CSE 573 FALL 1999 FINAL EXAM

Due: Wed. Dec 15 at 12:30pm

Turnin: Turn in your exam to Adam in his office at C109B. If he's not there, put it in his mailbox in room 127 and send him an email so he can get pick it up.

Instructions

This is an open book final. That means you can use your notes, the text or any other source of information short of searching the web for answers to the question.

This exam has 14 problems on 14 pages for a total of 100 points, plus 18 bonus points.

Don't discuss the problems with anyone except Pedro & Adam.

Write your name on every page.

For all problems, justify your answer. Please write clearly and concisely. Use the back of the page and additional pages if you need to.

Good luck.

Question 1

(8 + 4 bonus points)

Recall that the heuristic function in best-first search is $f(n) = g(n) + h(n)$, where $g(n)$ is the exact cost of getting to the current node n, and $h(n)$ is the estimated minimum cost of getting from n to a goal state.

- **a** (4 points) Suppose we run a greedy search algorithm with $h(n) = -g(n)$. What sort of search will the greedy search emulate?
- **b** (4 points) Prove that if the heuristic function h obeys the triangle inequality, then the f-cost along any path in the search tree is nondecreasing. (The triangle inequality says that the sum of the costs from A to B and B to C must not be less than the cost from A to C directly.)
- **c** (BONUS 4 points) Sometimes there is no good evaluation function for a problem, but there is a good comparison method: a way to tell if one node is better than another, without assigning numerical values to either. Show that this is enough to do a best-first search. What properties of best-first search do we give up if we only have a comparison method?

Question 2

(8 + 4 bonus points)

Consider the following map coloring problem:

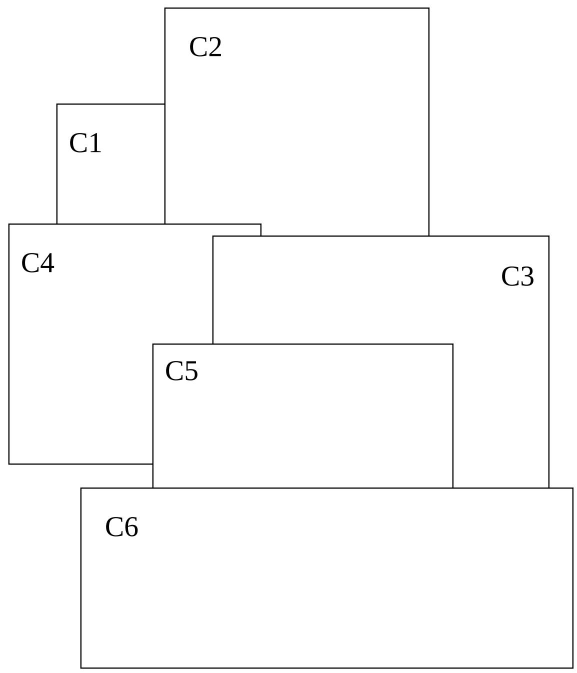


Figure 1: Assign each region one of the three colors (Red, Green or Blue) so that no two adjacent regions have the same colors.

- **a** (8 points) In class, we studied two heuristics for CSPs, *least-constrining-value* and *most-constrained-variable*. Solve the graph coloring problem above using these two heuristics and forward checking. (Show work on back)
- **b** (BONUS 4 points) Describe an additional heuristic that would be useful in solving this problem.

Question 3

(6 points)

Let us consider the problem of search in a *three-player* game. (You can assume no alliances are allowed.) We will call the players 0, 1, and 2 for convenience. Assume you have an evaluation function that returns a list of three values, indicating (say) the likelihood of winning for players 0, 1 and 2, respectively. Complete the following game tree by filling in the backed-up values for all remaining nodes including the root.

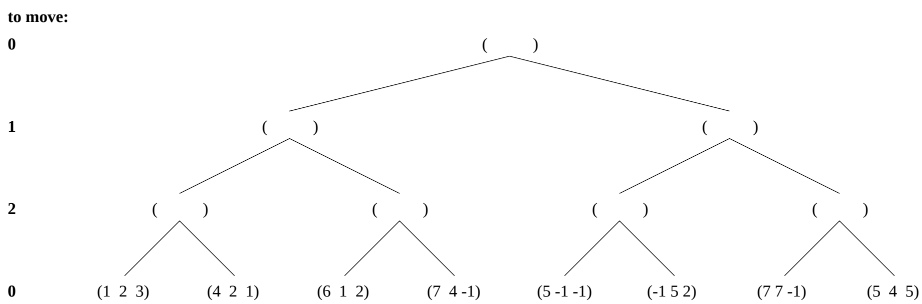


Figure 2: The first three ply of a game tree with three players (0, 1, and 2).

Question 4

(4 points)

Given the following, can you prove that the unicorn is mythical? How about magical? Horned?

> If the unicorn is mythical, thin it is immortal, but if it is not mythical, then it is a mortal mammal. If the unicorn is either immortal or a mammal, then it is horned. The unicorn is magical if it is horned.

