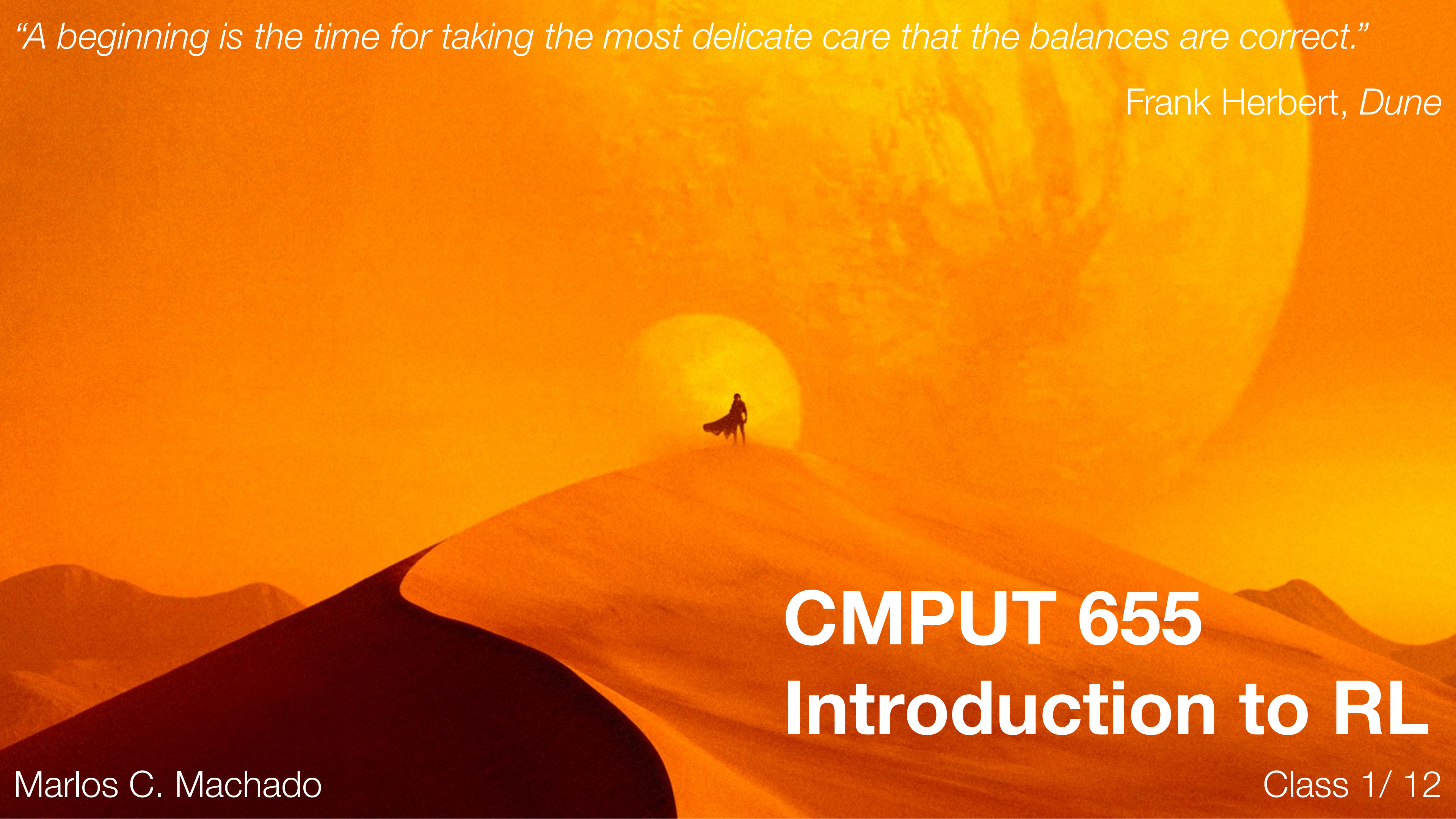


"A beginning is the time for taking the most delicate care that the balances are correct."

Frank Herbert, *Dune*

CMPUT 655
Introduction to RL

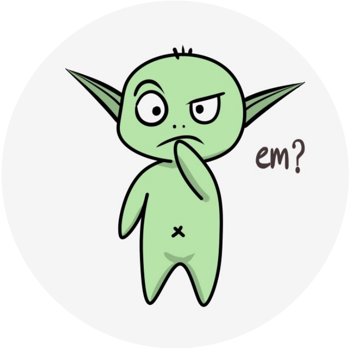
Marlos C. Machado

Class 1/ 12

Plan

- Introduction
- Course logistics
  - Instruction team
  - Pre-requisites
  - Flipped classroom
  - Textbook
  - Coursera
  - Academic integrity
  - Evaluation
- What is reinforcement learning?
- Probability & statistics
- Linear algebra
- Calculus

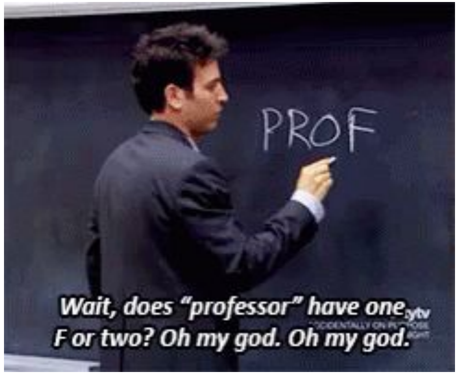
Please, interrupt me at any time!



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

About myself

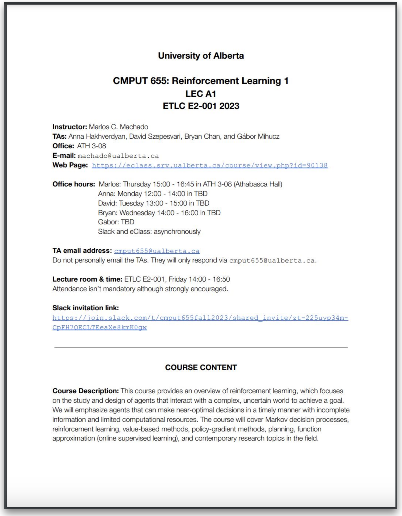
- Name: Marlos C. Machado
- I’m from Brazil
- I have been living in Edmonton for 10+ years
- I have 2 kids
- Ph.D. working on reinforcement learning
    - Interned at Microsoft Research, IBM Research, and DeepMind
- Worked 4 years at Google Brain and DeepMind
    - Among several other things, we deployed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



Course overview and logistics

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

Start here!



University of Alberta

CMPUT 655: Reinforcement Learning 1
LEC A1
ETLC E2-001 2023

Instructor: Marlos C. Machado
TAs: Anna Hakhverdyan, David Szepesvari, Bryan Chan, and Gábor Mihucz
Office: ATH 3-08
E-mail: machado@ualberta.ca
Web Page: https://eclass.srv.ualberta.ca/course/view.php?id=90138

Office hours: Marlos: Thursday 15:00 - 16:45 in ATH 3-08 (Athabasca Hall)
Anna: Monday 12:00 - 14:00 in TBD
David: Tuesday 13:00 - 15:00 in TBD
Bryan: Wednesday 14:00 - 16:00 in TBD
Gabor: TBD
Slack and eClass: asynchronously

TA email address: cmput655@ualberta.ca
Do not personally email the TAs. They will only respond via cmput655@ualberta.ca.

Lecture room & time: ETLC E2-001, Friday 14:00 - 16:50
Attendance isn't mandatory although strongly encouraged.

Slack invitation link:
https://join.slack.com/t/cmput655fall2023/shared_invite/zt-225uyp34m-CpFH7OECLTEeaXe8kmK0gw

COURSE CONTENT

Course Description: This course provides an overview of 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 processes, reinforcement learning, value-based methods, policy-gradient methods, planning, function approximation (online supervised learning), and contemporary research topics in the field.

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

Key resources

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

Anna

Bryan

David

Gabor

- TA email address: cmput655@ualberta.ca
- My email address: machado@ualberta.ca
- Slack invitation link: https://join.slack.com/t/cmput655fall2023/shared_invite/zt-225uyp34m-CpFH7QECLTEeaXe8kmK0gw

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 15:00 - 16:45 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: Wednesday 9:15-11:15 in CAB 3-13

Syllabus [eClass, Slack, website, Google Drive]

Pre-requisites

- 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.

Part of this class will be sort of a flipped classroom!

- I’m not going to assume you know the basics of reinforcement learning. But I’ll teach this first part as a flipped classroom (similar to CMPUT 365).

Part of this class will be sort of a flipped classroom!

- I'm not going to assume you know the basics of reinforcement learning. But I'll teach this first part as a flipped classroom (similar to CMPUT 365).
- 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

Part of this class will be sort of a flipped classroom!

- I'm not going to assume you know the basics of reinforcement learning. But I'll teach this first part as a flipped classroom (similar to CMPUT 365).
- 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

Part of this class will be sort of a flipped classroom!

- I'm not going to assume you know the basics of reinforcement learning. But I'll teach this first part as a flipped classroom (similar to CMPUT 365).
- 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

Part of this class will be sort of a flipped classroom!

- I’m not going to assume you know the basics of reinforcement learning. But I’ll teach this first part as a flipped classroom (similar to CMPUT 365).
- 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

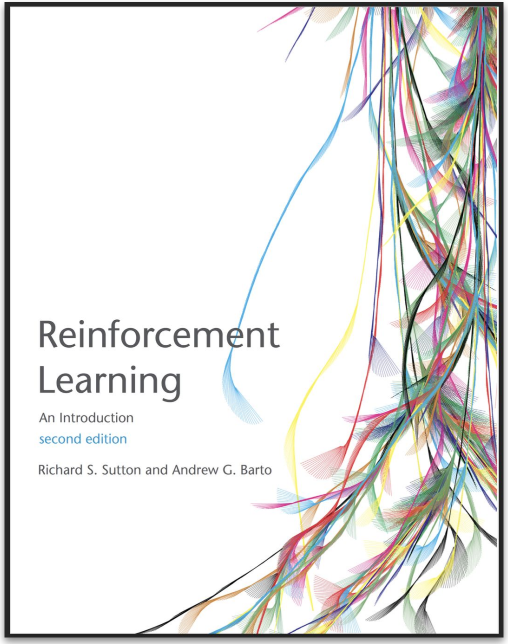
Required textbook

Reinforcement Learning: An Introduction

Richard S. Sutton & Andrew G. Barto

MIT Press. 2nd Edition.

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



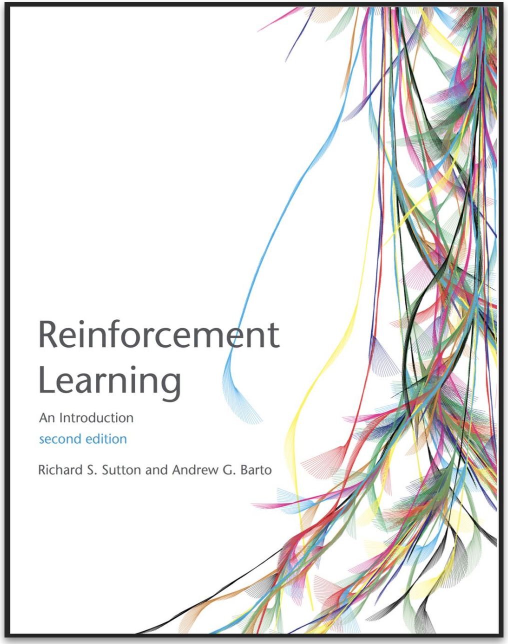
Required textbook

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!
- The book is really good!

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	9 x 3% = 21%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Project proposal	15 %	October 20, 2023 23:59:59
Midterm exam	20%	November 10, 2023
Final project	35%	December 15, 2023 *23:59:59*

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	**Weight**	**Date**
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	9 x 3% = 21%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Project proposal	15 %	October 20, 2023 23:59:59
Midterm exam	20%	November 10, 2023
Final project	35%	December 15, 2023 *23:59:59*

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	9 x 3% = 21%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Midterm exam	20%	November 10, 2023
Final project	35%	December 15, 2023 *23:59:59*

Coursera, almost every* week (<u>starting next week</u>): 30%

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	9 x 3% = 21%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Midterm exam	20%	November 10, 2023

Coursera, almost every* week (<u>starting next week</u>): 30%

Late submissions will not be accepted. There are 11 quizzes and 11 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.

