Machine Learning
Assignment 2

Mohit Jain
2014MCS2800

17 March 2015

Question 1

(a) Linear Kernel

Support vectors obtained from train-small and train have been provided separately along with the file.

(b) Linear Kernel

- **train-small** Support Vectors : 151
  Accuracy : 91.3%
  Filename : svSmallLinear.mat
- **train** Support Vectors : 452
  Accuracy : 98.7%
  Filename : svLargeLinear.mat

(c) Gaussian Kernel

- **train-small** Support Vectors : 536
  Accuracy : 89.2%
  Filename : svSmallGausian.mat
- **train** Support Vectors : 1775
  Accuracy : 98%
  Filename : svLargeGausian.mat

Accuracy using Gaussian SVM case is lower than accuracy using Linear SVM because switching to Gaussian kernel form linear kernel leads to over-fitting.

(d) LibSVM

- **train-small linear** Support Vectors : 151
  Accuracy : 91.3%
  Filename : svLsmallLinear.mat
- **train linear** Support Vectors : 452
  Accuracy : 98.7%
  Filename : scLlargeLinear.mat
- **train Gaussian** Support Vectors : 530
  Accuracy : 89.2%
  Filename : svLsmallGausian.mat
- **train-small Gaussian** Support Vectors :1985
  Accuracy : 98.7%
  Filename : svLlargeGausian.mat

CVX is general purpose software which is used to find alpha values for dual svm problem. Whereas LibSVM is highly optimized for solving svm problem only. Therefore LibSVM always return higher or equal accuracy than SVM problem trained through CVX package.

Question 2

(b)

Stoping Criteria : Change in value of error function $\Delta(J(\theta)) < 0.1$

(c)

Accuracy for 3 class : 99.21%
Accuracy for 8 class : 98.67%
Cumulative Accuracy : 98.94%
Training Time = 138.80 secs
Number of iterations = 27

(d)

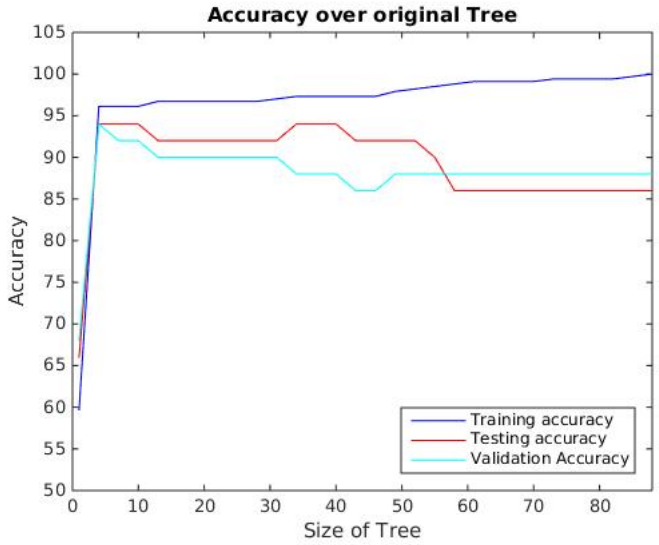
Minimum number of output units required in this case is 4 because at least 4 bits is required to represent 10 different classes. But 10 output units lead to better accuracy and hence 10 output units have been taken.
Stoping Criteria : Change in value of error function $\Delta(J(\theta)) < 50$
Accuracy for 0 class : 98.47%
Accuracy for 1 class : 98.41%
Accuracy for 2 class : 96.03%
Accuracy for 3 class : 95.45%

Accuracy for 4 class : 95.42%
Accuracy for 5 class : 96.30%
Accuracy for 6 class : 96.97%
Accuracy for 7 class : 96.40%
Accuracy for 8 class : 95.59%
Accuracy for 9 class : 94.15%
Cumulative Accuracy = 96.34%
Training Time = 659.06 secs
Number of iterations = 26

Training time for training neural network using original multiclass MNSIT data-set was significantly higher than training time for training neural network for 2 classes(3&8) because of larger data-set and more number of output units.

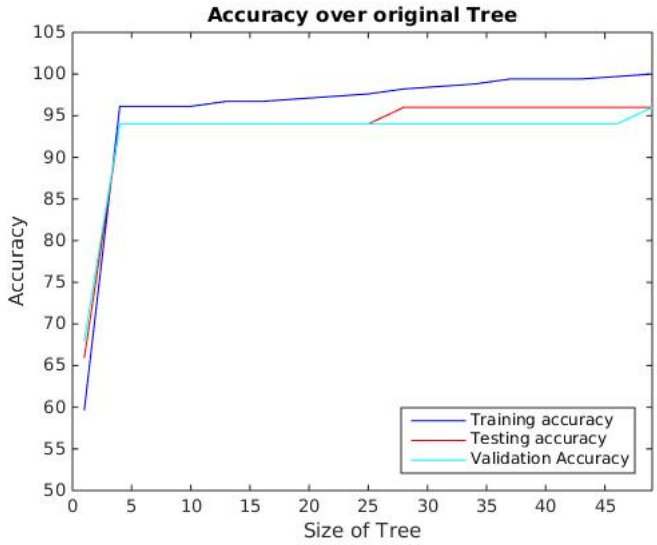
Question 3

(a) Decision Tree - Net error



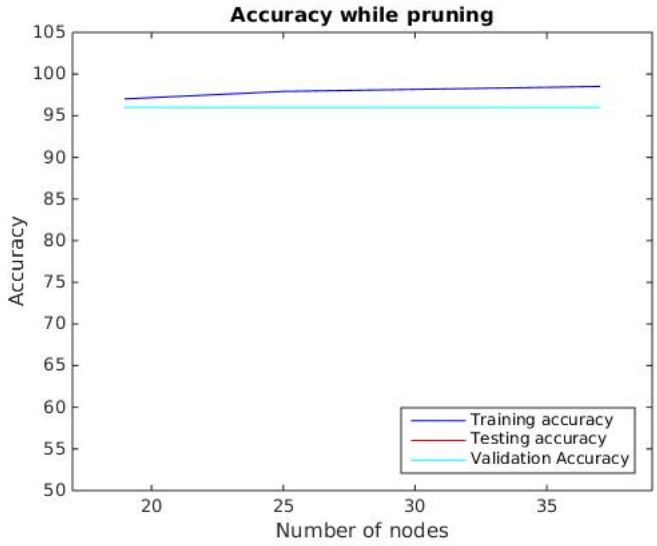
Increasing number of nodes in decision tree leads to over-fitting. Therefore training accuracy increases and testing &validation accuracy reduces as number of nodes in decision tree increases.

(b) Decision Tree - Information Gain



Decision tree formed by using information gain as criteria has significantly lesser nodes than decision tree formed by using net error and also feature used to split the data at particular node is different. Information gain is better estimate than net error in determining best feature to split at particular node and hence leads to lower number of nodes.

(c) Pruned Decision Tree - Information Gain



Pruning decision tree leads to lower training accuracy but higher or equal testing and validation accuracy because pruning decreases over-fitting in decision tree.

(d) Missing values

Missing values can be replaced by mean/mode of their corresponding feature values in other training examples. Other way could be to take mean/mode of only those training examples which have output class same as that of output class of current training example.