867002105_HW10

Charles Dotson

November 28, 2022

2.

We start by showing two partial simplifications for $q_\pi(s,a)$ and $v_\pi(s)$.

$$
\begin{aligned}
q_\pi(s,a) &= \mathbb{E}_\pi[G_t|S_t = s, A_t = a] \\
&= \mathbb{E}_\pi[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ... |S_t = s, A_t = a] \\
&= \mathbb{E}_\pi[R_{t+1} + \gamma G_{t+1}|S_t = s, A_t = a] \\
&= \sum_{s'} \sum_r \Pr(s', r|s, a)[r + \gamma \mathbb{E}_\pi[G_{t+1}|S_{t+1} = s', S_t = s, A_t = a]]
\end{aligned}
$$

$$
\begin{aligned}
v_\pi(s) &= \mathbb{E}_\pi[G_t|S_t = s] \\
&= \mathbb{E}_\pi[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ... |S_t = s] \\
&= \mathbb{E}_\pi[R_{t+1} + \gamma G_{t+1}|S_t = s] \\
&= \sum_a \pi(a|s) \sum_{s'} \sum_r \Pr(s', r|s, a)[r + \gamma \mathbb{E}_\pi[G_{t+1}|S_{t+1} = s']]
\end{aligned}
$$

Since the state value function v_π is not a function of a, we include the other two conditions, $S_t = s$ and $A_t = a$, as being apart of the conditional expection to arrive at,

$$
v_\pi(s) = \sum_a \pi(a|s) \sum_{s'} \sum_r \Pr(s', r|s, a)[r + \gamma \mathbb{E}_\pi[G_{t+1}|S_{t+1} = s', S_t = s, A_t = a]]
$$

substituting in $q_\pi(s,a)$ by definition we arrive at,

$$
v_\pi(s) = \sum_{a \in A(s)} \pi(a|s) q_\pi(s,a)
$$

$\pi(a|s)$ being in the final summation is due to the fact that it is a leftover from starting at $v_\pi(s)$ where it must be included due to the fact that $v_\pi(s)$ is not a function of a.

3.

The cummulative return for starting at u going left to $w = R_1^{(1)} = 1$.

The cummulative return for starting at u and going right to $v, z = R_1^{(2)} + \gamma R_2 = -1 + 10\gamma$.

To make going to left have a higher discounted return as compared to going right, we find the upper bound for γ by the following inequality.

$$R_1^{(1)} > R_1^{(2)} + \gamma R_2$$

$$1 > -1 + 10\gamma$$

$$\gamma < \frac{1}{5}$$

4.

From problem 1. we proved that

$$v_\pi(s) = \sum_{a \in A(s)} \pi(a|s) q_\pi(s,a)$$

We now continue our simplification of $q_\pi(s,a)$

$$\begin{aligned}
q_\pi(s,a) &= \mathbb{E}_\pi[G_t | S_t = s, A_t = a] \\
&= \mathbb{E}_\pi[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ... | S_t = s, A_t = a] \\
&= \mathbb{E}_\pi[R_{t+1} + \gamma G_{t+1} | S_t = s, A_t = a] \\
&= \sum_{s'} \sum_{r} \Pr(s', r | s, a)[r + \gamma \mathbb{E}_\pi[G_{t+1} | S_{t+1} = s', S_t = s, A_t = a]] \\
&= \sum_{s'} \sum_{r} \Pr(s', r | s, a)[r + \gamma \sum_{a'} \pi(a'|s') q_\pi(s', a')]
\end{aligned}$$

By definition, we substitute $v_\pi(s')$, and arrive at

$$q_\pi(s,a) = \sum_{s'} \sum_{r} \Pr(s', r | s, a)[r + \gamma v_\pi(s')]$$

We now move to the second part of the question. Using the above formulation we solve for the optimal action for state $s = 5$ on the $k = 2$ iteration using our formulation for $q_\pi(s,a)$ after $v_\pi(s)$ has been solved for all s. Note we are excluding the probabilties which would result in 0.

$$q_\pi(5, \mathrm{up}) = \Pr(1, -1 | 5, \mathrm{up})[-1 + v_\pi(1)] = -2.7$$

$$q_\pi(5, \mathrm{down}) = \Pr(9, -1 | 5, \mathrm{down})[-1 + v_\pi(9)] = -3$$

$$q_\pi(5, \mathrm{right}) = \Pr(6, -1 | 5, \mathrm{right})[-1 + v_\pi(6)] = -3$$

$$q_\pi(5, \mathrm{left}) = \Pr(4, -1 | 5, \mathrm{left})[-1 + v_\pi(4)] = -2.7$$

Since at $k = 2$ iterations the optimal action is solved for at states $s = 1$ and $s = 4$, we can say, that from state $s = 5$, the 2 best actions in either order are Up and Left.