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	9 x 3% = 21%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end,

One midterm, worth 20%. Closed book.

Assessment	Weight	Date
Midterm exam	20%	November 10, 2023
Final project	35%	December 15, 2023 *23:59:59*

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Project proposal	15 %	October 20, 2023 23:59:59
Midterm exam	20%	November 10, 2023
Final project	35%	December 15, 2023 *23:59:59*

The course project, proposal and final mansucript, sums to 50%.

You should be working on it the whole term. Late submissions will not be accepted for the final project (just like a conference deadline).

Syllabus [eClass, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	**Weight**	**Date**
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	9 x 3% = 21%	Day of the class on the topic(s) of the week at 11:59:59 (see Course schedule, at the end, for details)
Project proposal	15 %	October 20, 2023 23:59:59
Midterm exam	20%	November 10, 2023
Final project	35%	December 15, 2023 *23:59:59*

Syllabus [eClass, Slack, website, Google Drive]

Course project

- Each project should be done by at least three people (**no exceptions**).
    - The more people in a group, the higher the bar will be (but that's a good thing).

Course project

- Each project should be done by at least three people (**no exceptions**).
  - The more people in a group, the higher the bar will be (but that's a good thing).
- The project is not supposed to be a paper you write by yourselves.
  - I want to make it more grounded, less subjective, more useful.

Course project

- Each project should be done by at least three people (**no exceptions**).
    - The more people in a group, the higher the bar will be (but that's a good thing).
- The project is not supposed to be a paper you write by yourselves.
    - I want to make it more grounded, less subjective, more useful.
- The project is not supposed to be a regular paper you write by yourselves.
    - It will be about you coming up with a clear hypothesis or question and answering it (empirically).
    - The goal is for you to learn how to motivate a question, practice on how to ask it clearly, and for you to think carefully about empirical design.
    - *I couldn't care less if you just build something.*
    - You need to tread carefully when using the "My number is bigger than yours" argument.

Course project

- Each project should be done by at least three people (**no exceptions**).
  - The more people in a group, the higher the bar will be (but that's a good thing).
- The project is not … urselves.
  - I want to make it n…
- The project is no… by yourselves.
  - It will be about you… nswering it (empirically).
  - The goal is for you… to ask it clearly, and for you to think carefully a…
  - *I couldn't care less…*
  - You need to tread carefully when using the "My number is bigger than yours" argument.

Projects, meaning the hypothesis (or the question), need to be distinct enough! I'll be the judge.

When you have a group and a project, let me know. "Register" your project.

Syllabus [eClass, Slack, website, Google Drive]

Course project

- The project proposal really matters (~⅓ of the marks in the project). *I plan to give you feedback!*

Course project

- The project proposal really matters (~⅓ of the marks in the project). *I plan to give you feedback!*
- I *might* come up with a list of questions one can use as a project. They can come from all sorts of *undisclosed* sources at first.
    - I would like to avoid those with a supervisor benefiting from them.

Course project

- The project proposal really matters (~⅓ of the marks in the project). *I plan to give you feedback!*
- I *might* come up with a list of questions one can use as a project. They can come from all sorts of *undisclosed* sources at first.
    - I would like to avoid those with a supervisor benefiting from them.
- You should start thinking about the project now! Find a group soon. *Do not leave it for the last minute*.

Course project

- The project proposal really matters (~⅓ of the marks in the project). *I plan to give you feedback!*
- I *might* come up with a list of questions one can use as a project. They can come from all sorts of *undisclosed* sources at first.
    - I would like to avoid those with a supervisor benefiting from them.
- You should start thinking about the project now! Find a group soon. *Do not leave it for the last minute*.
- Do not overlook computation. Lack of computational resources is not a good justification for poor empirical practices.

Course project – Example

The role of grounded experience in offline learning. *Does grounded experience allows us to leverage other people's experience better?*

Course project – Example

The role of grounded experience in offline learning. *Does grounded experience allows us to leverage other people's experience better?*

Given a dataset and a fixed budget on the number of interactions the agent can have with the environment; is it better for the agent to learn from the dataset first and then fine tune their policy in the environment, or is it better to first interact with the environment and then fine tune a policy with data the agent hasn't seen yet? Or something in the middle?

Course project – Example

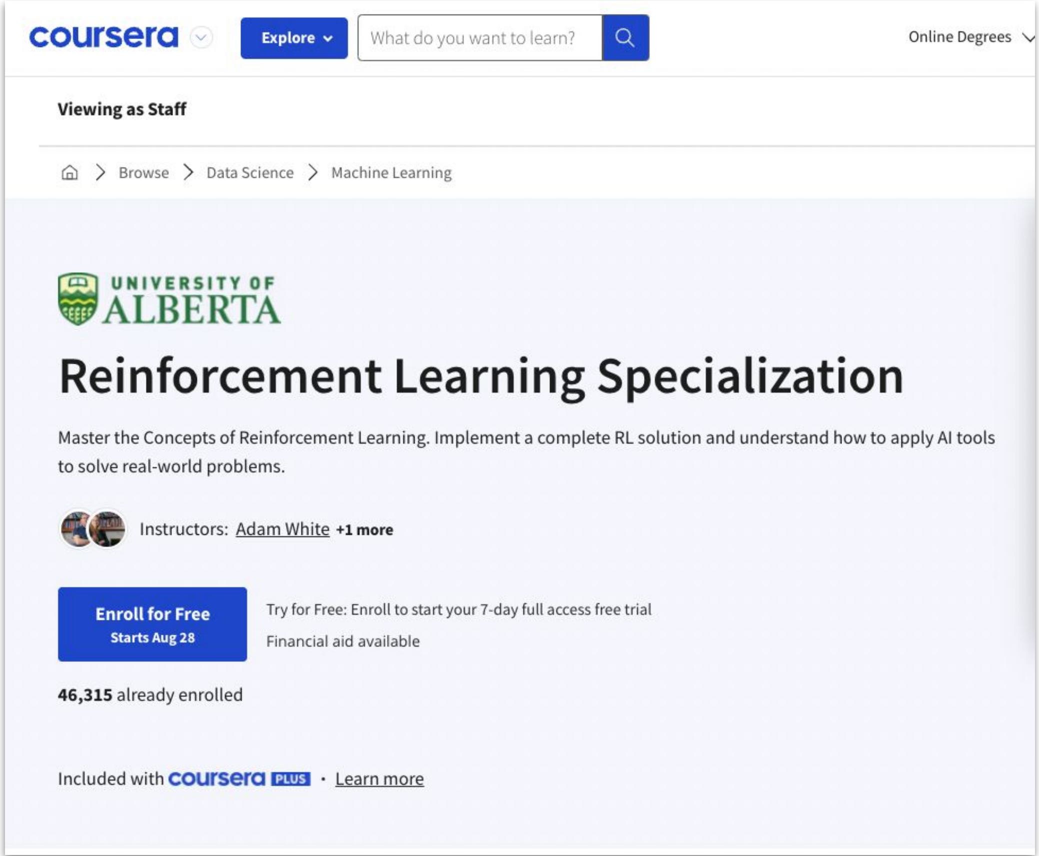
The role of grounded experience in offline learning. *Does grounded experience allows us to leverage other people's experience better?*

Given a dataset and a fixed budget on the number of interactions the agent can have with the environment; is it better for the agent to learn from the dataset first and then fine tune their policy in the environment, or is it better to first interact with the environment and then fine tune a policy with data the agent hasn't seen yet? Or something in the middle?

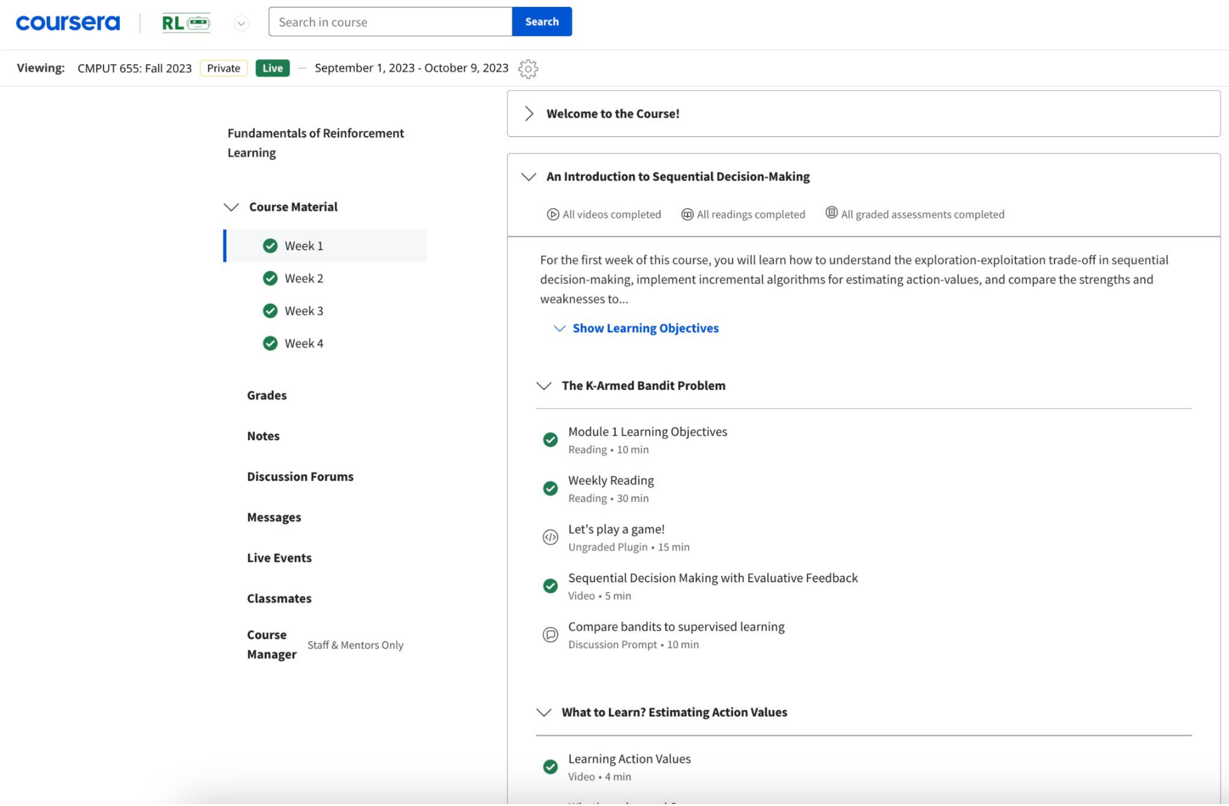
Exploration? Representation learning? Value-based vs. policy gradient methods? Quality of demonstration? And many many more…

Coursera

- Coursera will be <u>essential</u> to CMPUT 655
- You should have been added to a private session of the RL courses (we used your <u>university's email</u>)
  - 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
Viewing: CMPUT 655: Fall 2023 Private Live — September 1, 2023 - October 9, 2023
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

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 and, more broadly, machine learning. Thus, in order to achieve such learning outcomes, students can submit each practice quiz and graded assignment multiple times, which allows for many learning opportunities. Therefore, the use of advanced AI-tools based on large-language models such as ChatGPT or Bard 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."

Students are also allowed to use advanced AI-tools such as ChatGPT or Bard to proofread their manuscripts, but only after having written a first complete draft of the text to be proofread. Organizing ideas in writing is an essential part of the research process, and shortcutting this process will likely hinder a student's development. One is prohibited from using advanced AI-tools for help with related work. All interactions with an advanced AI-tool are to be submitted as Appendix in the project proposal and final project manuscript. The Appendix does not count toward the pre-specified page limit.

Schedule (tentative)

- This course is supposed to be an overview of “everything” reinforcement learning.
  - Other courses in the department can give you more “depth” (e.g., theory, policy gradient algorithms, etc).
- Each week we will cover 1 or 2 whole Chapters of the textbook.
- The initial (and ambitious) plan is to cover Chapters 1–13, 16, and 17.
- Topics covered in the MOOC will not be my main focus in class.
  - One, two, or three practice quizzes and graded assignments will be due every week until the midterm.
  - The deadline for submitting quizzes and assignments is 11:59:59.
- I’ll talk about relevant papers as we go along as well.

