15-381/781: Artificial Intelligence **Name:** ____________________

Fall 2016

Practice Midterm

80 minutes

This is a practice midterm examination. It was designed to be completed in 80 minutes. **The points assigned to the questions are arbitrary and should be ignored.**

1. (10 points) Consider a heuristic search instance. For any state x, let $h^*(x)$ be the shortest distance between x and goal state t. Let $h(\cdot)$ be a heuristic that overestimates $h^*(\cdot)$ by *exactly* a constant $c > 0$, that is for any state x, $h(x) = h^*(x) + c$.

   Assume that the search instance has a unique optimal path between start state s and goal node t. Show that A* *tree search* only expands the nodes that are on the optimal s-t path.

   Solution: Assume on the contrary that a node x that is not on the optimal path was expanded by A*. Let f^* denote the length of the optimal s-t path. Note that because x is not on the optimal path to t, any s-t path that uses is sub-optimal, so $f^* < \mathrm{cost}(s, x) + \mathrm{cost}(x, t)$.

   Consider the time-step when x was expanded. Let y be the node on the optimal s-t path that was on the frontier at that time. Since, x was chosen to be expanded, then $f(x) \leq f(y)$. That is

$$\begin{aligned} f(x) &\leq f(y) \\ g(x) + h^*(x) + c &\leq g(y) + h^*(y) + c \\ \mathrm{cost}(s, x) + \mathrm{cost}(x, t) + c &\leq f^* + c \\ \mathrm{cost}(s, x) + \mathrm{cost}(x, t) &\leq f^*, \end{aligned}$$

   which contradicts the the earlier statement about $f^* < \mathrm{cost}(s, x) + \mathrm{cost}(x, t)$.

2. (10 points) Consider a robot that is moving in an environment. The goal of the robot is to move from an initial point to a destination point as fast as possible. However, the robot has the limitation that if it moves fast, its engine can overheat and stop the robot from moving. The robot can move with two different speeds: slow and fast. If it moves fast, it gets a reward of 10; if it moves slowly, it gets a reward of 4. We can model this problem as an MDP by having three states: cool, warm, and off. The transitions are shown below. Assume that the discount factor is $\gamma = 0.9999$ and also assume that when we reach the state off, we remain there without getting any reward.

s	a	s'	$T(s, a, s')$
cool	slow	cool	1
cool	fast	cool	1/2
cool	fast	warm	1/2
warm	slow	cool	1/2
warm	slow	warm	1/2
warm	fast	warm	1/2
warm	fast	off	1/2

1. Consider the conservative policy π where the robot always moves slowly. What is the value of $V^\pi(\text{cool})$? Remember that $V^\pi(s)$ is the expected discounted sum of future rewards when starting at state s and following policy π.

   Solution: $V^\pi(\text{cool}) = 4 + 0.9999 * V^\pi(\text{cool})$

   $0.0001 V^\pi(\text{cool}) = 4$

   $V^\pi(\text{cool}) = 40000$

2. What is the optimal policy for each state? Justify your answer.

   Solution: In state Warm, we should go Slow because there is a 50% chance that we reach off. Since the discount factor is near 1, we would be much better off continuing the run and temporarily receiving a lower payoff rather than risking turning off.

   In state Cool, we should go fast because it gives a higher payoff, and given our policy in Warm there is no risk of turning off.

3. Is it possible to change the discount factor to get a different optimal policy? If yes, give such a change and the new optimal policy. If no justify your answer in at most two sentences.

   Solution: Yes, by decreasing the discount factor. For example by choosing the discount factor equal to zero the robot always chooses an action that gives the highest immediate reward.

3. (10 points) Consider an arbitrary MDP M_1 with states S, transition model $T(s, a, s')$, reward model $R(s, a, s')$ and discount factor γ. Now define MDP M_2 as exactly the same MDP, except that its reward model is $R(s, a, s')+d$, where d is a constant (in other words, all the rewards in M_2 are exactly the same as in M_1, except with an additional d added on). Prove that after running to convergence over an infinite horizon, M_2 has exactly the same optimal policy (or optimal policies) as M_1. You can assume that there are no terminal states in M_1 and M_2.

