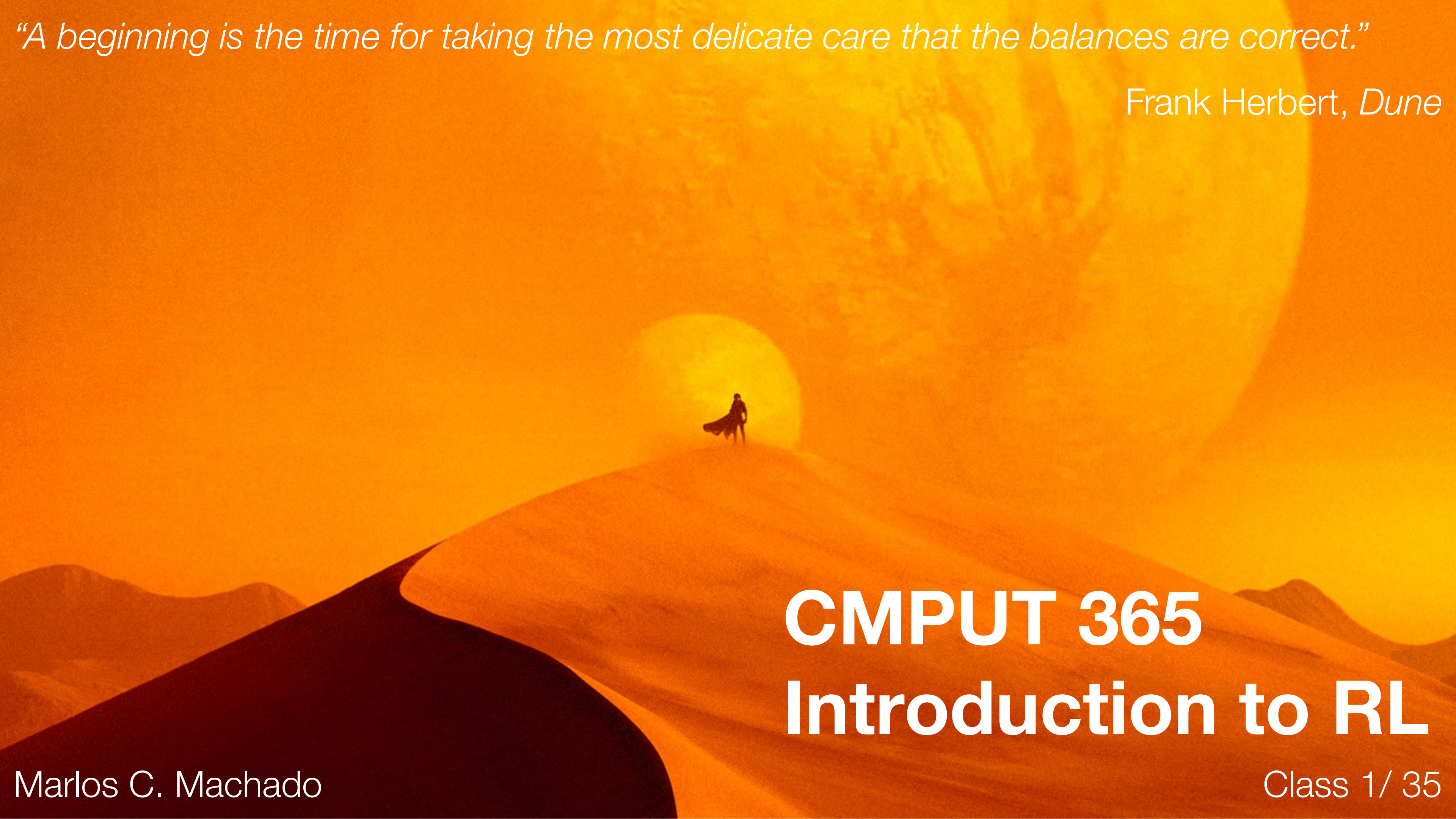


"A beginning is the time for taking the most delicate care that the balances are correct."

Frank Herbert, *Dune*

CMPUT 365
Introduction to RL

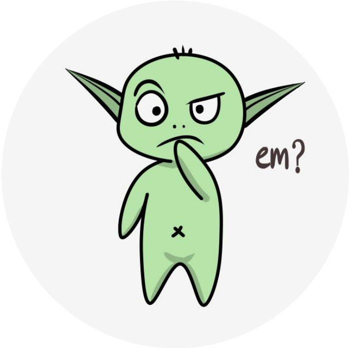
Marlos C. Machado

Class 1/ 35

Plan

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

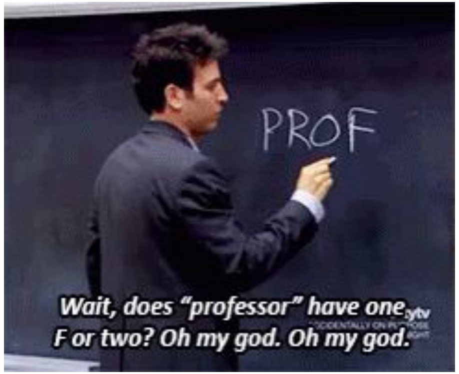
Please, interrupt me at any time!



