DS 200 Introduction to Data Sciences

Spring 2018

Homework 2 (30 points)

Due Feb 28, 10 pm

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. (20 points) The figure below shows three possible feature test for the root node of a decision tree to predict spam e-mail messages (based on the example discussed in class). (a) Calculate the expected information gain for each feature test (please show your formula). (b) Which feature test will be selected by the Decision Tree Learning algorithm?



1. (10 points) Is the entropy of a node in a Decision Tree learned always lower than the entropy of its parent node? Explain the rationale of your answer.