

CSE472: Machine Learning Sessional

Assignment 1:

Decision Tree Learning for Cancer Diagnosis

Mazharul Islam
1105013

September 5, 2016

1 Statistics.

Evaluation Criterion	Accuracy	Precision	Recall
Information Gain	0.837245508982	0.807218969273	0.397551216674

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 separate 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 If the dataset is split into 50 % then the accuracy increases which is the evidence that the data is overfitted in some sense.