CMPUT 365
Introduction to RL

Instructor: Scott Jordan

Some slides borrowed from Marlos C. Machado

Plan

- Introduction
- Course logistics
  - Instruction team
  - Pre-requisites
  - Flipped classroom
  - Textbook
  - Coursera
  - Academic integrity
  - Evaluation
- What is reinforcement learning?

About myself

- Name: Scott Jordan
- I’m from the United States (Oregon)
- Ph.D. from University of Massachusetts Amherst
  - Where Rich Sutton and Andy Barto did a lot of the early research on RL
- I am a postdoc (so don’t call me professor)
- Fun Fact: I have pet a shark
- Edmonton Goal: see a moose

Not like this →

Course overview and logistics

Key resources

- Syllabus
  - eClass, my website
- Teaching assistants


Shibhansh


Blanca


David


Gabor


Esraa

- TA email address: cmput365@ualberta.ca
- My email address: sjordan@ualberta.ca

I want to make this course is a **safe** and **inclusive** environment, for everyone.

It is ok to make mistakes.

We should all strive to be **respectful** to each other.

Syllabus

Read it!

Office: ATH 3-21
E-mail: sjordan@ualberta.ca
Web Page: https://scottjordan.github.io/courses/CMPUT365-W24/

Office hours: Scott: Monday 14:00 - 16:00 in ATH 3-31 (Athabasca Hall)
Blanca: Tuesday 14:00 - 16:00 in CAB 313
Gábor: Wednesday 09:00 - 11:00 in CAB 313
David: Wednesday 14:00 - 16:00 in CSC 3-50
Esra: Thursday 11:00 - 13:00 in CSC 3-26
Shibhansh: Friday 09:00 - 11:00 in CAB 3-13

Come to office hours
We are here to help

TA email address: cmput365@ualberta.ca
Do not personally email the TAs. They will only respond via cmput365@ualberta.ca.

Lecture room & time: CCIS 1-160, MWF 13:00 - 13:50
Attendance isn't mandatory although strongly encouraged.

Pre-requisites

- CMPUT 175 or CMPUT 275
- CMPUT 267

- Python
- Probability (e.g., expectations of random variables, conditional expectations)
- Calculus (e.g., partial derivatives)
- Linear algebra (e.g., vectors and matrices)

You should either be familiar with these topics or be ready
to pick them up quickly as needed by consulting outside resources.

This will be a flipped classroom!

- Roughly, you are initially introduced to **new topics outside** the classroom, so we can use the classroom time to explore topics in greater depth
    - A lecture is not necessarily the best use of class time

This will be a flipped classroom!

- Roughly, you are initially introduced to **new topics outside** the classroom, so we can use the classroom time to explore topics in greater depth
  - A lecture is not necessarily the best use of class time
- This is about creating meaningful learning opportunities for you, with more personalized interactions – to create **engaged** learning experiences

This will be a flipped classroom!

- Roughly, you are initially introduced to **new topics outside** the classroom, so we can use the classroom time to explore topics in greater depth
  - A lecture is not necessarily the best use of class time
- This is about creating meaningful learning opportunities for you, with more personalized interactions – to create **engaged** learning experiences
- I'm not doing this because it is easy, but because I think it is right
  - This is much much more work for me

https://giphy.com/gifs/usnationalarchives-moon-jfk-we-choose-to-go-the-1ULCj5IePUc8bI2

This will be a flipped classroom!

- Roughly, you are initially introduced to **new topics outside** the classroom, so we can use the classroom time to explore topics in greater depth
  - A lecture is not necessarily the best use of class time
- This is about creating meaningful learning opportunities for you, with more personalized interactions – to create **engaged** learning experiences
- I'm not doing this because it is easy, but because I think it is right
  - This is much much more work for me

- This **does not** mean lack of proper guidance, or that you have to teach yourself
- But you do have to become an **active** learner, instead of a passive learner

