Quiz: An Iterative Method

So far in this lesson, we have discussed how an agent might obtain the state-value function v_π corresponding to a policy π.

In the dynamic programming setting, the agent has full knowledge of the Markov decision process (MDP). In this case, it's possible to use the one-step dynamics $p(s', r|s, a)$ of the MDP to obtain a system of equations corresponding to the Bellman expectation equation for v_π.

In the gridworld example, the system of equations corresponding to the equiprobable random policy was given by:

$$v_\pi(s_1) = \frac{1}{2}(-1 + v_\pi(s_2)) + \frac{1}{2}(-3 + v_\pi(s_3))$$

$$v_\pi(s_2) = \frac{1}{2}(-1 + v_\pi(s_1)) + \frac{1}{2}(5 + v_\pi(s_4))$$

$$v_\pi(s_3) = \frac{1}{2}(-1 + v_\pi(s_1)) + \frac{1}{2}(5 + v_\pi(s_4))$$

$$v_\pi(s_4) = 0$$

In order to obtain the state-value function, we need only solve the system of equations.

While it is always possible to *directly* solve the system, we will instead use an *iterative* solution approach.

An Iterative Method

The iterative method begins with an initial guess for the value of each state. In particular, we began by assuming that the value of each state was zero.

Then, we looped over the state space and amended the estimate for the state-value function through applying successive update equations.

Recall that V denotes the most recent guess for the state-value function, and the update equations are:

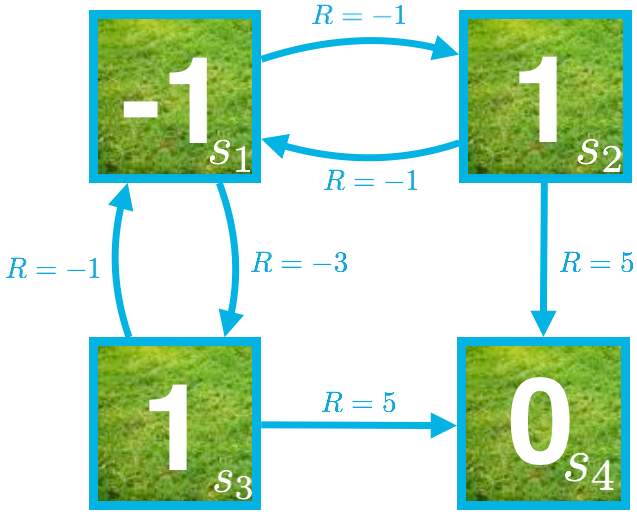
Quiz: An Iterative Method

$V(s_2) \leftarrow \frac{1}{2}(-1 + V(s_1)) + \frac{1}{2}(5)$

$V(s_3) \leftarrow \frac{1}{2}(-1 + V(s_1)) + \frac{1}{2}(5)$

Quiz Question

Say that the most recent guess for the state-value function is given in the figure below.



Currently, the estimate for $v_\pi(s_2)$ is given by $V(s_2) = 1$.

Say that the next step in the algorithm is to update $V(s_2)$.

What is the new value for $V(s_2)$ after applying the update step once?

QUIZ QUESTION

Select the appropriate value.

- -1.5
- .5
- 0

Quiz: An Iterative Method

1.5

SUBMIT

NEXT