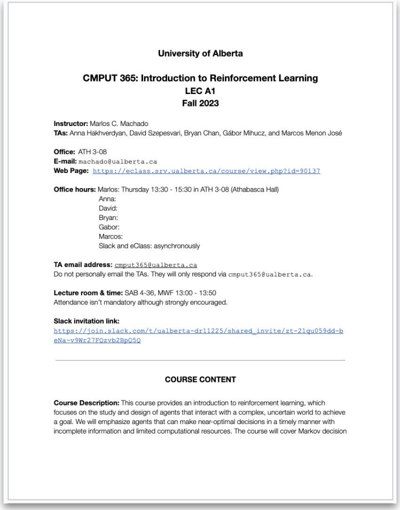
https://pngtree.com/freepng/question-expression-cartoon-illustration_4545209.ht

About myself

- Name: Marlos C. Machado
- I’m from Brazil
- I have been living in Edmonton for 10+ years
- I have 2 kids
- Ph.D. working on reinforcement learning
    - Interned at Microsoft Research, IBM Research, and DeepMind
- Worked 4 years at Google Brain and DeepMind
    - Among several other things, we deployed RL to fly balloons in the stratosphere
- I’m now a full-time professor at the University of Alberta



Course overview and logistics



University of Alberta

CMPUT 365: Introduction to Reinforcement Learning
LEC A1
Fall 2023

Instructor: Marlos C. Machado
TAs: Anna Hakhverdyan, David Szepesvari, Bryan Chan, Gábor Mihucz, and Marcos Menon José

Office: ATH 3-08
E-mail: machado@ualberta.ca
Web Page: https://eclass.srv.ualberta.ca/course/view.php?id=90137

Office hours: Marlos: Thursday 13:30 - 15:30 in ATH 3-08 (Athabasca Hall)
Anna:
David:
Bryan:
Gabor:
Marcos:
Slack and eClass: asynchronously

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

Lecture room & time: SAB 4-36, MWF 13:00 - 13:50
Attendance isn't mandatory although strongly encouraged.

Slack invitation link:
https://join.slack.com/t/ualberta-drl1225/shared_invite/zt-21qu059dd-beNa~v9Wr27FQzvb2BpQ5Q

COURSE CONTENT

Course Description: This course provides an introduction to reinforcement learning, which focuses on the study and design of agents that interact with a complex, uncertain world to achieve a goal. We will emphasize agents that can make near-optimal decisions in a timely manner with incomplete information and limited computational resources. The course will cover Markov decision

- eClass: ***link***
- Slack: ***link***
- My website: ***link***
- Google drive: ***link***

Start here!

http://www.phdcomics.com/comics/archive/phd051013s.gi

Key resources

- Syllabus
  - eClass, Slack, my website, Google Drive.
- Teaching assistants

Anna Bryan David Gabor Marcos

- TA email address: cmput365@ualberta.ca
- My email address: machado@ualberta.ca
- Slack invitation link: https://join.slack.com/t/ualberta-dr11225/shared_invite/zt-21qu059dd-beNa~v9Wr27FQzvb2BpQ5Q

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.

Office hours

- Slack and eClass: Asynchronous
- Marlos: Thursday 13:30 - 15:30 in ATH 3-08 (Athabasca Hall, 3-08)
- Anna: Monday 12:00 - 14:00 in CAB 3-13
- Bryan: Wednesday 14:00 - 16:00 in CAB 3-13
- David: Tuesday 13:00 - 15:00 in CSC 3-50
- Gabor: TBD
- Marcos: Friday 10:00 - 12:00 in CAB 3-13

Syllabus [eClass, Slack, website, Google Drive]

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

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

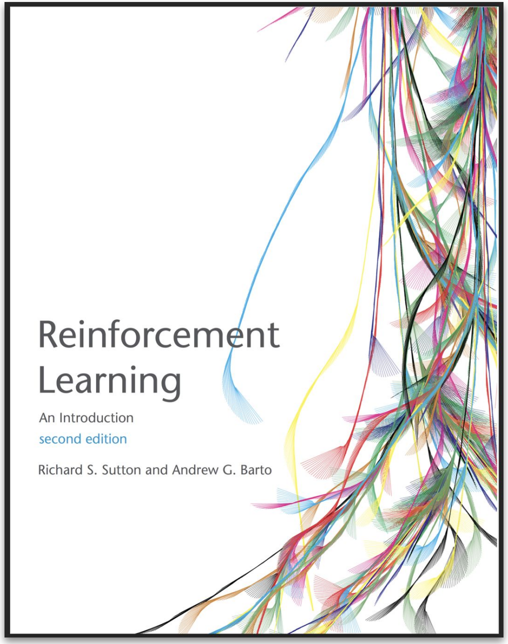
Required textbook

Reinforcement Learning: An Introduction

Richard S. Sutton & Andrew G. Barto

MIT Press. 2nd Edition.