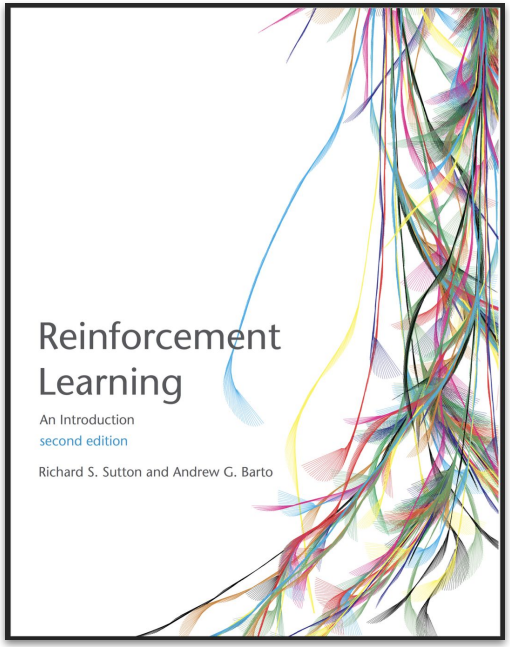
Required textbook

Reinforcement Learning: An Introduction

Richard S. Sutton & Andrew G. Barto

MIT Press. 2nd Edition.

http://www.incompleteideas.net/book/the-book-2nd.html



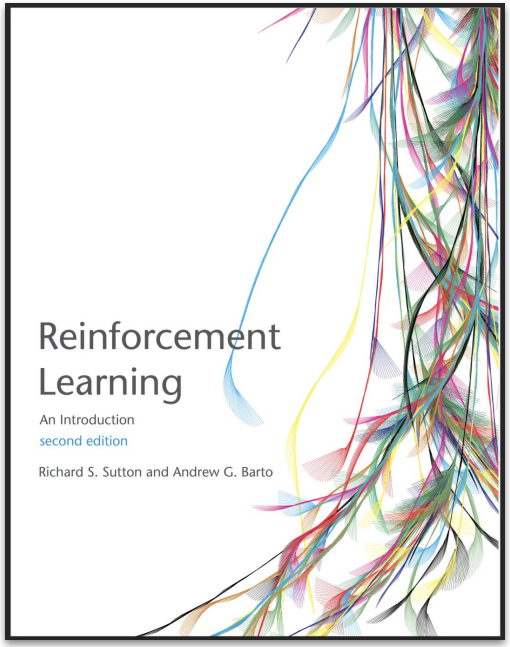
Required textbook

Reinforcement Learning: An Introduction

Richard S. Sutton & Andrew G. Barto

MIT Press. 2nd Edition.

http://www.incompleteideas.net/book/the-book-2nd.html



- You will need to read the book!

  (This is a flipped classroom, remember?)
- The book is really good!

GRADE EVALUATION

Assessment	Weight	Date
Quizzes (80% pass)	20%	Various
Assessments (notebooks on Coursera)	30%	Various
Midterm exam	20 %	Feb 12, 2024
Final exam	30%	Apr 25, 2024*

GRADE EVALUATION

Assessment	Weight	Date
Quizzes (80% pass)	20%	Various
Assessments (notebooks on Coursera)	30%	Various
Midterm exam	20 %	Feb 12, 2024
Final exam	30%	Apr 25, 2024*

GRADE EVALUATION

Assessment	Weight	Date
Quizzes (80% pass)	20%	Various
Assessments (notebooks on Coursera)	30%	Various
Midterm exam	20 %	Feb 12, 2024
Final exam	30%	Apr 25, 2024*

Late Policy

No late assignments will be accepted.

There are 12 quizzes and 8 programming assignments

The 2 quizzes and 1 programming assignment with the lowest grade will be dropped.

Academic integrity

- Code of Student Behaviour
- Student Conduct Policy
- Academic Integrity website

- **Appropriate collaboration:** You are allowed to discuss the quizzes and assignments with your classmates. Note, however, that you are not allowed to exchange any written text, code, or to give and/or receive detailed step-by-step instructions on how to solve the proposed problems.
- **Cell phones:** Cell phones are to be turned off during lectures, labs and seminars.
- **Recording and/or Distribution of Course Materials:** Audio or video recording, digital or otherwise, by students is allowed only with my prior written consent as a part of an approved accommodation plan.

Academic integrity – **Expectations for AI use**

The primary goal of this course is to foster *individual* critical, creative thinking, and problem-solving skills related to reinforcement learning. Thus, in order to achieve such learning outcomes, you can submit each practice quiz and graded assignment multiple times, which allows for many learning opportunities.

The use of advanced AI-tools based on large-language models such as ChatGPT is **strictly prohibited** for all quizzes and graded assignments. The only exception is their use for Python-related queries (but the use of such tools to help with the programming assignments themselves is still strictly prohibited).

