Reinforcement Learning formula

Lee Meng Oon

April 16, 2021

Discounted return at time step t:

$$G_t = \sum_{k=0} \gamma^k R_{t+k+1}, \quad \gamma \in [0,1] \tag{1}$$

One-step dynamics:

$$\begin{gathered} p(s', r|s, a) = \mathbb{P}(S_{t+1} = s', R_{t+1} = r|S_t = s, A_t = a), \\ s, s' \in \mathcal{S}, \quad a \in \mathcal{A}, \quad r \in \mathcal{R} \end{gathered} \tag{2}$$

Deterministic policy:

$$\pi : \mathcal{S} \to \mathcal{A}, \quad s \in \mathcal{S}, \quad a \in \mathcal{A} \tag{3}$$

Stochastic policy:

$$\pi : \mathcal{S} \times \mathcal{A} \to \pi(a|s) \in [0,1], \quad s \in \mathcal{S}, \quad a \in \mathcal{A} \tag{4}$$

State-value function:

$$v_\pi(s) = \mathbb{E}_\pi[G_t|S_t = s], \quad s \in \mathcal{S} \tag{5}$$

Bellman expectation equation:

$$v_\pi(s) = \mathbb{E}_\pi[R_{t+1} + \gamma v_\pi(S_{t+1})|S_t = s], \quad s \in \mathcal{S}, \quad \gamma \in [0,1] \tag{6}$$

Action-value function:

$$q_\pi(s, a) = \mathbb{E}_\pi[G_t|S_t = s, A_t = a], \quad s \in \mathcal{S}, \quad a \in \mathcal{A} \tag{7}$$

Optimal policy:

$$\pi_*(s) = \underset{a \in A(s)}{\arg\max}\, q_*(s, a), \quad s \in \mathcal{S} \tag{8}$$