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



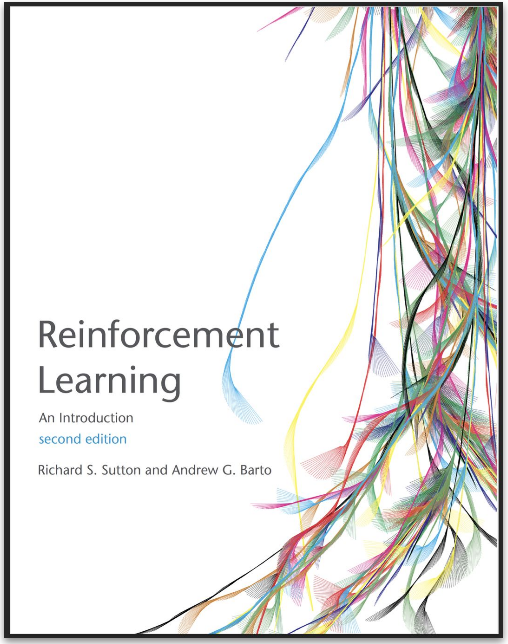
Required textbook

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
Practice quizzes (80% pass)	8 x 1.5% = 12%	Day of the 1st class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	8 x 3.5% = 28%	Day of the 2nd class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Midterm 1 exam	15 %	October 11, 2023
Midterm 2 exam	15%	November 10, 2023
Extra (practice quiz and notebook on Coursera's M3W4: Policy Gradient)	1.5 + 3.5% = 5%	December 6, 2023 *23:59:59*
Final exam	30%	December 14, 2023*

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	8 x 1.5% = 12%	Day of the 1st class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	8 x 3.5% = 28%	Day of the 2nd class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Midterm 1 exam	15 %	October 11, 2023
Midterm 2 exam	15%	November 10, 2023
Extra (practice quiz and notebook on Coursera's M3W4: Policy Gradient)	1.5 + 3.5% = 5%	December 6, 2023 *23:59:59*
Final exam	30%	December 14, 2023*

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	8 x 1.5% = 12%	Day of the 1st class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	8 x 3.5% = 28%	Day of the 2nd class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Extra (practice quiz and notebook on Coursera's M3W4: Policy Gradient)	1.5 + 3.5% = 5%	December 6, 2023 *23:59:59*
Final exam	30%	December 14, 2023*

Coursera, almost every* week (starting Monday): 40% + 5%

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	8 x 1.5% = 12%	Day of the 1st class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	8 x 3.5% = 28%	Day of the 2nd class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Extra (practice quiz and notebook on Coursera's M3W4: Policy Gradient)	1.5 + 3.5% = 5%	December 6, 2023 *23:59:59*

Coursera, almost every* week (<u>starting Monday</u>): 40% + 5%

Late submissions will not be accepted. There are 10 quizzes and 10 graded assignments. You're expected to do all of them, but s**t happens, so you can miss 2 of each and still get full marks.

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	8 x 1.5% = 12%	Day of the 1st class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	8 x 3.5% = 28%	Day of the 2nd class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Midterm 1 exam	15 %	October 11, 2023
Midterm 2 exam	15%	November 10, 2023
Final exam	30%	December 14, 2023*

Two midterms, summing to 30%. Closed book.

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	8 x 1.5% = 12%	Day of the 1st class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	8 x 3.5% = 28%	Day of the 2nd class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Midterm 1 exam	15 %	October 11, 2023
Midterm 2 exam	15%	November 10, 2023

Two midterms, summing to 30%. Closed book.
If you miss the midterm, you can can apply for an excused absence.
If granted, the weight of the missed midterm will be deferred to the final.

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	8 x 1.5% = 12%	Day of the 1st class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	8 x 3.5% = 28%	Day of the 2nd class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Midterm 1 exam	15 %	October 11, 2023
Midterm 2 exam	15%	November 10, 2023
Extra (practice quiz and notebook on Coursera's M3W4: Policy Gradient)	1.5 + 3.5% = 5%	December 6, 2023 *23:59:59*
Final exam	30%	December 14, 2023*

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	8 x 1.5% = 12%	Day of the 1st class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	8 x 3.5% = 28%	Day of the 2nd class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Midterm 1 exam	15 %	October 11, 2023
Midterm 2 exam	15%	November 10, 2023
Final exam	30%	December 14, 2023*

The final, worth 30%, will be about the whole course.

If you miss the final, you can apply to a deferred final examination.

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	**Weight**	**Date**
Practice quizzes (80% pass)	8 x 1.5% = 12%	Day of the 1st class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	8 x 3.5% = 28%	Day of the 2nd class on the topic of the week at 23:59:59 (see Course schedule, at the end, for details)
Midterm 1 exam	15 %	October 11, 2023
Midterm 2 exam	15%	November 10, 2023
Extra (practice quiz and notebook on Coursera's M3W4: Policy Gradient)	1.5 + 3.5% = 5%	December 6, 2023 *23:59:59*
Final exam	30%	December 14, 2023*

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	8 x 1.5% = 12%	
Assessments (graded quizzes / notebooks on Coursera)	8 x 3.5% = 28%	
Midterm 1 exam	15 %	
Midterm 2 exam	15%	
Extra (practice quiz and notebook on Coursera's M3W4: Policy Gradient)	1.5 + 3.5% = 5%	
Final exam	30%	

