Introduction

CMPUT 267: Basics of Machine Learning

Chapter 1

What is machine learning?

- Mitchell: "The field of machine learning is concerned with the question of how to construct computer programs that **automatically** improve with **experience**."
- Russell & Norvig: "... the subfield of AI concerned with programs that learn from **experience**."
- Murphy: "The goal of machine learning is to develop methods that can **automatically** detect patterns in **data**, and then to use the uncovered patterns to predict future data of other outcomes of interest."

What is this course about?

We want to be able to have good rules (or functions) for predicting outcomes
e.g., predict the temperature tomorrow, based on weather over the last five days

You could construct these rules by hand, or learn them from **data:**

- But the data are often **incomplete**:
  - Partial observability: Incomplete knowledge of environment
  - Incomplete knowledge of other agents' actions
- **Machine learning algorithms** are one way to learn from incomplete data

Course goal:
Understand machine learning algorithms by **deriving them** from the beginning.

- with a focus on prediction on new data

Example: Predicting house prices

- **Goal**: we want to predict house prices, given only the age of the house

  $f(age) =$ price of the house

- **Dataset**: house sales this year, with attributes **age** and target value **price**

  $\{(age_1, price_1), (age_2, price_2), \ldots, (age_9, price_9)\}$

- **Idea**: A function that accurately outputs price from age for these specific pairs might also provide good predictions for new houses

Formalizing the problem

Definitions:

Let x be **age** and y be **price**

Let $D = \{(x_1, y_1), \ldots, (x_9, y_9)\}$ be our dataset

Objective:

We want to make the **difference** between $f(x_i)$ and y_i **small**

$$\min_{f \text{ in function space}} \sum_{i=1}^{9} (f(x_i) - y_i)^2$$

Questions:

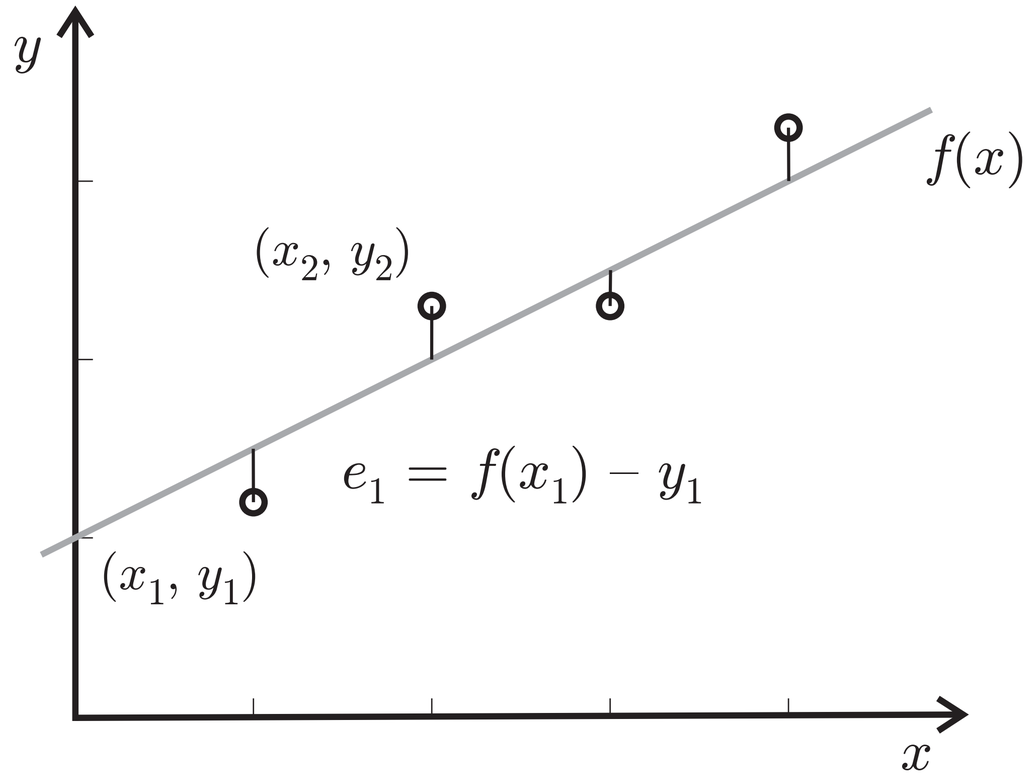
1. Why are we **squaring** the difference?
2. Why are we **summing** up the errors?
3. What could we consider for the **function space**?

