Introduction to Reinforcement Learning and AI

Will Guss
James Bartlett

Overview

Agenda

1. Background
2. Algorithms
3. Questions

Markov Decision Process (MDP)

Environment, $E = (\mathcal{S}, \mathcal{A}, \mathcal{R}, \rho, r)$.

1. State space, $\mathcal{S} = \mathbb{R}^n$
2. Action space, $\mathcal{A} = \mathbb{R}^m$
3. Reward space, $\mathcal{R} = \mathbb{R}$
4. Transition function, $\rho(s' \mid s, a)$. Given a previous state s and action a, environment gives s'.
5. Reward function $r(s, a) \in \mathcal{R}$.

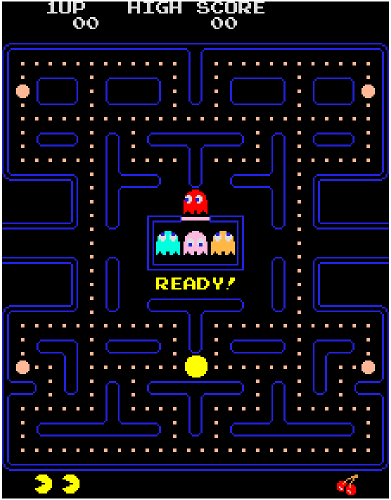
Reinforcement Learning Agent

Deterministic agent $\pi : \mathcal{S} \to \mathcal{A}$ acts in E.

$$s_1 \xrightarrow{\pi} a_1 \xrightarrow{\rho, r} s_2, r_2 \xrightarrow{\pi} a_2 \xrightarrow{\rho, r} \cdots$$

Eg. Pacman sees the screen, and decides to move $\uparrow, \downarrow, \rightarrow, \leftarrow$ and then gets a reward for eating food.

Reinforcement Learning Agent



Value in Reinforcement Learning

The value of a given state for an agent π is defined as

$$V^{\pi}(s_t) = \sum_{n=t+1}^{\infty} \gamma^{n} r(s_n, \pi(s_n))$$

1. γ is the discount factor
2. $\pi(s_n)$ is the action the agent π makes after seeing state s_n.
3. $r(s_n, \pi(s_n))$ is the reward the agent gets from taking that action.

Expected Future Reward

The expected future reward of an agent π, also known as the Q function, is

$$Q^\pi(s_t, a_t) = \underbrace{r(s_t, a_t)}_{\text{reward for } a_t} + V^\pi(s_t)$$

Bellman Equation

The Bellman equation says

$$Q^{\pi}(s_t, a_t) = r_t + \gamma Q^{\pi}(s_{t+1}, \pi(s_{t+1}))$$

Given some state s_t, the **best** agent, π^* is one that take action

$$a_t = \arg\max_a Q(s_t, a).$$

Questions?