CSE 473 Midterm Exam – Feb 8, 2018

Name:

This exam is take home and is due on **Wed Feb 14 at 1:30 pm**. You can submit it online (see the message board for instructions) or hand it in at the beginning of class. This exam should not take significantly longer than 3 hours to complete if you have already carefully studied all of course material. Studying while taking the exam may take longer. :)

This exam is open book and open notes, but you must complete all of the work yourself with no help from others. Please feel free to post clarification questions to the class message board, but please do not discuss solutions there.

Partial Credit: If you show your work and *briefly* describe your approach to the longer questions, we will happily give partially credit, where possible. We reserve the right to take off points for overly long answers. Please do not just write everything you can think of for each problem.

Name: Please do not forget to write your name in the space above!

Question 1 – True/False – 30 points

Circle the correct answer each True / False question.

1. True / False – Reflex agents cannot act optimally (in terms of maximizing total expected reward over time). (3 pt)

2. True / False – Minimax is optimal against perfect opponents. (3 pt)

3. True / False – Greedy search can take longer to terminate than uniform cost search. (3 pt)

4. True / False – Uniform cost search with costs of 1 for all transitions is the same as depth first search. (3 pt)

5. True / False – Alpha-Beta pruning can introduce errors during mini-max search. (3 pt)

6. True / False – Each state can only appear once in a state graph. (3 pt)

7. True / False – Policy Iteration always find the optimal policy, when run to convergence. (3 pt)

8. True / False – Higher values for the discount (γ) will, in general, cause value iteration to converge more slowly. (3pt)

9. True / False – For MDPs, adapting the policy to depend on the previous state, in addition to the current state, can lead to higher expected reward. (3pt)

10. True / False – Graph search can sometimes expand more nodes than tree search. (3pt)

Question 2 – Short Answer – 30 points

These short answer questions can be answered with a few sentences each.

1. Short Answer – Briefly describe the relationship between admissible and consistent heuristics. When would you use each, and why? (5 pts)

2. Short Answer – Briefly describe when you would use Alpha-beta pruning in minimax search. (5 pts)

3. Short Answer – For Q-learning, when would you prefer to use linear function approximation and when would you just use the tabular version? Is there ever any drawback to using the linear version? (5 pts)

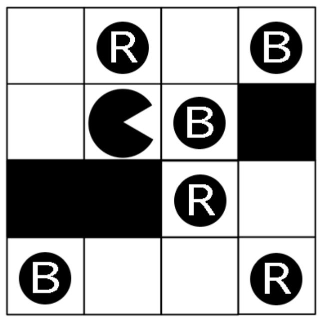
4. Short Answer – Briefly describe the difference between UCS and A* search. When would you prefer to use each, and why? (5 pts)

5. Short Answer – For Q-learning, briefly describe the conditions needed to ensure convergence. Is it guaranteed for any exploration policy? (5 pts)

6. Short Answer – Briefly describe the difference between value iteration and policy iteration. Describe conditions under which one algorithm might be preferred to the other, in practice. (5 pts)

Question 3 – Ordered Pacman Search – 25 points

Consider a new Pacman game where there are two kinds of food pellets, each with a different color (red and blue). Pacman has peculiar eating habits; he strongly prefers to eat all of the red dots before eating any of the blue ones. If Pacman eats a blue pellet while a red one remains, he will incur a cost of 100. Otherwise, as before, there is a cost of 1 for each step and the goal is to eat all the dots. There are K red pellets and K blue pellets, and the dimensions of the board are N by M.



$K = 3$, $N = 4$, $M = 4$

1. Give a non-trivial upper bound on the size of the state space required to model this problem. Briefly describe your reasoning. [10 pts]

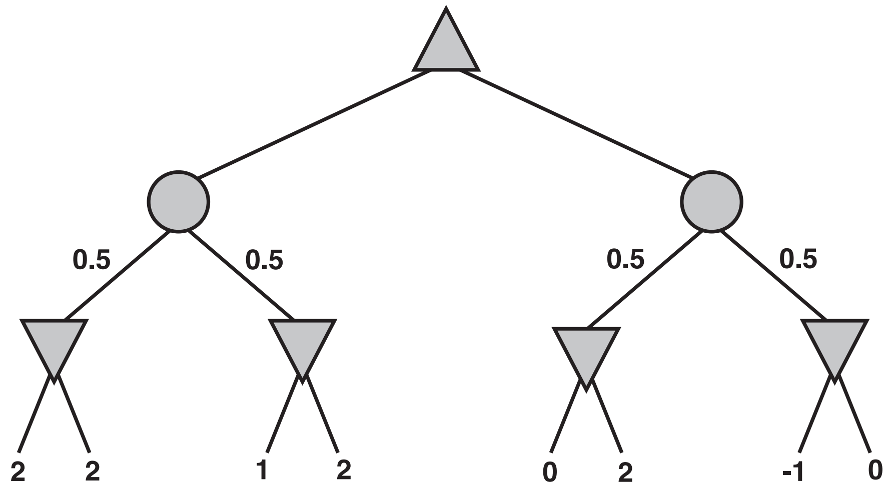
2. Give a non-trivial upper bound on the branching factor of the state space. Briefly describe your reasoning. [5 pts]

3. Name a search algorithm pacman could execute to get the optimal path? Briefly justify your choice (describe in one or two sentences) [5 pts]

4. Give an admissible heuristic for this problem. [5 pts]

Question 4 – Game Trees – 30 points

Consider the following game tree, which has min (down triangle), max (up triangle), and expectation (circle) nodes:



1. In the figure above, label each tree node with its value (a real number). [7 pts]

2. In the figure above, circle the edge associated with the optimal action at each choice point. [7 pts]

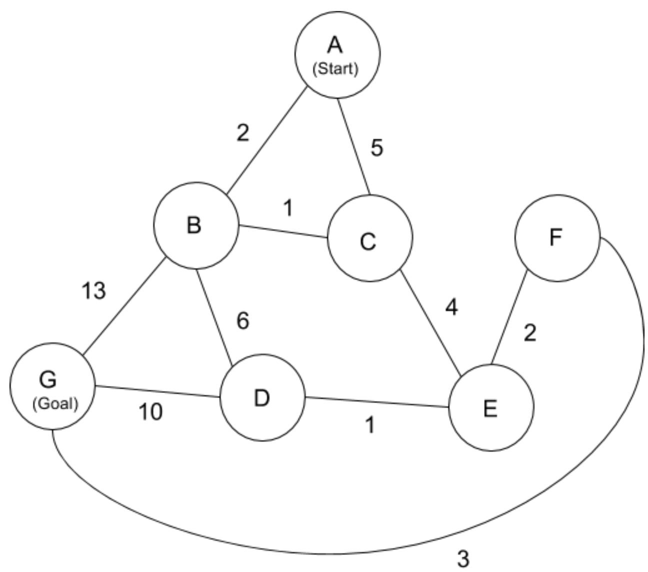
3. If we knew the values of the first six leaves (from left), would we need to evaluate the seventh and eighth leaves? Why or why not? [5 pts]

4. Suppose the values of leaf nodes are known to be in the range $[-2, 2]$, inclusive. Assume that we evaluate the nodes from left to right in a depth first manner. Can we now avoid expanding the whole tree? If so, why? Circle all of the nodes that would need to be evaluated (include them all if necessary). [11 pts]

Question 5 – Tree Search – 30 points

Given the state graph below, run each of the following algorithms and list the order that the nodes are expanded (a node is considered expanded when it is dequeued from the fringe). The values next to each edge denote the cost of traveling between states.

Use alphabetical ordering to break ties (i.e. A should be before B in the fringe, all of things being equal). It is also possible that a state may expanded more than once. However, you should use cycle checking to ensure you do not go into an infinite loop (e.g. never expand the same state twice in a single plan from the root to a leaf node). Every ordering should always start with the start node and end with the goal node.



1. Breadth first search [5 pts]

2. Depth first search [5 pts]

3. Iterative deepening [5 pts]

4. Uniform cost search [5 pts]

Now, consider the following two heuristics:

State s	H1(s)	H2(s)
A (start)	10	12
B	8	11
C	7	8
D	4	4
E	3	4
F	2	3
G (goal)	0	0

5. Provide the expansion ordering for A* search with heuristic H2 (again breaking ties alphabetically). [5 pts]

6. List which, if any, of the two heuristics are admissible [2.5 pts]

7. List which, if any, of the two heuristics are consistent [2.5 pts]

Question 6 – Stutter Step MDP and Bellman Equations – 25 points

Consider the following special case of the general MDP formulation we studied in class. Instead of specifying an arbitrary transition distribution $T(s, a, s')$, the stutter step MDP has a function $T(s, a)$ that returns a next state s' deterministically. However, when the agent actually acts in the world, it often stutters. It only actually reaches s' half of the time, and it otherwise stays in s. The reward $R(s, a, s')$ remains as in the general case.

1. Write down a set of Bellman equations for the stutter step MDP in terms of $T(s, a)$, by defining $V^*(s)$, $Q^*(s, a)$ and $\pi^*(s)$. Be sure to include the discount γ. [25 pts]

2. Consider the special case of the stutter step MDP where $R(s, a, s')$ is zero for all states except for a single good terminal state, which has reward 1, and a single bad terminal state, with reward -100. Furthermore, assume all states s are connected to both terminal states (there exists some sequence of actions that will go from s to the terminal state with non-zero probability).

   If $\gamma = 1$, briefly describe what the optimal values $V^*(s)$ for all states would look like. [5 pts]

3. Again, set the rewards as in the previous question, but now consider $\gamma = 0.1$ and describe $V^*(s)$. Would the optimal policy $\pi^*(s)$ change? [5 pts]