Advanced Topics in Reinforcement Learning

Lecture 1: Course Introduction

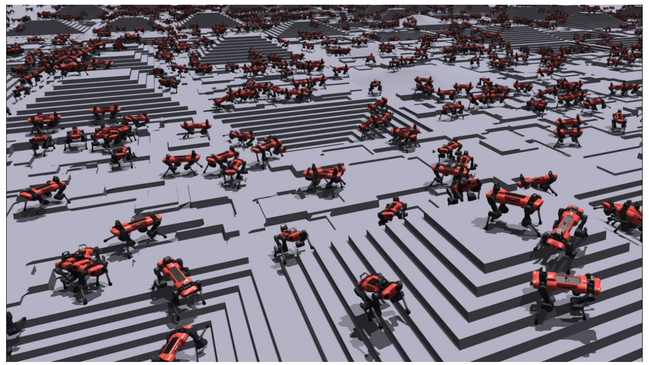
Josiah Hanna
University of Wisconsin — Madison

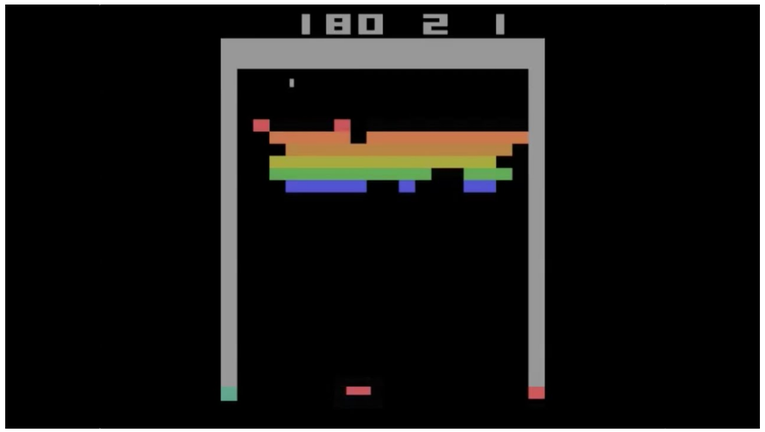
What is Reinforcement Learning?

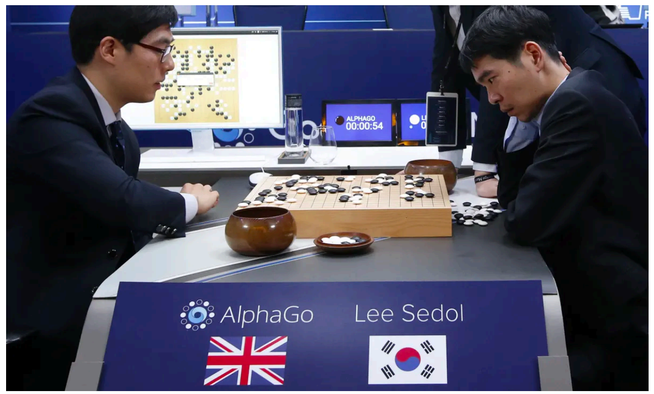
- Type of machine learning that focuses on learning from rewards and trial and error interaction.
- The learning agent takes actions, receives rewards, and over time learns to take actions that lead to the most reward.
- Think: training a dog to do tricks.

What Can RL Do?

- Play video games
- Play board games
- Control robots
- Recommend ads and web content
- Trade stocks
- Recommend medical treatments
- Control home thermostat systems
- Cooling of data centers
- Networking
- Databases
- Program Synthesis







