CSE 572: Data Mining
Spring 2017
Assignment 5 / Mini Project 3

<u>**Team members:-**</u>
1. Saurabh Jagdhane (1209572595)
2. Himanshu Tyagi (1209283279)
3. Rachita Gupta (1209247724)

<u>**Problem 1**</u>
In this problem we have a multi-label classification in which a data sample may belong to more than 1 class. In the dataset provided we have feature vector of dimension 29 and there are 6 classes. We have to find the classes to which each of 2407 data samples represent. To solve this problem we use SVM separately for each class. Using *fitcsvm* function we create 6 SVM's. SVM training is done for both polynomial and gaussian kernels with parameter 2.
Next part is to calculate the accuracy, to calculate that we use 2 parameters true class labels and predicted class labels.
Then using *bsxfun* we find and, or values of the 2 parameters and divide them. This is done for both polynomial and gaussian kernels. The accuracy is calculated by ratio of the number of intersection to the total sum of the true and predicted class.

SVM with polynomial kernel with parameter 2, **Accuracy= 64.3513%**
SVM with Gaussian kernel with parameter 2, **Accuracy= 64.3146%**

It can be observed that both classification models have performed equally.

<u>**Problem 2**</u>
Handwritten digits dataset is classified using KNN, SVM and ANN algorithms and their accuracy is reported as follows:

1) k-nearest neighbor with k = 7, **Accuracy= 95.2322%**

2) SVM with a polynomial kernel of degree 2, **Accuracy= 96.5347%**

3) Feedforward neural network with a single hidden layer with 25 neurons, **Accuracy= 93.4174%**

4) Ensemble, **Accuracy= 96.7415%**

Ensemble accuracy on the entire test set is obtained by taking a majority vote on the predictions of the each model which is done using mode() function.

Functions used:
1. mode(): This function is used to calculate majority of the predicted class from three different models whose labels are stored in row wise using
      ensemble = [label_KNN';label;classes];
   And then mode function is used to obtain ensemble label for the test data.
           mode(ensemble(:,i))

2. fitcsvm(): This function trains binary support vector machine classifier for two-class binary classification for a moderate-dimensional predictor data set
   fitcsvm(x_train,y_train(:,i),'KernelFunction','polynomial','Polynomialorder',2);
   fitcsvm(x_train,y_train(:,i),'KernelFunction','Gaussian','KernelScale', 'auto');
   In this situation we have 6 classes so to use svm we use technique 1-r and train 1 against the rest using kernel function.

3. bsxfun(): This function is used to operate 2 vectors element by element.
                 C = bsxfun(fun,A,B)
   A and B are 2 vectors and fun is the operator(in this case we have used 'and' and 'or').

.