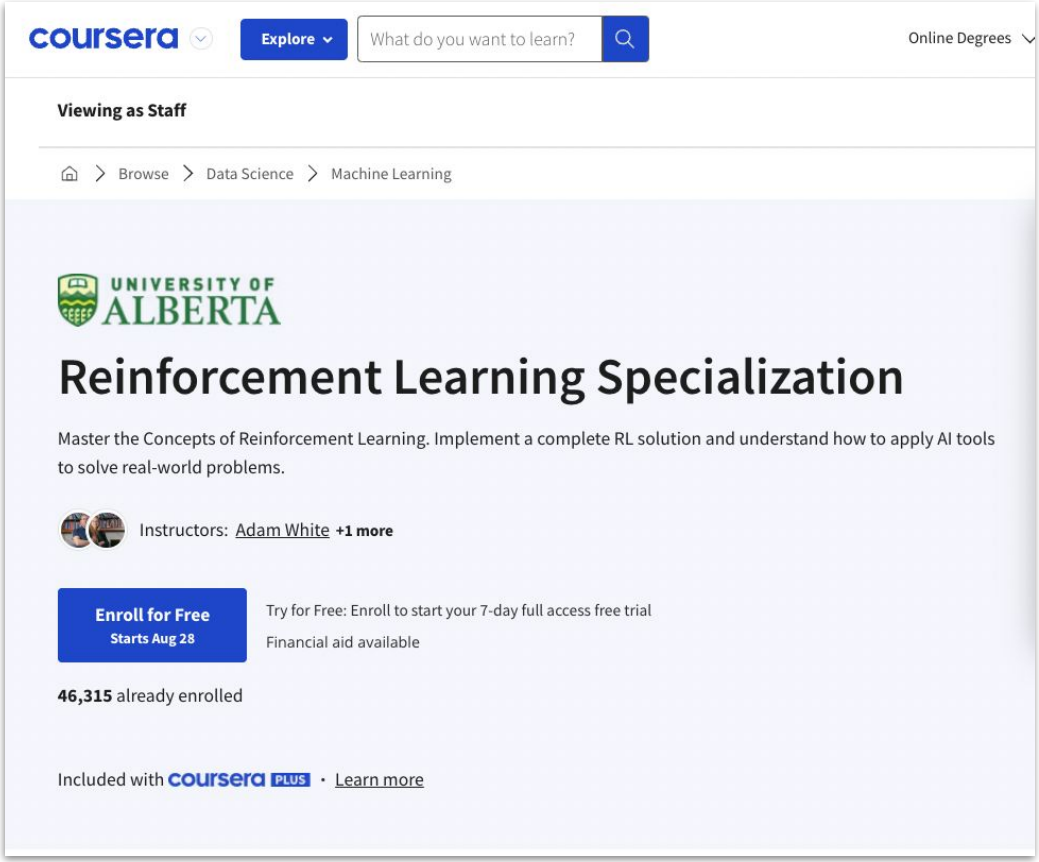
Total: **105%.**
You can **not-submit 2 quizzes** and **2 assessments.**

Grades **will not** be rounded at the end, and **no more extra marks** will be given. **No exceptions.**

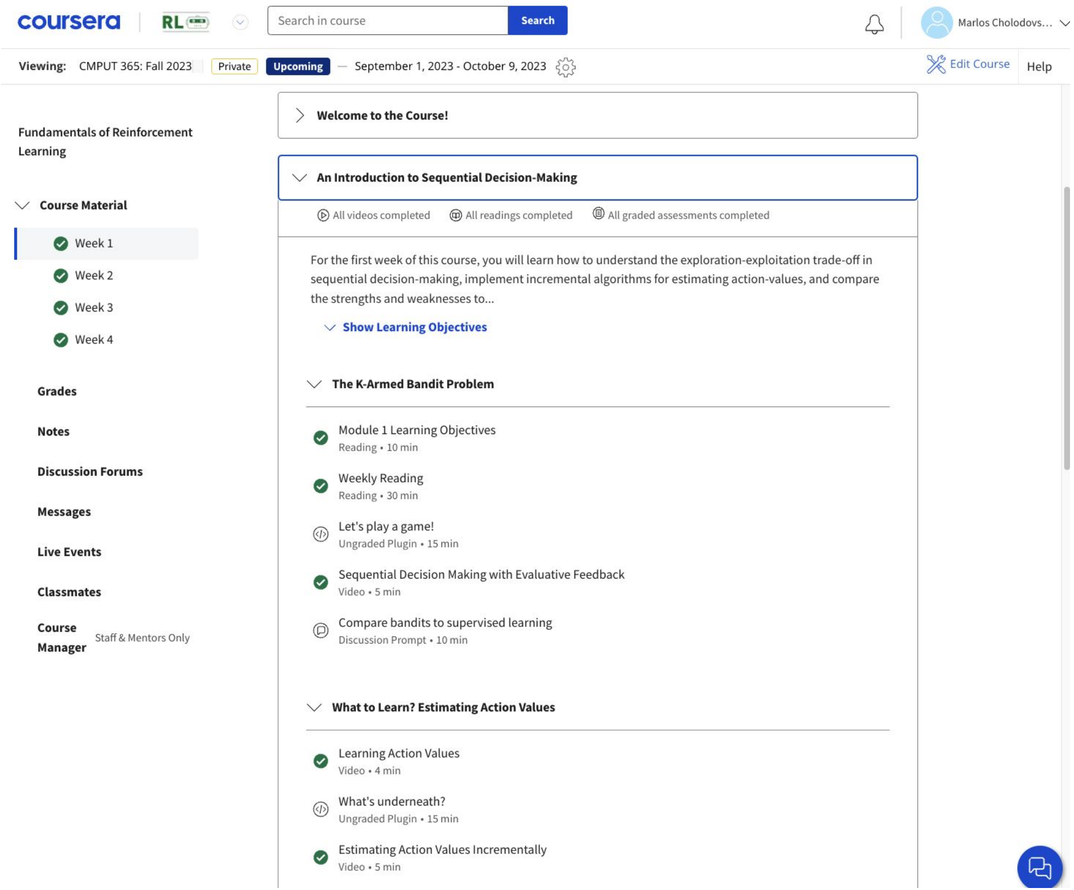
Syllabus [eClass, Slack, website, Google Drive]