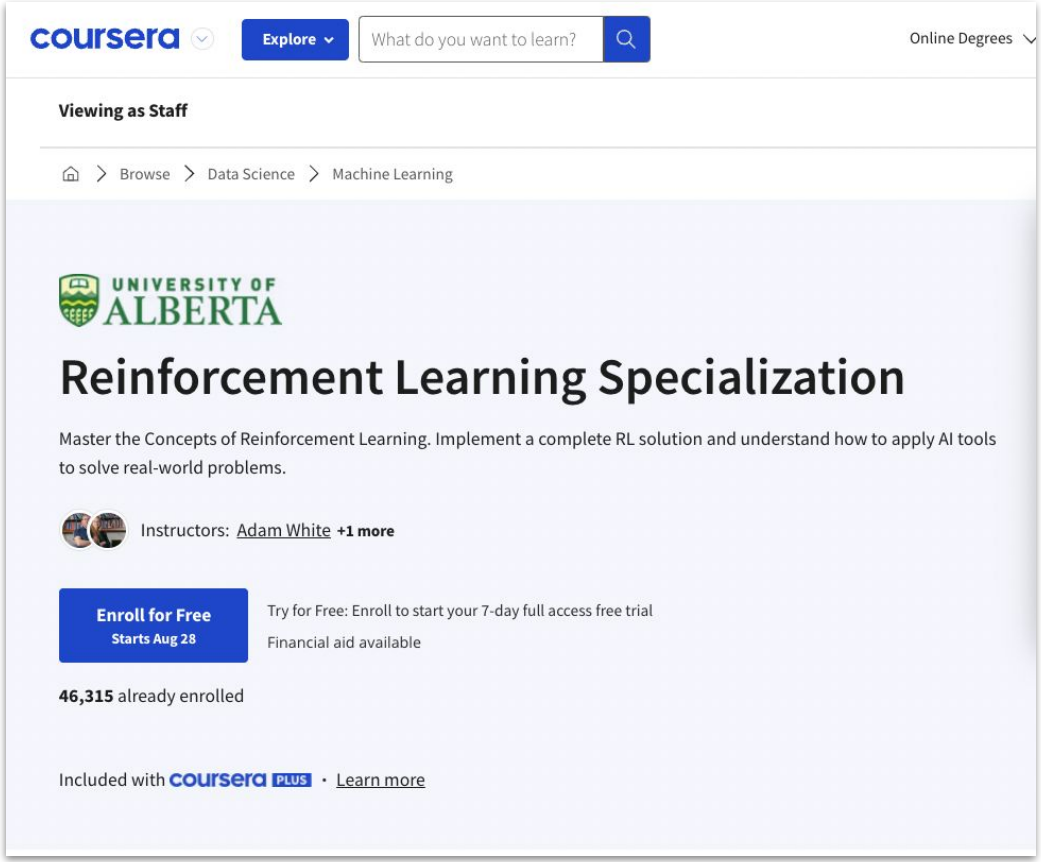
As stated in the university's AI-Squared - Artificial Intelligence and Academic Integrity webpage, *"learning is not only about the product; learning is also about the process of acquiring new knowledge or learning ways to think and reason."*

A request

Please don’t cheat

Coursera

- Coursera will be essential to CMPUT 365
- You should have been added to a private session of the RL courses (we used your university's email)
  - If you don't have access you should let me know!
  - **IMPORTANT: If you don't use the private session you won't get credit for submitted work!**



Schedule

- The course will be structured in “weeks”. **Not every week starts on Monday**
- Fundamentals of RL: modules 2-5
- Sample-based Learning methods: modules 2-4
- Prediction and Control with Function Approximation: modules 2-5

Schedule

- The course will be structured in “weeks”. Not every week starts on Monday
- Fundamentals of RL: modules 2-5
- Sample-based Learning methods: modules 2-4
- Prediction and Control with Function Approximation: modules 2-5

General Plan:

- Watch all videos for a module before the first class
- Quiz due on night of first class of the module (First Quiz is due Friday night)
- Programming assignment due night of last day of class
- Deadlines are midnight (listed as 12:00 AM the next day on Coursera)

