NAME: ....

CSCI S-89c Deep Reinforcement Learning

Part I of Assignment 8

Please consider a Markov Decision Process (MDP) with $\mathcal{S} = \{s^A, s^B, s^C\}$.

Given a particular state $s \in \mathcal{S}$, the agent is allowed to either try staying there or switching to any of the other states. Let's denote an intention to move to state s^A by a^A, to state s^B by a^B, and to state s^C by a^C. The agent does not know transition probabilities, including the distributions of rewards. There is, however, some evidence that the agent gets rewards only at the entrance to s^C; and transition MDP probabilities to/from s^A appear to be same (or nearly same) as to/from s^B.

Suppose the agent chooses policy $\pi(a^A|s) = 0.05$, $\pi(a^B|s) = 0.05$, $\pi(a^C|s) = 0.90$ for all $s \in \{s^A, s^B, s^C\}$. Because of the apparent symmetry between s^A and s^B, it makes sense to assume that $v_\pi(s^A) \approx v_\pi(s^B)$ and approximate the state-values as follows:

$$v_\pi(s) \approx \hat{v}(s, \mathbf{w}) = w_1 \cdot \mathbb{1}_{(s=s_A)} + w_1 \cdot \mathbb{1}_{(s=s_B)} + w_2 \cdot \mathbb{1}_{(s=s_C)}.$$

Please notice that $\hat{v}(s^A, \mathbf{w}) = \hat{v}(s^B, \mathbf{w})$ for any choice of weights.

Assume the agent runs the TD(λ) with Approximation for estimating v_π:

$$\begin{aligned}
\mathbf{z}_{-1} &\doteq (0,0)^T, \\
\mathbf{z}_t &\doteq \gamma\lambda \mathbf{z}_{t-1} + \nabla \hat{v}(S_t, \mathbf{w}_t) \quad \text{for } t \geq 0, \\
\mathbf{w}_{t+1} &\doteq \mathbf{w}_t + \alpha \left[R_{t+1} + \gamma \hat{v}(S_{t+1}, \mathbf{w}_t) - \hat{v}(S_t, \mathbf{w}_t)\right] \mathbf{z}_t \quad \text{for } t \geq 0,
\end{aligned}$$

where $\lambda = 0.2$, $\alpha = 0.1$, $\gamma = 0.9$, and weights $\mathbf{w}_t$ are set to zero at time $t = 0$.

If the agent observes the following sequence of states, actions, and rewards:

$$\begin{aligned}
&S_0 = s^A, A_0 = a^C, R_1 = 20, \\
&S_1 = s^C, A_1 = a^B, R_2 = 0, \\
&S_2 = s^B, A_2 = a^C, R_3 = 20, \\
&S_3 = s^C, A_3 = a^C, R_4 = 20, \\
&S_4 = s^C, A_4 = a^B, R_5 = 0, \\
&S_5 = s^B,
\end{aligned}$$

find (a) weights $\mathbf{w}_t$ and (b) corresponding approximations $\hat{v}(s, \mathbf{w}_t)$ for $t = 1, 2, \ldots, 5$. Specifically, please fill the tables in below:

SOLUTION:

(a) weights $\mathbf{w}_t = (w_{1,t}, w_{2,t})^T$:

	$t=0$	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
$w_{1,t}$	0					
$w_{2,t}$	0					

(b) approximations $\hat{v}(s, \mathbf{w}_t)$:

	$t=0$	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
$\hat{v}(s^A, \mathbf{w}_t)$	0					
$\hat{v}(s^B, \mathbf{w}_t)$	0					
$\hat{v}(s^C, \mathbf{w}_t)$	0					