Amrit Sahoo – 1211075109

Vahini Mohan – 1210414124

Data Mining Assignment – 3

1. The Euclidean distance measure was used to build the knn model with k = 7. The feature and class label files of both training and test data was imported into the code using the load method. Fitcknn() was used to construct the knn model followed by the predict() method to predict the class label of the X_test feature data. The predicted result is compared with the actual class labels in y_test to calculate the accuracy of the built model.
   Accuracy achieved = 90.23%

2. (2a) Similar to the 1st question, the Euclidean distance measure was used to build the knn model with k = 7. The feature and class label files of both training and test data was imported into the code using the load method. Fitcknn() was used to construct the knn model followed by the predict() method to predict the class label of the X_test feature data. The predicted result is compared with the actual class labels in y_test to calculate the accuracy of the built model.
   Accuracy achieved = 98.4%

3. (2b) The training set X_train, y_train is used to train a feedforward neural network with 1 hidden layer containing 25 neurons using patternnet and train function one time. The y_train is converted to a target matrix using ind2vec() before utilizing it in the train() function. The class labels for the test set is fetched from the generated neural network. The result is compared with the actual y_test to compute the accuracy.
   Accuracy achieved = 98.0% The k-nearest neighbor and feedforward neural network have performed similarly with very minimal difference in their accuracies.

   Functions used for the above questions:

   1. Fitcknn() - A nearest-neighbor classification object, where both distance metric ("nearest") and number of neighbors can be altered. The object classifies new observations using the predict method.
   2. ind2vec() - allows indices to be represented either by themselves, or as vectors containing a 1 in the row of the index they represent.
   3. vec2ind() - Convert vectors to indices
   4. patternnet(`hiddenSizes,trainFcn,performFcn`) – pattern recognition networks are feedforward networks that can be trained to classify inputs according to target classes. The target data for pattern recognition networks should consist of vectors of all zero values except for a 1 in element i, where i is the class they are to represent.
   5. Train (`net,X,T`) – Trains a neural network according to the topology specified. Net – network, X – network inputs, T – network targets

   References:

   1. https://www.mathworks.com/help/matlab/
   2. http://matlab.mathworks.com