Reinforcement Learning

7

States $s \in \mathcal{S}$

Actions $a \in \mathcal{A}$

Transition function $P(s_{t+1} \mid s_t, a_t)$

Reward function $r_t = r(s_t, a_t)$

Start state s_0

Discount factor $\gamma \in [0,1]$

Policy $\pi : \mathcal{S} \rightarrow \mathcal{A}$

Goal:

$$\max_{\pi} \mathbb{E}_{\pi,P}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, \pi\right]$$

Value function:

$$V_\pi(s) = \mathbb{E}_{\pi,P}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, \pi\right]$$

Action-value function (aka Q-function):

$$Q_\pi(s, a) = \mathbb{E}_{\pi,P}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, a_0 = a, \pi\right]$$

Markov Decision Process (MDP)

Reinforcement Learning

Markov Decision Process (MDP)

States $s \in \mathcal{S}$

Actions $a \in \mathcal{A}$

Transition function $P(s_{t+1} \mid s_t, a_t)$

Reward function $r_t = r(s_t, a_t)$

Start state s_0

Discount factor $\gamma \in [0,1]$

Policy $\pi : \mathcal{S} \rightarrow \mathcal{A}$

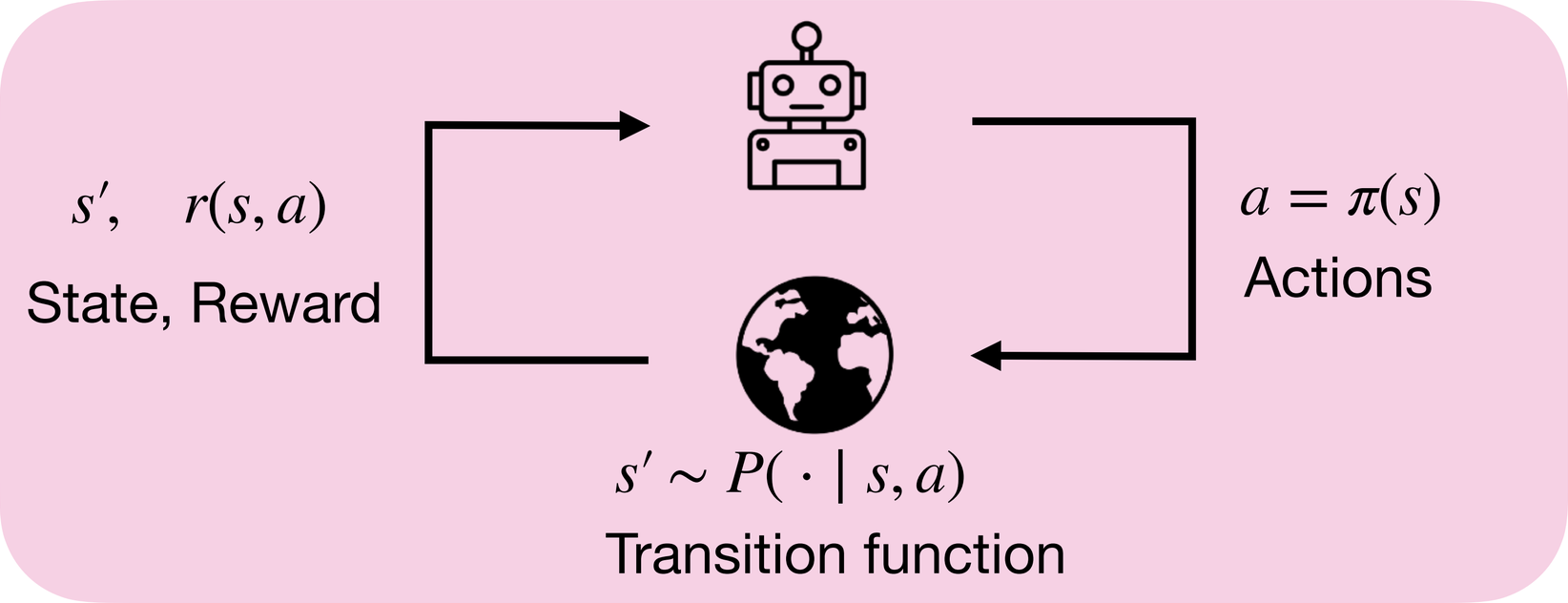
Goal:

$$\max_{\pi} \mathbb{E}_{\pi,P}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, \pi\right]$$

Value function:

$$V_\pi(s) = \mathbb{E}_{\pi,P}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, \pi\right]$$

Action-value function (aka Q-function):

$$Q_\pi(s, a) = \mathbb{E}_{\pi,P}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, a_0 = a, \pi\right]$$

Reinforcement Learning