Artificial Intelligence Qualifying Example

Fall 2019

ID:

1 Search (15 points)

For this question, we will consider the behavior of the A^* algorithm, and we will consider two different admissible heuristics, h_1 and h_2. For convenience, you may want to use the notation $c^*(s)$ to represent to the cost of getting from state s to the lowest cost goal state reachable from s under the optimal (lowest cost) path from s. You may assume that the state space has a tree structure.

1.1 Combinations of Heuristics (5 points)

Prove that for any $0 \leq \alpha_1 \leq 1$ and $0 \leq \alpha_2 \leq 1$ with $\alpha_1 + \alpha_2 = 1$, $h_3 = \alpha_1 h_1 + \alpha_2 h_2$ must be an admissible heuristic.

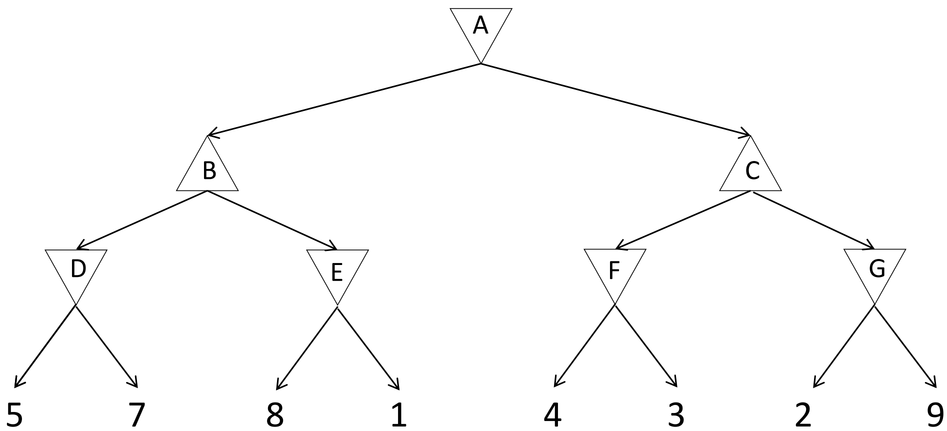
1.2 Choosing Heuristics (10 points)

Assume that $\forall s : h_2(s) \geq h_1(1)$. Prove that it is *impossible* for A^* to expand more nodes using h_2 than it would have expanded if it used h_1. Hint: Try doing a proof by contradiction. Assume that there exists some state that is expanded by A^* using h_2 that is *not* expanded by A^* using h_1. Use your knowledge about the performance and operation of the A^* algorithm, as well as the assumptions stated in this question, to show that this is impossible.

Note: There is a slight caveat to the claim we have asked you to prove - in the case of ties, it may require additional assumptions about how ties are broken. To deal with this, you can assume that if there is a goal node in the priority queue that has the same f value as other nodes in the queue, A^* *luckily* pops the node that satisfies the goal test first.

2 Game Search (15 points)

For this question, we will consider the following game tree. **Note that the root node is a min node and that max nodes are interleaved between min nodes!** As is customary, we will assume that this tree is searched in a depth first, left to right order.



2.1 Minimax Value I (2 points)

Write the minimax value of the root node in the box below:

2.2 Minimax Value II (3 points)

Suppose you could change the value of the leaf node which currently has value 3 to any value (even ∞). What is the the largest value the root node could take assuming that all other leaf nodes remain unchanged. Write your answer in the box below.

2.3 Minimax Value III (3 points)

Suppose you could change the value of the leaf node which currently has value 3 to any value (even $-\infty$). What is the the smallest value the root node could take assuming that all other leaf nodes remain unchanged. Write your answer in the box below.

2.4 Alpha-Beta pruning I (3 points)

Write a list of edges that would be pruned by alpha-beta in the box below. Use the node labels or a combination of node label and leaf values to indicate which edges are pruned. For example, AB would indicate that the edge between the root and $\max$ node labeled B would be pruned, and D3 would indicate that the edge between the $\min$ node and the leaf with value 3 would be pruned. *If an edge is pruned, you should NOT list edges beneath it.*

2.5 Alpha-Beta pruning II (4 points)

Although it cannot change the value of the root node, the order in which siblings are expanded can significantly influence the efficiency of alpha-beta pruning. Propose *one or more* reorderings of siblings that will maximize the amount of pruning achieved by alpha-beta. Indicate swaps writing down pairs of siblings that are reordered. For example, if you want to reorder the children of the G node, would you indicate $(1, 9)$ to indicate that the 1 and 9 are swapped. Similarly, if you wanted to reorder the children of the B node, you would indicated (D, E) to indicated that D and E are swapped. Write your answer in the box below:

3 Bayes Nets (15 points)

Consider a Bayesian network with *binary* random variables $X_0 \ldots X_n$. Variable X_0 has no parents. For $1 \leq i \leq n$, variable X_i has variable X_{i-1} as its only parent.

3.1 Atomic Events (2 points)

How many atomic events (sometimes also called states) are representable by the variables in this network? Write your answer as a function of n in the box below:

3.2 Size of the Joint Distribution (2 points)

Without taking advantage of the conditional independence assumptions encoded in the network structure, how many numbers would be required to represent the joint distribution over $X_0 \ldots X_n$? Note that your answer should take into account that some probabilities can be inferred from others. For example, a coin toss corresponds to two random events, *heads* and *tails*, but you need only one number to specify the probability distribution because $P(tails) = 1 - P(heads)$. Write your answer as a function of n in the box below:

3.3 Exploiting Conditional Independence (2 points)

Taking full advantage of the conditional independence assumptions encoded in the network structure, how many numbers would be required to represent the joint distribution over $X_0 \ldots X_n$? (In other words, how many numbers are required to specify all of the conditional probabilities in the Bayesian network?) Note that your answer should take into account that some probabilities can be inferred from others. For example, a coin toss corresponds to two random events, *heads* and *tails*, but you need only one number to specify the probability distribution because $P(tails) = 1 - P(heads)$. Write your answer as a function of n in the box below.