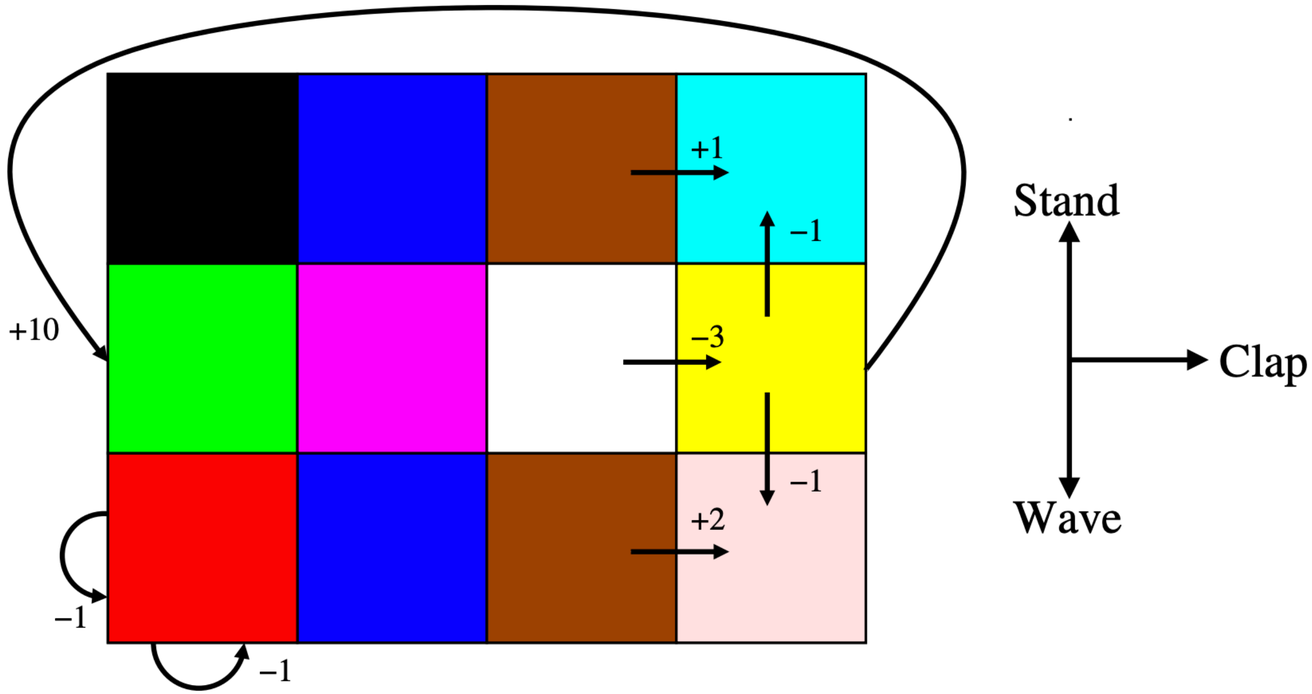
Be an RL Agent*

- You (as a class) are the learning agent.
- Three actions: stand, clap, or wave
- Observations: colors $\in \{$red, blue, orange, pink$\}$
- Rewards: depends on color you see and action you take.
- Goal: find the optimal policy.
  - Policy: mapping from colors to actions.
  - Optimal policy: policy that gives you the most reward.

*** Activity credit to Peter Stone.**

Be an RL Agent

- How did you learn?
- What structure does the world have?

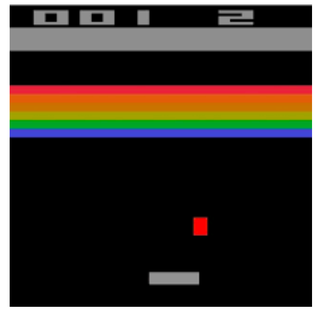


Reinforcement Learning Problems

- States: 3x4 grid
- Observations: colors
  - In this class (and the course textbook), states and observations will be treated the same.
- Actions: stand, clap, wave
- Rewards: +1, +2, -1, or +10
- State transitions dependent on action chosen.