   Solution: We first prove that the optimal Q values in MDP 2 are exactly the same as MDP 1, with an additional fixed offset

$$\begin{aligned}
Q^{M_2}(s,a) &= \sum_{s'} T(s,a,s')(R^{M_2}(s,a,s') + \gamma \max_{a'} Q^{M_2}(s',a')) \\
&= \sum_{s'} T(s,a,s')(R^{M_1}(s,a,s') + d + \gamma \max_{a'} Q^{M_2}(s',a')) \\
&= \sum_{s'} T(s,a,s')(R(s,a,s') + d \\
&+ \gamma \max_{a'} \left[\sum_{s''} T(s',a',s'')(R(s,a,s') + d + \gamma \max_{a''} Q^{M_2}(s'',a'')) \right] \\
&= d \sum_{i=0}^{\infty} \gamma^i + Q^{M_1}(s,a),
\end{aligned}$$

   where the last equality holds by taking out all the terms to do with d.
   Add the same constant to all Q(s,a) values cannot change the relative ordering of actions for a given state (e.g. if $Q(s,a1) > Q(s,a2)$, then $Q(s,a1) + d\sum_{i=0}^{\infty} > Q(s,a2) + d\sum_{i=0}^{\infty}$. Therefore the argmax action (the optimal action) for a state in M2 must be the same as in M1.

4. (5 points) *For this question, recall from class that a heuristic $h(\cdot)$ is said to dominate heuristic $g(\cdot)$, if for all states x, $h(x) \geq g(x)$.*

   Let $h_1(\cdot)$ and $h_2(\cdot)$ be two arbitrary admissible heuristics. Let $h_3 = \frac{2}{3}h_1 + \frac{1}{3}h_2$.

   1. Prove or disprove: h_3 is an admissible heuristic.

      Solution: Prove. h_3 is a convex combination of h_1 and h_2, so $h_3 \leq \max\{h_1, h_2\}$, which is still admissible.

   2. Prove or disprove: h_3 dominates one or both of h_1 and h_2.

      Solution: Disprove. Let there be two states a and b, such that $h_1(a) = 3$, $h_1(b) = 9$, and $h_2(a) = 9$ and $h_2(b) = 3$. Then $h_3(a) = 5$ and $h_3(b) = 7$ does not dominate either of h_1 or h_2.

5. (5 points) A patient has tested positive for the disease x. Let pos/neg indicate testing positive or negative for the disease respectively, and let $x/\neg x$ indicate having disease x or not. You are given the following data:

   - If a person has the disease, the probability of testing positive is $p(pos|x) = 0.99$.
   - If a person does not have the disease, the probability of testing negative is $p(neg|\neg x) = 0.99$.
   - One out of ten thousand people have x, so the probability of having x prior to testing is $p(x) = 0.001$.

   Write the formula for computing $p(x|pos)$, the probability of having the disease x given that the patient tested positive. You do not have to calculate the final value, but your answer should be in terms of raw numbers and mathematical operators like plus, minus, times and division, so that we could immediately evaluate it using a calculator.

   Solution:

$$p(x|pos) = \frac{p(pos|x) \cdot p(x)}{p(pos|x) \cdot p(x) + p(pos|\neg x)p(\neg x)}$$

6. (6 points) Prove or disprove: Consider a STRIPS planning problem where every operation has no preconditions, and for every condition there is some operation that has it as a postcondition. Then in this problem's planning graph, all of the conditions will appear in level S_1 (the second level of conditions), and there are no mutex relations on that level except between pairs of conditions such that one is the negation of the other.

   Solution: False. Suppose there are two conditions a, b and both are false initially. There is one operation that makes both of the conditions true. Then all four conditions (a, b, $\neg a$, $\neg b$) appear on level S_1, but a is mutex with $\neg b$, and b is mutex with $\neg a$ (due to the inconsistent support condition).