NAME: Andrew Caide

CSCI S-89c Deep Reinforcement Learning

Part I of Assignment 2

Please consider a Markov Decision Process with two states: s^A and s^B.

Assume that the sets of admissible actions in states s^A and s^B are $\mathcal{A}(s^A) = \{a_1^A, a_2^A\}$ and $\mathcal{A}(s^B) = \{a_1^B, a_2^B\}$, respectively. Further, assume that the transition probabilities are given by:

$$p(s', r|s^A, a_1^A) = \begin{cases} 1, & \text{if } s' = s^A, r = r_1^A, \\ 0, & \text{otherwise,} \end{cases}$$

$$p(s', r|s^A, a_2^A) = \begin{cases} 1, & \text{if } s' = s^A, r = r_2^A, \\ 0, & \text{otherwise,} \end{cases}$$

$$p(s', r|s^B, a_1^B) = \begin{cases} 1, & \text{if } s' = s^B, r = r_1^B, \\ 0, & \text{otherwise,} \end{cases}$$

$$p(s', r|s^B, a_2^B) = \begin{cases} 1, & \text{if } s' = s^B, r = r_2^B, \\ 0, & \text{otherwise,} \end{cases}$$

where r_1^A, r_2^A, r_1^B, and r_2^B are known.

If policy $\pi(a|s)$ is to always take action a_1^A in state s^A and action a_1^B in state s^B, find

(a) $v_\pi(s^A)$

(b) $q_\pi(s^A, a_1^A)$

(c) $q_\pi(s^A, a_2^A)$

note on quiz, question 4

- I thought I was solving for $\pi(a|s)$ where

$\pi(a|s) \overset{?}{=} \sum P(s|a) = P(s_A|a_1) \cdot P(s_A|a_2)$

$= 0.1(1-0.4) = 0.35$

* obviously, however $\sum \pi(a|s_A) = 1$ because ~~all~~ $\sum$ actions $= 1$

SOLUTION:

a - $V_\pi(s^A) = \dfrac{r_1^A}{1-\gamma}$

b - $q_\pi(s^A, a_1^A) = \mathbb{E}_\pi[G_T | S_T = s, A_t = a]$

$= r_1^A + \gamma V_\pi(s^A) = \boxed{r_1^A + \gamma \dfrac{r_1^A}{1-\gamma}}$

c - $q_\pi(s^A, a_2^A) = r_1^B + \gamma V_\pi(s^A) = \boxed{r_1^B + \gamma \dfrac{r_1^A}{1-\gamma}}$