Schedule

Course Schedule & Assigned Readings

Week	Date	Topic	Deadlines (all due at 11:59:59)	Readings
1	Fri, Sep 8	Course overview Discussion about what is reinforcement learning Background review: Probability, statistics, linear algebra, and calculus		Chapter 1: Introduction
2	Fri, Sep 15	Fundamentals of RL: An introduction to sequential decision-making Optimality of UCB	Practice quiz (Sequential decision-making) Program. assignment (Bandits & exploration / exploitation)	Chapter 2: Multi-armed Bandits
3	Fri, Sep 22	Fundamentals of RL: Markov decision processes (MDPs) Fundamentals of RL: Value functions & Bellman equations Fundamentals of RL: Dynamic programming	Practice quiz (MDPs) Practice quiz (Value functions & Bellman equations) Graded quiz (Value functions & Bellman equations) Practice quiz (Dynamic programming) Program. Assignment (Optimal policies with dynamic programming)	Chapter 3: Finite Markov Decision Processes Chapter 4: Dynamic Programming Ross's Chapter 2
4	Fri, Sep 29	Sample-based learning methods: MC methods for Prediction & Control	Graded quiz (Off-policy Monte Carlo)	Chapter 5: Monte-Carlo Methods
5	Fri, Oct 6	Sample-based learning methods: TD learning for prediction	Practice quiz (Advantages of TD) Program. Assignment (Policy	Chapter 6: Temporal-Difference Learning Chapter 7: n-step Bootstrapping
		Sample-based learning methods: TD learning for control	evaluation with TD learning) Practice quiz (Expected Sarsa) Program. assignment (Q-learning & Expected Sarsa)	
	Mon, Oct 9	Thanksgiving		
6	Fri, Oct 13	Sample-based learning methods: Planning, learning, & acting	Practice quiz (Dealing with inaccurate models) Program. assignment (Dyna-Q & Dyna-Q+)	Chapter 8: Planning and Learning with Tabular Methods
7	Fri, Oct 20	Prediction and Control with FA: On-policy prediction with approx. Prediction and Control with FA: Constructing features for prediction Prediction and Control with FA: Control with approximation	Practice quiz (On-policy prediction with approximation) Program. assignment (Semi-gradient TD(0) with state aggregation) Practice quiz (Constructing features for prediction) Program. assignment (Semi-gradient TD with a neural network) Practice quiz (Control with approximation) Program. assignment (Function approximation & control)	Chapter 9: On-policy Prediction with Approximation Chapter 10: On-policy Control with Approximation
Fri, Oct 20 at 23:59		**Project proposal**		
8	Fri, Oct 27			Chapter 12: Off-policy Methods with Approximation Chapter 12: Eligibility Traces
9	Fri, Nov 3	Prediction and Control with FA: Policy Gradient	Practice quiz (Policy gradient methods) Program. assignment (Average reward softmax Actor-Critic with tile-coding)	Chapter 13: Policy Gradient Methods
	Fri, Nov 10	**Midterm**		
	Mon, Nov 13	Remembrance day holiday in lieu		
	Nov 14 - Nov 17	Reading week		
10	Fri, Nov 23	Deep Reinforcement Learning		
11	Fri, Dec 1	Major Successes of Reinforcement Learning		Chapter 16: Applications and Case Studies
12	Fri, Dec 8	Frontiers		Chapter 17: Frontiers
Fri, Dec 15 at 23:59		**Final course project**		

Syllabus [eClass, Slack, website, Google Drive]

Schedule

Course Schedule & Assigned Readings

Week	Date	Topic	Deadlines (all due at 11:59:59)	Readings
1	Fri, Sep 8	Course overview Discussion about what is reinforcement learning Background review: Probability, statistics, linear algebra, and calculus		Chapter 1: Introduction
2	Fri, Sep 15	Fundamentals of RL: An introduction to sequential decision-making Optimality of UCB	Practice quiz (Sequential decision-making) Program. assignment (Bandits & exploration / exploitation)	Chapter 2: Multi-armed Bandits
3	Fri, Sep 22	Fundamentals of RL: Markov decision processes (MDPs) Fundamentals of RL: Value functions & Bellman equations Fundamentals of RL: Dynamic programming	Practice quiz (MDPs) Practice quiz (Value functions & Bellman equations) Graded quiz (Value functions & Bellman equations) Practice quiz (Dynamic programming) Program. Assignment (Optimal policies with dynamic programming)	Chapter 3: Finite Markov Decision Processes Chapter 4: Dynamic Programming Ross's Chapter 2
4	Fri, Sep 29	Sample-based learning methods: MC methods for Prediction & Control	Graded quiz (Off-policy Monte Carlo)	Chapter 5: Monte-Carlo Methods

Syllabus [eClass, Slack, website, Google Drive]

What is reinforcement learning?

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** “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)	**The “laws of thought” approach**
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.

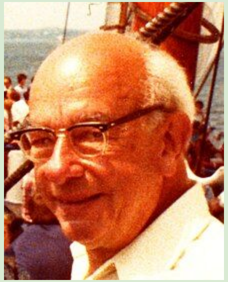
(Russell & Norvig; 2010)

Artificial intelligence

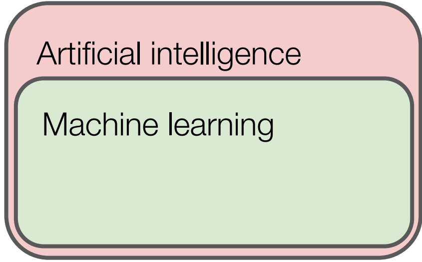
"AI is the ability of machines to perform tasks that are typically associated with human intelligence, such as learning and problem-solving." –Wikipedia

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

– W. V. Quine

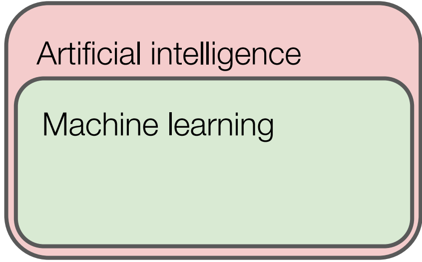


https://upload.wikimedia.org/wikipedia/commons/c/cd/Willard_Van_Orman_Quine_on_Bluenose_II_in_Halifax_NS_harbor_1980.



Machine learning

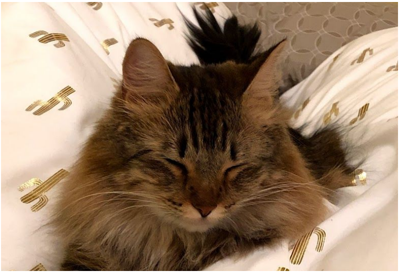
Machine learning is a subfield of AI in which the system’s desired behavior is not explicitly programmed, instead it is *learned* from data



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

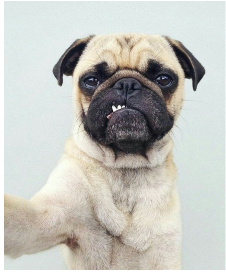
- "*Supervised learning* is learning from a training set of labeled examples provided by a knowledgeable external supervisor" (Sutton & Barto; 2018)



Cat

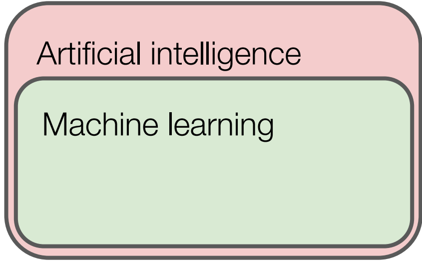
Cat

Not cat



Cat or not cat?

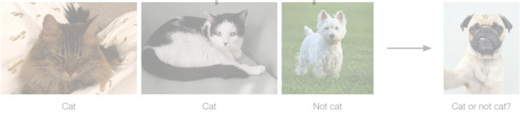
https://i0.statig.com.br/bancodeimagens/0r/yu/ls/0ryulsnl6uvkyvy6vv98bkms9.j
https://www.dogue.com.au/cdn/shop/articles/8447610bb0578436f1faa531b6564ab5.jpg?v=1687241



Machine learning

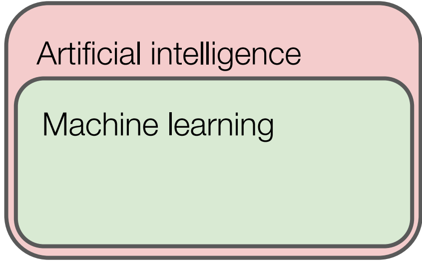
Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data

- "*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)

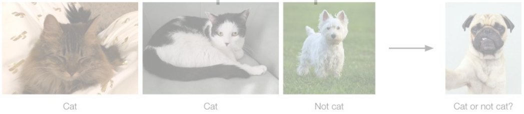
https://www.researchgate.net/publication/309515576/figure/fig1/AS:960348551122945@1605976345743/uits-can-be-clustered-in-different-ways-images-are-from-Internet-a-Fruits-b.gif



Machine learning

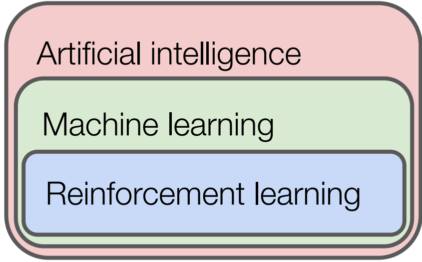
Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data

- "*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)

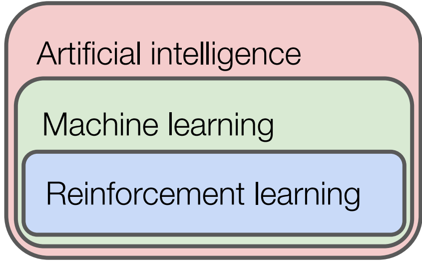
… and *reinforcement learning*!



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

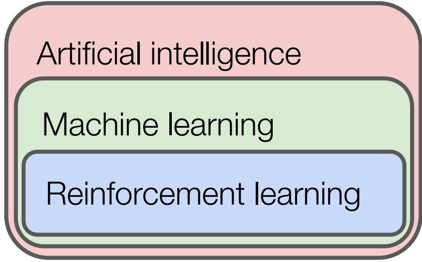


Reinforcement learning

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... 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:

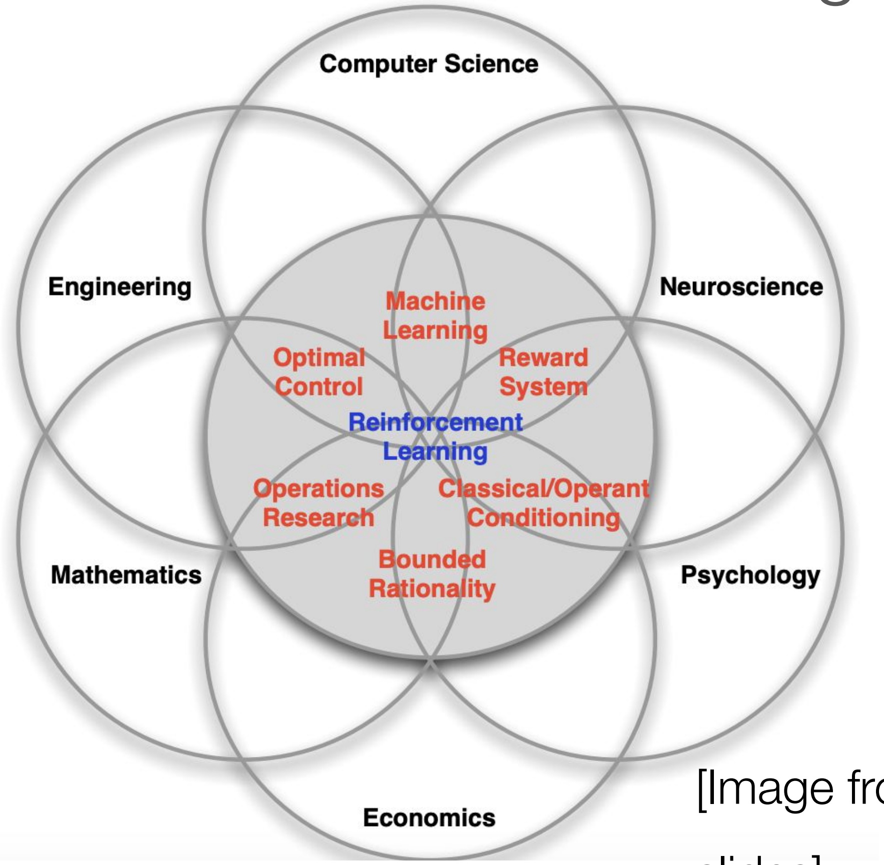
- Trial-and-er...
- The trade-... oration and exploitation
- The delayed c... gnment / delayed reward problem

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
  - ...

