**CSE 572: Data Mining Spring 2017**

**Assignment 3 / Mini Project 1**

**GROUP MEMEBRS:**

ADITI SONIK 1211009433

KANDHAN SEKAR 1209331327

PRIYADHARSHINI GUNARAJ 1211075356

**KNN Algorithm:**

Implemented the k-nearest neighbor algorithm with ‘k = 5’ and distance measure ‘Euclidean distance’.

The toolbox ‘Bioinformatics’ is used to access the functions. The function ‘KNNCLASSIFY’ takes test data, train data, training class, ‘k’ value as input parameters. The test data and train data must have same number of dimensions/features. KNNCLASSIFY classifies test data by finding ‘k’ nearest neighbors (k=5 here) in train data (for each row/record of test data, 5 nearest train data records are found out). Then it takes the class of those nearest neighbors and predicts the test class as the highest number of occurring class.

The predicted class is then compared with actual class of test data using ‘CLASSPERF’ function and the performance rate is found out. Then the accuracy is calculated.

The accuracy for “Human Activity Recognition” dataset is **90.0238%** and “VidTIMIT” dataset is **98.7000%**

**FEED FORWARD Neural Network:**

Trained a feedforward neural network with 1 hidden layer containing 25 neurons. The toolbox ‘Neural Networks Toolkit’ is used here.

Each record has 100 attributes corresponding to 100 inputs to the network. And, we have designed the neural network with 25 outputs of each being value 0 or 1, with 1 corresponding to the class of that record.

The function ‘ind2vec’ is used to convert the training class to 25\*3500 matrix and the network is created using ‘FEEDFORWARDNET’ and trained using ‘train’ with training data and training class.

The training function used here is ‘trainlm’ with maximum epoch 1000. Then the network is tested against test data and the class is predicted. The predicted class is converted to indices using ‘vec2ind’ to map it to the correct class. Finally, by comparing predicted and actual class, the accuracy is found out.

The training took on average 29 iterations and terminated by validation stop with accuracy **97%**

**SVM Classifier:**

Training an SVM classifier with a polynomial kernel with parameter 2 on the training set and testing on the test set for “Human Activity Recognition” and “VidTIMIT” datasets.

In “Human Activity Recognition” there are 6 class labels and for each class label an SVM classifier is trained using one against the rest approach. Likewise, in “VidTIMIT” dataset there are 25 class labels and similar approach has been followed for training the SVM classifiers.

Matlab’s inbuilt function ‘fitcsvm’ is used for training the SVM with train data and similarly, ‘predict’ function is used for assigning label to each data sample after training.

For training a classifier at a time, the main class label is taken as it is and the other class labels are taken as ‘-1’. In case of multiple labels resulting on a single sample by training all the classifiers, the label with the maximum value is selected.

For “Human Activity Recognition” dataset, an accuracy of **95.9281%** is observed. And for “VidTIMIT” dataset, an accuracy of **98.8000%** is observed.