CSE472: Machine Learning Sessional
Assignment 1:
Decision Tree Learning for Cancer Diagnosis

Mazharul Islam
1105013

September 8, 2016

1 Statistics.

Evaluation Criterion	Accuracy	Precision	Recall
Information Gain(Testing Data)	0.945939849624	0.941352088035	0.33101078168
Information Gain(Training Data)	1.000000	1.000000	0.346679104478

2 Questions

2.1 Why are you dividing the dataset 80 % into training and 20 % into test data rather than using 100 % data for training?

Because we need to have some data in order to test the evaluated ID3 tree and how much it is correct. Since we are not given any seperate testing dataset we need to split the given dataset into training set & testing set.

2.2 Do you see evidence of overfitting in some experiments Explain.

Yes For 90 % spliting the precession of train data is 1.0 but for test data it is 0.938787878788. which is less than when the training data is split into 80 %