Many Faces of Reinforcement Learning



[Image from David Silver's slides]

https://www.davidsilver.uk/wp-content/uploads/2020/03/intro_RL.pd

On intelligence, AGI, etc etc…

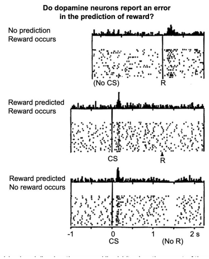
- People in the field have different, non-competing, perspectives and motivations
  - Some study RL to learn about / develop tools for solving sequential decision-making problems
  - Some look at RL as a computational model of intelligence

On intelligence, AGI, etc etc…

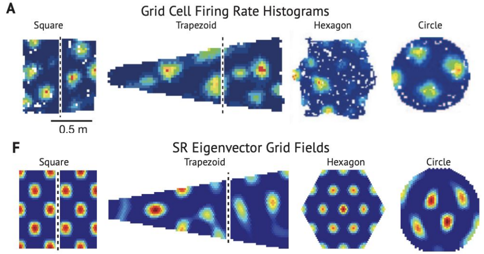
- People in the field have different, non-competing, perspectives and motivations
  - Some study RL to learn about / develop tools for solving sequential decision-making problems
  - Some look at RL as a computational model of intelligence
- I'll steer away from philosophical discussions and I'll focus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

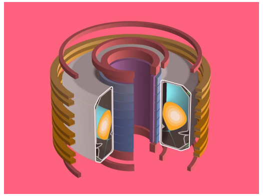


(Stachenfeld, Botvinick, & Gershman; 2017)

(Silver et al.; 2016)


(Bellemare et al.; 2020)

(Degrave et al.; 2022)



5-minute break

Probability and statistics

Definitions

- **Probability** is about predicting the likelihood of future events.
- **Statistics** is about estimating a model (rule) from past events.

We'll need to understand probability to do statistics.

Probability – The basics

A probability is a function that associates a number between 0 and 1 to an event, with this number being a measure of the likelihood of that set of outcomes.

Example

Dungeons & Dragons!

1
2
3
...
18
19
20

Pr(rolling 20) = 1/20 = 5%

Pr(rolling 19 **or** 20) =
Pr(rolling 19) + **Pr**(rolling 20) =
1/20 + 1/20 = 10%

Pr(rolling 20 **and** 20) =
Pr(rolling 20) x **Pr**(rolling 20) =
1/20 x 1/20 = 1/400 = 0.25%

https://i.etsystatic.com/8800676/r/il/6e938b/3226455710/il_794xN.3226455710_64zr.jp
https://cdn.britannica.com/71/241671-050-B5227EBD/Game-of-Dungeons-and-Dragons.jp

Probability – Somewhat more formally

A probability is a function that associates a number between 0 and 1 to an event, with this number being a measure of the likelihood of that set of outcomes.

Probability – Somewhat more formally

A probability is a function that associates a number between 0 and 1 to an ***event****, with this number being a measure of the likelihood of that* ***set of outcomes****.*

- A **set** is collection of disjoint elements.
- A **sample space** is the set of all possible outcomes of an experiment.
- An **event** is any subset of the sample space.

Example

Sample space. {1, 2, …, 20}

Event. Rolling higher than 16: {17, 18, 19, 20}

Probability – Somewhat more formally

*A probability is a **function** that **associates** a number between 0 and 1 to an event, with this number being a measure of the likelihood of that set of outcomes.*

- A **set** is collection of disjoint elements.
- A **sample space** is the set of all possible outcomes of an experiment.
- An **event** is any subset of the sample space.
- A **function**, f: $A \rightarrow B$, is a map, a rule, that maps every element of the set A to a unique element in the set B. We call A the *domain*, and B the *codomain*, or the *range*, of the function. Given x $\in A$, the element it is associated with in the set B is called its *image* under f.

Probability – Somewhat more formally

A probability is a function that associates a number between 0 and 1 to an event, with this number being a ***measure of the likelihood*** *of that set of outcomes.*

- A probability distribution is defines how the probability is distributed among the outcomes.

Example

For an unbiased dice, each number is equally likely (i.e., uniform probability distribution). Thus, for each outcome e ∈ S, **Pr**(e) = 1/|S|.

Probability – Somewhat more formally

A probability is a function that associates a number between 0 and 1 to an event, with this number being a ***measure of the likelihood*** *of that set of outcomes.*

- A probability distribution is defines how the probability is distributed among the outcomes.
- A way of calculating the probability of a specific event is a matter of identifying the sample space (set of all possible outcomes) and the probability distribution.

Example 1

For an unbiased dice, the probability of rolling a 20 is **Pr**(rolling 20) = 1/20.

Example 2

For an unbiased dice, the probability of rolling higher than 18 is **Pr**(rolling 19 or 20) = 1/20 + 1/20 = 1/10.

Probability – Properties

A probability is a function that associates a number between 0 and 1 to an event, with this number being a measure of the likelihood of that set of outcomes.

- Nonnegativity: **Pr**(A) ≥ 0.
- Normalization: $\sum_{e \in S}$ **Pr**(e) = 1.
- Additivity: **Pr**($A \cup B$) = **Pr**(A) + **Pr**(B); $A \cap B = \{\ \}$.

Example

For an unbiased dice, the probability of rolling higher than 18 is **Pr**(rolling 19 or 20) = 1/20 + 1/20 = 1/10.

Probability – Considering all possible events

How many distinct events are possible in a dice rolling experiment?

Probability – Considering all possible events

How many distinct events are possible in a dice rolling experiment?

- The number of all possible subsets of the sample space.
- The power set of the sample space S, denoted 2^S.

Example

2^S = {S, {}, {1}, {2}, {3}, …, {20}, {1, 2}, {1, 3}, … {18, 19, 20}, … {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13}, …, {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19}, …}

Number of elements in 2^S:
2^{20} = 1,048,576

Probability – Considering all possible events

How many distinct events are possible in a dice rolling experiment?

The number of all possible subsets of the sample space.

The power set of the sample space S, denoted 2^S.

Pr(S) = 1.

Pr({}) = 0.

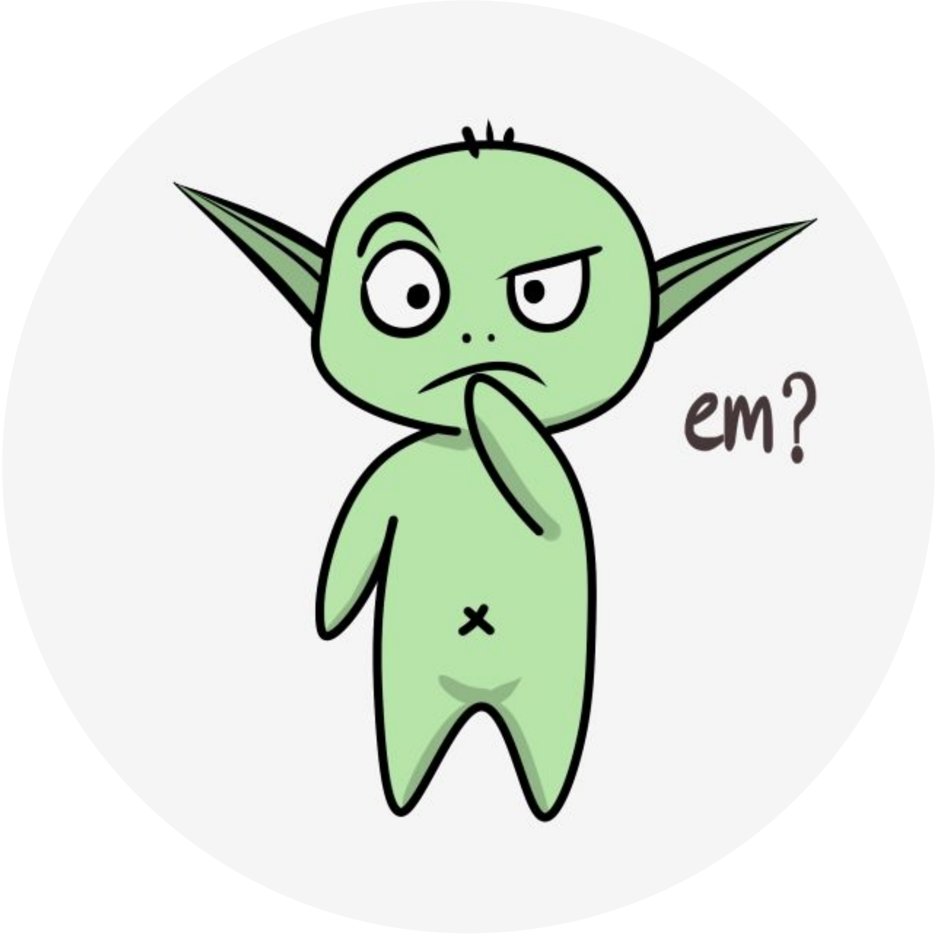
Formally, **Pr**: $2^S \rightarrow [0, 1]$.

Example 1

For an unbiased dice, the probability of rolling a 20 is **Pr**(rolling 20) = 1/20.

Example 2

For an unbiased dice, the probability of rolling higher than 18 is **Pr**(rolling 19 or 20) = 1/20 + 1/20 = 1/10.



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

Random variables and expectations

Random variables

Random variables are ways to map outcomes of random processes to real numbers.

They are not a traditional variable, nor random 😂

Capital letter!

Example 1

$$X = \left\{ \begin{array}{l} 1 \text{ if roll } 1 \\ 2 \text{ if roll } 2 \\ \ldots \\ 19 \text{ if roll } 19 \\ 20 \text{ if roll } 20 \end{array} \right\}$$

Example 2

$$Y = \left\{ \begin{array}{l} 1 \text{ if heads} \\ 0 \text{ if tails} \end{array} \right\}$$

Example 3

$$Z = \{\text{sum of 2 dice}\}$$

We can write **Pr**(X = 20) to represent **Pr**(rolling 20).

We can write **Pr**(X ≥ 19) to represent **Pr**(rolling 19 or 20).

Examples

When rolling a d20 dice, let X be the random variable denoting the outcome of the roll.

$\mathbf{Pr}(1 \leq X \leq 20) = 1$

$\mathbf{Pr}(X = 15) = 1/20$

$\mathbf{Pr}(X = 0) = 0$

$\mathbf{Pr}(2X = 1) = 0$

$\mathbf{Pr}(X < 19)$?

$\mathbf{Pr}([X = 19] \cup [X = 20] \cup [1 \leq X \leq 18]) = 1$
$\mathbf{Pr}([X = 19] \cup [X = 20] \cup [X < 19]) = 1$
$\mathbf{Pr}(X = 19) + \mathbf{Pr}(X = 20) + \mathbf{Pr}(X < 19) = 1$
$\mathbf{Pr}(X < 19) = 1 - \mathbf{Pr}(X = 19) - \mathbf{Pr}(X = 20)$
$\mathbf{Pr}(X < 19) = 1 - 1/20 - 1/20$
$\mathbf{Pr}(X < 19) = 18/20$
$\mathbf{Pr}(X < 19) = 9/10$

Conditional probabilities

Chain rule:
$\mathbf{Pr}(A \cap B) = \mathbf{Pr}(A, B) = \mathbf{Pr}(A \mid B)\, \mathbf{Pr}(B)$

The probability of an event A given another event B is defined as:

$$\mathbf{Pr}(A \mid B) \doteq \frac{\mathbf{Pr}(A \cap B)}{\mathbf{Pr}(B)} .$$

In a classroom with 100 students, out of those 100, 20 students play tabletop RPG, and 30 students have read *The Lord of the Rings* books. There are 15 students who play tabletop RPG who have read LOTR. What is the probability that a student has read LOTR given that the student plays tabletop RPG?

Let X be the random variable denoting the probability that a student plays tabletop RPG, and let Y be the random variable denoting the probability that a student has read LOTR.

