# CS 540-1: Introduction to Artificial Intelligence

*Midterm Exam: 7:15-9:15pm, March 22, 1999*

CLOSED BOOK

(one page 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.

**Name** _____

**Student ID** _____

| **Problem** | **Score** | **Max Score** |
|---|---|---|
| 1 | _____ | 25 |
| 2 | _____ | 25 |
| 3 | _____ | 8 |
| 4 | _____ | 20 |
| 5 | _____ | 10 |
| 6 | _____ | 12 |
| Total | _____ | 100 |

## PROBLEM 1 - Decision Trees (25 points)

Assume you are given the following features with the possible values shown.

```
F1  ∈  {A, B, C}
F2  ∈  {D, E}
F3  ∈  {G, H, I}
F4  ∈  {J, K}
```

a) Consider the following training examples.

| | | | | | |
|---|---|---|---|---|---|
| ex1 | F1 = A | F2 = E | F3 = H | F4 = J | category = + |
| ex2 | F1 = A | F2 = E | F3 = G | F4 = K | category = - |
| ex3 | F1 = A | F2 = D | F3 = G | F4 = J | category = + |
| ex4 | F1 = B | F2 = D | F3 = G | F4 = J | category = + |
| ex5 | F1 = B | F2 = E | F3 = H | F4 = J | category = + |
| ex6 | F1 = B | F2 = E | F3 = H | F4 = K | category = - |
| ex7 | F1 = C | F2 = D | F3 = H | F4 = K | category = - |
| ex8 | F1 = C | F2 = D | F3 = G | F4 = J | category = - |

i. What score would the *information gain* formula assign to each of the features?
***Be sure to show all your work.***
(*Some* calculations that might be of use: I(x,y)=I(y,x), I(1,0)=0, I(1/2,1/2)=1, I(1/3,2/3)=0.92, I(1/4,3/4)=0.81, I(1/5,4/5)=0.72, I(1/6,5/6)=0.65, I(1/7,6/7)=0.59, I(1/8,7/8)=0.54)

ii. Which feature would be chosen as the root? _____
(Break any ties in favor of *F1* over *F2* over *F3* over *F4*, and '+' over '-'.)

iii. Show the next *interior* node, if any, that ID3 would select. Again, be sure to show all your work (use the back of this or the previous sheet if necessary).

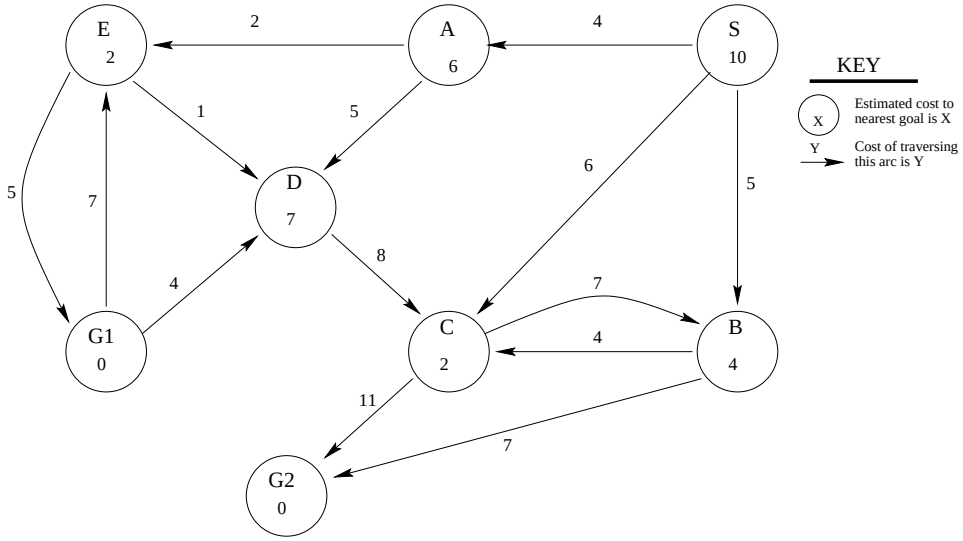
Note that, to avoid spending excessive time, you are only being asked to create at most the first *two* (2) interior nodes. Do create as many the leaf nodes as possible, given your choice of interior nodes.

b) Imagine that we used in our training examples a discrete-valued feature that had a unique value for each example, such as including in examples describing medical patients, each patient's social-security number. (Notice that we are *not* treating this as a continuously-valued feature, but just as a discrete-valued feature with a very large number of possible values.)

What would be the information gain of such a feature? Briefly discuss the implications of your answer.

## PROBLEM 2 - Search Strategies (25 points)

a) Assume you have the following search graph, where *S* is the start node and *G1* and *G2* are goal nodes. Arcs are labeled with the cost of traversing them and the estimated cost to a goal is reported inside nodes.




