ML@B Bootcamp

James Bartlet & Will Guss

Algorithms

Questions

## Introduction to Reinforcement Learning and Al

Will Guss James Bartlett

#### Overview



ML@B Bootcamp

James Bartle & Will Guss

Algorithms

## Agenda

- 1 Background
- 2 Algorithms
- 3 Questions

#### Markov Decision Process (MDP)



ML@B Bootcamp

James Bartlet & Will Guss

#### Background

Algorithms

Questions

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

- **1** State space,  $S = \mathbb{R}^n$
- **2** Action space,  $A = \mathbb{R}^m$
- $oxed{3}$  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



ML@B Bootcamp

James Bartlet & Will Guss

Background Algorithms

Questions

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



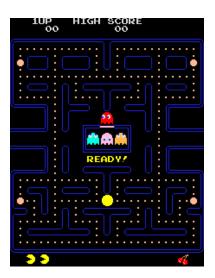
ML@B Bootcamp

James Bartlet

#### Background

Algorithms

Questions



### Value in Reinforcement Learning



ML@B Bootcamp

James Bartle & Will Guss

Background

Algorithms

The value of a given state for an agent  $\pi$  is defined as

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

- $oldsymbol{1}$   $\gamma$  is the discount factor
- $\mathbf{Z}$   $\pi(s_n)$  is the action the agent  $\pi$  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**



ML@B Bootcamp

James Bartlet & Will Guss

Background

-Questions The expected future reward of an agent  $\pi$ , 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



ML@B Bootcamp

James Bartle & Will Guss

Background

Algorithms Questions 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,  $\pi^*$  is one that take action

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

ML@B Bootcamp

James Bartlett & Will Guss

Backgroun

Algorithms

Questions

# Questions?