BE a reinforcement learner
BE a reinforcement learner

- You act as a learning agent
BE a reinforcement learner

- You act as a learning agent
- **Actions**: Wave, Stand, Clap
BE a reinforcement learner

- You act as a learning agent
- **Actions**: Wave, Stand, Clap
- **Observations**: colors, reward
BE a reinforcement learner

- You act as a learning agent
- **Actions**: Wave, Stand, Clap
- **Observations**: colors, reward
- **Goal**: Find an optimal *policy*
BE a reinforcement learner

- You act as a learning agent

- **Actions**: Wave, Stand, Clap

- **Observations**: colors, reward

- **Goal**: Find an optimal *policy*
  - Way of selecting actions that gets you the most reward
How did you do it?
How did you do it?

- What is your policy?
- What does the world look like?
How did you do it?

- What is your policy?
- What does the world look like?
Formalizing What Just Happened

Knowns:
Formalizing What Just Happened

Knowns:

- $\mathcal{O} = \{\text{Blue, Red, Green, Black, ...}\}$
- Rewards in $\mathbb{R}$
- $A = \{\text{Wave, Clap, Stand}\}$
Formalizing What Just Happened

Knowns:

• $\mathcal{O} = \{\text{Blue, Red, Green, Black, ...}\}$
• Rewards in $\mathbb{R}$
• $\mathcal{A} = \{\text{Wave, Clap, Stand}\}$

\[ o_0, a_0, r_0, o_1, a_1, r_1, o_2, \ldots \]
Formalizing What Just Happened

Knowns:
- \( \mathcal{O} = \{ \text{Blue, Red, Green, Black, \ldots} \} \)
- Rewards in \( \mathbb{R} \)
- \( \mathcal{A} = \{ \text{Wave, Clap, Stand} \} \)
- \( o_0, a_0, r_0, o_1, a_1, r_1, o_2, \ldots \)

Unknowns:
Formalizing What Just Happened

Knowns:
- $O = \{\text{Blue, Red, Green, Black, \ldots}\}$
- Rewards in $\mathbb{R}$
- $A = \{\text{Wave, Clap, Stand}\}$

Unknowns:
- $S = 4 \times 3$ grid
- $R : S \times A \mapsto \mathbb{R}$
- $T : S \mapsto O$
- $P : S \times A \mapsto S$
Formalizing What Just Happened

Knowns:

- $\mathcal{O} = \{\text{Blue, Red, Green, Black, . . .} \}$
- Rewards in $\mathbb{R}$
- $\mathcal{A} = \{\text{Wave, Clap, Stand}\}$

Unknowns:

- $s = 4 \times 3 \text{ grid}$
- $\mathcal{R} : S \times A \mapsto \mathbb{R}$
- $\mathcal{T} = S \mapsto \mathcal{O}$
- $\mathcal{P} : S \times A \mapsto S$
Formalizing What Just Happened

Knowns:

- $O = \{\text{Blue, Red, Green, Black, \ldots}\}$
- Rewards in $\mathbb{R}$
- $A = \{\text{Wave, Clap, Stand}\}$

Unknowns:

- $S$ = 4x3 grid
- $R : S \times A \rightarrow \mathbb{R}$
- $T = S \rightarrow O$
- $P : S \times A \rightarrow S$

$o_i = T(s_i)$
Formalizing What Just Happened

Knouns:
- $O = \{\text{Blue, Red, Green, Black, \ldots}\}$
- Rewards in $\mathbb{R}$
- $A = \{\text{Wave, Clap, Stand}\}$

Unknowns:
- $S = 4 \times 3$ grid
- $R : S \times A \mapsto \mathbb{R}$
- $T = S \mapsto O$
- $P : S \times A \mapsto S$

\[ o_0, a_0, r_0, o_1, a_1, r_1, o_2, \ldots \]

\[ s_0, o_0, a_0, r_0, s_1, o_1, a_1, r_1, s_2, o_2, \ldots \]

\[ o_i = T(s_i) \quad r_i = R(s_i, a_i) \]
Formalizing What Just Happened

Knouns:
- $\mathcal{O} = \{\text{Blue, Red, Green, Black, ...}\}$
- Rewards in $\mathbb{R}$
- $\mathcal{A} = \{\text{Wave, Clap, Stand}\}$

\[
\begin{array}{c}
o_0, a_0, r_0, o_1, a_1, r_1, o_2, \ldots
\end{array}
\]