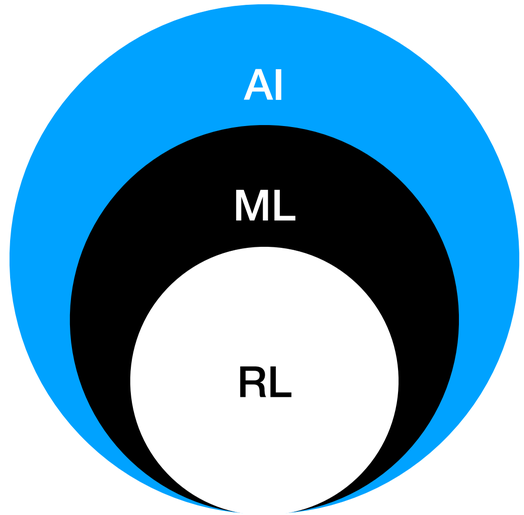
Example RL Problems

- What are the states? Actions? Rewards?
- Atari Breakout
- Home thermostat
- Stock trading



RL within Artificial Intelligence

- Supervised learning: learn from labelled examples.
  - Given a data set of {(X,Y)}, learn to map new instances of X to appropriate Y.
  - Ex: image classification, object detection, spam filtering.
- Unsupervised learning: discover structure in unlabelled data.
  - Ex: clustering, synthesizing images, language modeling
- Reinforcement learning: learn from rewarded interaction.
- Reinforcement learning also relates to AI planning.



Why is RL hard?

- Credit Assignment:
  - May take many actions before reward is received. Which ones were most important?
  - Example: you study 15 minutes a day all semester. The morning of the final exam, you eat a bowl of yogurt. You receive an A on the final. Was it the studying or the yogurt that led to the A?
  - Not trivial for people and animals either!
- Exploration vs. Exploitation
  - Should you keep trying actions that led to reward in the past or try new actions that might lead to even more reward?

Course Goals

After taking this course, you will be able to:

1. Explain fundamental RL concepts and apply fundamental RL algorithms.
2. Explain distinctions between advanced topics in RL research and the problems the research aims to address.
3. Complete an RL research project including implementation and experimental analysis of that implementation.

10,000 Foot Preview

- RL Fundamentals (~ 2 months)
  - Tabular methods
  - RL with function approximation
  - Research papers in RL fundamentals
- Advanced Topics in RL Research (~ 1 month)

Schedule Overview

- See course webpage: https://pages.cs.wisc.edu/~jphanna/teaching/2022fall_cs839/schedule.html

Classroom Environment

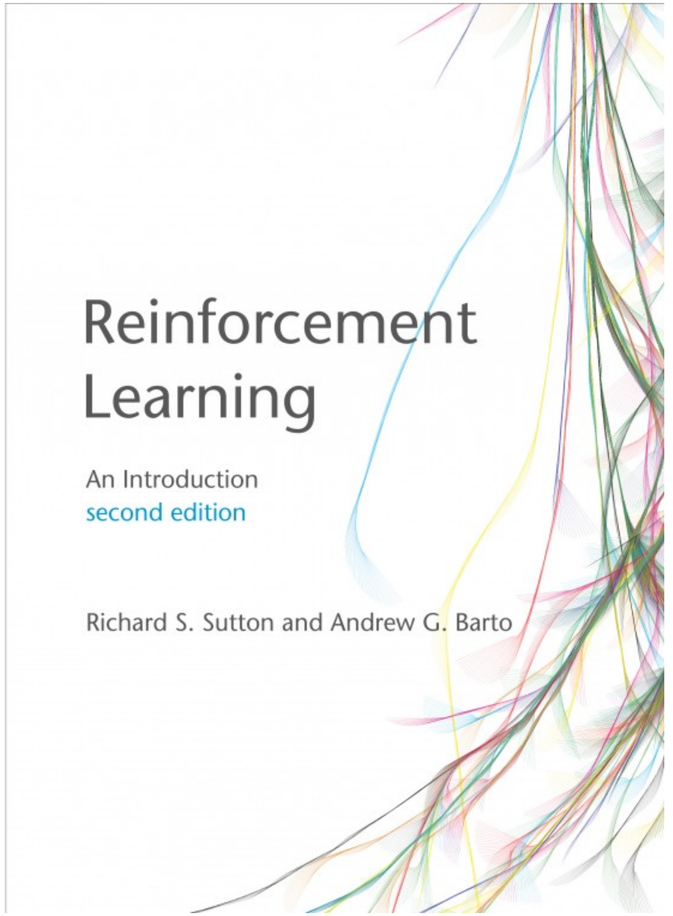
- Please, commit to helping create a climate where we treat everyone with dignity and respect.
- Listening to different viewpoints and approaches enriches our experience, and it is up to us to be sure others feel safe to contribute.
- Creating an environment where we are all comfortable learning is everyone's job: offer support and seek help from others if you need it, not only in class but also outside class while working with classmates.
- If class meetings conflict with your religious events, please let me know so that we can make arrangements for you.