Linear function space

Definition:

A function f is a **linear function** of x if it can be written as $f(x) = w_0 + w_1 x$

$$\min_{f \text{ in linear functions}} \sum_{i=1}^{9} (f(x_i) - y_i)^2$$

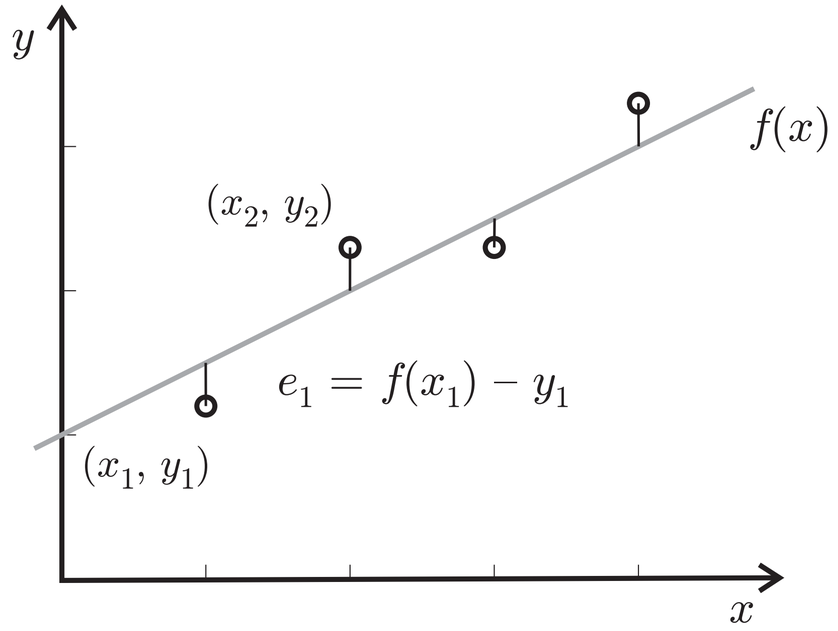


Solving for the optimal function

Objective then becomes:

$$\min_{f \text{ in function space}} \sum_{i=1}^{9} (f(x_i) - y_i)^2$$

$$= \min_{w_0, w_1} \sum_{i=1}^{9} (\underbrace{w_0 + w_1 x_i}_{f(x_i)} - y_i)^2$$



Questions:

