NAME: ID: GRAD or UNDERGRAD

CS 412 - 512 Machine Learning

In-Class Quiz

Not Graded

Oct 6, 2015

A) What is the entropy of a three sided dice with p(1) = 0.5, p(2)=p(3) = 0.25
Note: log2 0.5 = -1 and log2 0.25 = -2 *(you should be able to compute simple logs in general).*

Entropy = …………………………………………………………………………………..

      = …………………………………………….

B) Assume that you have trained two separate decision tree classifiers A and B and C using the same training set containing 1,000 training instance.

The performance of the classifiers are as follows, where the test set is a separate set of 100 instances. Answer the following questions accordingly:

Classifier	Train Set Performance	Test Set Performance
A	%99	%70
B	%85	%80

1) Can we say that A overfit to the training data? Circle Yes/No

2) How many classification errors is made by A, in the test set? Answer: …………………………

3) Can we say that B overfit to the training data? **Circle** Yes/No

4) If you have 10 times more data,
    • would you **expect** that train and test performances would be closer? **Circle** Yes/No

    • is it guaranteed that train error will be lower? **Circle** Yes/No