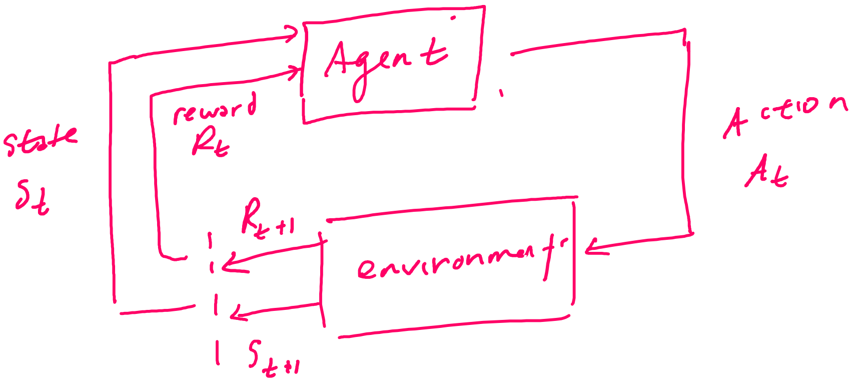
# Summary of The RL Framework: The Problem

Saturday, November 28, 2020 1:28 PM



## Setting

- RL Framework: Agent learning witin environment
- Each step: → Agent recieves environment's State
  → Agent performs action based on state
- Next step: → Reward recieved from last step
  → New environment state recieved
  → Agent performs next action

## Episodic vs. Continuing Tasks

- task: instance of RL Framework
- Continuing Tasks: Continue forever
- Episodic Tasks: well-defined starting & ending
  ↳ under when agent reaches terminal state

**Reward Hypothesis:** All goals are aimed to maximize expected culmative reward.

Culmative Reward: $G_t \doteq R_{t+1} + R_{t+2} \ldots$

Discounted Return:

$G_t \doteq \gamma^k \sum_{k=0}^{\infty} R_{k+1}$ , where $\gamma \in [0,1]$

↳ $\gamma \Rightarrow 0$ imedi ate reward
$\gamma \Rightarrow 1$ , long-term
if $\gamma = 1$ , no discount

## MDP (Merkov Decision Process):

→ state space: all non-terminal states ( $S$ )
↳ $S^+$ includes terminal states

→ action space : set of possible actions ( $A$ )

→ one-step Dynamic:

$$p(s', r | s, a) = P(S_{t+1} = s', R_{t+1} = r / S_t = s, A_t = a)$$

→ Finite MDP:

- finite $S$
- finite $A$
- set of Rewards
- one-step dynamics of environment
- discount rate $\gamma \in [0,1]$