

# Report on Assignment 2: Ensemble Learning

A. S. M. Ahsan-Ul-Haque

1205021

1. The performance measures for 5-fold cross validation is given below:

Cross Validation	Accuracy
1	87.7005%
2	93.9759%
3	94.2857%
4	97.0238%
5	95.3333%

The performance measures for 10-fold cross validation is given below:

Cross Validation	Accuracy
1	81.9277%
2	88.7574%
3	97.0238%
4	96.6292%
5	96.4072%
6	34.6786%
7	100%
8	94.6429%
9	95.5%
10	34.6786%

The performance measures for 20-fold cross validation is given below:

Cross Validation	Accuracy
1	89.9281%
2	90.3448%
3	90.8333%
4	94.0594%
5	70.098%
6	34.6786%
7	97.2067%
8	34.6786%
9	93.1579%
10	93.4066%
11	34.6786%
12	34.6786%
13	99.1736%
14	97.1963%
15	95.6522%
16	95.6522%
17	99.1597%
18	93.913%
19	34.6786%
20	34.6786%

The accuracy for leave one out cross validation (10 samples) is: 30%

2. Compare and analyze the accuracies obtained by different learners: decision stump alone, boosting with 30 rounds, your ID3 implementation.

Ans.

For 5-fold cross validation, the accuracies for ID3 implementation were 91.9776%, 93.4579%, 89.9065%, 92.1495% and 91.6045% respectively.

Whereas, the accuracies for boosting with 30 rounds were 88.7701%, 93.2099%, 94.7368%, 97.0238% and 94.0828% respectively.

On average the boosting performed better than ID3. This is expected because for  $T=30$ , boosting makes use of 30 decision trees, each learning entirely on a single feature, and learns to combine their results to make future predictions.

The decision stumps alone, however, performs very poorly and has an accuracy of 65.2695%. Only when they are combined with boosting, we get a good classifier.