

Temporal-Difference Learning Study Guide (AIMA 22, RLAI 6)

Artificial Intelligence

1 Reinforcement Learning

1. What is reinforcement learning?

Solution: Reinforcement learning is learning how to map states to actions by interacting with an environment that provides numerical feedback signals, which we call *rewards*. A reinforcement learning agent tries to maximize long-term reward, not just one-step reward.

2. What is the *reward hypothesis*?

Solution: All of what we mean by goals and purposes can be well thought of as the maximization of the expected value of the cumulative sum of a received scalar signal (called reward).

3. What does “model-free” mean in the context of reinforcement learning algorithms that assume there is an underlying MDP model of the environment?

Solution: Model-free RL algorithms don’t know the underlying MDP, so they learn policies or value functions directly by sampling interactions with the environment.

4. How many time steps does TD(0) learning use in its target to update to the value of a state?

Solution: 1