CS 540: Introduction to Artificial Intelligence

Midterm Exam: 7:15-9:15 pm, October 29, 2002
Room 1240 CS & Stats

CLOSED BOOK
(one sheet of notes and a calculator allowed)

Write your answers on these pages and show your work. If you feel that a question is not fully specified, state any assumptions you need to make in order to solve the problem. You may use the backs of these sheets for scratch work.

Write your name on this and all other pages of this exam. Make sure your exam contains *six* problems on *twelve* pages.

Name _____

Student ID _____

Problem	Score	Max Score
1	_____	15
2	_____	20
3	_____	5
4	_____	20
5	_____	10
6	_____	30
TOTAL	_____	100

Problem 1 – Decision Trees (15 points)

Assume that you are given the set of labeled *training examples* below, where each feature has three possible values *a, b,* or *c*. You choose to apply the C5 algorithm to this data and select "-" as the default output.

F1	F2	F3	Output
a	a	a	+
c	b	c	+
c	a	c	+
b	a	a	-
a	b	c	-
b	b	c	-

a) What score would the *information gain* calculation assign to each of the features?
   Be sure to show all your work (use the back of this or the previous sheet if needed).

b) Which feature would be chosen as the root of the decision tree being built? _____
   (Break ties in favor of *F1* over *F2* over *F3*.)

Name: _____

c) Show the *next* interior node, if any, that the C5 algorithm would add to the decision tree. Again, be sure to show all your work. (Even if this secod interior node does not completely separate the training data, stop after adding this second node.)

Be sure to label all the leaf nodes in the decision tree that you have created.

d) Assuming you have the following *tuning set*, what would the first round of HW1's pruning algorithm do to the decision tree produced in Part c? (If you need to break any ties, default to "-".) Justify your answer.

F1	F2	F3	Output
b	a	b	+
a	b	a	+
c	c	a	-

Problem 2 – Search (20 points)

a) Consider the search space below, where *S* is the start node and *G1*, *G2,* and *G3* satisfy the goal test. Arcs are labeled with the cost of traversing them and the estimated cost to a goal is reported inside nodes.

For each of the following search strategies, indicate which goal state is reached (if any) and list, *in order*, all the states *popped off of the OPEN list*. When all else is equal, nodes should be removed from OPEN in alphabetical order.
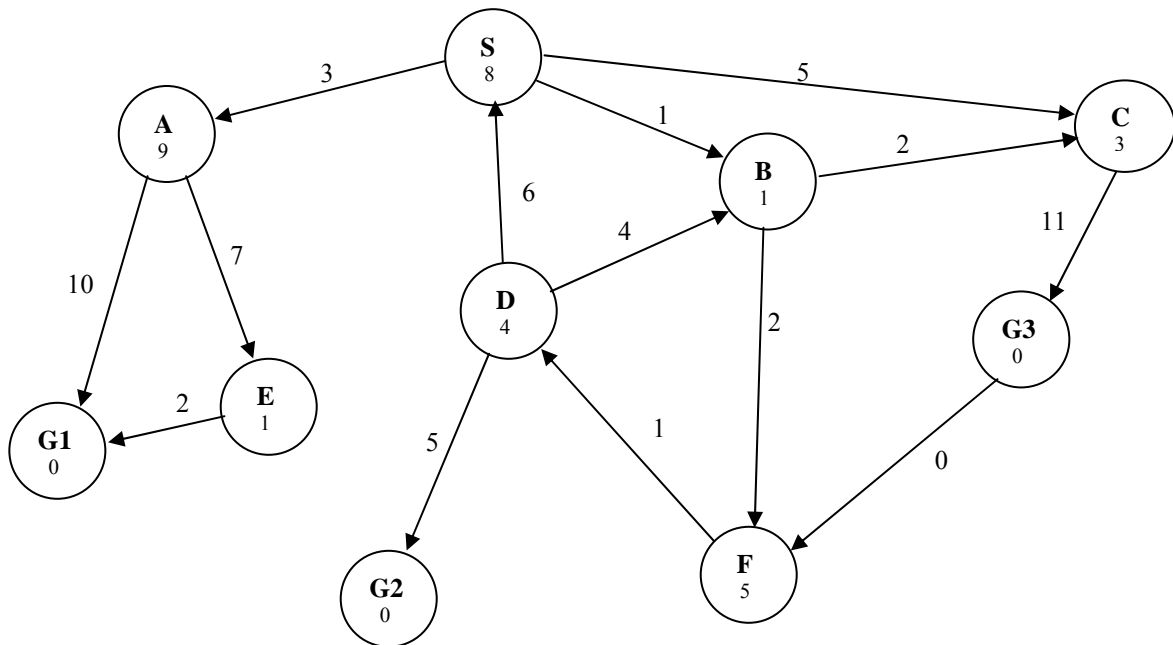
Iterative Deepening

