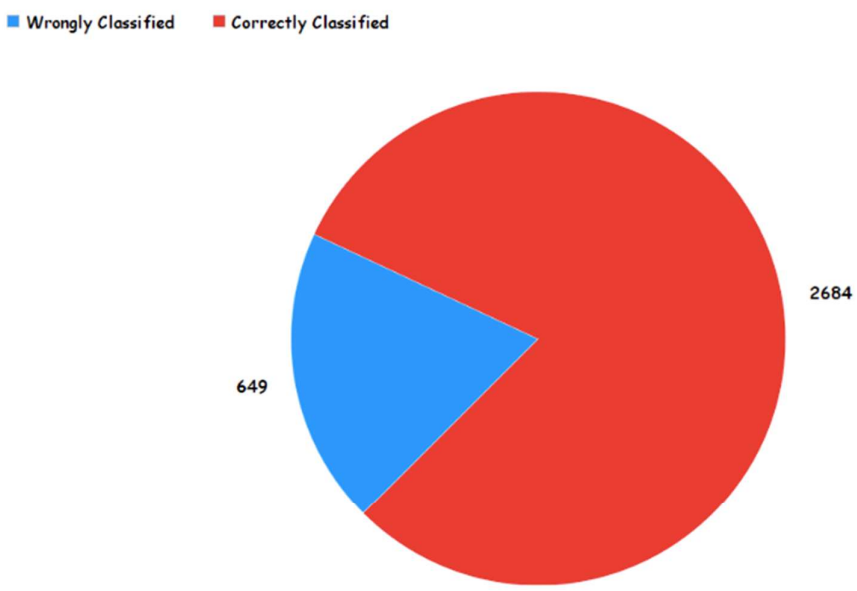
Further, the output of the C-code (KNN_Classifier.c) is:

```
Initializing... Done
Reading Training set from File... Done
Precomputing distance between every neighbours... Done

Cross Validation
1st fold as test set from 0 to 2222
Finding nearest k-nodes for 1<=k<=25... Done
2nd fold as test set from 2223 to 4445
Finding nearest k-nodes for 1<=k<=25... Done
3rd fold as test set from 4446 to 6670
Finding nearest k-nodes for 1<=k<=25... Done
Finding value of k... Done
Value of k : 3
Training Over

Reading the test set... Done
Calculating distances from every sets and perform K-NNC...
Done
Total correct classes : 3077
Accuraccy : 92.319232%
```

For task 2(Naïve Bayes Classifier),
The training set was used to find the probability matrix using the standard Bayes formulas. Using the probability matrix, the test set was tested and the final accuracy was found to be between 79% to 80%.

The output of the C-code (NaïveBayes.c) is:

```
Reading training set... Done
Training Started
Calculating the probability matrix... Done

Testing Started
Reading test file and Calculating their class... Done
Total correct classes: 2684
Accuracy in Percentage: 80.528053
```


Wrongly Classified
Correctly Classified
649
2684