$\mathbf{Pr}(X) = 0.2$ $\mathbf{Pr}(Y) = 0.3$ $\mathbf{Pr}(X \cap Y) = 0.15$

$\mathbf{Pr}(Y \mid X) = 0.15/0.2 = 0.75$

Conditional probabilities

The probability of an event A given another event B is defined as:

$$\mathbf{Pr}(A \mid B) \doteq \frac{\mathbf{Pr}(A \cap B)}{\mathbf{Pr}(B)} \; .$$

When playing D&D, Tristan needs to roll 17 or higher on a d20 to successfully hit the troll. Tristan gets a critical hit when they roll a 20. Knowing that Tristan has successfully hit the target, what's the likelihood that Tristan got a critical hit?

Let X be the random variable denoting the number Tristan rolled on a d20, and Y a binary random variable denoting whether Tristan rolled a 20 (Y=1) or not (Y=0).

$\mathbf{Pr}(X \geq 17) = 1/5$ $\qquad$ $\mathbf{Pr}(Y = 1 \cap X \geq 17) = 1/20$

$$\frac{\mathbf{Pr}(Y = 1 \cap X \geq 17)}{\mathbf{Pr}(X \geq 17)} = \frac{1/20}{1/5} = \frac{5}{20} = 25\%$$

Independence

Two events are independent when the likelihood of an event does not change after knowing the other event. A is independent of B if and only if

$$\mathbf{Pr}(A \mid B) = \mathbf{Pr}(A).$$

$\mathbf{Pr}(A \mid B) = \mathbf{Pr}(A \cap B) \,/\, \mathbf{Pr}(B)$
$\mathbf{Pr}(A \cap B) = \mathbf{Pr}(A \mid B)\, \mathbf{Pr}(B)$
$\mathbf{Pr}(A \cap B) = \mathbf{Pr}(A)\, \mathbf{Pr}(B)$
$\mathbf{Pr}(B \mid A) = \mathbf{Pr}(B \cap A) \,/\, \mathbf{Pr}(A)$
$= \mathbf{Pr}(B)\, \mathbf{Pr}(A) \,/\, \mathbf{Pr}(A)$
$= \mathbf{Pr}(B)$

Example

Tristan now rolls two d20 dice. Given that they rolled a 1 on the first dice, what's the likelihood of them running a 20 on the second dice?

Let X be the random variable denoting the roll on the first dice, and Y be the equivalent for the second dice.

$\mathbf{Pr}(X = 1) = 1/20$ $\quad \mathbf{Pr}(Y = 20) = 1/20$ $\quad \mathbf{Pr}(X = 1 \cap Y = 20) = 1/400$

$\mathbf{Pr}(Y = 20 \mid X = 1) = (1/400)/(1/20) = 1/20$

Conditional probabilities with more than 2 variables

The probability of an event A given another event B is defined as:

$$\mathbf{Pr}(A \mid B) \doteq \frac{\mathbf{Pr}(A \cap B)}{\mathbf{Pr}(B)}.$$

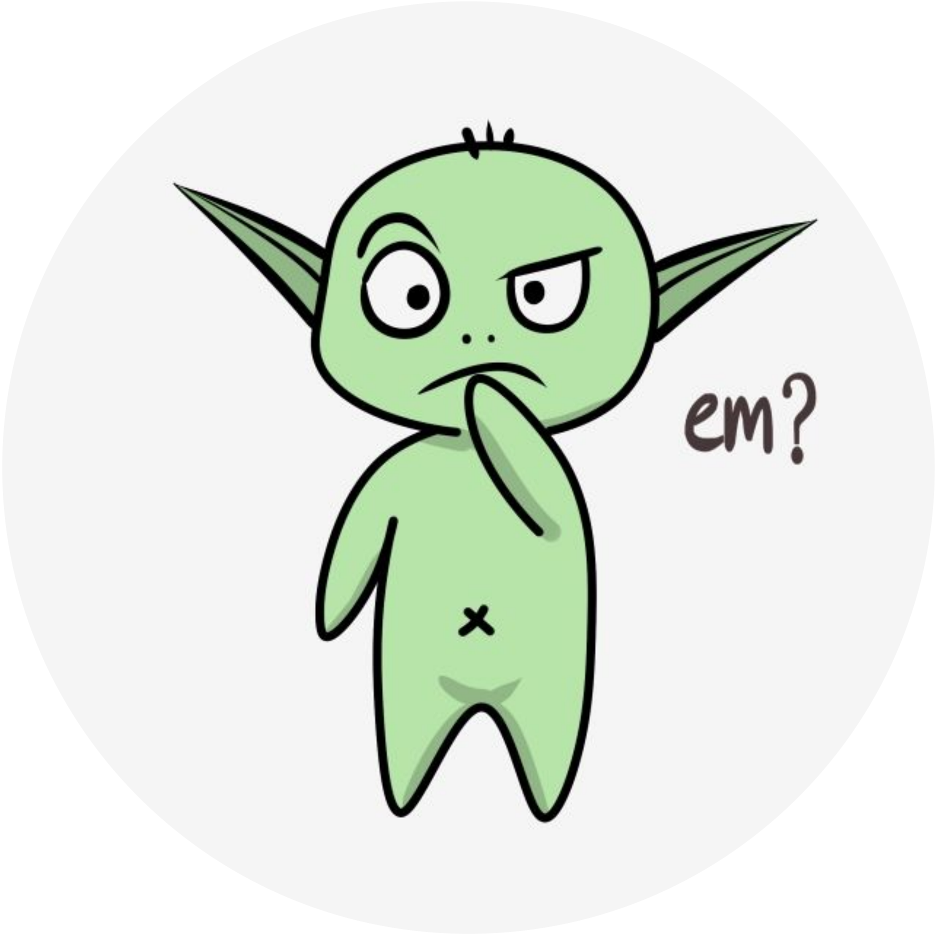
Chain rule:
$\mathbf{Pr}(A \cap B) = \mathbf{Pr}(A, B) = \mathbf{Pr}(A \mid B)\, \mathbf{Pr}(B)$

What's $\mathbf{Pr}(A, B \mid C)$?

Let $D = A \cap B$. Then, $\mathbf{Pr}(D \mid C) = \mathbf{Pr}(D, C) \,/\, \mathbf{Pr}(C)$. Thus $\mathbf{Pr}(A, B \mid C) = \mathbf{Pr}(A, B, C) \,/\, \mathbf{Pr}(C)$.

Now, let $E = B \cap C$, and recall, by the chain rule, that $\mathbf{Pr}(A, E) = \mathbf{Pr}(A \mid E)\, \mathbf{Pr}(E)$.
We then have $\mathbf{Pr}(A, B, C) = \mathbf{Pr}(A \mid B, C)\, \mathbf{Pr}(B, C) = \mathbf{Pr}(A \mid B, C)\, \mathbf{Pr}(B \mid C)\, \mathbf{Pr}\,(C)$.

Putting these two together: $\mathbf{Pr}(A, B \mid C) = \mathbf{Pr}(A \mid B, C)\, \mathbf{Pr}(B \mid C)\, \mathbf{Pr}\,(C) \,/\, \mathbf{Pr}(C)$.
Assuming $\mathbf{Pr}(C) \neq 0$, $\mathbf{Pr}(A, B \mid C) = \mathbf{Pr}(A \mid B, C)\, \mathbf{Pr}(B \mid C)$.



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

Example – Probabilities with two random variables

Let X be the random variable denoting the outcome of the roll of a d20, and let Y be the random variable denoting the outcome of the roll of a d6. What's $\mathbf{Pr}(X + Y \geq 25)$?

$$
\begin{aligned}
\mathbf{Pr}(X + Y \geq 25) &= \mathbf{Pr}([X = 20] \cap [Y = 5]) + \mathbf{Pr}([X = 20] \cap [Y = 6]) + \mathbf{Pr}([X = 19] \cap [Y = 6]) \\
&= \mathbf{Pr}(X = 20)\ \mathbf{Pr}(Y = 5) + \mathbf{Pr}(X = 20)\ \mathbf{Pr}(Y = 6) + \mathbf{Pr}(X = 19)\ \mathbf{Pr}(Y = 6) \\
&= 1/20 \times 1/6 + 1/20 \times 1/6 + 1/20 \times 1/6 \\
&= 1/120 + 1/120 + 1/120 \\
&= 3/120 \\
&= 1/40
\end{aligned}
$$

Marginalization

- The marginal probability is the probability of a single event occurring, independent of other events.
- If we have the joint distribution **Pr**(x, y), we can find the marginals **Pr**(x) and **Pr**(y).

$$\mathbf{Pr}(X = x) = \sum_{y \in \mathcal{Y}} \mathbf{Pr}(X = x, Y = y) \qquad \mathbf{Pr}(Y = y) = \sum_{x \in \mathcal{X}} \mathbf{Pr}(X = x, Y = y)$$

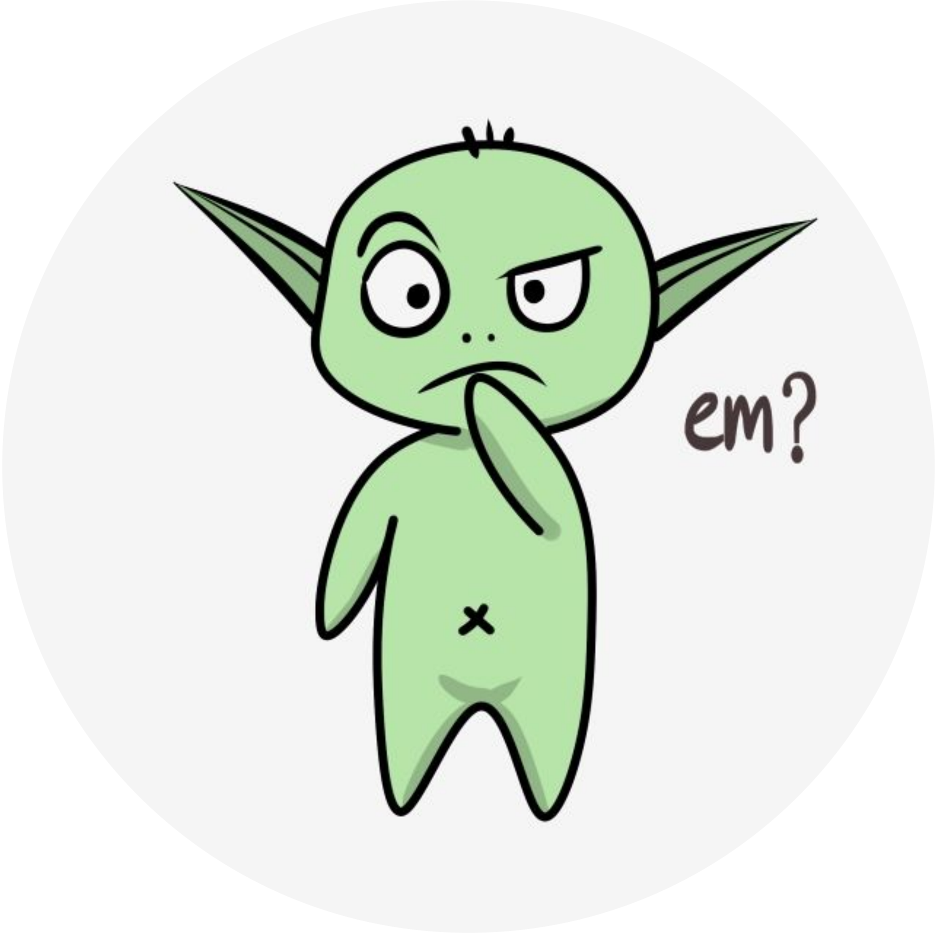
Example

Animal's favourite activity (columns); Type of pet (rows)

	Sleep	Play
Cat	0.3	0.2
Dog	0.1	0.4

Pr(Sleep) = **Pr**(Sleep, Cat) + **Pr**(Sleep, Dog) = 0.3 + 0.1 = 0.4
Pr(Play) = **Pr**(Play, Cat) + **Pr**(Play, Dog) = 0.2 + 0.4 = 0.6

Pr(Cat) = **Pr**(Sleep, Cat) + **Pr**(Play, Cat) = 0.3 + 0.2 = 0.5
Pr(Dog) = **Pr**(Sleep, Dog) + **Pr**(Play, Dog) = 0.1 + 0.4 = 0.5



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

Expectations

The expectation of a numeric random variable is the weighted average of its possible numeric outcomes, where the weights are the prob. of the outcome occurring:

$$\mathbb{E}[Y] \doteq \sum_{y \in \mathcal{Y}} y \, \mathbf{Pr}(Y = y).$$

Example

$\mathbb{E}[Y] \quad = ¼ \times 1 + ¼ \times 2 + ¼ \times 3 + ¼ \times 4$

$= 10/4 = 2.5.$

We can also compute the expectation of a function of a random variable:

$$\mathbb{E}[f(Y)] \doteq \sum_{y \in \mathcal{Y}} f(y) \, \mathbf{Pr}(Y = y).$$

Properties of expectations

The expectation of a numeric random variable is the weighted average of its possible numeric outcomes, where the weights are the prob. of the outcome occurring:

$$\mathbb{E}[Y] \doteq \sum_{y \in \mathcal{Y}} y \, \mathbf{Pr}(Y = y).$$

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, if X and Y are independent in nature.
- $\mathbb{E}[X Y] = \mathbb{E}[X] \, \mathbb{E}[Y]$, if X and Y are independent in nature.
- $\mathbb{E}[X + c] = \mathbb{E}[X] + c$, where c is not a random variable of the model.
- $\mathbb{E}[cX] = c \, \mathbb{E}[X]$, where c is not a random variable of the model.