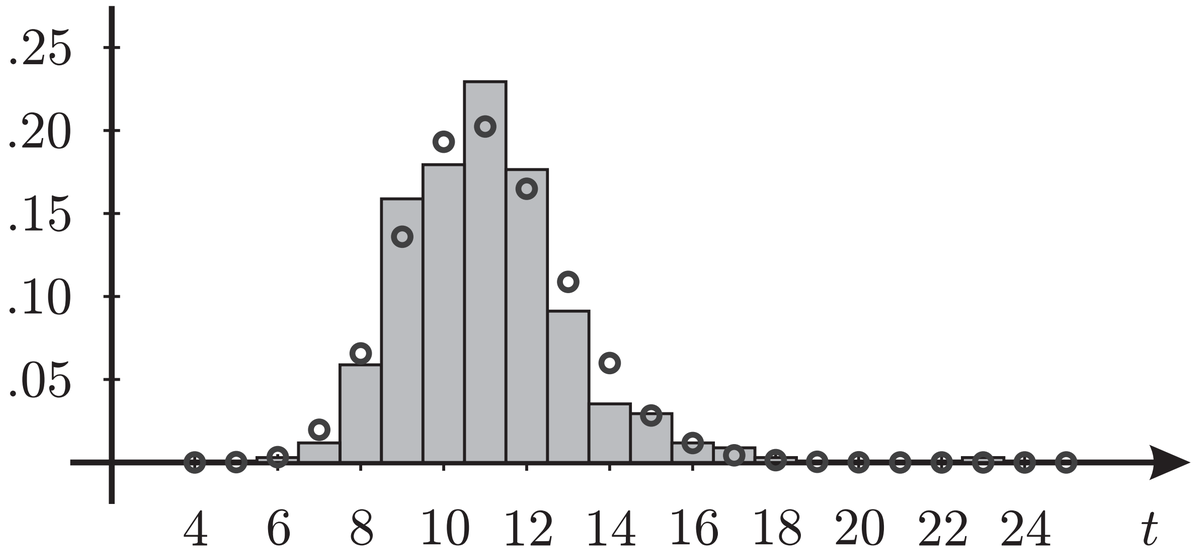
1. Would you use this to predict the value of a house? Why/why not?
2. Will this predict well? How do we know?
3. What is missing to make these assessments?

Probabilities!

- **Question:** Is it likely that there is a **deterministic** function from **age** to **price**?
  - Many houses will have the same **age** but different **price**...
- We can instead use a probabilistic approach:
  - Learn a function that gives a **distribution** over **targets** (**price**) given **attributes** of the item (**age**)
- **Question:** Does this mean that we think the world is stochastic rather than deterministic?
  - Stochasticity can come from **partial observability**
  - Maybe the outcome *really is* deterministic if we knew **age**, and **size**, and **number of rooms**, and **distance to airport**, and **whether the queen lives there**, and ...

Probabilities let us specify our uncertainty in predictions

- Imagine we have a distribution over the prices for a given age of 40 years
  - The x-axis is the prices (in 10s of thousands)
  - The y-axis is the probability of that price, for the age of 40 years
- We might reasonably **predict** our house price is something like 100000 or 120000, based on this distribution
- But we also know that the spread is quite large, and that there is **uncertainty** in our prediction



Probabilities let us specify our uncertainty in predictions

- We might reasonably **predict** our house price is something like 100000 or 120000, based on this distribution
- But we also know that the spread is quite large, and that there is **uncertainty** in our prediction
- A further nuance: this distribution itself is an **estimate** (from data) and so we actually have **uncertainty** about it too!
  - We will also use **probabilities** to quantify this **uncertainty**

Course topics

1. Probability background (ch.2)
2. Estimation with sample averages (ch.3)
   - Concentration inequalities: how confident should we be in our estimates?
   - Sample complexity and convergence rate
3. Optimization (ch.4)
4. Parameter estimation (ch.5)
   - Maximum likelihood and MAP
   - Beyond point estimates: Bayesian estimation
5. Stochastic Gradient Descent and Handling Big Datasets (ch. 6)

Course topics #2

5. Prediction (ch.7)
   - Formalizing the prediction objective
6. Linear & polynomial regression (ch.8)
7. Generalization error and evaluating models (ch.9)
8. Regularization and constraining the function space (ch.10)
9. Logistic regression and linear classifiers (ch.11)
10. Bayesian linear regression (ch.12)

Course essentials

- **Course information:** https://marthawhite.github.io/mlbasics/
  - Schedule and readings
- **Access-controlled course information:** eClass
  - Getting Started and FAQ (please visit this today!)
  - Video recordings, links to lecture meetings and assignment submission

Hybrid Teaching

- Some students are still remote
  - and maybe recordings are a good idea anyway
- **Lectures will be in-person in a classroom AND on Zoom**
- In class, I will project my screen. I and a TA will also connect to Zoom. My screen will be shared in Zoom too.
- The TA will monitor the Zoom chat and ask any questions posted there
- The lectures will be recorded.

