Deterministic Bellman Equations

Bodhi Nguyen

January 21th, 2021

1 Four Bellman Equations

A deterministic policy π_D has the following property relating to a stochastic policy:

$$\pi(s, \pi_D(s)) = 1$$

and so the stochastic policy equals 0 for any other action not equal to $\pi_D(s)$. Thus we can rewrite the first bellman equation as

$$V^{\pi_D}(s) = R(s, \pi_D(s)) + \gamma \sum_{s'} P(s, \pi_D(s), s') V^{\pi_D}(s') \tag{1}$$

This comes from the fact that the first and seconds sums over different actions can be removed as only one action has a probability of 1.

The second equation simplifies similarly, removing the sum over the probabilities of different possible actions that could be taken:

$$V^{\pi_D}(s) = Q^{\pi_D}(s', \pi_D(s)) \tag{2}$$

The third equation comes from the first two:

$$Q^{\pi_D}(s, \pi_D(s)) = R(s, \pi_D(s)) + \gamma \sum_{s'} P(s, \pi_D(s), s') V^{\pi_D}(s') \tag{3}$$

Note that $Q^{\pi_D}(s, a)$ for any other a is similar:

$$Q^{\pi_D}(s, a) = R(s, a) + \gamma \sum_{s'} P(s, \pi_D(s), s') V^{\pi_D}(s') \tag{4}$$

This is because this is the value of taking action a in state s and following the deterministic policy π_D thereafter.

The last bellman equation is derived by replacing the value function in the previous equation with (2)

$$Q^{\pi_D}(s, a) = R(s, a) + \gamma \sum_{s'} P(s, \pi_D(s), s') Q^{\pi_D}(s', \pi_D(s)) \tag{5}$$