Schedule

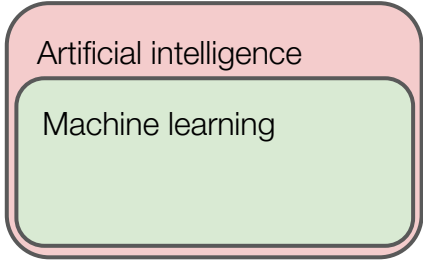
- First day of module: Short review, extra lecture/context to learning content
- Second day of module: mostly answering questions
  - Submit via Google Form (before day of class): https://forms.gle/uz68a3YxsqRMRkux8
  - You can bring your own questions to class in person
  - This is where the class can be tailored to your experience
- Third day of module: examples and in class exercises

Schedule

Week	Date	Topic	Deadlines (all due at 12:59:59)	Readings
1	Mon, Jan 8	Course overview Discussion about what is reinforcement learning		
1	Wed, Jan 10	Background review: Probability, statistics, linear algebra, and calculus		
1	Fri, Jan 12	Fundamentals of RL: An introduction to sequential decision-making	Quiz: Sequential decision-making	Chapter 2, up to §2.7 (pp. 25-36), and §2.10 (pp. 42-44)
2	Mon, Jan 15	Fundamentals of RL: An introduction to sequential decision-making		
2	Wed, Jan 17	Fundamentals of RL: An introduction to sequential decision-making	Program. assignment (Bandits & exploration / exploitation)	
2	Fri, Jan 19	Fundamentals of RL: Markov decision processes (MDPs)	Quiz: MDPs	Chapter 3, up to §3.3 (pp. 47-56)
3	Mon, Jan 22	Fundamentals of RL: Markov decision processes (MDPs)		
3	Wed, Jan 24	Fundamentals of RL: Markov decision processes (MDPs)		
3	Fri, Jan 26	Fundamentals of RL: Value functions & Bellman equations	Quiz: Value functions & Bellman equations 1	Chapter 3, §3.5-§3.8 (pp. 58-69)
4	Mon, Jan 29	Fundamentals of RL: Value functions & Bellman equations		
4	Wed, Jan 31	Fundamentals of RL: Value functions &	Quiz: Value functions & Bellman	

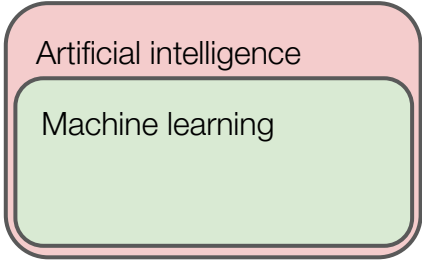
Questions?

What is reinforcement learning?



Machine learning

Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data



Machine learning

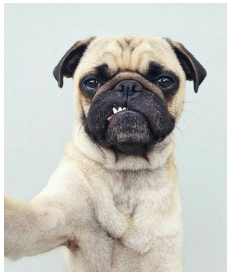
Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data

- "*Supervised learning* is learning from a training set of labeled examples provided by a knowledgeable external supervisor" (Sutton & Barto; 2018)

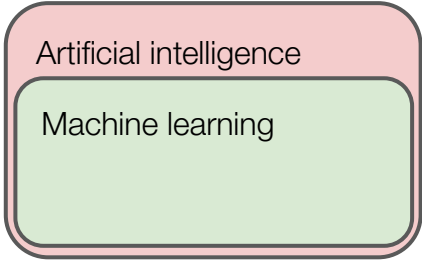
Cat

Cat

Not cat



Cat or not cat?



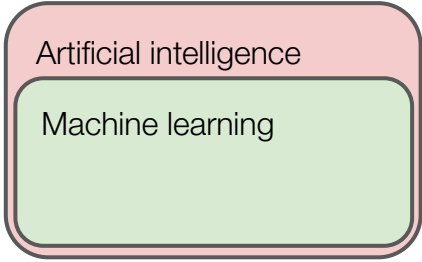
Machine learning

Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data

- "*Supervised learning* is learning from a training set of labeled examples provided by a knowledgeable external supervisor" (Sutton & Barto; 2018)

- "*Unsupervised learning* is typically about finding structure hidden in collections of unlabeled data" (Sutton & Barto; 2018)

https://www.researchgate.net/publication/309515576/figure/fig1/AS:960348551122945@1605976345743/uits-can-be-clustered-in-different-ways-images-are-from-Internet-a-Fruits-b.gif



