

Reinforcement Learning Training 2025

Bellman's Equation

- Foundation to solving MDP and RL problems.

Recall (1)

- Markov decision process has transition probabilities

$$\mathbb{P}\left[S_{t+1} = s', R_{t+1} = r | S_t = s, A_t = a\right]$$

which transitions the agent to state S_{t+1} and a reward of R_{t+1}

- Cumulative reward at time t

$$G_t = R_{t+1} + \gamma R_{t+2} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+1+k}$$

Recall (2)

- A value function is an expected cumulative reward

$$v_{\pi}(s) = \mathbb{E}_{\pi}[G_t | S_t = s]$$

- An action-value function is an expected cumulative reward from taking action a

$$q_{\pi}(s, a) = \mathbb{E}_{\pi}[G_t | S_t = s, A_t = a]$$

■ Note that v and q depend on the policy π .

Bellman Equation

- Allows relationships among v and q .

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

$$q_{\pi}(s, a) = \sum_{s', r} p(s', r | s, a) \left[r + \gamma \sum_{a'} \pi(a'|s') q(s', a') \right]$$