Course essentials

- **Course information:** https://marthawhite.github.io/mlbasics/
  - Schedule and readings
- **Access-controlled course information:** eClass
  - Getting Started and FAQ (please visit this today!)
  - Video recordings, links to lecture meetings and assignment submission
- **Lectures:** Tues. Thurs 12:30-1:20pm in BS M 1-45 and on Zoom
  - Lectures will be recorded and posted on eClass
- **Office hours:** Wednesday morning (on Zoom and in-person)

Teaching Assistants

Deepak Mamillapalli
Farzane Aminmansour
Hager Radi
Kamran Janjua
Robert Joseph
Thang Chu

- **Office hours:** see eClass for times and Zoom link
  - Typically question/answer sessions
  - Please no arguing for marks (see more in the FAQ)
- **No TA office hours this week**

Lab

- There is no official lab
- One of the office hours will be a 2 hour chunk in a large classroom, to facilitate asking about assignments
- If there are seminar or tutorials during this time, TAs will record them
- Two TAs will be available for the first hour.
- For the second hour, only one TA will be present
- **Please complete the poll on eClass to find a good time**

Readings

- Readings are from the Basics of ML textbook
  - Available on course site and written by myself
  - It's a fast read
  - I have modified these notes each time I have taught this course. They are stabilizing, but I have a few more clarifications to add
- See the schedule for sections and for reading deadlines
- Readings have an associated marked component called Thought Questions

Lectures

- Lectures will mostly involve me writing on my iPad (like a whiteboard)
- I highly encourage you to ask any question
  - You can raise your hand and then ask outloud
  - You can type questions in Zoom chat
  - We will use Discord for any questions you think of outside of class, that I will address in class
- We will have small exercise breaks in class
  - usually a small derivation or where you write a question for a neighbor
  - sometimes jumping jacks or sit ups (:D)
- I will post my written notes afterwards (and videos will be published)

You should still take your own notes

- The notes I upload will be whatever is left at the end of class. This means
  - It won't include things I said outloud
  - Sometimes I might modify a picture as we go, and it gets a bit messy by the end
  - Sometimes it might be useful to erase a couple of things, and so that will also not be in the uploaded notes
- Further, there is a known phenomena where **writing things down helps you remember and learn them**
  - If you write it, then you might realize that something is confusing

Course Discussion

- We have create a **Discord group; please sign up!**
- I want to generate as much class discussion as possible
- Please go there first to ask questions
- Please answer your classmates questions!
  - We'll step in if there is misinformation, but in many cases you can all help each other faster than we can get to the question
  - Peer discussions can very beneficial
- **Details in FAQ and Getting Started linked on eClass**

Course Pre-reqs

- Official pre-reqs
  - Calculus 1, Computing 1 (CMPUT 174/274)
- Co-requisites include
  - Stats 1, Linear Algebra 1 (Math 125/127), CMPUT 272
- Co-requisites are actually ideally taken as Pre-reqs, but are not strictly necessary as pre-reqs
  - This gives you the flexibility in scheduling
  - But you still have those concepts reinforced either before or during this course, so that by the middle of the course it is helping

More on Prerequisites

- Basic mathematics
  - Some calculus
  - Some probabilities
  - Some linear algebra (vectors and dot products mostly)
- Motivation to learn
- Motivation to think **beyond the material**
  - This is what thought questions are meant to practice
- I hope you have exposure to these topics and I re-teach them, from the perspective of their utility to Machine Learning

Course Expectations

- This course does not have many official pre-reqs, and I do not expect you to be skilled yet with math
  - Some students will have more math background; don’t worry to much about it
- But, I expect that you want to learn these foundations, even if they are hard
- I expect you to ask questions and come to me during the course if you are struggling
- I expect some amount of struggle and a somewhat steep learning curve; as the course progresses, it will become easier
- We will all make mistakes (me, you, the TAs)
- I welcome feedback, both during and outside of lecture

A brief interlude with some unsolicited advice
(after all, I’m here to impart my knowledge to you)

"Why is there so much math?"