For each of the search strategies listed below, indicate which goal state is reached (if any) and list, in order, all the states expanded. (Recall that a state is *expanded* when it is *removed* from the OPEN list.) *When all else is equal, nodes should be expanded in alphabetical order.*

### Breadth-First

Goal state reached: _____ States expanded: _____

### Iterative Deepening

Goal state reached: _____ States expanded: _____

### Hill Climbing (using the *h* function only)

Goal state reached: _____ States expanded: _____

### A*

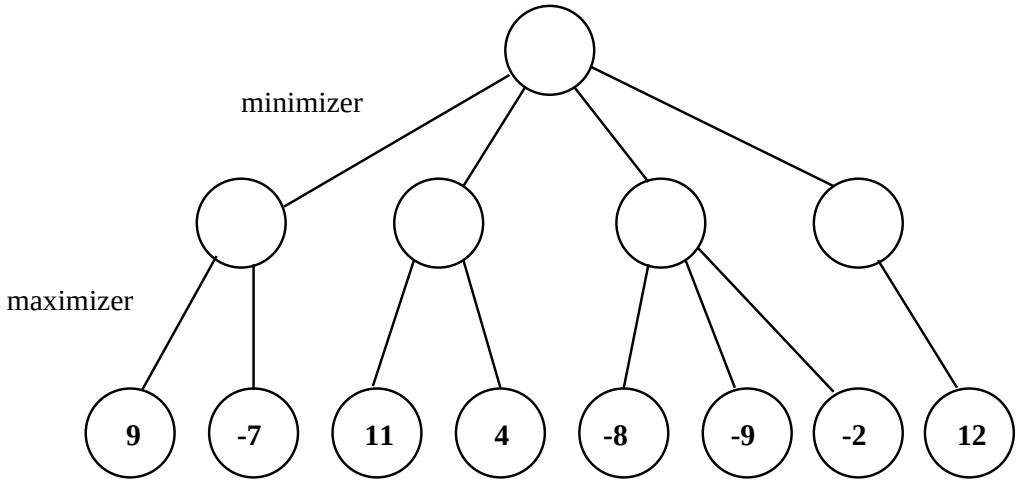
Goal state reached: _____ States expanded: _____

b) Imagine that the simulated annealing algorithm is at node A and has randomly chosen node D is the candidate next state. Assuming the temperature equals 5, what is the probability that node D will be accepted as the next state?

Repeat the above, but this time assume node E is the candidate next state.

## PROBLEM 3 - Game Playing (8 points)

a) Apply the *minimax* algorithm to the partial game tree below, where it is the *minimizer's* turn to play. The values estimated by the static-board evaluator (SBE) are indicated in the leaf nodes. Write the estimated values of the intermediate nodes inside their circles, and indicate the proper move of the minimizer by circling one of the root's outgoing arcs.




b) Indicate, by crossing out, *one* (1) unnecessary call to the static-board evaluator. Explain why this call to the SBE is unnecessary.

## PROBLEM 4 - Propositional Logic (20 points)

a) Given the *interpretation P=true and Q=false*, determine the truth value of the following:

$$[(P \vee Q) \Rightarrow Q] <=> [P \wedge \neg Q]$$

b) Using a truth table, show that the following is *valid* (i.e., is true for all possible interpretations for *P* and *Q*).

$$(P \wedge Q) <=> \neg(\neg P \vee \neg Q)$$

c) Given the well-formed formulae (wff's) below, show that *(S ∨ R)* logically follows (don't do more than 10 deductive steps):

| | WFF | Justification |
|---|---|---|
| 1 | P | given |
| 2 | (Q ∧ W) | given |
| 3 | (P ∧ Q) ⇒ (S ∨ Z) | given |
| 4 | (A ∨ R ∨ ¬ Z) | given |
| 5 | ¬ A | given |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | | |
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |

## PROBLEM 5 - Miscellaneous Short Answers (10 points)

Provide brief answers to the following questions.

---

A. Why do we use *test sets* in machine learning? What is wrong with using the *same* subset of data for one's *pruning set* and *test set* in ID3?

---

B. If a solution of length $N$ is known to exist, is *beam-search* with beam-width=$2N$ guaranteed to find it?

---

C. Express the following sentence in predicate calculus:

*Every red ball is made of rubber.*

---

**The End!**

## Problem 6 - Information Theory and Expected Values (12 points)

Assume that in your morning email each day you get a message from Amazing.com. The last line of each message always says how many free dollars you have for use at their site that day. You're told that 95% of the time, you'll get $1 and 5% of the time you'll get $10. (Be sure to show your work when answering the questions below.)

a) What is the expected dollar value of your daily message?

b) How many bits of information (about the free dollars awarded) do you expect to receive in each day's email from Amazing.com?

c) You've discovered that whenever the phrase "big sale" appears in the email's subject header, which happens half the time, you will always receive only the $1 award. How much information (in bits), about the award contained in the email, does this discovery provide you?

**The End!**