Unknowns:
- $S = 4 \times 3$ grid
- $\mathcal{R} : S \times \mathcal{A} \mapsto \mathbb{R}$
- $\mathcal{T} = S \mapsto \mathcal{O}$
- $\mathcal{P} : S \times \mathcal{A} \mapsto S$

\[
\begin{array}{c}
s_0, o_0, a_0, r_0, s_1, o_1, a_1, r_1, s_2, o_2, \ldots
\end{array}
\]

\[
o_i = \mathcal{T}(s_i) \\
r_i = \mathcal{R}(s_i, a_i) \\
s_{i+1} = \mathcal{P}(s_i, a_i)
\]
This Course

- Reinforcement Learning theory (start)
This Course

- Reinforcement Learning theory (start)
- Reinforcement Learning in practice (end)
The Big Picture

- AI
The Big Picture

- AI $\rightarrow$ ML
The Big Picture

- AI $\rightarrow$ ML $\rightarrow$ RL
The Big Picture

- AI $\rightarrow$ ML $\rightarrow$ RL
- Types of Machine Learning
The Big Picture

- AI $\rightarrow$ ML $\rightarrow$ RL

- Types of Machine Learning

  **Supervised learning:** learn from labeled examples
The Big Picture

- AI $\rightarrow$ ML $\rightarrow$ RL

- Types of Machine Learning
  
  **Supervised learning:** learn from labeled examples
  
  **Unsupervised learning:** cluster unlabeled examples
The Big Picture

- AI $\rightarrow$ ML $\rightarrow$ RL

- Types of Machine Learning

  **Supervised learning:** learn from labeled examples
  **Unsupervised learning:** cluster unlabeled examples
  **Reinforcement learning:** learn from interaction
The Big Picture

- AI $\rightarrow$ ML $\rightarrow$ RL

- Types of Machine Learning
  
  **Supervised learning:** learn from labeled examples
  
  **Unsupervised learning:** cluster unlabeled examples
  
  **Reinforcement learning:** learn from interaction

    - Defined by the problem:
The Big Picture

- AI $\rightarrow$ ML $\rightarrow$ RL

- Types of Machine Learning
  - **Supervised learning**: learn from labeled examples
  - **Unsupervised learning**: cluster unlabeled examples
  - **Reinforcement learning**: learn from interaction
    - Defined by the problem:
      - closed-loop
      - select own actions
      - sequential (time-delayed)
The Big Picture

- AI $\rightarrow$ ML $\rightarrow$ RL

- Types of Machine Learning

  **Supervised learning:** learn from labeled examples
  **Unsupervised learning:** cluster unlabeled examples
  **Reinforcement learning:** learn from interaction

  - Defined by the problem:
    - closed-loop
    - select own actions
    - sequential (time-delayed)

  - Many approaches possible (including evolutionary)
The Big Picture

- AI $\rightarrow$ ML $\rightarrow$ RL

Types of Machine Learning

**Supervised learning**: learn from labeled examples

**Unsupervised learning**: cluster unlabeled examples

**Reinforcement learning**: learn from interaction

- Defined by the problem:
  - closed-loop
  - select own actions
  - sequential (time-delayed)

- Many approaches possible (including evolutionary)
- Book focusses on a particular class of approaches
Reduced Formalism

Knowns:

- $S = \{\text{Blue, Red, Green, Black, \ldots}\}$
- Rewards in $\mathbb{R}$
- $A = \{\text{Wave, Clap, Stand}\}$

$s_0, a_0, r_0, s_1, a_1, r_1, s_2, \ldots$
Reduced Formalism

Knowns:
• $S = \{\text{Blue, Red, Green, Black, \ldots}\}$
• Rewards in $\mathbb{R}$
• $A = \{\text{Wave, Clap, Stand}\}$

Unknowns:

$s_0, a_0, r_0, s_1, a_1, r_1, s_2, \ldots$
Reduced Formalism

**Knowns:**
- \( S = \{ \text{Blue, Red, Green, Black, \ldots} \} \)
- Rewards in \( \mathbb{R} \)
- \( A = \{ \text{Wave, Clap, Stand} \} \)

\[
\begin{align*}
S_0, a_0, r_0, s_1, a_1, r_1, s_2, & \ldots \\
\end{align*}
\]

**Unknowns:**
- \( R : S \times A \rightarrow \mathbb{R} \)
- \( P : S \times A \rightarrow S \)
Reduced Formalism

**Knowns:**
- $S = \{\text{Blue, Red, Green, Black,} \ldots\}$
- Rewards in $\mathbb{R}$
- $A = \{\text{Wave, Clap, Stand}\}$

\[ s_0, a_0, r_0, s_1, a_1, r_1, s_2, \ldots \]

**Unknowns:**
- $R : S \times A \mapsto \mathbb{R}$
- $P : S \times A \mapsto S$

\[ r_i = R(s_i, a_i) \quad s_{i+1} = P(s_i, a_i) \]
This course

- Agent’s perspective: only policy under control
  - State representation, reward function pre-exist
This course

- Agent’s perspective: only **policy** under control
  - State representation, **reward function** pre-exist
  - **value function**: how good a state is in the long run
This course

- Agent’s perspective: only policy under control
  - State representation, reward function pre-exist
  - value function: how good a state is in the long run
  - model: reward function + state transition function
This course

- Agent’s perspective: only **policy** under control
  - State representation, **reward function** pre-exist
  - **value function**: how good a state is in the long run
  - **model**: reward function + state transition function
  - Focus on policy-learning algorithms, theoretical analyses
This course

- Agent’s perspective: only policy under control
- State representation, reward function pre-exist
- value function: how good a state is in the long run
- model: reward function + state transition function
- Focus on policy-learning algorithms, theoretical analyses
- Appeal: program by just specifying goals
This course

- Agent’s perspective: only policy under control
  - State representation, reward function pre-exist
  - value function: how good a state is in the long run
  - model: reward function + state transition function
  - Focus on policy-learning algorithms, theoretical analyses
  - Appeal: program by just specifying goals
  - Practice: need to pick the representation, reward
This course

- Agent’s perspective: only policy under control
  - State representation, reward function pre-exist
  - value function: how good a state is in the long run
  - model: reward function + state transition function
  - Focus on policy-learning algorithms, theoretical analyses
  - Appeal: program by just specifying goals
  - Practice: need to pick the representation, reward
  - videos
This course

- Agent’s perspective: only **policy** under control
  - State representation, **reward function** pre-exist
  - **value function**: how good a state is in the long run
  - **model**: reward function + state transition function
  - Focus on policy-learning algorithms, theoretical analyses
  - Appeal: program by just specifying goals
  - Practice: need to pick the representation, reward
  - videos

- Methodical approach
  - Solid foundation rather than comprehensive coverage
This course

- Agent’s perspective: only policy under control
  - State representation, reward function pre-exist
  - value function: how good a state is in the long run
  - model: reward function + state transition function
  - Focus on policy-learning algorithms, theoretical analyses
  - Appeal: program by just specifying goals
  - Practice: need to pick the representation, reward
  - videos

- Methodical approach
  - Solid foundation rather than comprehensive coverage
  - RL reading group
Syllabus

• Available on-line
Assignments

- Join piazza and edX!
Assignments

- Join piazza and edX!
- Read Chapter 2 (and 1 if you haven’t)
Assignments

- Join piazza and edX!
- Read Chapter 2 (and 1 if you haven’t)
- Send a reading response by 5pm Wednesday
Assignments

• Join piazza and edX!
• Read Chapter 2 (and 1 if you haven’t)
• Send a reading response by 5pm Wednesday
• Start on first short answer and programming assignments
Assignments

- Join piazza and edX!
- Read Chapter 2 (and 1 if you haven’t)
- Send a reading response by 5pm Wednesday
- Start on first short answer and programming assignments