CS 188 Fall 2014 — Introduction to Artificial Intelligence — Written HW8

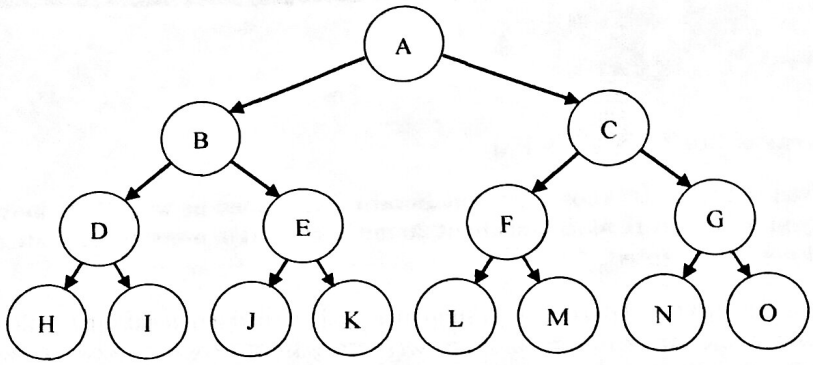
INSTRUCTIONS

- **Due:** Monday, November 3rd, 2014 11:59 PM
- **Policy:** Can be solved in groups (acknowledge collaborators) but must be written up individually. However, we strongly encourage you to first work alone for about 30 minutes total in order to simulate an exam environment. Late homework will not be accepted.
- **Format:** You must solve the questions on this handout (either through a pdf annotator, or by printing, then scanning; we recommend the latter to match exam setting). Alternatively, you can typeset a pdf on your own that has answers appearing in the same space (check edx/piazza for latex templating files and instructions). **Make sure that your answers (typed or handwritten) are within the dedicated regions for each question/part. If you do not follow this format, we may deduct points.**
- **How to submit:** Go to www.pandagrader.com. Log in and click on the class CS188 Fall 2014. Click on the submission titled Written HW 8 and upload your pdf containing your answers. If this is your first time using pandagrader, you will have to set your password before logging in the first time. To do so, click on "Forgot your password" on the login page, and enter your email address on file with the registrar's office (usually your @berkeley.edu email address). You will then receive an email with a link to reset your password.

Last Name	Lee
First Name	Charles
SID	23728976
Email	charles.lee.94@berkeley.edu
Collaborators	daniel he

Q1. [20 pts] Occupy Cal

You are at Occupy Cal, and the leaders of the protest are deciding whether or not to march on California Hall. The decision is made centrally and communicated to the occupiers via the "human microphone"; that is, those who hear the information repeat it so that it propagates outward from the center. This scenario is modeled by the following Bayes net:



A	$P(A)$
$+m$	0.5
$-m$	0.5

$\pi(X)$	X	$P(X\|\pi(X))$
$+m$	$+m$	0.9
$+m$	$-m$	0.1
$-m$	$+m$	0.1
$-m$	$-m$	0.9

Each random variable represents whether a given group of protestors hears instructions to march ($+m$) or not ($-m$). The decision is made at A, and both outcomes are equally likely. The protestors at each node relay what they hear to their two child nodes, but due to the noise, there is some chance that the information will be misheard. Each node except A takes the same value as its parent with probability 0.9, and the opposite value with probability 0.1, as in the conditional probability tables shown.

(a) [4 pts] Compute the probability that node A sent the order to march ($A = +m$) given that both B and C receive the order to march ($B = +m$, $C = +m$).

Put your answer to 1a here:

$$\frac{.5 \cdot .9^2}{.5 \cdot .9^2 + .5 \cdot .1^2} = \frac{.9^2}{.9^2 + .1^2} = .988$$

(b) [4 pts] Compute the probability that D receives the order $+m$ given that A sent the order $+m$.

