Northeastern University
CS 7180 – Special Topics in AI (Reinforcement Learning)
Fall 2018, Robert Platt

MDPs Assignment

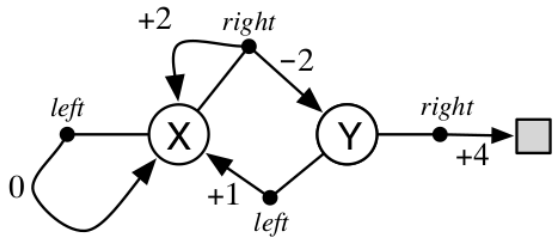
Name: _____

Problem	Points
Trajectories, returns, and values.	/15
Maze running.	/5
Broken vision system.	/5
Understanding value functions	/5
Calculating total returns.	/5
Bellman, part 1	/15
Bellman, part 2	/5
Adding a constant to rewards.	/5
Bellman, part 3	/10
Bellman, part 4	/20
Total	/90

Instructions

- Don't cheat!

(15 pts.) Trajectories, returns, and values.



Consider the MDP above, in which there are two states, X and Y, two actions, *right* and *left*, and the deterministic rewards on each transition are as indicated by the numbers. Note that if action *right* is taken in state X, then the transition may be either to X with a reward of +2 or to Y with a reward of −2. These two possibilities occur with probabilities 2/3 (for the transition to X) and 1/3 (for the transition to state Y).
Consider two deterministic policies, π_1 and π_2:

$$\pi_1(\mathsf{X}) = \textit{left} \qquad \pi_2(\mathsf{X}) = \textit{right}$$
$$\pi_1(\mathsf{Y}) = \textit{right} \qquad \pi_2(\mathsf{Y}) = \textit{right}$$

a. (2 pts) Show a typical trajectory (sequence of states, actions and rewards) from X for policy π_1:

b. (2 pts) Show a typical trajectory (sequence of states, actions and rewards) from X for policy π_2:

c. (2 pts) Assuming the discount-rate parameter is $\gamma = 0.5$, what is the return from the initial state for the second trajectory?

d. (2 pts) Assuming $\gamma = 0.5$, what is the value of state Y under policy π_1?

e. (2 pts) Assuming $\gamma = 0.5$, what is the action-value of $(Y, left)$ under policy π_1?

f. (5 pts) Assuming $\gamma = 0.5$, what is the value of state (X) under policy π_2?

(5 pts.) Maze running.

Exercise 3.7 in the SB textbook.

(5 pts.) Broken vision system.

Imagine that you are a vision system. When you are first turned on for the day, an image floods into your camera. You can see lots of things, but not all things. You cant see objects that are occluded, and of course you cant see objects that are behind you. After seeing that first scene, do you have access to the Markov state of the environment? Suppose your camera was broken that day and you received no images at all, all day. Would you have access to the Markov state then?

(5 pts.) Understanding value functions

Suppose we have two problems with the same state and action spaces. Let the optimal action-value functions of the two problems be denoted Q^* and $\bar{Q}^*$, and suppose it happens to be the case that $\bar{Q}^*(s,a) = Q^*(s,a) + f(s), \forall s, a$ for some function $f : S \rightarrow \mathbb{R}$. What is the relationship between the optimal policies π^* and $\bar{\pi}^*$ for the two problems?

(5 pts.) Calculating total returns.

Exercise 3.9 in the SB textbook.

(15 pts.) Bellman, part 1

Exercise 3.11, 3.13 in the SB textbook.

(5 pts.) Bellman, part 2

Exercise 3.14 in the SB textbook.

(5 pts.) Adding a constant to rewards.

Exercise 3.15 in the SB textbook.

(10 pts.) Bellman, part 3

Exercise 3.19 in the SB textbook.

(20 pts.) Bellman, part 4

Exercise 3.26, 3.27, 3.28 in the SB textbook.