Name: Aniket Bajpai

Entry No: 2014CS10209

A) Cross validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set.

In this exercise, as we had limited data, it made sense to use 5-fold cross validation, as it gives a lower variance than conventional validation. It will also help in checking overfitting.

As there were only 5 files, cross validation was performed by manually changing the names of the 4 training and 1 test file.

Results of 5-fold cross validation -

1. Training files : 1,2,3,4

Test file : 5

Accuracy : 94.88 %

2. Training files : 1,2,3,5

Test file : 4

Accuracy : 94.88 %

3. Training files : 1,2,4,5

Test file : 3

Accuracy : 95.64 %

4. Training files : 1,3,4,5

Test file : 2

Accuracy : 94.88 %

5. Training files : 2,3,4,5

Test file : 1

Accuracy : 97.92 %

Avg. accuracy = 95.62 %

B) Accuracy in random prediction = 12.62 %

C) Impovement over random predicton = 83.00 %