Put your answer to 1b here:

$$P(D=+ \mid A=+) \quad P(D=+, B=b, A=+)$$

$$\frac{.9^2 + .1^2}{.1^2 + .1^2 + 2(.9)(.1)} \qquad \frac{P(D+ \mid B=b)\, p(B=b \mid A+)}{P(D+ \mid B=b)\, p(B=b \mid A=+)}$$

$$\geq .82$$

You are at node D, and you know what orders have been heard at node D. Given your orders, you may either decide to march (*march*) or stay put (*stay*). (Note that these actions are distinct from the orders $+m$ or $-m$ that you hear and pass on. The variables in the Bayes net and their conditional distributions still behave exactly as above.) If you decide to take the action corresponding to the decision that was actually made at A (not necessarily corresponding to your orders!), you receive a reward of $+1$, but if you take the opposite action, you receive a reward of -1.

(c) [4 pts] Given that you have received the order $+m$, what is the expected utility of your optimal action? (Hint: your answer to part (b) may come in handy.)

Put your answer to 1c here:

$.82 + (1-.82)\cdot -1 = .82 - .18 = \boxed{.64}$ march

Now suppose that you can have your friends text you what orders they have received. (Hint: for the following two parts, you should not need to do much computation due to symmetry properties and intuition.)

(d) [4 pts] Compute the VPI of A given that $D = +m$.

Put your answer to 1d here:

$1 - .64 = .36$

(e) [4 pts] Compute the VPI of A given that $D = +m$ and $B = -m$.

Put your answer to 1e here:

1.09

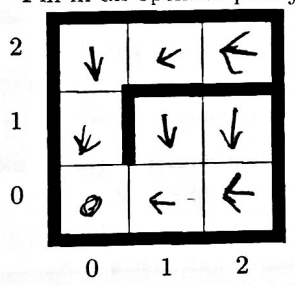
Q2. [18 pts] The nature of discounting

Pacman is stuck in a friendlier maze where he gets a reward every time he takes any action from state (0,0). This setup is a bit different from the one you've seen before: Pacman can get the reward multiple times; these rewards do not get "used up" like food pellets and there are no "living rewards". As usual, Pacman can not move through walls and may take any of the following actions: go North ($\uparrow$), South ($\downarrow$), East ($\rightarrow$), West ($\leftarrow$), or stay in place ($\circ$). Actions give deterministic results (taking action East will always move Pacman East). State (0,0) gives a total reward of 1 every time Pacman takes an action in that state regardless of the outcome, and all other states give no reward. The precise reward function is: $R((0,0), a, s') = 1$ for any action a and $R(s, a, s') = 0$ for all $s \neq (0,0)$.

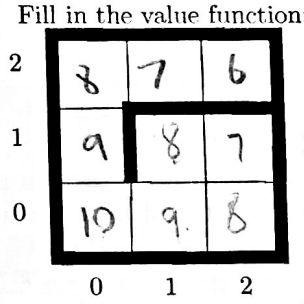
You should not need to use any other complicated algorithm/calculations to answer the questions below. We remind you that geometric series converge as follows: $1 + \gamma + \gamma^2 + \cdots = 1/(1-\gamma)$.

(a) [6 pts] Assume finite horizon of $h = 10$ (so Pacman takes exactly 10 steps) and no discounting ($\gamma = 1$).

Fill in an optimal policy:



Fill in the value function:



(available actions: $\uparrow, \downarrow, \rightarrow, \leftarrow, \circ$)

(b) The following Q-values correspond to the value function you specified above.

(i) [2 pts] The Q value of state-action $(0,0)$, (East) is: 9

(ii) [2 pts] The Q value of state-action $(1,1)$, (East) is: 6

(c) Assume finite horizon of $h = 10$, no discounting, but the action to stay in place is temporarily (for this sub-point only) unavailable. Actions that would make Pacman hit a wall are not available. Specifically, Pacman can not use actions North or West to remain in state $(0,0)$ once he is there.

(i) [2 pts] [*true* or ~~*false*~~] There is just one optimal action at state $(0,0)$.

