

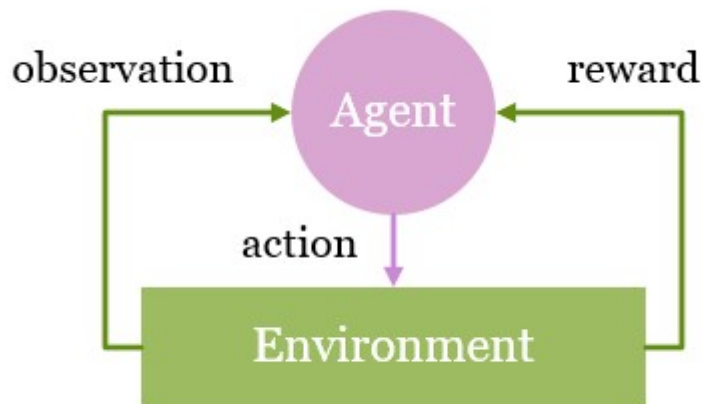
# Reinforcement Learning

---

Reinforcement Learning (RL) is a standard framework to achieve target in Markov Decision Process.

In a MDP  $\langle O, A, P, \gamma, R \rangle$ ,

- At each time period, the environment is in a state  $s$ , and agent in environment receives local observation  $o$  based on  $s$ .
- Agent takes action  $a$ , and receives a local reward from environment  $r$  ( $R : S \times A \rightarrow \mathbb{R}$ ). Then environment moves to the next state. The process repeats.
- $P$  is the state transition function,  $P(s, a, s')$
- $\gamma$  is the discount factor
- MDP can be represented as  $\langle o_0, a_0, r_1, o_1, a_1, \dots, o_{t-1}, a_{t-1}, r_t \rangle$ .
- In RL process, agent gets a series of sample and improve its policy to get better reward.



We can define the total discounted reward:

$$\mathcal{R} = \sum_{t=0}^{\infty} \gamma^t R_{t+1}$$