- **This course is very mathematical**, with detailed derivations
  - This is **absolutely necessary**
- "But I just want to use machine learning to solve Problem X!"
  1. **Applying algorithms correctly** is much easier when you understand their development and assumptions
     - You will be more effective at solving Problem X if you **understand the algorithms** that you apply
     - This means understanding their derivation
  2. **Formalizing the problem** is often half the battle to solving it effectively!
     - Comfort with math is an important part of being a computer scientist

Problem solving

- CS is about problem solving through the medium of computing
  - Not about becoming an expert programmer
- Primary goal is carefully designing solutions to problems, by:
  - **Formalizing** the problem
  - **Understanding** different potential approaches
  - **Evaluating** the solution
- Comfort with mathematical concepts enables **clarity** through logical thinking

While I am giving you unsolicited advice…

- **Writing** and **clear thinking** are extremely important skills for CS
- You learn a lot of really useful things in CS, that will serve you well afterwards, but only if you take the **learning** seriously
  - CS is not a professional degree where you need the piece of paper
  - You can get jobs in CS without a CS degree, and CS degree does not guarantee you get a job in CS
  - If you get a CS degree with the bare minimum, then …
- Our goal is to teach you how to approach problems and continue learning in the future (not necessarily to teach you the tools or languages of today)

Some career advice

- **Internships** are a good idea
- We are not a professional organization, but you are skilled. Be **confident** in yourself (aka, you're pretty great)
- Practice some **professional skills:** writing resumes, giving talks (toastmasters), general writing clarity (take a writing course)
- If you want to go into **graduate studies**, you need good marks
  - Or you need to start working a prof in undergraduate research early

Back to the course organization

Grading

- 30%: Assignments
  - Mixture of mathematical problems and programming exercises
- 5%: Quiz on **October 14**
- 20%: Midterm exam on **November 18**
- 35%: Final exam **December 21**
- 10%: Thought questions

More on Grading

- There will be no arguing for marks, for assignments or exams
  - Arguing introduces a lot of bias: it benefits those willing to argue and punishes those that are not willing to argue
  - We will of course fix incorrectly tabulated marks
- We have a double marking system for exams, to try to ensure consistency
- Assignments are meant to be challenging. Exams are much simpler and meant to test if you understood the fundamental ideas.
- Final letter grades are given based on relative performance, but they are not on a Gaussian curve. The cut-offs are adjusted each year

More on Grading

- Final letter grades are given based on relative performance, but they are not on a Gaussian curve. The cut-offs are adjusted each year
- Last year's announcement on letter grades: "A+: mark > 97, A: mark > 90, A-: mark > 83, B+: mark > 77, B: mark > 71, B-: mark > 67, C+: mark > 61, C: mark > 58, C-: mark > 52, D: mark > 48
  You might wonder where this spacing comes from. It gets adjusted each year due to performance. There is a slightly more generous interval for As, than Bs; and a slightly more generous interval for Bs than Cs. The reason for this is that in some sense grades compress as you get closer to the 50% mark. It is easier to get at least 50%, and harder to get 90%. Grade allocation is tough, since there are always some near the boundary. But I hope you remember that I have already erred on the side of being somewhat generous (rather than strict) for the grade decisions."