Question 5

(6 points)

Consider a world in which there are only four propositions, A, B, C, and D. How many models are there for the following sentences?

a (2 points) $A \wedge B$

b (2 points) $A \vee B$

c (2 points) $A \wedge B \wedge C$

Question 6

(10 + 2 bonus points)

Here are two sentences in the language of first-order logic:

(A) $\forall x \exists y (x \geq y)$

(B) $\exists y \forall x (x \geq y)$

a (2 point) Assume that the variables range over all natural numbers $0, 1, 2, \ldots, \infty$, and that the "$\geq$" predicate means "greater than or equal to." Under this interpretation, translate the these sentences into English.

b (1 point) Is (A) true under this interpretation?

c (1 points) Is (B) true under this interpretation?

d (2 points) Does (A) logically entail (B)?

e (2 points) Does (B) logically entail (A)?

f (2 points) Try to prove that (A) follows from (B) using resolution. Do this even if you think that (B) does not logically entail (A); continue until the proof breaks down and you cannot proceed (if it does break down). Show the unifying substitution for each resolution step. If the proof fails, explain exactly where, how and why it breaks down.

g (BONUS 2 points) Now try to prove that (B) follows from (A).

Question 7

(15 + 5 bonus points)

Two astronomers, in different parts of the world, make measurements M_1 and M_2 of the number of stars N in some small region of the sky, using their telescopes. Normally, there is a small possibility of error by up to one star. Each telescope can also (with a slightly smaller probability) be badly out of focus (events F_1 and F_2), in which case, the scientist will undercount by three or more stars. Consider the three networks shown below.

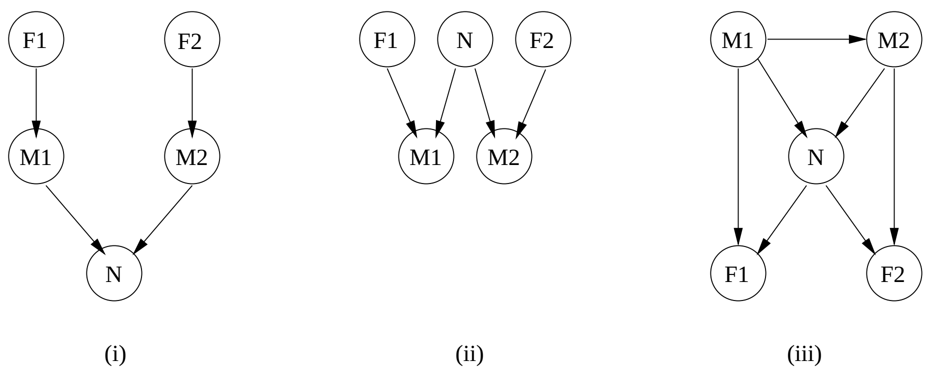


Figure 3: Three possible networks for the telescope problem.

a (5 points) Which of these belief networks correctly (but not necessarily efficiently) represent the above information?

b (5 points) Which is the best network?

c (5 points) Give a reasonable conditional probability table for the values of $\mathbf{P}(M_1|N)$. (For simplicity, consider only the possible values 1, 2, and 3 in this part.)

d (BONUS 5 points) Suppose $M_1 = 1$ and $M_2 = 3$. What are the possible numbers of stars?

Question 8

(10 + 3 bonus points)

Let the instance space be $X = \{0,1\}^4$, the training set be $D = \{(< 0,0,0,0 >, 1)\}$, and the hypothesis space H be the set of conjunctions over X.

a (4 points) Compute the cardinality of the version space of H over D, $|VS_{H,D}|$.

b (3 points) Derive the S and G frontiers using the candidate elimination algorithm.

c (3 points) Suppose you see the additional example $(< 1,1,1,1 >, 0)$. Derive the new S and G frontiers.

d (BONUS 3 points) Suppose you see one more example, $(< 0,1,1,1 >, 1)$. Derive the new S and G frontiers.

Question 9

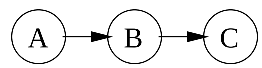
(4 points)

Suppose a training set is made up of 16 examples of class A, 8 examples of class B, 32 examples of class C, and 8 examples of class D. When growing a decision tree from this training set, what is the maximum information gain that any attribute can have?

Question 10

(5 points)

Consider the following Bayesian network, in which variables A, B and C are Boolean:



Suppose you want to learn the parameters for this network using the training set $\{< 0, 1, 1 >, < 1, 0, 0 >, < 1, 1, 1 >, < 1, ?, 0 >\}$, where examples are in the form $< A, B, C >$, and "?" indicates a missing value. Show the sequence of filled-in values and parameters produced by the EM algorithm, assuming the parameters are initialized by ignoring missing values. (Hint: EM converges very quickly on this problem.)

Question 11

(3 points)

Let H be the set of hypotheses of the form $x = x_0 \vee x = x_1$, where $x_0, x_1 \in X$ (i.e., x_0 and x_1 are arbitrary elements of the instance space). What is the VC dimension of H?

Question 12

(5 points)

Suppose you want to learn to recognize digits in a 7-segment LED display from noisy examples (i.e., each segment has been flipped with 10% probability). Which of the learning algorithms you studied would you use?

Question 13

(8 points)

How would you modify Graphplan to handle uncertainty (i.e., the initial state is a set of worlds) assuming that actions are still STRIPS (i.e., no sensing actions are available)?

Question 14

(8 points)

Describe how you would design a probabilistic part-of-speech tagger. You may assume you have a tagged corpus (that is, a large body of text in which each word is already labelled with its correct part of speech.) Describe the objects you will compute probabilities over, and how you will use those probabilities to predict the tags for unseen text. (Your scheme can be as simple as you want, but describe its shortcomings.)