CS 540: Introduction to Artificial Intelligence

Midterm Exam: 7:15-9:15 pm, October 26, 2011
Room B371 Chemistry

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 ***five*** problems on ***eight*** pages.

Name ____

Student ID ____

Problem	**Score**	**Max Score**
1	______	**25**
2	______	**25**
3	______	**13**
4	______	**25**
5	______	**12**
TOTAL	______	**100**

Initials: ________________________________

Problem 1 – Decision Trees (25 points)

Assume that you are given the set of labeled *training examples* below, where each of three features has three possible values: *a, b,* or *c.* You choose to learn a decision tree from this data.

	F1	*F2*	*F3*	*Output*
ex1	*c*	*b*	*b*	+
ex2	*a*	*a*	*c*	+
ex3	*b*	*c*	*c*	+
ex4	*b*	*c*	*a*	-
ex5	*a*	*b*	*c*	-
ex6	*c*	*a*	*b*	-

a) What score would the *information gain* calculation assign to feature *F1,* when deciding which feature to use as the root node for a decision tree being built?
Be sure to show all your work (on this and all other questions).

b) Consider the following *new* candidate test for a node in the decision tree. What would be its information gain on the above data set when considered as the root node?

Does F2 = F3?

c) In HW1, a tree-pruning algorithm addressed the overfitting issue in machine learning. Which *search strategy* did that algorithm implicitly employ? Explain your answer.

d) Assume you want to exploit the fact that you have lots of computer cycles available in order to learn a better decision tree for a given set of training examples. Briefly describe one way you might use a more extensive search strategy. Be sure to state a specific search strategy. Justify your choice. (Note: you cannot use *decision forests* as an answer since it is the topic of the next question.)

e) You want to produce an ensemble of 11 learned decision trees (a *decision forest*). Describe one way to *reduce* the chances that all 11 learned decisions trees are exact copies of one another. Informally explain how your answer encourages different decision trees to be learned from the same training set.

Initials: ______________________________

Problem 2 – Search (25 points)

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

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

Best-First (using f = h)

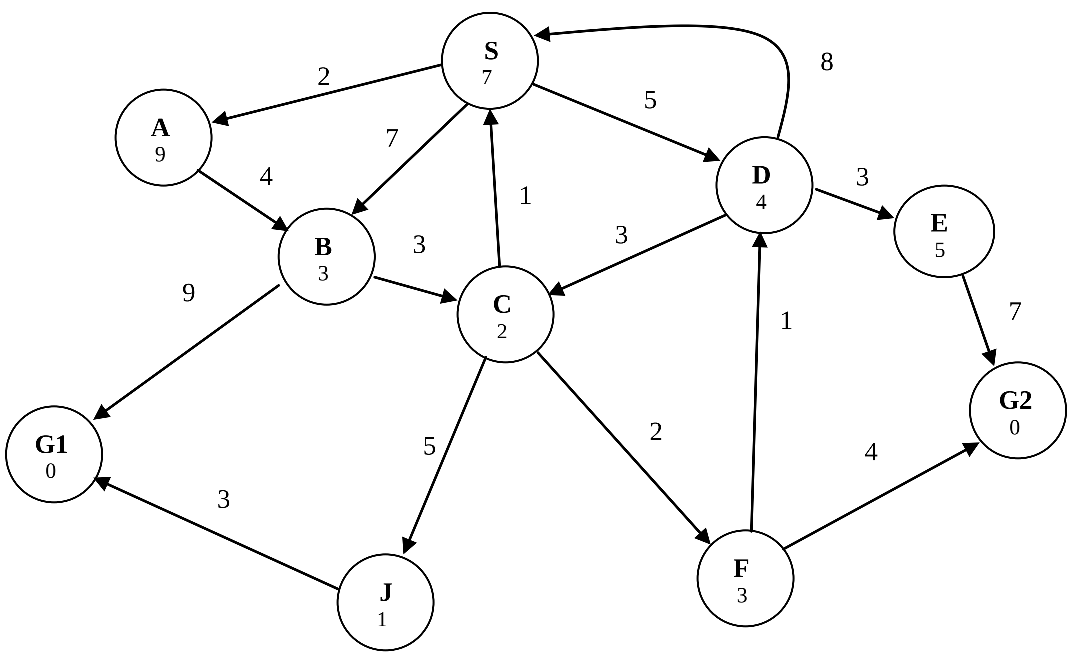
Goal state reached: ________ States popped off OPEN: ________________________________

Iterative Deepening

Goal state reached: ________ States popped off OPEN: ________________________________

A*

Goal state reached: ________ States popped off OPEN: ________________________________



b) Using the same search space as in Part *a*, consider using *Simulated Annealing* as your search strategy. Assume the current *temperature* is 6.

