CS 188 Spring 2014 Introduction to Artificial Intelligence Written HW4

INSTRUCTIONS

- **Due:** Monday, February 24th, 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 Spring 2014. Click on the submission titled Written HW 4 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	Chen
First Name	Jianzhong
SID	23478230
Email	chenjianzhong@berkeley.edu
Collaborators	None

For staff use only

Q. 1	Q. 2	Total
/22	/8	/30

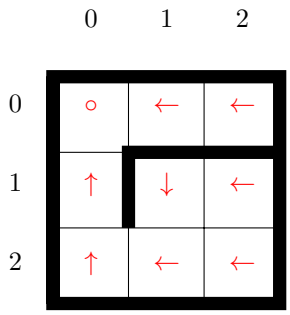
Q1. [22 pts] The nature of discounting

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 (↑), South (↓), East (→), West (←), or stay in place (○). 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: $R_{(0,0),a} = 1$ for any action a and $R_{s',a} = 0$ 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: $1 + \gamma + \gamma^2 + \cdots = 1/(1-\gamma)$.

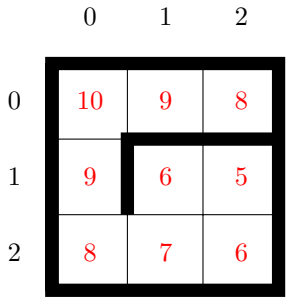
(a) [6 pts] Assume finite horizon of $h = 10$ (so Pacman takes exactly 10 steps) and no discounting ($\gamma = 1$).

Fill in an optimal policy:



(available actions: ↑, ↓, →, ←, ○)

Fill in the value function:



(b) Assume finite horizon of $h = 10$. 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: 4

(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 also not available. For example, if Pacman is in state state (0,0), he cannot take actions Stay, West, or North.

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

false

(ii) [2 pts] The value of state $(0,0)$ is: 4

(d) [4 pts] Assume infinite horizon, discount factor $\gamma = 0.9$.

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

(e) [4 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: _____

Q2. [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.

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