Bellman Equations

In this gridworld example, once the agent selects an action,

- it always moves in the chosen direction (contrasting general MDPs where the agent doesn't always have complete control over what the next state will be), and
- the reward can be predicted with complete certainty (contrasting general MDPs where the reward is a random draw from a probability distribution).

In this simple example, we saw that the value of any state can be calculated as the sum of the immediate reward and the (discounted) value of the next state.

Alexis mentioned that for a general MDP, we have to instead work in terms of an *expectation*, since it's not often the case that the immediate reward and next state can be predicted with certainty. Indeed, we saw in an earlier lesson that the reward and next state are chosen according to the one-step dynamics of the MDP. In this case, where the reward r and next state s' are drawn from a (conditional) probability distribution $p(s', r|s, a)$, the **Bellman Expectation Equation (for v_π)** expresses the value of any state s in terms of the *expected* immediate reward and the *expected* value of the next state:

$$v_\pi(s) = \mathbb{E}_\pi[R_{t+1} + \gamma v_\pi(S_{t+1})|S_t = s].$$

Calculating the Expectation

In the event that the agent's policy π is **deterministic**, the agent selects action $\pi(s)$ when in state s, and the Bellman Expectation Equation can be rewritten as the sum over two variables (s' and r):

$$v_\pi(s) = \sum_{s' \in \mathcal{S}^+, r \in \mathcal{R}} p(s', r|s, \pi(s))(r + \gamma v_\pi(s'))$$

In this case, we multiply the sum of the reward and discounted value of the next state $(r + \gamma v_\pi(s'))$ by its corresponding probability $p(s', r|s, \pi(s))$ and sum over all possibilities to yield the expected value.

If the agent's policy π is **stochastic**, the agent selects action a with probability $\pi(a|s)$ when in state s, and the Bellman Expectation Equation can be rewritten as the sum over three variables (s', r, and a):

$$v_\pi(s) = \sum_{s' \in \mathcal{S}^+, r \in \mathcal{R}, a \in \mathcal{A}(s)} \pi(a|s)p(s', r|s, a)(r + \gamma v_\pi(s'))$$

In this case, we multiply the sum of the reward and discounted value of the next state $(r + \gamma v_\pi(s'))$ by its corresponding probability $\pi(a|s)p(s', r|s, a)$ and sum over all possibilities to yield the expected value.

There are 3 more Bellman Equations!

In this video, you learned about one Bellman equation, but there are 3 more, for a total of 4 Bellman equations.

> All of the Bellman equations attest to the fact that *value functions satisfy recursive relationships*.

For instance, the **Bellman Expectation Equation (for v_π)** shows that it is possible to relate the value of a state to the values of all of its possible successor states.

After finishing this lesson, you are encouraged to read about the remaining three Bellman equations in sections 3.5 and 3.6 of the **textbook**. The Bellman equations are incredibly useful to the theory of MDPs.

NEXT