Bias

The bias of an estimator $\hat{w}$, of the true parameters w, is $\mathbb{E}[\hat{w} - w] = \mathbb{E}[\hat{w}] - w$. An estimator is unbiased if its bias is zero for all w.

Covariance (and variance)

The covariance of two random variables, X and Y, is defined as:

$$cov(X, Y) \doteq \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Notice cov(X, X) = var(X). Also, if X and Y are independent, then cov(X, Y) = 0.

The bias-variance trade-off

If you output only a constant *c*, you potentially have a lot of *bias*, but your output is not spread out. If you only look at samples, you might have a lot of variance, depending on the process, but no bias. Sometimes it is best to be in-between.

The **variance** captures how spread out the random variable X is from its mean.

Conditional expectations

Law of total expectation:

$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$

$\mathbb{E}[X] = \sum_{y \in \mathcal{Y}} \mathbb{E}[X \mid Y = y] \, \mathbf{Pr}(Y = y)$

A conditional expectation of a random variable is the expected value of the variable given that an event is already known to have happened.

$$\mathbb{E}[X \mid Y = y] \doteq \sum_{x \in \mathcal{X}} x \, \mathbf{Pr}(X = x \mid Y = y).$$

Example

Consider a D&D player who needs to roll 16 or higher to hit the target. When they hit the target, they cause 1d8 of damage. What's the expected damage this player will cause during such a battle?

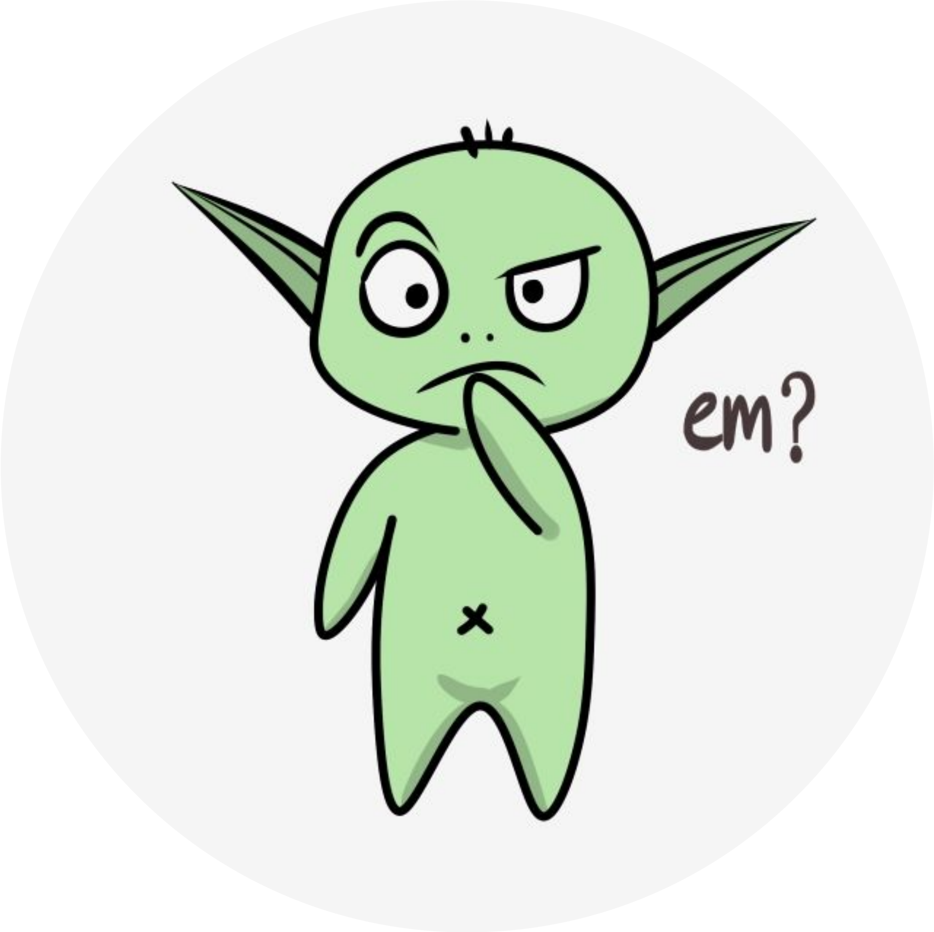
Let X be the random variable denoting the 1d8 damage roll, and Y be the r.v. denoting the d20 roll.

$\mathbb{E}[X \mid Y < 15] = 0$

$\mathbb{E}[X \mid Y > 16] = 4.5$

$\mathbb{E}[X \mid Y = 16] = 1 \, \mathbf{Pr}\,(X = 1 \mid Y = 16) + \ldots + 8 \, \mathbf{Pr}\,(X = 8 \mid Y = 16) = 36/8 = 4.5$

$\mathbb{E}[X] = 15/20 \times 0 + 5/20 \times 4.5 = ¾ \times 0 + ¼ \times 4.5 = 1.125.$



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

Linear algebra

Vectors and matrices

A vector can be thought as a list of numbers.

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

A matrix can be thought as a table of numbers.

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$$

Products

A vector is a matrix with one of its dimensions 1. Same rules apply.

A dot product between two vectors, $\mathbf{v}$ *and* $\mathbf{w} \in \mathbb{R}^d$*, is defined as:*

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^\top \mathbf{w} = \textstyle\sum_i v_i w_i.$$

A product between a matrix $\mathbf{M} \in \mathbb{R}^{n \times d}$ *and a matrix* $\mathbf{P} \in \mathbb{R}^{d \times p}$ *is defined such that:*

$$\boldsymbol{MP} = \boldsymbol{R},$$

$$\text{where } r_{ij} = m_{i1} p_{1j} + m_{i2} p_{2j} + \ldots + m_{id} p_{dj} = \textstyle\sum_{k=1}^{d} m_{ik} p_{kj}$$

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \\ p_{31} & p_{32} \end{bmatrix} = \begin{bmatrix} m_{11}p_{11} + m_{12}p_{21} + m_{13}\, p_{31} & m_{11}p_{12} + m_{12}p_{22} + m_{13}\, p_{32} \\ m_{21}p_{11} + m_{22}p_{21} + m_{23}\, p_{31} & m_{21}p_{12} + m_{22}p_{22} + m_{23}\, p_{32} \\ m_{31}p_{11} + m_{32}p_{21} + m_{33}\, p_{31} & m_{31}p_{12} + m_{32}p_{22} + m_{33}\, p_{32} \end{bmatrix}$$

Expectations in vector form

When dealing with more than one random variable, sometimes it is useful to use vector and/or matrix notation.

We can list n random variables into a vector $\mathbf{x} \in \mathbb{R}^n$, getting a vector of random variables. We call such a vector a *random vector*.

$$\mathbb{E}[\mathbf{x}] = \begin{bmatrix} \mathbb{E}[x_1] \\ \dots \\ \mathbb{E}[x_n] \end{bmatrix}$$

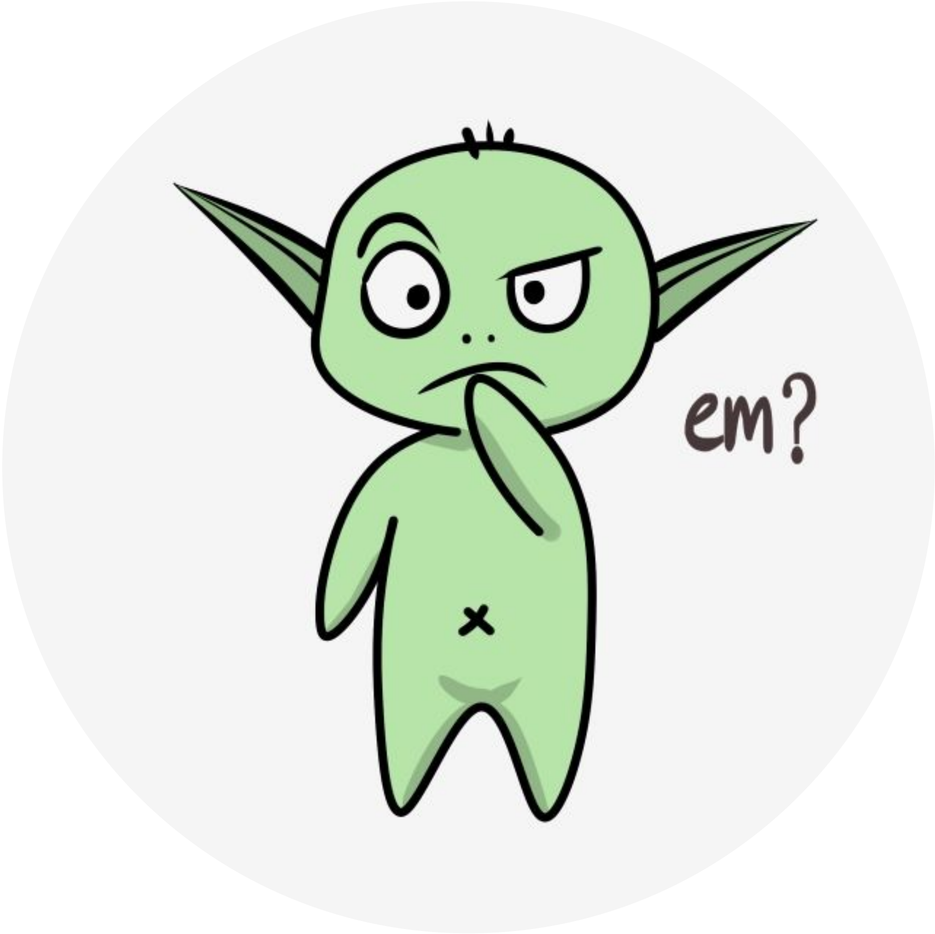
Several properties still apply, such as $\mathbb{E}[\mathbf{A} + \mathbf{B}] = \mathbb{E}[\mathbf{A}] + \mathbb{E}[\mathbf{B}]$.

Norm of a vector

The norm of a vector $\mathbf{v}$, $||\mathbf{v}||$, can be seen as a measure of the size of a vector. It has properties that behave like distances:

1. $||\mathbf{v}|| > 0$ when $\mathbf{v} \neq \mathbf{0}$ and $||\mathbf{v}|| = 0$ iff $\mathbf{v} = \mathbf{0}$ (non-negativity and definiteness).
2. $||c\mathbf{v}|| = |c|\ ||\mathbf{v}||$ for any scalar c (homogeneity).
3. $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$ (triangle inequality).

The p-norm of a vector is defined as $||\mathbf{v}||_p \doteq (\sum_i |v_i|^p)^{1/p}$, with $||\mathbf{v}||_\infty = \max_i |v_i|$.