Machine learning

Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data

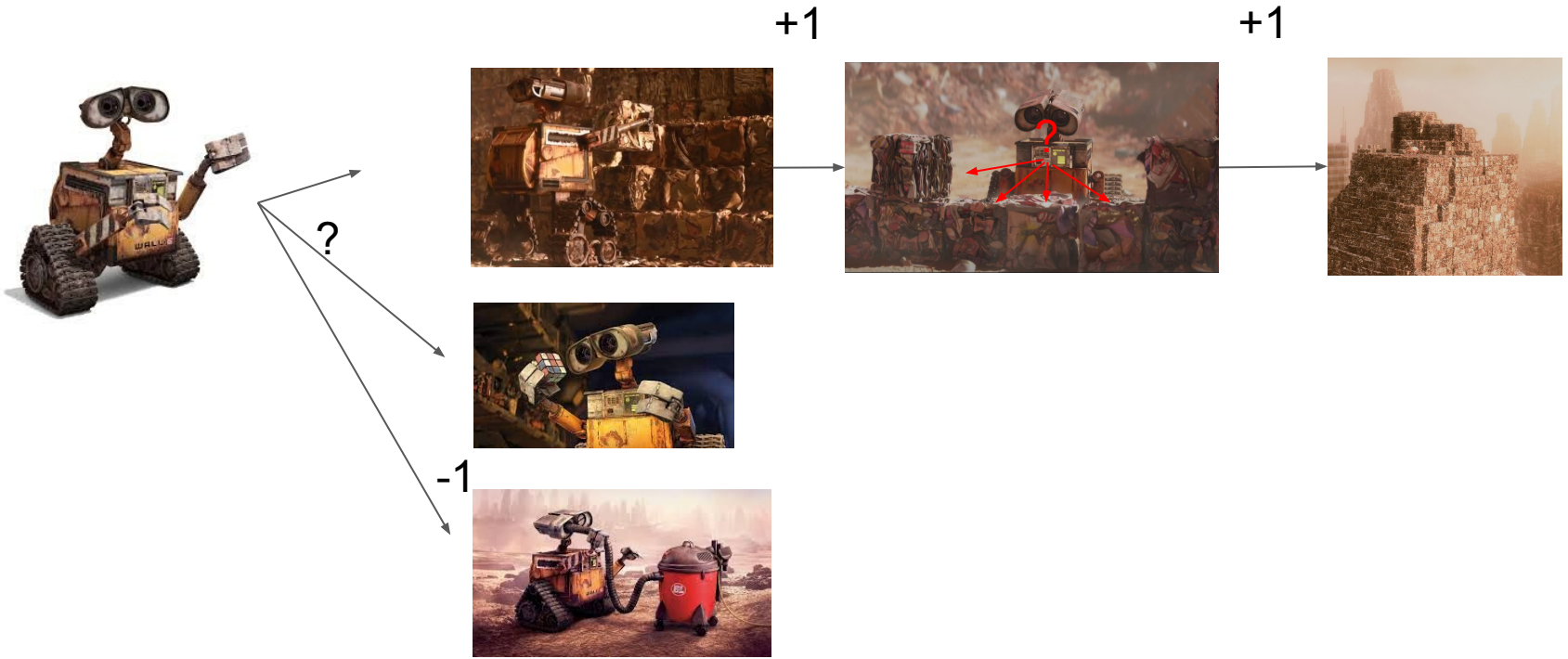
- "*Supervised learning* is learning from a training set of labeled examples provided by a knowledgeable external supervisor" (Sutton & Barto; 2018)

- "*Unsupervised learning* is typically about finding structure hidden in collections of unlabeled data" (Sutton & Barto; 2018)

… and *reinforcement learning*!

