**AI Assignment 03**

**Report**

**J Sowmya Vasuki – 201601035**

**Parkhi Mohan - 201601061**

Assignment 3 includes Data Sets (Train and Test sets given separately) in OCR. Handwritten data is available at <https://sites.google.com/site/viswanathpulabaigari/data-sets>. It is a 192-dimensional, 10 class problem with pre-assigned label.

**TASK 1 - Find k to be used in k-nearest neighbour classifier (k-NNC) using a 3-fold cross validation. Give the results in your report. You can draw a plot. With k value thus found employ k-NNC and find the error rate. Report this in your report.**

**Outline and procedure:**

Implementation of K-Nearest-Neighbours needs the value of optimum K.

Cross validation method is used to calculated the optimum value of K.

3-fold (r = 3) cross validation technique is applied and the training data is taken in one-thirds. Value of K is varied from 1 to 25 and error is found for each fold. After calculating mean error, ceil of minimum of all values found is the value of K.

As the data is uniformly distributed, optimum value of comes to be K=1. Having found the value of K-Optimum, k-NNC algorithm is applied on the given Test Data.

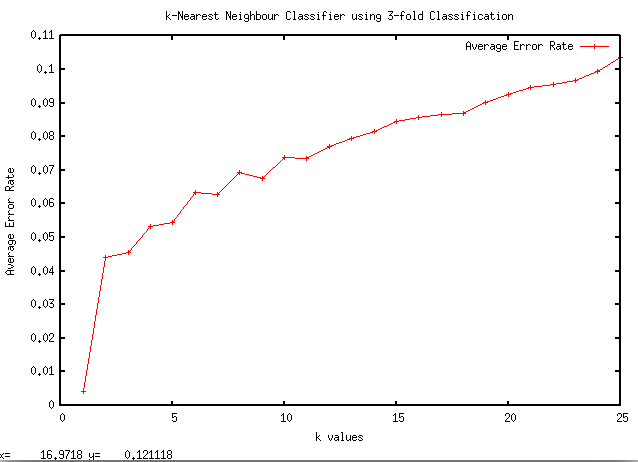
**Observations and Results:**

When the assigned labels are checked with the original label assigned in the Test Data, correct and incorrect classifications are calculated and appropriate accuracy is obtained.

**Optimum value of K: 1**

**Accuracy: 92.019203 -> 3067 entries were correctly classified**

**Error: 7.980797**



**TASK 2 - Employ the Naive Bayes classifier with the above-mentioned data set. You can use log of the posterior (to overcome the small numbers problem). Report about your observations in your report.**

**Outline and procedure:**

Conditional probabilities are calculated and assignment to the classes using Naïve Bayes Classifier.

P(w0), P(w1), P(w2), … , P(w9) -> the apriori probabilities are obtained for all 10 classes.

Posterior probabilities are calculated **Bayes rule**: with the help of apriori probabilities and class conditional distributions

**P(wi/x) = ( P(x/wi) \* P(wi) ) / P(x)**

Where P(wi/x): posterior probability, P(x/wi): likelihood probability, P(wi): prior, P(x): evidence.

Having calculated the respective probabilities, testing process is implemented.

For each entry in the validation set, probabilities for each class are multiplied and stored. Log() of the values is obtained to take into account the smaller values.

The class with the highest probability is assigned that particular label entry.

**Observations and Results:**

When the assigned labels are checked with the original label assigned in the Test Data, correct and incorrect classifications are calculated and appropriate accuracy is obtained.

**Accuracy: 81.728173 -> 2723 entries were correctly classified**

**Error: 18.271827**