Name: ....
CSCI E-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) $q_\pi(s^A, a_2^A)$

SOLUTION:

(a) Policy is to always take action a_1^A in state s^A with reward r_1^A

$v_\pi(s^A) \doteq r_1^A + \gamma r_1^A + \gamma^2 r_1^A + ... \gamma^{n-1} r_1^A = \sum_{k=0}^{\infty} \gamma^k r_1^A = \frac{r_1^A}{1-\gamma}$

(b) Action value of $q_\pi(s^A, a_1^A) \doteq v_\pi(s^A)$

Therefore: $q_\pi(s^A, a_1^A) = \frac{r_1^A}{1-\gamma}$ also

(c) $q_\pi(s^A, a_2^A) = r_2^A + \gamma r_1^A + \gamma^2 r_1^A + ... \gamma^{n-1} r_1^A = \gamma \frac{r_1^A}{1-\gamma}$