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!**



Coursera


coursera
RL
Search in course
Search
Marlos Cholodovs...
Viewing: CMPUT 365: Fall 2023
Private
Upcoming
September 1, 2023 - October 9, 2023
Edit Course
Help
Fundamentals of Reinforcement Learning
Course Material
Week 1
Week 2
Week 3
Week 4
Grades
Notes
Discussion Forums
Messages
Live Events
Classmates
Course Manager
Staff & Mentors Only
Welcome to the Course!
An Introduction to Sequential Decision-Making
All videos completed
All readings completed
All graded assessments completed
For the first week of this course, you will learn how to understand the exploration-exploitation trade-off in sequential decision-making, implement incremental algorithms for estimating action-values, and compare the strengths and weaknesses to...
Show Learning Objectives
The K-Armed Bandit Problem
Module 1 Learning Objectives
Reading • 10 min
Weekly Reading
Reading • 30 min
Let's play a game!
Ungraded Plugin • 15 min
Sequential Decision Making with Evaluative Feedback
Video • 5 min
Compare bandits to supervised learning
Discussion Prompt • 10 min
What to Learn? Estimating Action Values
Learning Action Values
Video • 4 min
What's underneath?
Ungraded Plugin • 15 min
Estimating Action Values Incrementally
Video • 5 min

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."*

Schedule

- The course will be structured in “weeks”. **Not every week starts on Monday**
- We have 12 weeks of content classes and we’ll cover 12 weeks of the MOOC
  - This corresponds to 8 chapters of the textbook (+1 for extra marks)

Schedule

- The course will be structured in “weeks”. **Not every week starts on Monday**
- We have 12 weeks of content classes and we’ll cover 12 weeks of the MOOC
  - This corresponds to 8 chapters of the textbook (+1 for extra marks)
- My overall (and tentative) plan for each one of the 3 days of the course-week:
  - 1st day: Non-comprehensive summary of the topic of the week (and additional content)
  - 2nd day: Mostly respond to questions asked live, on eClass, and Slack, about the content
  - 3rd day: Additional exercises, some in class activities

Schedule

- The course will be structured in "weeks". **Not every week starts on Monday**
- We have 12 weeks of content classes and we'll cover 12 weeks of the MOOC
  - This corresponds to 8 chapters of the textbook (+1 for extra marks)
- My overall (and tentative) plan for each one of the 3 days of the course-week:
  - 1st day: <u>Non-comprehensive summary</u> of the topic of the week (and additional content)
  - 2nd day: <u>Mostly respond to questions</u> asked live, on eClass, and Slack, about the content
  - 3rd day: <u>Additional exercises</u>, some in class activities
- A practice quiz is due in the 1st day of almost every course-week
- A graded assignment is due in the 2nd day of almost every course-week
- The deadline for submitting assignments and quizzes is 23:59:59

