Reinforcement Learning

Gong Ying 2023.11.17

Contents

- Basic terminologies
- Return and value

Basic terminologies

Agent

State: **s**

Action : **a**

Reward: **r**

Policy: **π(a|s)**

State transition: **p(s'|s,a)**

Basic terminologies

Markov property

In state transitions, S_{t+1} only depends on S_t and A_t, but does not depends on history states or actions.

Return and value

- **Return and discounted return**

$$\mathrm{U_t} = \mathrm{R_t} + \gamma R_{t+1} + \gamma^2 \mathrm{R_{t+2}} + \gamma^3 \mathrm{R_{t+3}} + \cdots$$

- **Value function**

action-value function

optimal action-value function

state-value function

Return and value

- **Action-value function:**

Evaluates the value of a certain action a in state s.

$Q_\pi(s_t, a_t)$ is conditional expectation of U_t in terms of S_{t+1}, A_{t+1}, ..., S_n, A_n. It is affected by π.

- **Optimal action-value function**

- **State-value function**

Return and value

- **Action-value function**

- **Optimal action-value function:**

$Q_*(s_t, a_t) = \max_{\pi} Q_\pi(s_t, a_t)$, excluding the effect of policy π, only depending on s_t and a_t.

- **State-value function**

Return and value

- **Action-value function**
- **Optimal action-value function**
- **State-value function:**

Evaluates the value of state s with policy π, excluding the effect of a.

$V_\pi(s_t)$ is expectation of A_t.

Thank you.