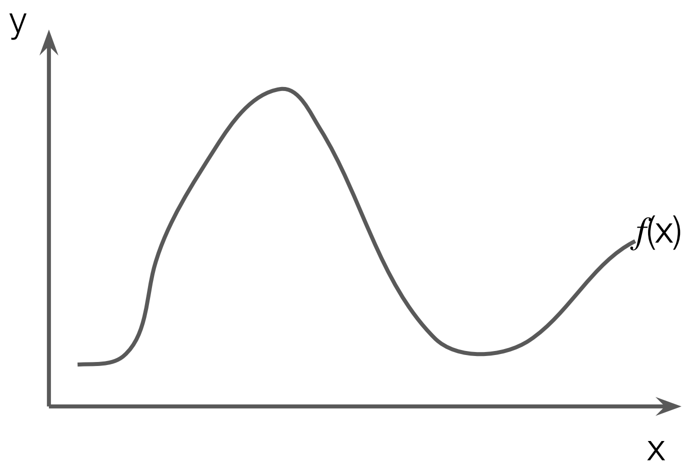


https://pngtree.com/freepng/question-expression-cartoon-illustration_4545209.ht

Calculus

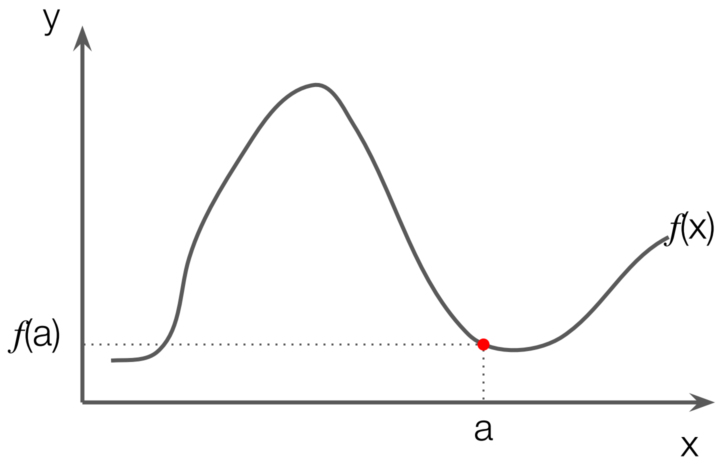
Derivatives

The derivative $df(a)/dx$ of a function f is the instantaneous rate of change of $y = f(x)$ with respect to x when $x = a$.



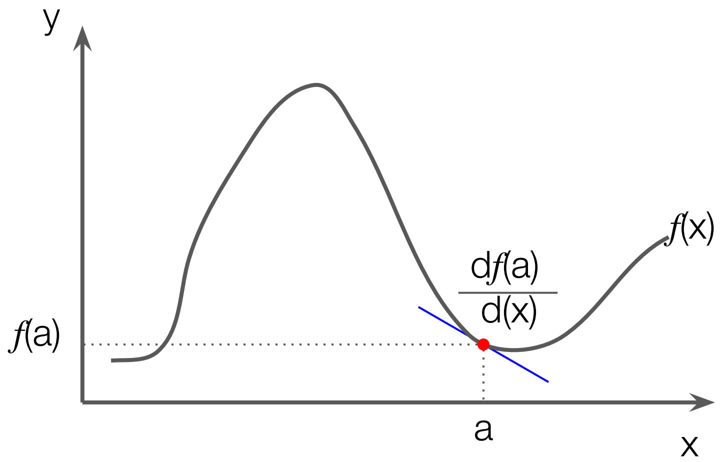
Derivatives

The derivative $df(a)/dx$ of a function f is the instantaneous rate of change of $y = f(x)$ with respect to x when $x = a$.



Derivatives

The derivative $df(a)/dx$ of a function f is the instantaneous rate of change of $y = f(x)$ with respect to x when $x = a$.

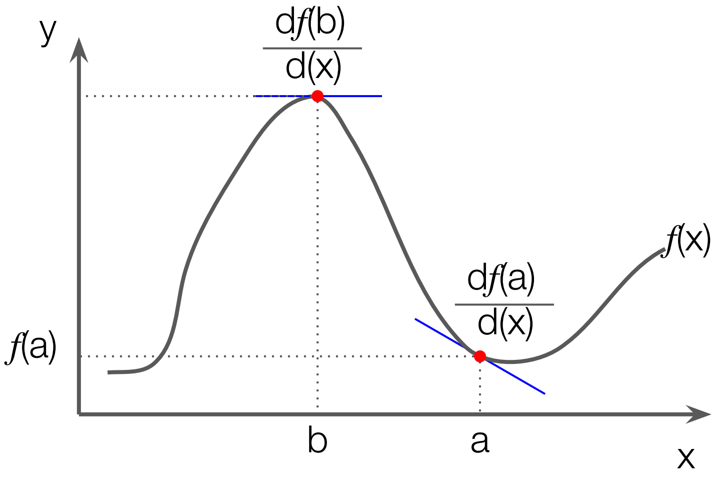


Derivatives

The derivative $df(a)/dx$ of a function f is the instantaneous rate of change of $y = f(x)$ with respect to x when $x = a$.

Useful property

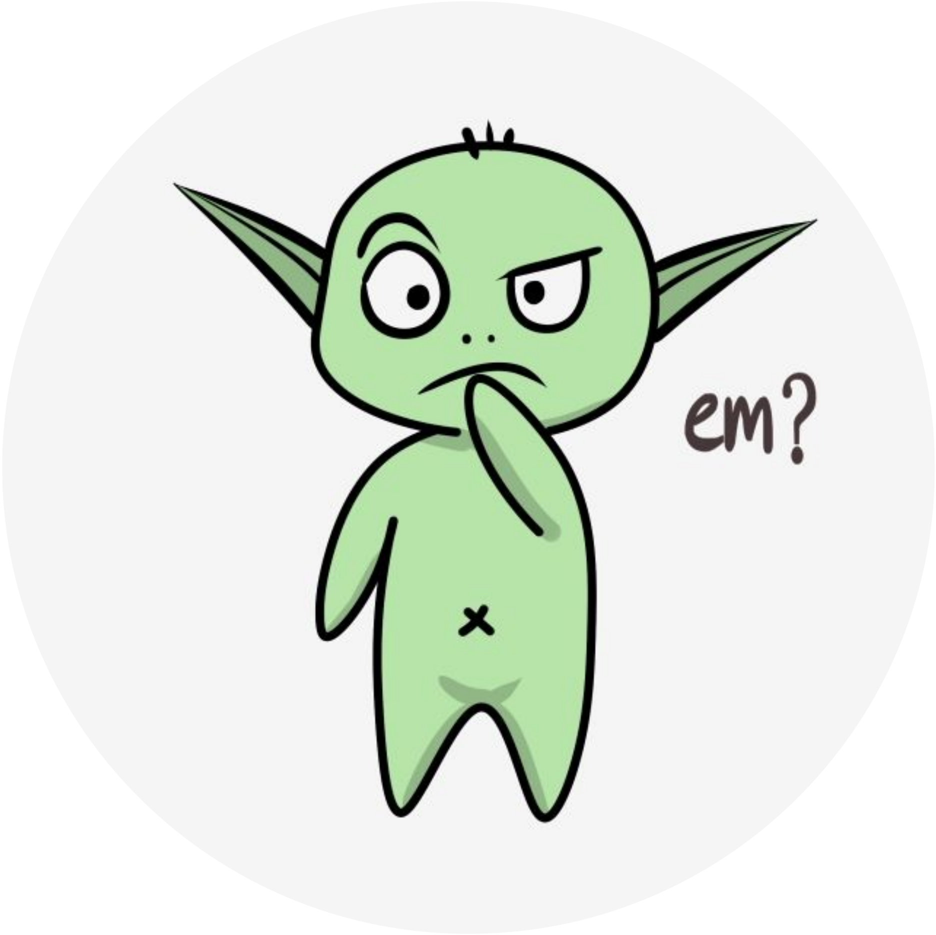
The derivative of a function is zero at its local minima and its local maxima.



Useful property

We can sample from f and we can use its gradient to find a local minimum or a local maximum. That's stochastic gradient descent / ascent:

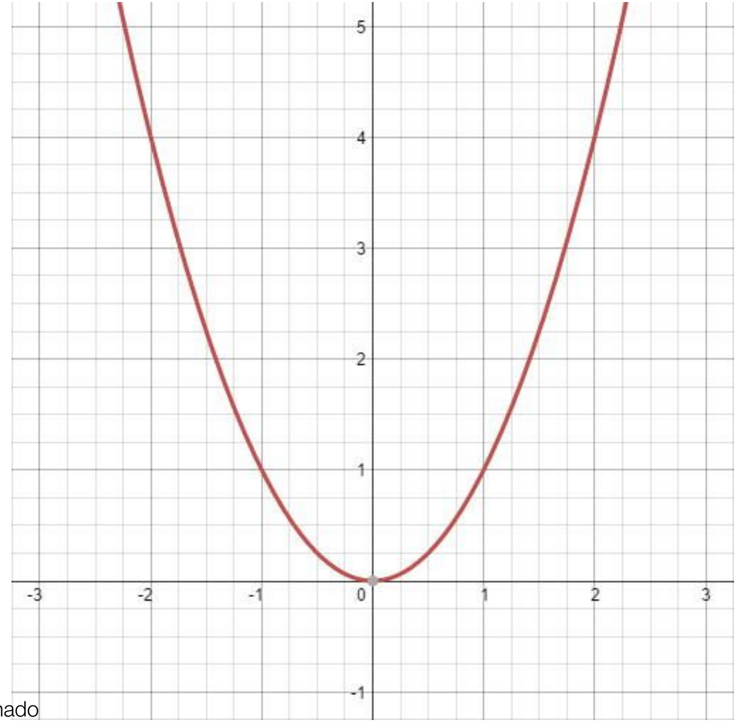
$$x' \leftarrow x \pm \alpha \nabla_x f(x).$$



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

Example – Stochastic gradient descent

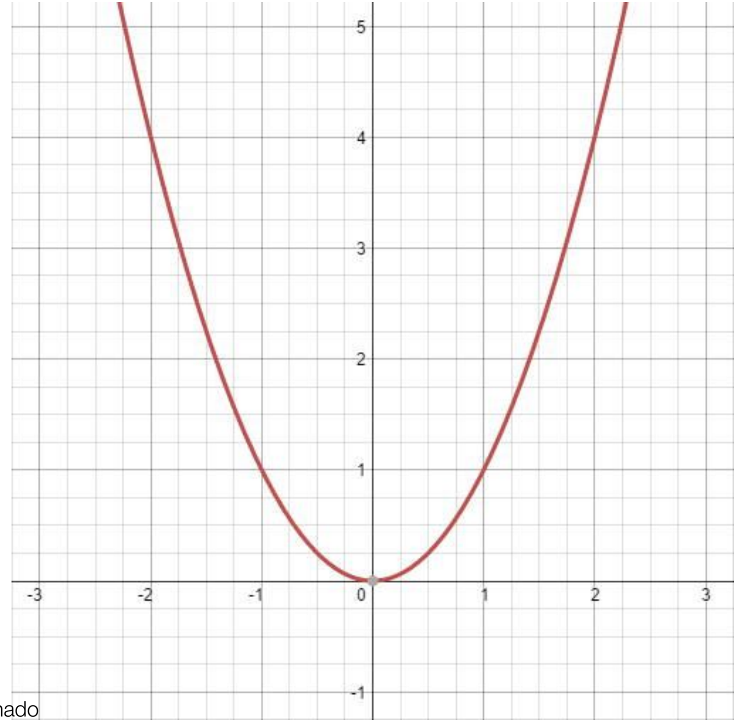
Say we have a function $f(z) = z^2$, and we want to find the z that minimizes its value.



$z' \leftarrow z \pm \alpha \nabla_z f(z)$

Example – Stochastic gradient descent

Say we have a function $f(z) = z^2$, and we want to find the z that minimizes its value.

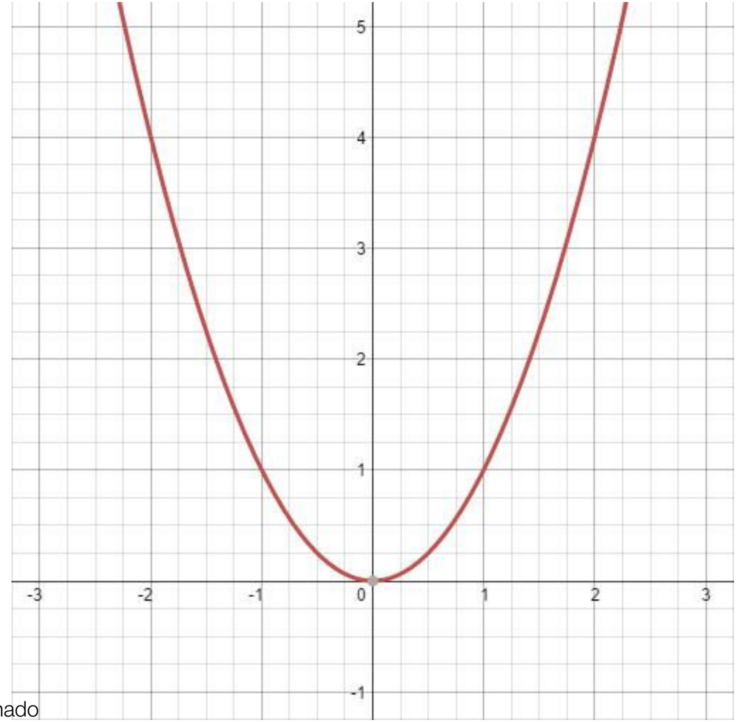


$$\frac{df(z)}{dz} =$$

$$\boxed{z' \leftarrow z \pm \alpha \nabla_z f(z)}$$

Example – Stochastic gradient descent

Say we have a function $f(z) = z^2$, and we want to find the z that minimizes its value.