Schedule

Course Schedule & Assigned Readings

Week	Date	Topic	Deadlines (all due at 23:59:59)	Readings
1	Wed, Sep 6	Course overview Discussion about what is reinforcement learning		
1	Fri, Sep 8	Background review: Probability, statistics, linear algebra, and calculus		
2	Mon, Sep 11	Fundamentals of RL: An introduction to sequential decision-making	Practice quiz (Sequential decision-making)	Chapter 2, up to §2.7 (pp. 25-36), and §2.10 (pp. 42-44)
2	Wed, Sep 13	Fundamentals of RL: An introduction to sequential decision-making	Program. assignment (Bandits & exploration / exploitation)	
2	Fri, Sep 15	Fundamentals of RL: An introduction to sequential decision-making		
3	Mon, Sep 18	Fundamentals of RL: Markov decision processes (MDPs)	Practice quiz (MDPs)	Chapter 3, up to §3.3 (pp. 47-56)
3	Wed, Sep 20	Fundamentals of RL: Markov decision processes (MDPs)		
3	Fri, Sep 22	Fundamentals of RL: Markov decision processes (MDPs)		
4	Mon, Sep 25	Fundamentals of RL: Value functions & Bellman equations	Practice quiz (Value functions & Bellman equations)	Chapter 3, §3.5-§3.8 (pp. 58-69)
4	Wed, Sep 27	Fundamentals of RL: Value functions & Bellman equations	Graded quiz (Value functions & Bellman equations)	
4	Fri, Sep 29	Fundamentals of RL: Value functions & Bellman equations		
5	Mon, Oct 2	Fundamentals of RL: Dynamic programming	Practice quiz (Dynamic programming)	Chapter 4, §4.1-§4.4 (pp. 73-84); §4.6-§4.7 (pp. 86-89)
5	Wed, Oct 4	Fundamentals of RL: Dynamic programming	Program. Assignment (Optimal policies with dynamic programming)	
5	Fri, Oct 6	Fundamentals of RL: Dynamic programming		
	Mon, Oct 9	Thanksgiving		
	Wed, Oct 11	**Midterm exam 1**		
6	Fri, Oct 13	Sample-based learning methods: MC methods for Prediction & Control		Chapter 5, up to §5.5 (pp. 91-108); §5.10 (pp. 115-116)
6	Mon, Oct 16	Sample-based learning methods: MC methods for Prediction & Control	Graded quiz (Off-policy Monte Carlo)	
6	Wed, Oct 18	Sample-based learning methods: MC methods for Prediction & Control		
7	Fri, Oct 20	Sample-based learning methods: TD learning for prediction	Practice quiz (Advantages of TD)	Chapter 6, up to §6.3 (pp. 119-128)
7	Mon, Oct 23	Sample-based learning methods: TD learning for prediction	Program. Assignment (Policy evaluation with TD learning)	
7	Wed, Oct 25	Sample-based learning methods: TD learning for prediction		
8	Fri, Oct 27	Sample-based learning methods: TD learning for control	Practice quiz (Expected Sarsa)	Chapter 6, §6.4-§6.6 (pp. 129-134); §6.10 (p. 138)
8	Mon, Oct 30	Sample-based learning methods: TD learning for control	Program. assignment (Q-learning & Expected Sarsa)	
8	Wed, Nov 1	Sample-based learning methods: TD		
		learning for control		
9	Fri, Nov 3	Sample-based learning methods: Planning, learning, & acting	Practice quiz (Dealing with inaccurate models)	Chapter 8, up to §8.3 (pp. 159-168); §8.12-§8.13 (pp. 210-191)
9	Mon, Nov 6	Sample-based learning methods: Planning, learning, & acting	Program. assignment (Dyna-Q & Dyna-Q+)	
9	Wed, Nov 8	Sample-based learning methods: Planning, learning, & acting		
	Fri, Nov 10	**Midterm exam 2**		
	Mon, Nov 13	Remembrance day holiday in lieu		
	Nov 14 - Nov 17	Reading week		
10	Mon, Nov 20	Prediction and Control with FA: On-policy prediction with approx.	Practice quiz (On-policy prediction with approximation)	Chapter 9, up to §9.4 (pp. 197-209)
10	Wed, Nov 22	Prediction and Control with FA: On-policy prediction with approx.	Program. assignment (Semi-gradient TD(0) with state aggregation)	
10	Fri, Nov 24	Prediction and Control with FA: On-policy prediction with approx.		
11	Mon, Nov 27	Prediction and Control with FA: Constructing features for prediction	Practice quiz (Constructing features for prediction)	Chapter 9, §9.5 (pp. 210-222) and §9.12 (pp. 236-237)
11	Wed, Nov 29	Prediction and Control with FA: Constructing features for prediction	Program. assignment (Semi-gradient TD with a neural network)	
11	Fri, Dec 1	Prediction and Control with FA: Constructing features for prediction		
12	Mon, Dec 4	Prediction and Control with FA: Control with	Practice quiz (Control with	Chapter 10, up to §10.1 (pp.
		approximation	approximation)	243-248); §10.3 (pp. 249-252), §10.6 (p. 256)
12	Wed, Dec 6	Prediction and Control with FA: Control with approximation	Program. assignment (Function approximation & control)	
12	Fri, Dec 8	Prediction and Control with FA: Control with approximation		
Thu, Dec 14 at 2pm		**Final exam** (Students must verify this date on BearTracks when it is posted)		

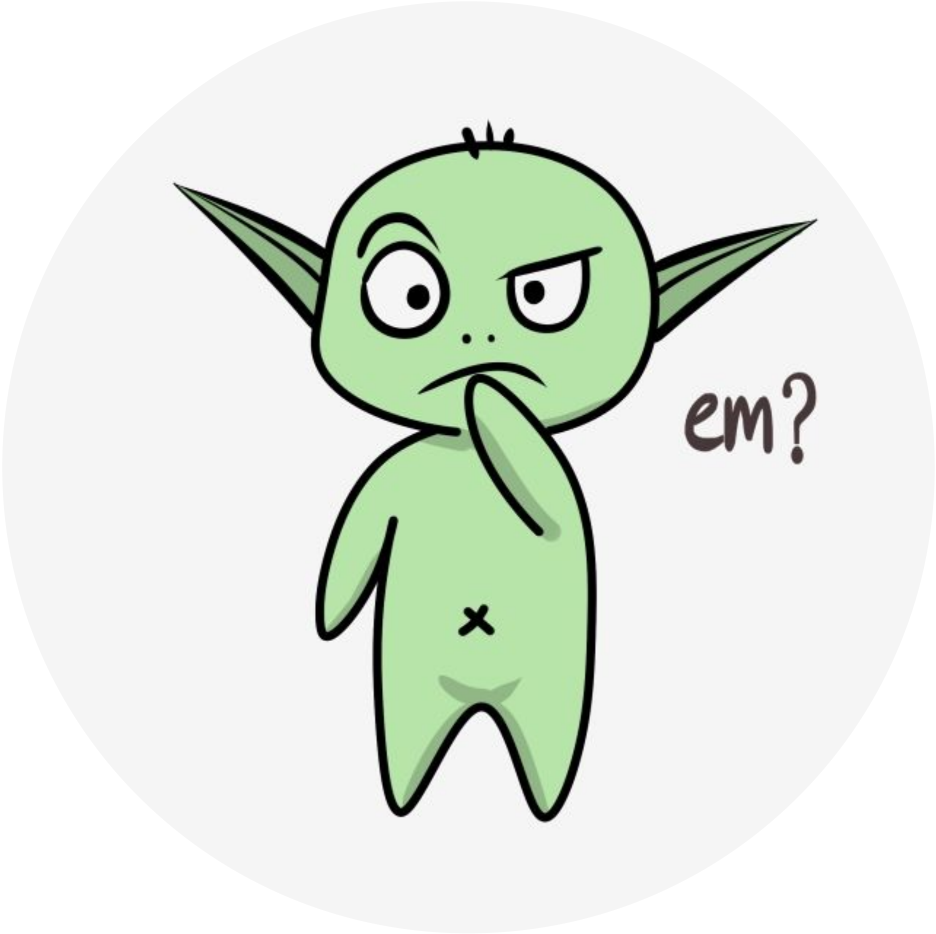
Syllabus [eClass, Slack, website, Google Drive]

