- Two reward type

Reward (Animation Image)

GIF1

GIF2

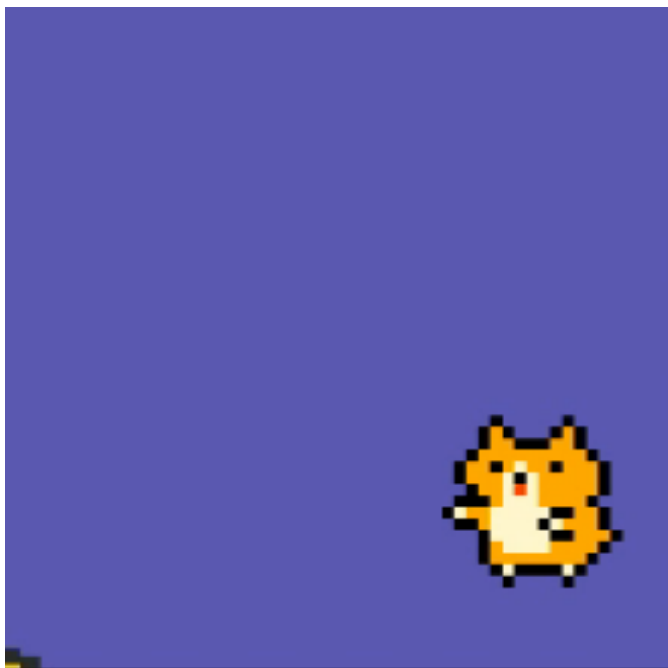GIF1

GIF2

-1 reward

Reward (Animation Image)

- Two reward type

+1 reward

-1 reward

Reinforcement Learning

- **MDP & Bellman equation**

  - Markov Decision Process

    - MDP is a model for sequential stochastic decision problems

    - with four tuples *<S, A, R, tr>* (State, Action, Reward, Transition probability)

  - Bellman Equation

    $$v_\pi(s) \doteq \sum_{a \in A} \pi(a|s) \sum_{s',r} p(s', r|s, a)[r + \gamma v_\pi(s')], \quad \text{for all} \quad s \in S$$

    - state - value function

    $$v_\pi(s) \doteq \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_t = s \right], \quad \text{for all } s \in \mathcal{S},$$

    - action - value function

    $$q_\pi(s, a) \doteq \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_t = s, A_t = a \right]$$