3.4 Marginalization I (3 points)

What is the computational complexity of computing $P(X_n)$? Write your answer in big O notation in the box below:

3.5 Marginalization II (3 points)

Suppose variable X_1 is observed (the value is known). What is the computational complexity of computing $P(X_n)$? Write your answer in big O notation in the box below:

3.6 Marginalization III (3 points)

Suppose variable X_{n-1} is observed (the value is known). What is the computational complexity of computing $P(X_n)$? Write your answer in big O notation the box below:

4 Value of Perfect Information (15 points)

You are given the opportunity to play the following game: There is a box with three visually identical coins inside. Coin A has 100% chance of coming up heads when flipped. Coin B has 100% chance of coming up tails when flipped. Coin C is unbiased, i.e., 50% chance of heads and 50% chance of tails. The rules of the game are that first a coin is chosen from the box uniformly at random, i.e, $P(A) = P(B) = P(C)$. Next, the coin is flipped. If it comes up heads you win \$10. If it comes up tails, you lose \$10. Note that for this question we will assume that there is a $1 : 1$ relationship between money and utility, so we can talk about expected utility or expected monetary value interchangeably.

For the questions below, **even if your answer is correct, you may not get full credit if you don't show your calculations explicitly.**

4.1 Expected Value (5 points)

What is the expected value (expected gain in utility) for playing this game? Write your answer in the box immediately below, but use the rest of the page to show your work.

4.2 Value of Perfect Information (10 points)

Suppose you are able to back out of playing the game after the coin is selected but before it is flipped. Without any ability to know which coin was selected, this option really is no different from the original description of the game. Assume, however, that you have a friend who is a physics major, and that this friend has invented a device that can estimate the type of a coin from a distance. Suppose the device is perfectly accurate. How much would you be willing to pay to use the device once? Write your answer in the box immediately below, but use the rest of the page to show your work.

5 Markov Decision Processes (20 points)

Consider the simple, 1D grid world as shown below. Each square is a state in an MDP. States without a reward shown have reward 0. There are just two actions in this problem, Right (R) and Left (L). Both move in the intended direction with probability 1. At the extreme right and extreme left, if the agent bounces into a wall, it stays in the same location. This means that the agent collect the rewards at the extreme states ad infinitum. Note that the reward function is defined over states only, i.e., $R(s)$, not $R(s, a)$ and not $R(s, a, s')$. The discount factor for this problem is 0.5. (This is much smaller than a typical discount factor, but was chosen to make the calculations simple.)

+1				+4

5.1 Policy Value I (5 points)

Using the empty grid below, write down the value of all states under the policy that goes *right* in all states.

5.2 Policy Value II (5 points)

Using the empty grid below, write down the value of all states under the policy that goes *left* in all states.

5.3 Optimal Value Function I (5 points)

Using the empty grid below, write down the value of all states under the optimal policy.

5.4 Optimal Value Function II (5 points)

In the space below, demonstrate your understanding of the Bellman equation by writing down the Bellman equation showing explicitly that the value you have computed for the starred state satisfies the Bellman equation, i. e., value iteration would not change the value of this state.

		*		

6 HMMs (20 points)

Consider a Hidden Markov Model with n states and m observations (sometimes called emissions or outputs). Assume that for each state, at most k next states have positive probability. For the following questions, if you are asked to use big O notation you should give the tightest possible bounds given the available information about the problem, but without making any additional assumptions. Unless otherwise stated, you may assume a uniform distribution on the states at time 0. Unless otherwise stated, you should assume that an observation is made for each time step after time 0.

6.1 Paths Through the State Space (2 points)

A *path* through the state space of length t is a sequence of states of the length t where the i^{th} element of the sequence corresponds to the state at time i. How many possible length t paths are there? Write your answer in big O notation in the box below:

6.2 Viterbi Path (4 points)

The Viterbi path is the highest probability path through the state space. What is the computational complexity of finding the Viterbi path given an initial state distribution and sequence of t time steps and t observations. Write your answer in big O notation in the box below:

6.3 State Distribution (2 points)

How many numbers are required to represent the state distribution at a particular time step? Note that your answer should take into account that some probabilities can be inferred from others. For example, a coin toss corresponds to two random events, *heads* and *tails*, but you need only one number to specify the probability distribution because $P(tails) = 1 - P(heads)$. Write your answer as a number (not big O) in the box below.

6.4 Tracking (4 points)

Given an initial state distribution, sequence of t time steps and t observations, what is the computational complexity of computing the distribution over states at time t? Write your answer in big O notation in the box below.

6.5 Hindsight (4 points)

Given an initial state distribution, sequence of t time steps and t observations, what is the computational complexity of computing the distribution over states at time 1, taking into account all observations through time step t. Write your answer in big O notation in the box below.

6.6 A Simpler Case (4 points)

Describe a condition on the observation distribution, $P(O|S)$, that would allow a constant time, i.e., $O(1)$, calculation of the state distribution at any time step.