

# CS472 Foundations of Artificial Intelligence

Fall 2002

## Assignment 4

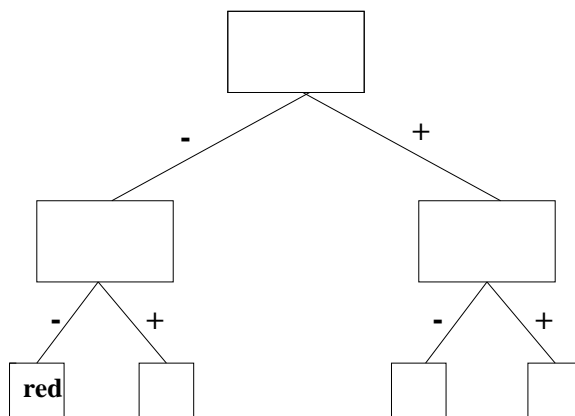
*Due Wednesday, November 27 at the beginning of class.*

Solutions must be typed, although equations, graphs, tables, etc., can be drawn in by hand.

1. **Decision Tree Learning (20 pts. total)** You are given a set of 8 training examples, and are asked to construct decision trees to classify them. The following table shows the examples with the value of their four attributes, *A*, *B*, *C*, and *D*, as well as the class to which they belong (*red* or *green*).

example	A	B	C	D	class
x1	+	-	-	-	red
x2	+	-	+	-	red
x3	-	+	-	-	green
x4	-	+	+	-	green
x5	+	-	-	+	green
x6	-	-	+	+	green
x7	+	+	+	+	red
x8	-	+	+	+	red

- (a) (5 pts.) Fill in the empty nodes in figure below to make a decision tree that correctly classifies the examples in the training set. No computations necessary.



- (b) (**10 pts.**) Consider the decision tree that will be induced by a decision tree learning algorithm using information gain. What is the attribute that will be selected as the root of the tree? Show all the entropy calculations involved. (Note: you are only asked to provide the top node, not to grow the full tree.)
- (c) (**5 pts.**) Will the tree found by the learning algorithm have as few nodes as the tree given in part (a)? Briefly explain why or why not. What general conclusion can you reach regarding the size of the trees produced by the decision tree learning algorithm with information gain?
2. **Boolean functions (10 pts. total)** How many distinct Boolean functions on  $n$  propositional letters are there? Explain where expression comes from. Can each of these functions be represented by a 3CNF (3SAT) formula on  $n$  propositional letters? Explain your answer.
3. **Learning, general. (5 pts. total)** Discuss the tradeoff between generalization error and training error for a machine learning method in the context of the expressiveness of the hypothesis class of the learning method.
4. (**15 pts**) Find the smallest decision tree that can be used to represent the concept

$$[p_1(x) \wedge p_2(x)] \vee [p_3(x) \wedge p_4(x) \wedge p_5(x)]$$

Why is this decision tree larger than the expression itself?

5. (**5 pts**) Exercise 18.4 of Russel & Norvig.
6. (**15 pts**) Exercise 18.13 of Russel & Norvig.