Schedule

Course Schedule & Assigned Readings

Week	Date	Topic	Deadlines (all due at 23:59:59)	Readings
1	Wed, Sep 6	Course overview Discussion about what is reinforcement learning		
1	Fri, Sep 8	Background review: Probability, statistics, linear algebra, and calculus		
2	Mon, Sep 11	Fundamentals of RL: An introduction to sequential decision-making	Practice quiz (Sequential decision-making)	Chapter 2, up to §2.7 (pp. 25-36), and §2.10 (pp. 42-44)
2	Wed, Sep 13	Fundamentals of RL: An introduction to sequential decision-making	Program. assignment (Bandits & exploration / exploitation)	
2	Fri, Sep 15	Fundamentals of RL: An introduction to sequential decision-making		

Syllabus [eClass, Slack, website, Google Drive]



https://pngtree.com/freepng/question-expression-cartoon-illustration_4545209.ht

What is reinforcement learning?

Artificial intelligence

"AI is the ability of machines to perform tasks that are typically associated with human intelligence, such as learning and problem-solving." –Wikipedia

Thinking Humanly "The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	**Thinking Rationally** "The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly "The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	**Acting Rationally** "Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998) "AI . . . is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Figure 1.1 Some definitions of artificial intelligence, organized into four categories.

(Russell & Norvig; 2010)

The cognitive modeling approach ← Thinking Humanly

The "laws of thought" approach ← Thinking Rationally

The Turing Test approach ← Acting Humanly

The rational agent approach ← Acting Rationally

Artificial intelligence

Artificial intelligence

"AI is the ability of machines to perform tasks that are typically associated with human intelligence, such as learning and problem-solving." –Wikipedia

The cognitive modeling approach ←	**Thinking Humanly**	**Thinking Rationally**	→ The "laws of thought" approach
	"The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)	
The Turing Test approach ←	**Acting Humanly** "The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Intelligence is the computational part of the ability to achieve goals in the world." (McCarthy, 1997)	→ **The rational agent approach**

Figure 1.1 Some definitions of artificial intelligence, organized into four categories.

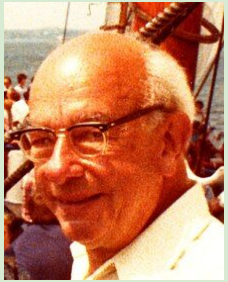
(Russell & Norvig; 2010)

Artificial intelligence

"AI is the ability of machines to perform tasks that are typically associated with human intelligence, such as learning and problem-solving." –Wikipedia

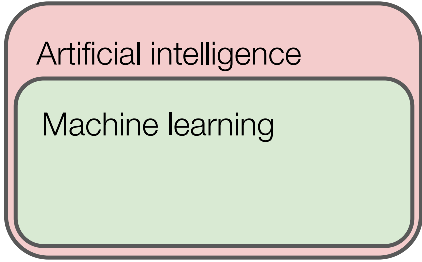
The less a science has advanced, the more its terminology tends to rest on an uncritical assumption of mutual understanding.

– W. V. Quine

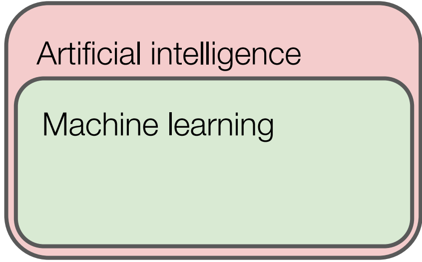


https://upload.wikimedia.org/wikipedia/commons/c/cd/Willard_Van_Orman_Quine_on_Bluenose_II_in_Halifax_NS_harbor_1980.