(ii) [2 pts] The value of state $(0,0)$ is: 5

(d) [2 pts] Assume infinite horizon, discount factor $\gamma = 0.9$.

The value of state $(0,0)$ is: 10

(e) [2 pts] Assume infinite horizon and no discount ($\gamma = 1$). At every time step, after Pacman takes an action and collects his reward, a power outage could suddenly end the game with probability $\alpha = 0.1$.

The value of state $(0,0)$ is: 10

Q3. [8 pts] The Value of Games

Pacman is the model of rationality and seeks to maximize his expected utility, but that doesn't mean he never plays games.

(a) [4 pts] **A Costly Game.** Pacman is now stuck playing a new game with only costs and no payoff. Instead of maximizing expected utility $V(s)$, he has to minimize expected costs $J(s)$. In place of a reward function, there is a cost function $C(s, a, s')$ for transitions from s to s' by action a. We denote the discount factor by $\gamma \in (0, 1)$. $J^*(s)$ is the expected cost incurred by the optimal policy. Which one of the following equations is satisfied by J^*?

- ○ $J^*(s) = \min_a \sum_{s'} [C(s, a, s') + \gamma \max_{a'} T(s, a', s') * J^*(s')]$
- ○ $J^*(s) = \min_{s'} \sum_a T(s, a, s')[C(s, a, s') + \gamma * J^*(s')]$
- ○ $J^*(s) = \min_a \sum_{s'} T(s, a, s')[C(s, a, s') + \gamma * \max_{s'} J^*(s')]$
- ○ $J^*(s) = \min_{s'} \sum_a T(s, a, s')[C(s, a, s') + \gamma * \max_{s'} J^*(s')]$
- ● $J^*(s) = \min_a \sum_{s'} T(s, a, s')[C(s, a, s') + \gamma * J^*(s')]$
- ○ $J^*(s) = \min_{s'} \sum_a [C(s, a, s') + \gamma * J^*(s')]$

(b) [4 pts] **It's a conspiracy again!** The ghosts have rigged the costly game so that once Pacman takes an action they can pick the outcome from all states $s' \in S'(s, a)$, the set of all s' with non-zero probability according to $T(s, a, s')$. Choose the correct Bellman-style equation for Pacman against the adversarial ghosts.

- ○ $J^*(s) = \min_a \max_{s'} T(s, a, s')[C(s, a, s') + \gamma * J^*(s')]$
- ○ $J^*(s) = \min_{s'} \sum_a T(s, a, s')[\max_{s'} C(s, a, s') + \gamma * J^*(s')]$
- ○ $J^*(s) = \min_a \min_{s'} [C(s, a, s') + \gamma * \max_{s'} J^*(s')]$
- ● $J^*(s) = \min_a \max_{s'} [C(s, a, s') + \gamma * J^*(s')]$
- ○ $J^*(s) = \min_{s'} \sum_a T(s, a, s')[\max_{s'} C(s, a, s') + \gamma * \max_{s'} J^*(s')]$
- ○ $J^*(s) = \min_a \min_{s'} T(s, a, s')[C(s, a, s') + \gamma * J^*(s')]$

Q5. [4 pts] Minimax MDPs

This exercise considers two-player MDPs that correspond to zero-sum minimax games. Let the players be A and B, and let $R(s)$ be the reward for player A in state s (the reward for B is always equal and opposite).

(a) [4 pts] Let $U_A(s)$ be the utility of state s when it is A's turn to move in s, and let $U_B(s)$ be the utility of state s when it is B's turn to move in s. All rewards and utilities are calculated from A's point of view (just as in a minimax game tree). Write down Bellman equations defining $U_A(s)$ and $U_B(s)$.

Put your answer to 1b here:

$$U_A(s) = \max_A \sum_{s'} P(s' | A, s') [U_A(s') + \gamma U_B(s')]$$

$$U_B(s) = \max_B \sum_{s'} P(s' | B, s') [U_B(s') + \gamma U_A(s')]$$