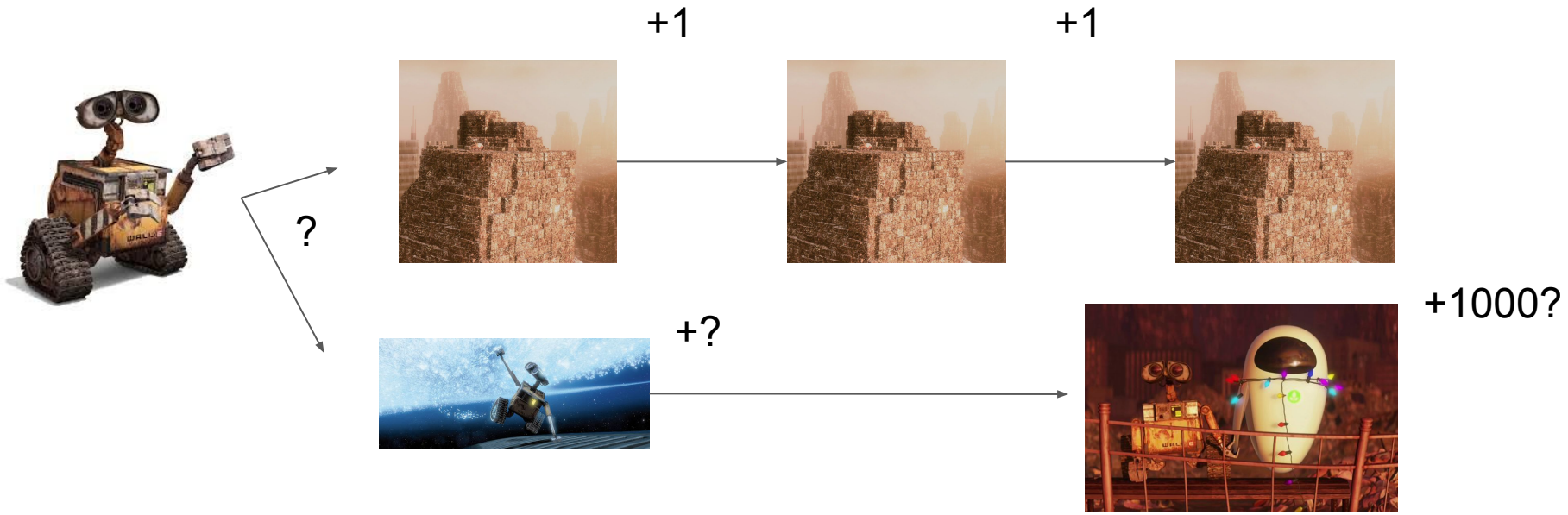
What problems can it solve?

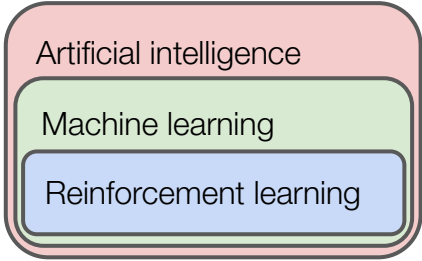
Sequential decision-making



What problems can it solve?

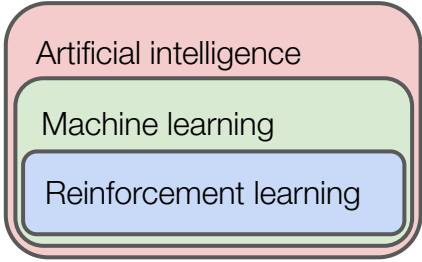
Sequential decision-making





Reinforcement learning

Reinforcement learning is a computational approach to learning from interaction to maximize a numerical reward signal (Sutton & Barto; 2018)

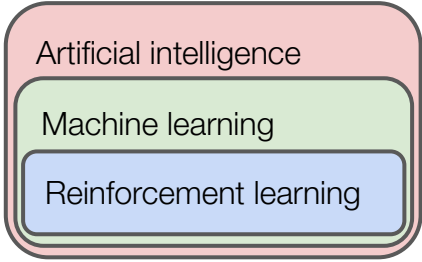


Reinforcement learning

Reinforcement learning is a computational approach to learning from interaction to maximize a numerical reward signal (Sutton & Barto; 2018)

- The idea of learning by interacting with our environment is very natural
- It is based on the idea of a learning system that wants something, and that adapts its behavior to get that

https://images.squarespace-cdn.com/content/v1/5bdf4a4d7c93279afce735f7/1548721628539-YNH1GBN9MXX3CQMCSYVR/treat+dog+in+background.jpg?format



Reinforcement learning

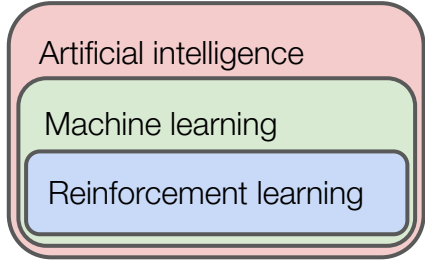
Reinforcement learning is a computational approach to learning from interaction to maximize a numerical reward signal (Sutton & Barto; 2018)