Machine learning



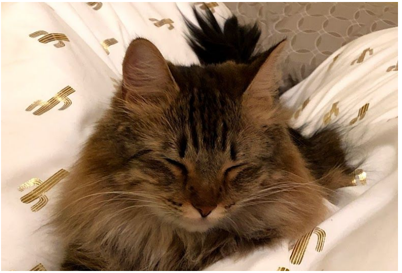
Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data



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

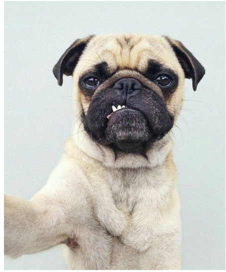
- "*Supervised learning* is learning from a training set of labeled examples provided by a knowledgeable external supervisor" (Sutton & Barto; 2018)



Cat

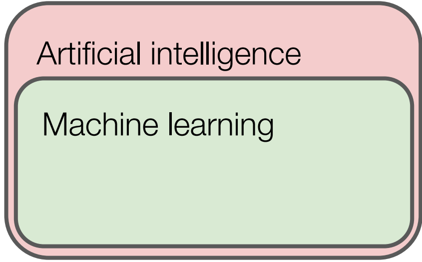
Cat

Not cat



Cat or not cat?

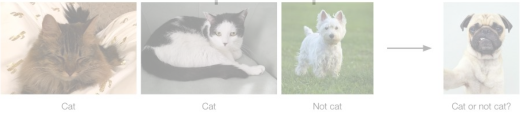
https://i0.statig.com.br/bancodeimagens/0r/yu/ls/0ryulsnl6uvkyvy6vv98bkms9.j
https://www.dogue.com.au/cdn/shop/articles/8447610bb0578436f1faa531b6564ab5.jpg?v=1687241



Machine learning

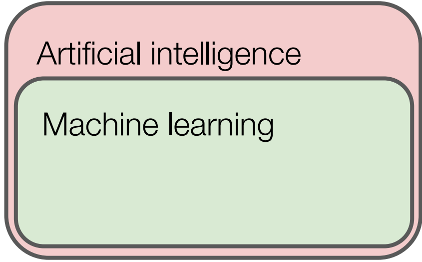
Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data

- "*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)

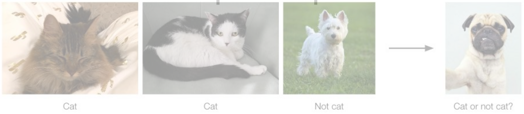
https://www.researchgate.net/publication/309515576/figure/fig1/AS:960348551122945@1605976345743/
uits-can-be-clustered-in-different-ways-images-are-from-Internet-a-Fruits-b.gif



Machine learning

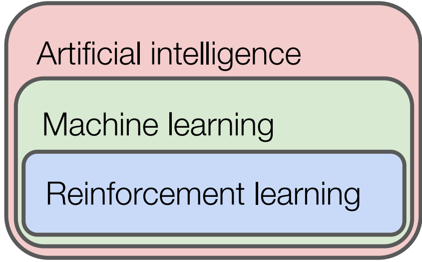
Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data

- "*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)

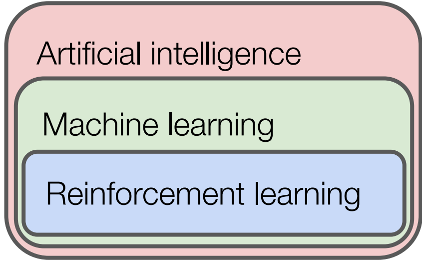
… and *reinforcement learning*!



Reinforcement learning

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

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

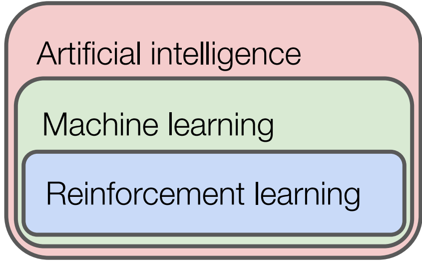


Reinforcement learning

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

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

Artificial intelligence

Machine learning

Reinforcement learning

Reinforcement learning

Problem or solution?

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

- The idea of learning by interacti… …nment 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

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
  - ...

On intelligence, AGI, etc etc…

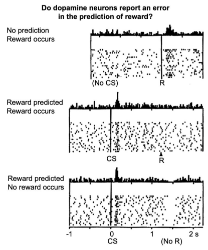
- People in the field have different, non-competing, perspectives and motivations
  - Some study RL to learn about / develop tools for solving sequential decision-making problems
  - Some look at RL as a computational model of intelligence

On intelligence, AGI, etc etc…

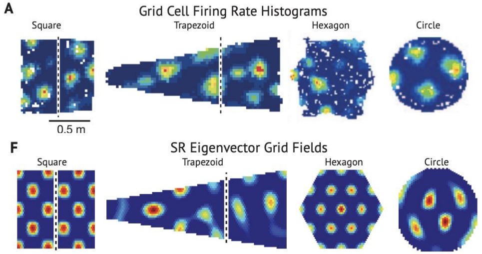
- People in the field have different, non-competing, perspectives and motivations
  - Some study RL to learn about / develop tools for solving sequential decision-making problems
  - Some look at RL as a computational model of intelligence
- I’ll steer away from philosophical discussions and I’ll focus 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)

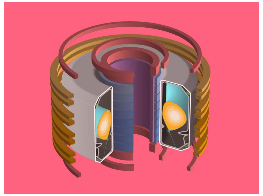


(Stachenfeld, Botvinick, & Gershman; 2017)

(Silver et al.; 2016)

(Bellemare et al.; 2020)

(Degrave et al.; 2022)



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)