Assignment 4

Machine Learning Sessional

Submitted By:

Mohammad Imam Hossain

Id: 1105031

L/T: 4/2

Dept. CSE, BUET.

1.1

Methods	Precision	Recall	Accuracy	F1 Score
Supervised Learning	0.961	0.928	0.962	0.944
Semi-supervised Learning	0.929	0.910	0.945	0.919

Analysis:

From the output, it is clear that the avg. precision, avg. recall, avg. accuracy and avg. f1 score has lower value for semi-supervised learning than supervised learning. As in semi-supervised case, we have some unleveled data, and we are leveling the unleveled data based on the leveled data's decision tree, so some error is included at the time of decision tree building.

1.2

Methods	Accuracy (out of 1)
Supervised Learning	0.962
K fold Cross Validation (k=5)	0.954
K fold Cross Validation (k=10)	0.953
K fold Cross Validation (k=20)	0.959
Leave-one-out Cross Validation	0.961

In supervised learning, we are just dividing the data set into two parts- training(80%) and testing(20%) sets randomly . So if we run this algorithm different times, then each time the output variance is high.

For the case of K fold cross validation, the output variance is low as we are dividing the dataset into k subsets and each set becomes testing set

1 time and training set k-1 times. So in this case the output variance becomes low. But it is costly rather than the supervised one.

Leave-one-out-cross validation can be thought as k fold cross validation with k=no of data points. So it has the lowest output variance.

1.3

Methods	Accuracy (out of 1)
Supervised Learning	0.962
Decision Stump	0.929
AdaBoost(30 rounds)	0.963

The accuracy of decision stump is low , as it is a 1 depth tree.
But the accuracy of Supervised learning is high as we are using ID3 for this.
For AdaBoost case the accuracy is also high although we are using decision stump for this case. But here we are using 30 rounds for wrong decision correction. As a result, the accuracy goes high.

Methods	Accuracy (out of 1)
AdaBoost(k=5)	0.951
AdaBoost(k=10)	0.955
AdaBoost(k=20)	0.962
AdaBoost(k=30)	0.963

With the increase of rounds number in adaboost the accuracy increases. As the more we increase the value of k,the more we can correct the weights correctly. So the accuracy increases.