Assignments

- Four assignments
- Coarse binned grading:
  - 80 - 100 → 100
  - 60 - 80 → 80
  - 40 - 60 → 60
  - **0 - 40 → 0**

Three exams

- Giving **clear** answers to short answer questions is a **skill**
  - It takes practice!
  - First quiz is your chance to practice this skill with low stakes
  - It's only 5% of the grade (less than one assignment)
- Practice questions will be available
- Exams will ***hopefully*** be in-person
  - May have to resort to eClass, depending on the situation
  - Some accommodations for those that are still remote by the quiz
- For all exams you are allowed a **two page cheat-sheet**

Collaboration policy

Detailed version on the syllabus section of the website

You are **encouraged to discuss assignments** with other students:

1. You must **list** everyone you talked with about the assignment.
2. You **may not** share or look at each other's **written work or code**.
3. You must **write up** your solutions individually

Individual work only on **exams**: No collaboration allowed

Academic conduct

- Submitting someone else's work as your own is **plagiarism**.
- So is helping someone else to submit your work as their own.
- We report **all cases** of academic misconduct to the university.
- The university takes academic misconduct **very seriously**.
  Possible consequences:
  - Zero on the assignment or exam (virtually guaranteed)
  - Zero for the course
  - Permanent notation on transcript
  - Suspension or expulsion from the university
- **If you are thinking of cheating, since you are stuck or doing poorly, please just talk to me instead. We'll figure it out.**

Additional Questions

- Any questions you have are likely answered in the FAQ and Getting Started document that we have linked on eClass
  - Policies like “No late assignments accepted”, “How to contact TAs”, “What to do if you are going to miss a deadline or exam”
  - “How can I get extra resources?” and “How can I brush up on my math background?”

Readings and Thought Questions

- **It is critical that you do the readings**
- I wrote the notes, and in class lectures essentially follow them quite closely
- If you read and understand the notes, you have learned a lot about ML
- Marked Thought Questions encourage you to actually do the readings

Thought questions

- Thought questions correspond to readings in the notes
- They should demonstrate that you have read **and thought about** the topics
- Needn't have an answer

General format:

1. First, show/explain how you understand a concept
2. Given this context, propose a follow-up question
3. Optional: Proposal an answer to the question, or the way you might find it

Example:
"Good" Thought Question

"After reading about independence, I wonder how one could check in practice if two variables are independent, given a database of samples? Is this even possible? One possible strategy could be to approximate their conditional distributions, and examine the effects of changing a variable. But it seems like there could be other more direct or efficient strategies."

Example:
"Bad" Thought Questions

- "I don't understand linear regression. Could you explain it again?"
  - i.e., a request for an explanation. If you want to request a clarification, please use slido. avoid any clarification requests from thought questions
- "Derive the maximum likelihood approach for a Gaussian."
  - i.e., an exercise question from a textbook. This is not showing your understanding
- "What is the difference between a probability mass function and a probability density function?"
  - i.e., a question that could be directly answered by reading definitions
  - *BUT* the following modification would be fine: "I understand that PMFs are for discrete random variables and PDFs are for continuous random variables. Is there a way we could define probabilities over both discrete and continuous random variables in a unified way, without having to define two different kinds of function?"

Thought Question marks (10%)

- Four Thoughts Question deadlines (TQ1, TQ2, TQ3, TQ4)
- For each, you need to submit two questions about different subsections in the readings
  - e.g., for TQ1, you might submit one for Section 2.1 and say one for Section 3.2 (please label the corresponding question in your submission)
  - Sometimes the question is more high-level and spans sections. That is fine too; you can write (Spans sections) as the section
- 9% of this mark is for the average of the best three of four
- 1% of this mark is for posting your question on Discord for feedback

Finally: Using Julia instead of Python

- Julia is a newish programming language, focused on numerical computing
  - syntax similar to Matlab
- **I know that this is likely a big deal for some of you**
- But, do not fret!
  - We have provided a tutorial notebook
  - We will release a video early next week going over the basics of Julia
  - The programming is through notebooks; the number of lines of code you implement is not that big
  - You will find Julia quite easy to use and the notebook interface nice

Let us test one poll

- Let’s keep it simple and use Discord
- I posted a poll to lecture-questions, where you can add your reaction.
- Note: we will never check who reacted how, we will just see the general reactions
- **It is effectively anonymous**