Assignment 5

Written Assignment - Machine Learning & Decision Trees

Max possible score:

- 4308: 50 Points
- 5360: 50 Points

Task 1

20 points

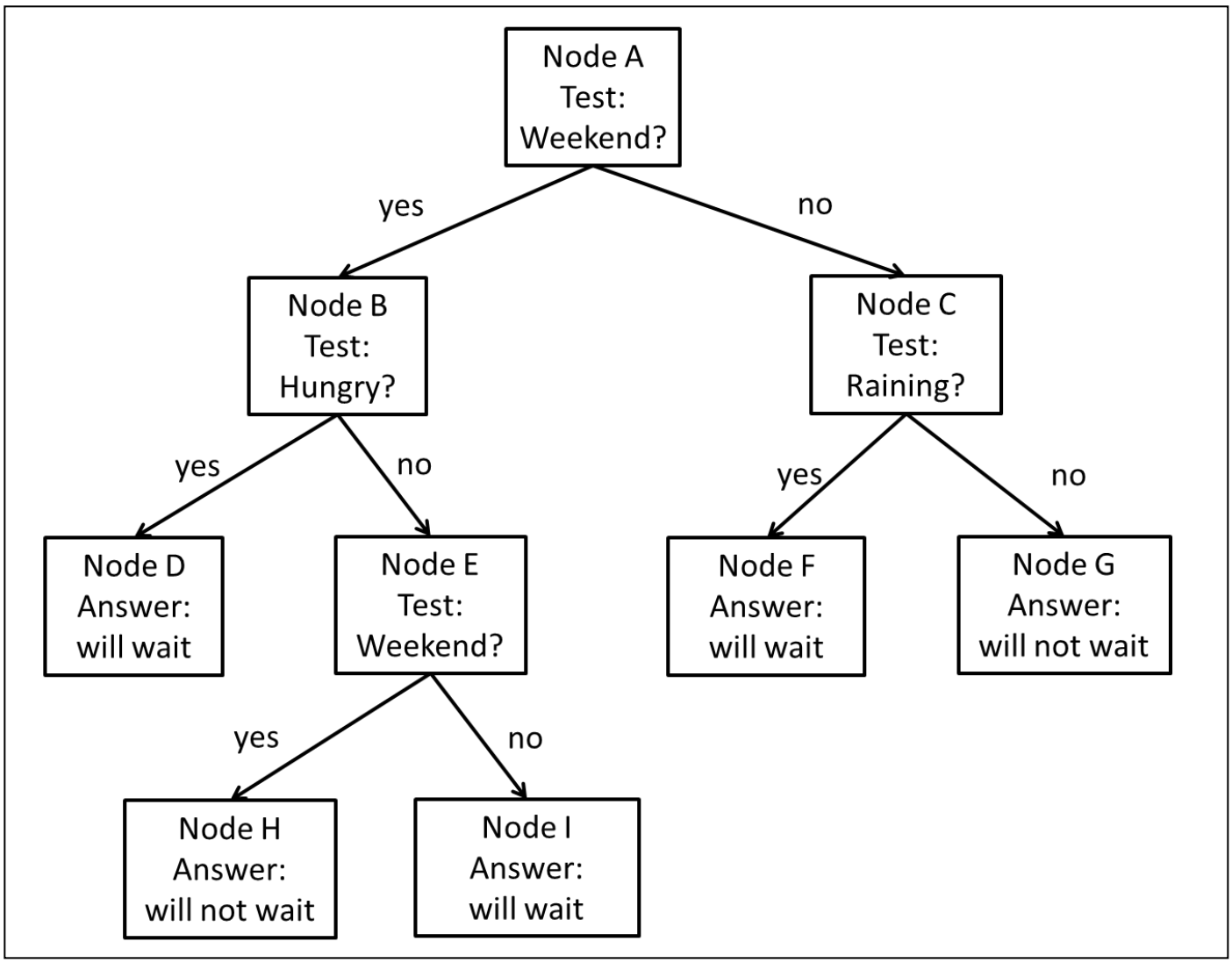

Figure 1: A decision tree for estimating whether the patron will be willing to wait for a table at a restaurant.

Part a (5 points): Suppose that, on the entire set of training samples available for constructing the decision tree of Figure 1, 65 people decided to wait, and 35 people decided not to wait. What is the initial entropy at node A

(before the test is applied)?

Part b (5 points): As mentioned in the previous part, at node A, 65 people decided to wait and 35 people decided not to wait.

- Out of the cases where people decided to wait, in 25 cases it was weekend and in 40 cases it was not weekend.
- Out of the cases where people decided not to wait, in 20 cases it was weekend and in 15 cases it was not weekend.

What is the information gain for the weekend test at node A?

Part c (5 points): In the decision tree of Figure 1, node E uses the exact same test (whether it is weekend or not) as node A. What is the information gain, at node E, of using the weekend test?

Part d (5 points): We have a test case of a hungry patron who came in on a rainy Sunday. Which leaf node does this test case end up in? What does the decision tree output for that case?

Task 2

15 points

Class	A	B	C
X	1	2	1
X	2	1	2
X	3	2	2
X	1	3	3
X	1	2	2
Y	2	1	1
Y	3	1	1
Y	2	2	2
Y	3	3	1
Y	2	1	1

We want to build a decision tree that determines whether a given pattern is class X or class Y. The decision tree can only use tests that are based on attributes A, B, and C. Each attribute has 3 possible values: 1, 2, 3. We have the 10 training examples, shown on the table (each row corresponds to a training example, the first column is the class label and the other 3 columns are the pattern). What is the information gain of each attribute at the root? Which attribute achieves the highest information gain at the root?

Task 3

10 points

Suppose that, at a node N of a decision tree, we have 1000 training examples. There are four possible class labels (A, B, C, D) for each of these training examples.

Part a: What is the highest possible and lowest possible entropy value at node N?

Part b: Suppose that, at node N, we choose an attribute K. What is the highest possible and lowest possible information gain for that attribute?

Task 4

5 points

Your boss at a software company gives you a binary classifier (i.e., a classifier with only two possible output values) that predicts, for any basketball game, whether the home team will win or not. This classifier has a 28% accuracy, and your boss assigns you the task of improving that classifier, so that you get an accuracy that is better than 60%. How do you achieve that task? Can you guarantee achieving better than 60% accuracy?