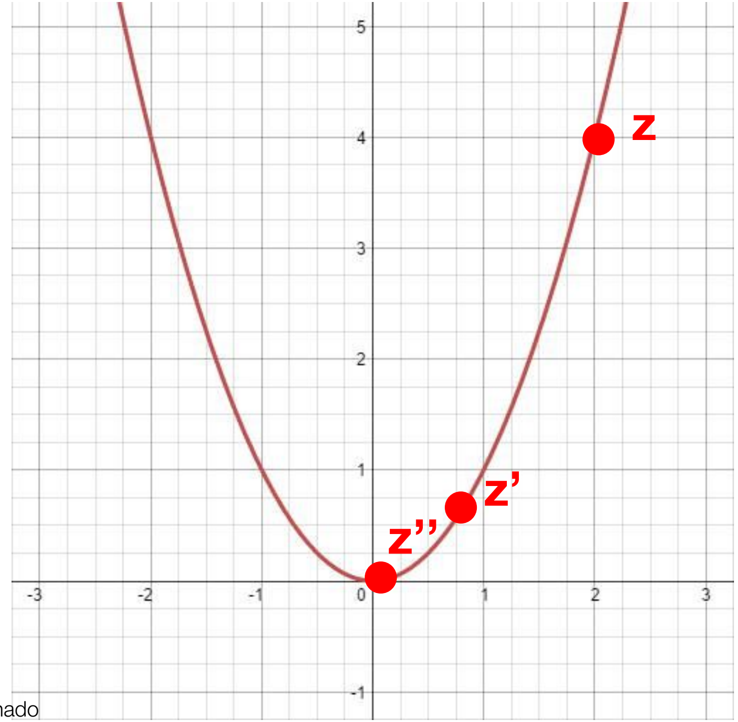


$$\frac{df(z)}{dz} = 2z$$

$$\boxed{z' \leftarrow z \pm \alpha \nabla_z f(z)}$$

Example – Stochastic gradient descent (intuition)

Say we have a function $f(z) = z^2$, and we want to find the z that minimizes its value.



$$\frac{df(z)}{dz} = 2z$$

$$\boxed{z' \leftarrow z \pm \alpha \nabla_z f(z)}$$

$$\alpha = 0.4$$

$\nabla f(4) = 2 \times 4 = 8$
$z' \leftarrow 4 - 0.4 \times 8$
$z' = 0.8$

$\nabla f(0.8) = 2 \times 0.8 = 1.6$
$z'' \leftarrow 0.8 - 0.4 \times 1.6$
$z'' = 0.16$

The gradient vector

The gradient of f, denoted by ∇f, is a generalization of derivatives to a multi-dimensional function (the collection of all of its partial derivatives).

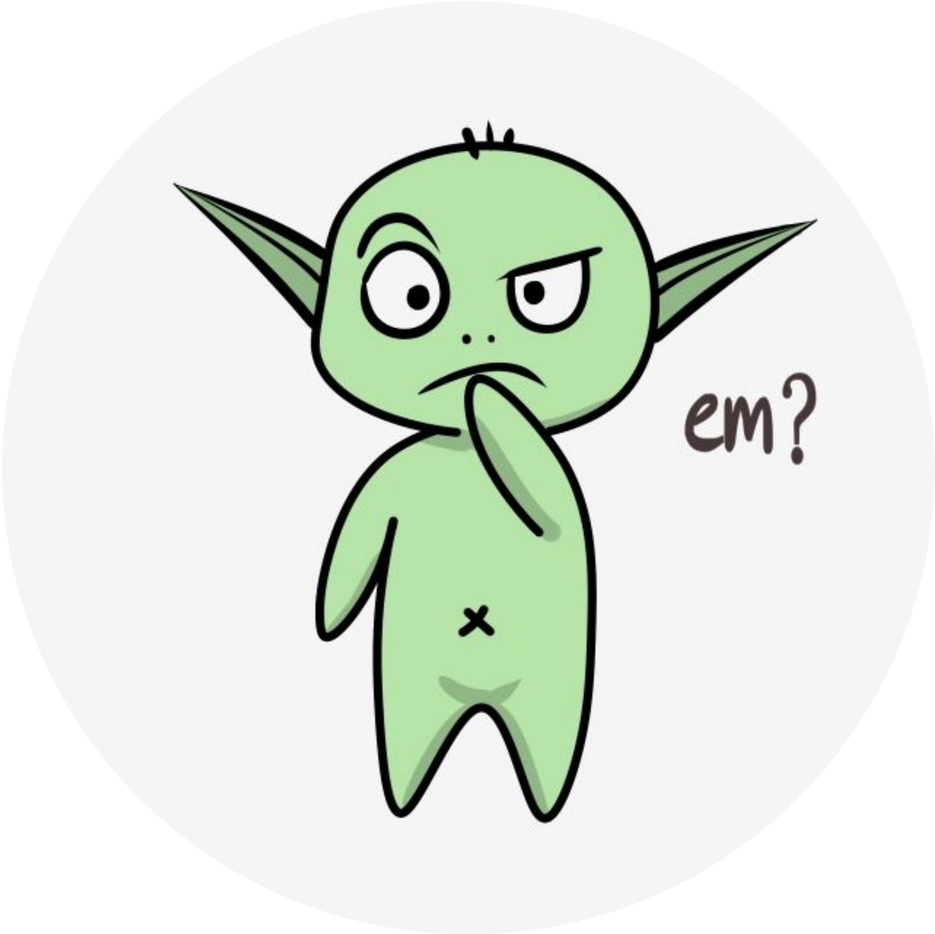
$$\nabla f(x_0, y_0, \ldots) = \left[\frac{\partial f}{\partial x}(x_0, y_0, \ldots), \ \frac{\partial f}{\partial y}(x_0, y_0, \ldots), \ \ldots \right]^\top$$

Example

If $f(x, y) = x^2 + x \ln y$, which one is the right ∇f?

a. $\begin{bmatrix} 2x + \ln y \\ x/y \end{bmatrix}$ b. $\begin{bmatrix} 2x + x\ln y \\ x^2 + x/y \end{bmatrix}$

∇f outputs a vector with all possible partial derivatives of f.



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

Matrix calculus (just in case)

The gradient is the transpose of the scalar-by-vector derivative, but there's more!

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}.$$

scalar-by-vector
(a.k.a. gradient)

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}$$

vector-by-scalar

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

vector-by-vector
(a.k.a. Jacobian)

matrix-by-scalar

$$\frac{\partial \mathbf{Y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix}$$

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{p1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{p2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1q}} & \frac{\partial y}{\partial x_{2q}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}$$

scalar-by-matrix

See Wikipedia [article](#) for details (I got these images from there).

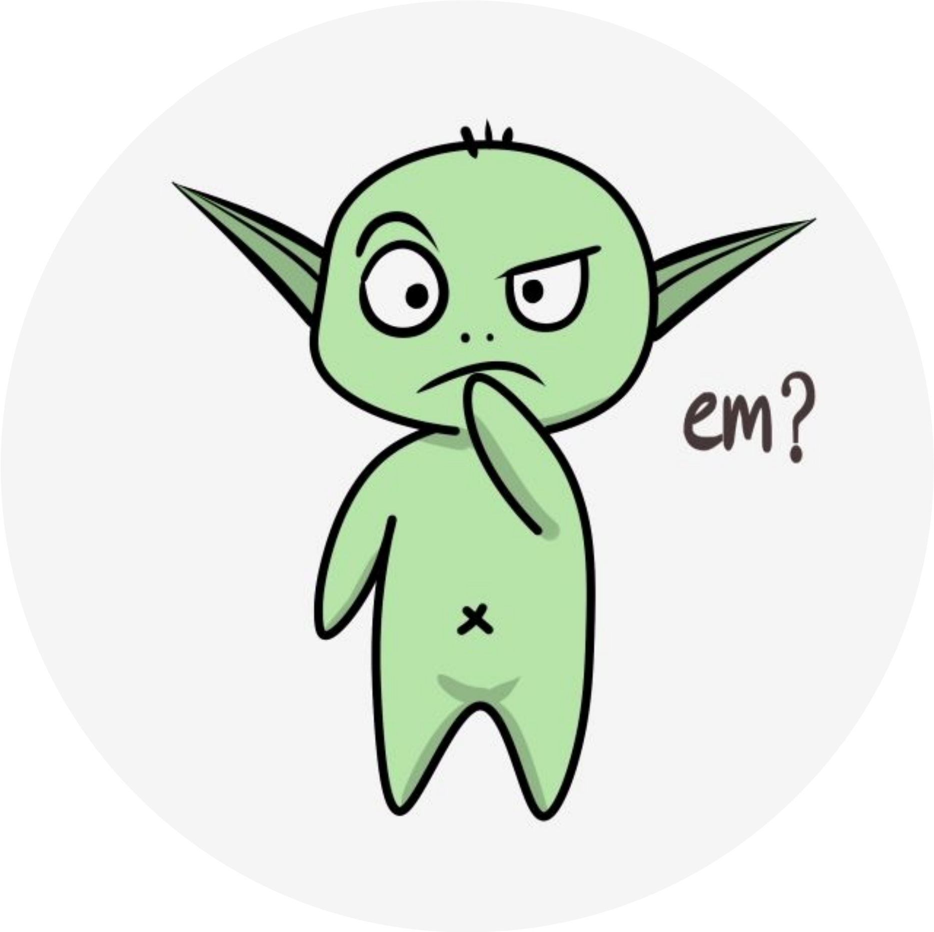
Matrix calculus (just in case)

“Traditional” way: write out objective as a sum, differentiate, find a matrix notation.

$$\begin{aligned}
\partial \mathbf{A} &= 0 \qquad\qquad (\mathbf{A} \text{ is a constant}) \\
\partial(\alpha \mathbf{X}) &= \alpha \partial \mathbf{X} \\
\partial(\mathbf{X} + \mathbf{Y}) &= \partial \mathbf{X} + \partial \mathbf{Y} \\
\partial(\mathrm{Tr}(\mathbf{X})) &= \mathrm{Tr}(\partial \mathbf{X}) \\
\partial(\mathbf{X}\mathbf{Y}) &= (\partial \mathbf{X})\mathbf{Y} + \mathbf{X}(\partial \mathbf{Y}) \\
\partial(\mathbf{X} \circ \mathbf{Y}) &= (\partial \mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (\partial \mathbf{Y}) \\
\partial(\mathbf{X} \otimes \mathbf{Y}) &= (\partial \mathbf{X}) \otimes \mathbf{Y} + \mathbf{X} \otimes (\partial \mathbf{Y}) \\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial \mathbf{X})\mathbf{X}^{-1}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} &= \frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \\
\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} &= \mathbf{a}\mathbf{b}^T \\
\frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{b}}{\partial \mathbf{X}} &= \mathbf{b}\mathbf{a}^T \\
\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{a}}{\partial \mathbf{X}} &= \frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T
\end{aligned}$$

See The Matrix Cookbook by Petersen & Pedersen (all these relationships are from there).



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

Next class

- What **I** plan to do: Fundamentals of RL: An introduction to sequential decision-making (Bandits)
  - Discuss, more in depth, things related to bandits (Chapter 2 of the textbook).

- What I recommend **YOU** to do for next class:
  - Make sure you have access to Coursera, eClass, and Slack.
  - Read Chapter 1 (not mandatory) and Chapter 2 of the textbook.
  - Finish weeks 1 and 2 of "Fundamentals of RL: An introduction to sequential decision-making".
  - Submit practice quiz and programming assignment for Coursera's M1 W2.
  - Start thinking about the course project and groups.