- The idea of learning by interacting with our environment is very natural
- It is based on the idea of a learning system that wants something, and that adapts its behavior to get that

Some features are unique to reinforcement learning:

- Trial-and-error
- The trade-off between exploration and exploitation
- The delayed credit assignment / delayed reward problem

https://images.squarespace-cdn.com/content/v1/5bdf4a4d7c93279afce735f7/1548721628539-YNH1GBN9MXX3CQMCSYVR/treat+dog+in+background.jpg?format



Reinforcement learning

Reinforcement learning is a computational a g from interaction to maximize a numerical reward signal (Sutto

- The idea of learning by interacti ment is very natural
- It is based on the idea of that wants something, and that to get that

Some features are u ent learning:

- Trial-and-er
- The trade-o oration and exploitation
- The delayed c gnment / delayed reward problem

Problem or solution?

https://images.squarespace-cdn.com/content/v1/5bdf4a4d7c93279afce735f7/1548721628539-YNH1GBN9MXX3CQMCSYVR/treat+dog+in+background.jpg?format

RL is now commonly deployed in the real-world

- **Recommendation systems**
  - Ads, news articles, videos, etc
- **General game playing**
  - Go, Chess, Shogi, Atari 2600, Starcraft, Minecraft, Gran Turismo
- **Industrial automation**
  - Cooling commercial buildings
  - Inventory management
  - Gas turbine optimization
  - Optimizing combustion in coal-fired power plants
- **Algorithms**
  - Video compression on YouTube
  - Faster matrix multiplication
  - Faster sorting algorithms
- **Control / Robotics**
  - Navigating stratospheric balloons
  - Plast control for nuclear fusion
- **And more (see Csaba's slides)**
  - COVID-19 border testing
  - Conversational agents
  - ...

Next class

- What **I** plan to do: A reminder about the required theoretical background
  - Probability (e.g., expectations of random variables, conditional expectations)
  - Calculus (e.g., partial derivatives)
  - Linear algebra (e.g., vectors and matrices)
  - I won't remind / teach you Python.

- What I recommend **YOU** to do for next class:
  - Make sure you have access to Coursera, eClass, and Slack
  - Brush up whatever you feel you are rusty on in terms of background
  - Read Chapter 1 of the textbook (not mandatory)
  - Start "Fundamentals of RL: An introduction to sequential decision-making" on Coursera (Week 1)

On intelligence, AGI, etc etc…

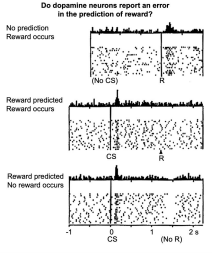
- People in the field have different, non-competing, perspectives and motivations
  - Some study RL to learn about / develop tools for solving sequential decision-making problems
  - Some look at RL as a computational model of intelligence

On intelligence, AGI, etc etc…

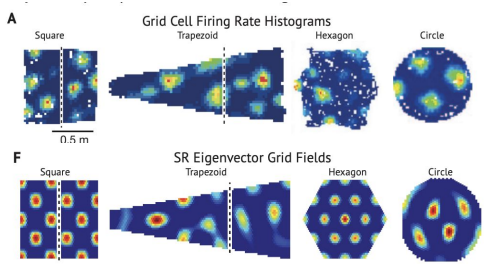
- People in the field have different, non-competing, perspectives and motivations
  - Some study RL to learn about / develop tools for solving sequential decision-making problems
  - Some look at RL as a computational model of intelligence
- I'll steer away from philosophical discussions and I'll focus on the algorithms
  - We should develop a critical view around these topics, and an ability to recognize hype / PR pieces

On intelligence, AGI, etc etc…

- People in the field have different, non-competing, perspectives and motivations
  - Some study RL to learn about / develop tools for solving sequential decision-making problems
  - Some look at RL as a computational model of intelligence
- I'll steer away from philosophical discussions and I'll focus on the algorithms
  - We should develop a critical view around these topics, and an ability to recognize hype / PR pieces
- Both perspectives are valid and both had had successes in the past



(Schultz, Dayan, & Montague; 1997)



(Stachenfeld, Botvinick, & Gershman; 2017)

(Silver et al.; 2016)

(Bellemare et al.; 2020)

(Degrave et al.; 2022)