If you are at Node *D* and simulated annealing has randomly selected node *S* for consideration, what is the probability that moving to this node is accepted?

c) Now imagine that you wish to run a *Genetic Algorithm* on Part *a*'s search space. You use four bits to represent nodes: *A* = 0001, *B* = 0010, *C* =0 011, *D* = 0100, *E* = 0101, *F* =0 110, *G1* =1 001, *G2* = 1110, *J* = 0111, and *S* = 0000.

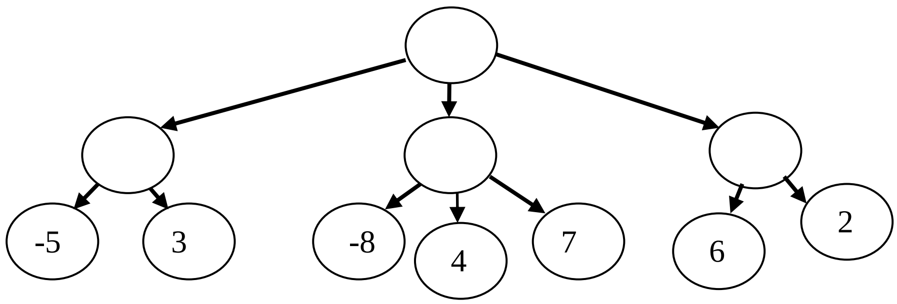
i. If nodes *B* and *E* are chosen for the cross-over operator, show two possible children that can be produced.

ii. With 4 bits one can represent 16 distinct nodes, but only 10 are in this task's search problem. What do you think should be done when a bit string is generated that matches none of the nodes?

Problem 3 – Game Playing and Expected Values (13 points)

a) Apply the *minimax* algorithm to the partial game tree below, where it is the **minimizer**'s turn to play and the game does not involve randomness. The values estimated by the static-board evaluator (SBE) are indicated in the leaf nodes (higher scores are better for the maximizer).

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) List one leaf node in the above game tree whose SBE score need not be computed. Explain why.

c) Consider the two scenarios below. Which would you prefer? Explain why.

i. You get one lottery ticket that wins $10 1% of the time, wins $2 10% of the time, and wins $0.50 (i.e., 50 cents) 89% of the time.

ii. You get two lottery tickets, *A* and *B*. Half the time ticket *A* will win you $1; otherwise you win nothing. Half the time ticket *B* says "you double your winnings from ticket A"; otherwise ticket *B* says you win $0.10 (i.e., 10 cents).

Initials: ______________________________

Problem 4 – Probably Miscellaneous Questions (25 points)

What is <u>one</u> strength of *best-first* search compared to *uniform-cost* search?
<u>One</u> weakness? Briefly explain your answers.

a) <u>ONE Strength (of best-first search compared to uniform-cost search)</u>

b) <u>ONE Weakness</u>

---

c) Briefly, why do we need <u>both</u> a *tuning* set and a *testing* set in machine learning?

Assume we are given this *joint probability distribution* involving random events *A* and *B*:

A	**B**	**Prob**
F	F	0.2
F	T	0.4
T	F	0.1
T	T	0.3

d) What is *P(A)*? ______________________

e) What is *P(B | A)*? ______________________

We are told *P(A)* = 0.4 and *P(B)* = 0.7.

f) If *A* and *B* are independent, what is *P(A and B)*? ______________________________

Now assume we do *not* know anything about the independence of *A* and *B*.

g) What is the <u>largest</u> possible value for *P(A and B)*? ______________________
Explanation (hint: think about Venn Diagrams):

h) What is the <u>smallest</u> possible value for *P(A and B)*? ______________________
Explanation (hint: again, think about Venn Diagrams):

Problem 5 – Key AI Concepts (12 points)

Briefly describe each of the following AI concepts and explain each's significance.

Admissibility

Description:

Significance:

Horizon Effect

Description:

Significance:

Temperature(in the Simulated Annealing Algorithm)

Description:

Significance:
