CSE 473 Midterm Exam – May 1, 2016

Name:

This exam is take home and is due on **Sunday May 8 at 11:45 pm**. You can submit it in the online DropBox or to one of the TAs. 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 answer 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 you name in the space above!

There are 12 pages in this exam (including this one).

Scores						
Q.1 (30)	**Q.2 (30)**	**Q.3 (30)**	**Q.4 (20)**	**Q.5 (30)**	**Q.6 (30)**	**Total (170)**

Question 1 – True/False – 30 points

Circle the correct answer each True / False question.

1. True / False – A* Tree Search requires a consistent heuristic for optimality. (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 search tree. (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 – Using the true cost from any state to the goal is a consistent heuristic. (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 explain the difference between value iteration and policy iteration. Name one advantage or disadvantage of each. (5 pts)

3. Short Answer – Both alpha-beta pruning and depth-limited search can improve the speed of minimax. Suppose you can only choose one or the other. Under what conditions would you prefer alpha-beta, and under what conditions would you prefer depth-limited search?

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

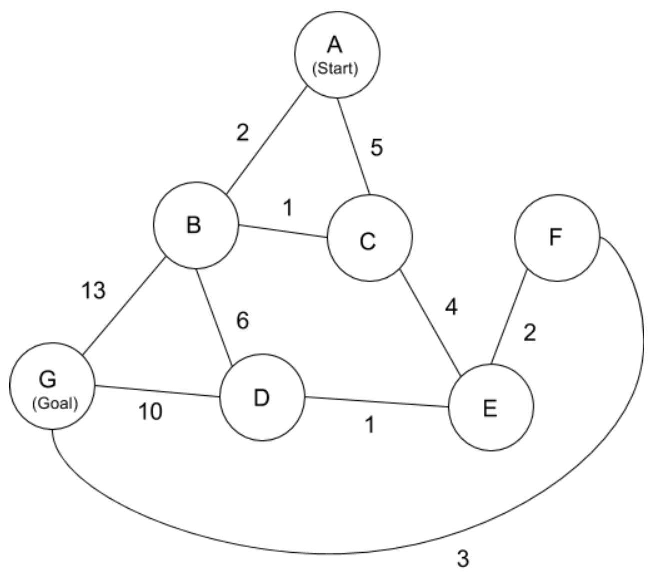
5. Short Answer – Value iteration and expectimax can both be used to solve MDPs. When would you prefer to use each, and why? (5 pts)

6. Short Answer – Can reflex agents act rationally? Please provide a brief example to justify your answer. (5 pts)

Question 3 – 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. 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

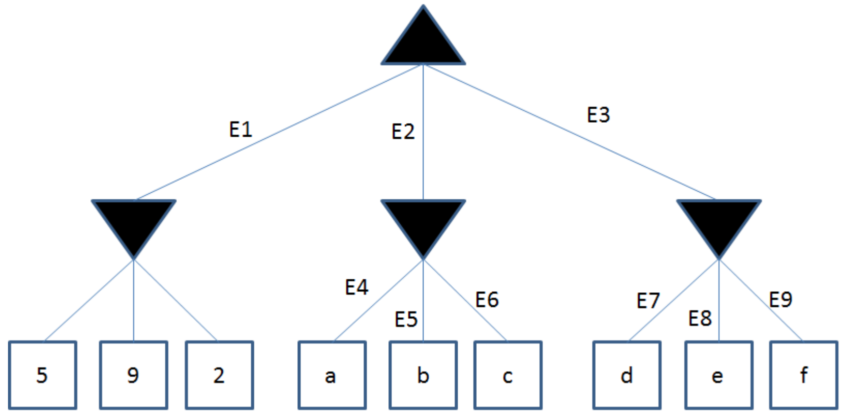
5. Provide the expansion ordering for A* search with heuristic H2 (again breaking ties alphabetically). [5 pts]

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 4 – Game Trees – 20 points

In the following game tree, triangles pointing up are maximizing nodes and triangles pointing down are minimizing nodes.



In this problem, you need to fill in numeric values for the letters (a, b, c, d, e, and f) to cause different subtrees to be pruned during alpha-beta search, or to say if no such values are possible (by writing "not possible" to the right of the blanks).

1. E7 is pruned (and no others) [5 pts]

   a:____ b:____ c:____ d:____ e:____ f:____

2. E5, E6, E8, and E9 are pruned (and no others) [5 pts]

   a:____ b:____ c:____ d:____ e:____ f:____

3. E6 and E9 are pruned (and no others) [5 pts]

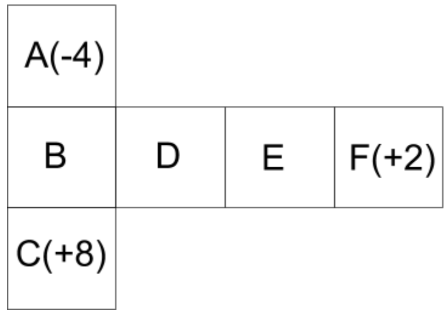
   a:____ b:____ c:____ d:____ e:____ f:____

4. E3 is pruned (along with everything below it, as always) [5 pts]

   a:____ b:____ c:____ d:____ e:____ f:____

Question 5 – Value Iteration – 30 points

Consider the following gridworld MDP shown below where each state is labelled by a letter, and all terminal states are denoted by a positive or negative reward value in parentheses (NOTE: The reward is given after taking the only action EXIT from that terminal state).



There are no living rewards and assume that each legal action moves the actor to the adjacent square in that given direction (North, East, South, West) with 100% success except for state B which has the transition function:

T(B, North, A) = 1.00
T(B, South, C) = 0.75
T(B, South, A) = 0.25
T(B, East, D) = 1.00
T(B, West, B) = 1.00

1. Given a discount factor $\gamma = 0.5$, fill in the values $V^n(s)$ below that the value integration algorithm would compute:[15pt]

n	A	B	C	D	E	F
V_0						
V_1						
V_2						
V_3						

Note: you do not need to show every step of your work and are welcome to write a program to do this or compute it by hand. The transition distribution is designed to minimize the overall computation involved. However, writing out the math for a few steps at the top of the next page will help us assign partial credit if your final numbers are not correct.

(extra space for part one)

2. Write the policy that would be executed given the values V_3 computed above [10pt].

3. The policy that you computed above should assign $\pi(E) = East$, that is it should prefer going east from state E. If this is not the case, please double check your work. What range of discount factors γ (if any) would change optimal action of state E to go West? Briefly explain your reasoning/show your work. As always, please assume $\gamma \in [0, 1]$ [5pt].

Question 6 – Stutter Step MDP and Bellman Equations – 30 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. [20 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]