Pre-requisites

- Probability and Statistics
  - Random variables, probability distributions, expectation, bias, variance, random sampling.
- Linear Algebra: dot-product, transpose, vector-matrix multiplication, matrix inverses.
- Calculus: basic differentiation with respect to scalar and vector variables.
- Machine Learning: neural networks, linear and logistic regression.
- Programming: assignments in Python. Possibly advantageous for final project to know Python.

Course Textbook

- Reinforcement Learning: an Introduction, 2nd edition
- Richard Sutton and Andrew Barto
- PDF link can be found on course website.



Class Periods

- **[Before class]** Required Weekly Readings
  - Most weeks will be from the course textbook.
  - **Submit reactions and questions by Monday at 4pm US central time.**
  - Send an email to jphanna@cs.wisc.edu with subject line “CS 839: Response for mm/dd” where mm/dd is the date of the Monday when the reading is due.
- Any questions?
- Lecture
- Student paper presentations

Reading Responses

- Please send in the email body, **not an attachment.**
- Credit is based on evidence that the reading was completed.
- Responses and questions will be used to shape the week's lecture.
- Possible responses:
  - Questions
  - Solutions to exercises in the book.
  - Critiques or suggestions for extensions.
  - What you want to learn about more.
  - Thoughts on what you find most important.

Student Presentations

- Goal: expose the entire class to a variety of RL research papers.
- One time during the semester, you will read and prepare a 10 minute presentation on an assigned paper.
- The number per class period may vary but in general there will be two speakers each class for the first part of the course.
- Sign-ups will be available soon.
- I have curated a list of papers to sign-up, however, feel free to email me with alternatives.

Grading

- Course Project (40%)
- Weekly Readings (10%)
- Class Participation (10%)
- Paper Presentations (10%)
- Programming Assignments (30%)

Late Policy

- 10% off for each 24 hour period late up to 5 days late.
    - Then zero for assignments aside from weekly reading responses.
- Weekly reading responses can be turned in up to the last day of class (late penalty of 50% after first five days).
- No late submission of final project.

Attendance Policy

- Class attendance and participation is necessary for participation component of grade.
- Absences will be approved if an email is sent before class starts.
- Please remain at home if unwell.
  - Classmates or others living with classmates may be immunocompromised.
  - Lectures and discussions will be recorded and uploaded to Canvas after class.

Logistics

- Course Webpage: https://pages.cs.wisc.edu/~jphanna/teaching/2022fall_cs839/
- Piazza: piazza.com/wisc/fall2022/cs839001
- Canvas: https://canvas.wisc.edu/courses/323656
- Presentation sign-ups: https://docs.google.com/spreadsheets/d/1-dce7-qzt8EVM4gYOLIl5WzYEGpioWM4x0VyA6QimzY/edit?usp=sharing
- Office Hours: Tuesdays @ 11am-12pm (after lecture) or by appointment.

Action Items

- Join Piazza!
- Read Chapters 2 and 3 of course textbook (skim 2.5-2.9).
- Send a reading response by 4pm on Monday.
- Presentation sign-ups (will be posted on Piazza).