Goal state reached: _____   States popped off OPEN: _____

Hill Climbing

Goal state reached: _____   States popped off OPEN: _____

A*

Goal state reached: _____   States popped off OPEN: _____

b) (Note: This question talks about search spaces in general and is *not* referring to the specific search space used in Part a.)

Consider the following scoring function for heuristic search:

$$score(node) = Wgt \times g(node) + (1 - Wgt) \times h(node) \quad \text{where} \quad 0 \leq Wgt \leq 1$$

      i. Which search algorithm do you get with *Wgt* set to 0?

      ii. Which search algorithm do you get with *Wgt* set to 1?

      iii. If *h(node)* is an admissible heuristic, for what range of *Wgt* values is the above scoring function guaranteed to produce an admissible search strategy? Explain your answer.

Problem 3 – Representation using First-Order Logic (5 points)

Convert each of the following English sentences into first-order predicate calculus (FOPC), using reasonably named predicates, functions, and constants. If you feel a sentence is ambiguous, clarify which meaning you're representing in logic. (Write your answers in the space *below* the English sentence.)

Someone in Madison likes pizza.

Everyone who likes ham also likes cheese.

Problem 4 – Representation & Reasoning with Propositional Logic (20 pts)

a) Is the following WFF *valid*? Justify your answer using a truth table.

$$[(P \lor Q) \land (Q \lor R)] \Rightarrow (P \lor R)$$

b) Use *propositional* logic to represent the following sentences.

   i)   *Fido is always either sleeping or barking.*

   ii) *When Fido is hungry Fido barks, but Fido barking does not necessarily mean that Fido is hungry.*

c) Formally show that S follows from the "given's" below.
   (Don't deduce more than 10 additional WFF's.)

Number	WFF	Justification
1	$P \wedge Z$	given
2	$(\neg R \wedge \neg W) \vee (\neg P)$	given
3	$(W \wedge Q) \Rightarrow P$	given
4	$Q \vee W$	given
5	$Q \Rightarrow (S \vee P)$	given
6	$(P \wedge Q) \Rightarrow (S \vee R)$	given
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

Problem 5 – Miscellaneous Short Answers (10 points)

Briefly describe each of the following AI concepts and explain each's significance.
(Write your answers *below* the phrases.)

Simulated Annealing

Expected-Value Calculations

Interpretations

Minimax Algorithm

CLOSED List

Problem 6 – Learning, Logic, and Search Combined (30 points)

Imagine that we have a training set of labeled "fixed-length feature vectors." However, instead of using C5 to induce a decision tree, this time we want to learn inference rules like the following sample:

(*Color=red* ∧ *Shape=round*) ⇒ *positive example*

Assume that each training example is described using features F_1 through F_3 and each feature has three possible values, *a, b,* or *c* Each training example is labeled is either in the positive (+) category or the negative (-) category. Also assume that you have set aside 10% of the training examples for a test set and another 10% for a tuning set.

Your task is to cast this as a search problem where your search algorithm's job is to find the best propositional-logic wff for this task. Your wff should use implication (⇒) as its main connective and the wff should imply when an example belongs to the positive category (e.g., as done in the sample wff above).

a) Describe a *search space* that you could use for learning such wff's. Show a sample node.

b) What would be the *initial state*?

c) Describe two (2) *operators* that your search algorithm would use. Be concrete.

      i)

      ii)

d) Would you prefer to use a goal test for this task or would you prefer to look for the highest scoring state you can find? Justify your answer.

e) What would be a good scoring function for this task? Why?

f) Which search strategy would you use? Why?

g) Show one (1) state (other than the initial state) that might arise during your search and show three (3) possible children of this state.

h) What would be the analog to *decision-tree pruning* for this task? Explain how and why "pruning" could be done in this style of machine learning.