Introduction

CMPUT 296: 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**

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

- **Question:** Does **age** give any information on selling **price**?

- **Question:** Do **these pairs** tell us anything about the relationship between **age** and **price** in **future** sales? Why?

- 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:

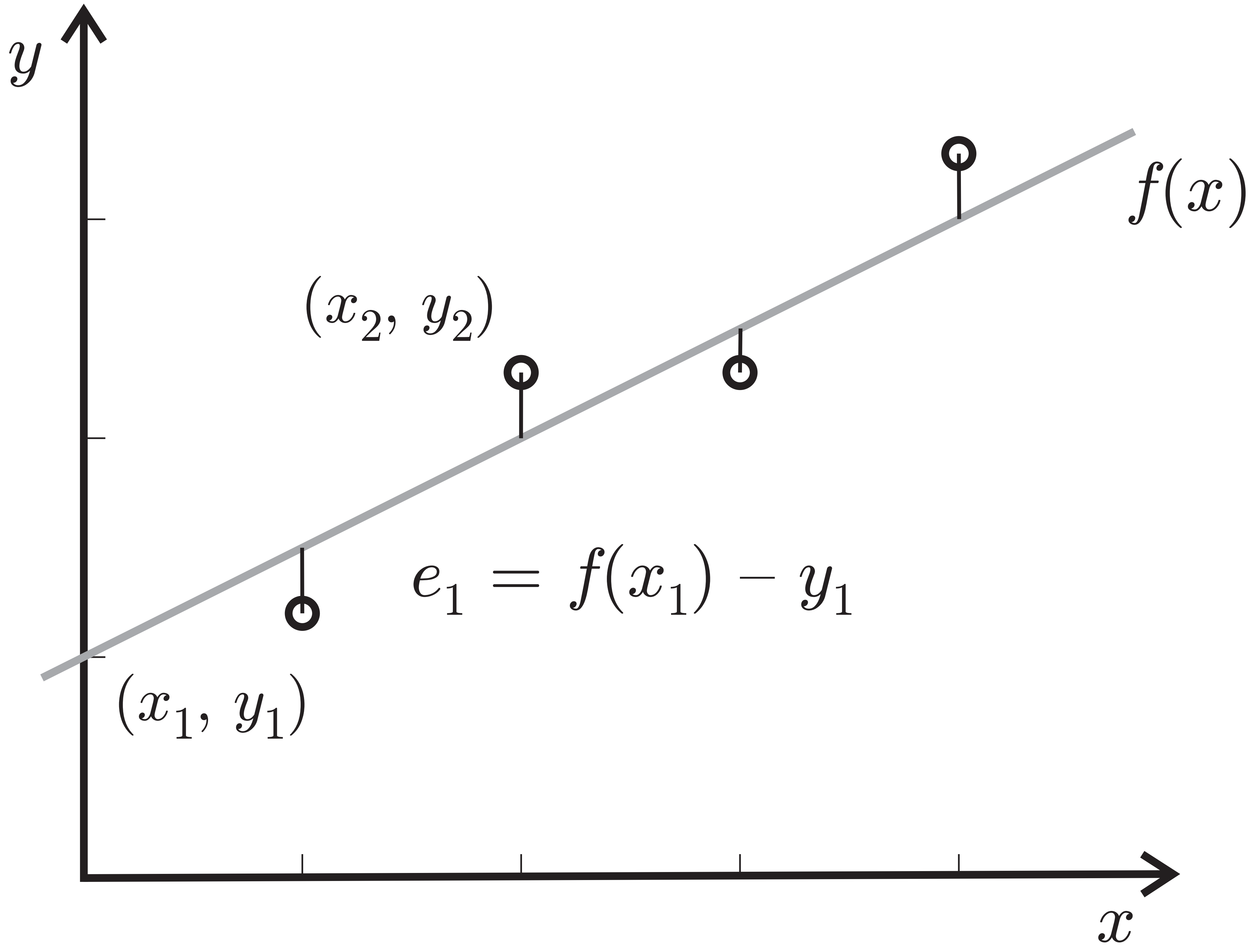
1. Why are we **squaring** the difference?
2. Why are we **summing** up the errors?
3. What could we consider for the **function space**?

Linear function space

Definition:

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

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

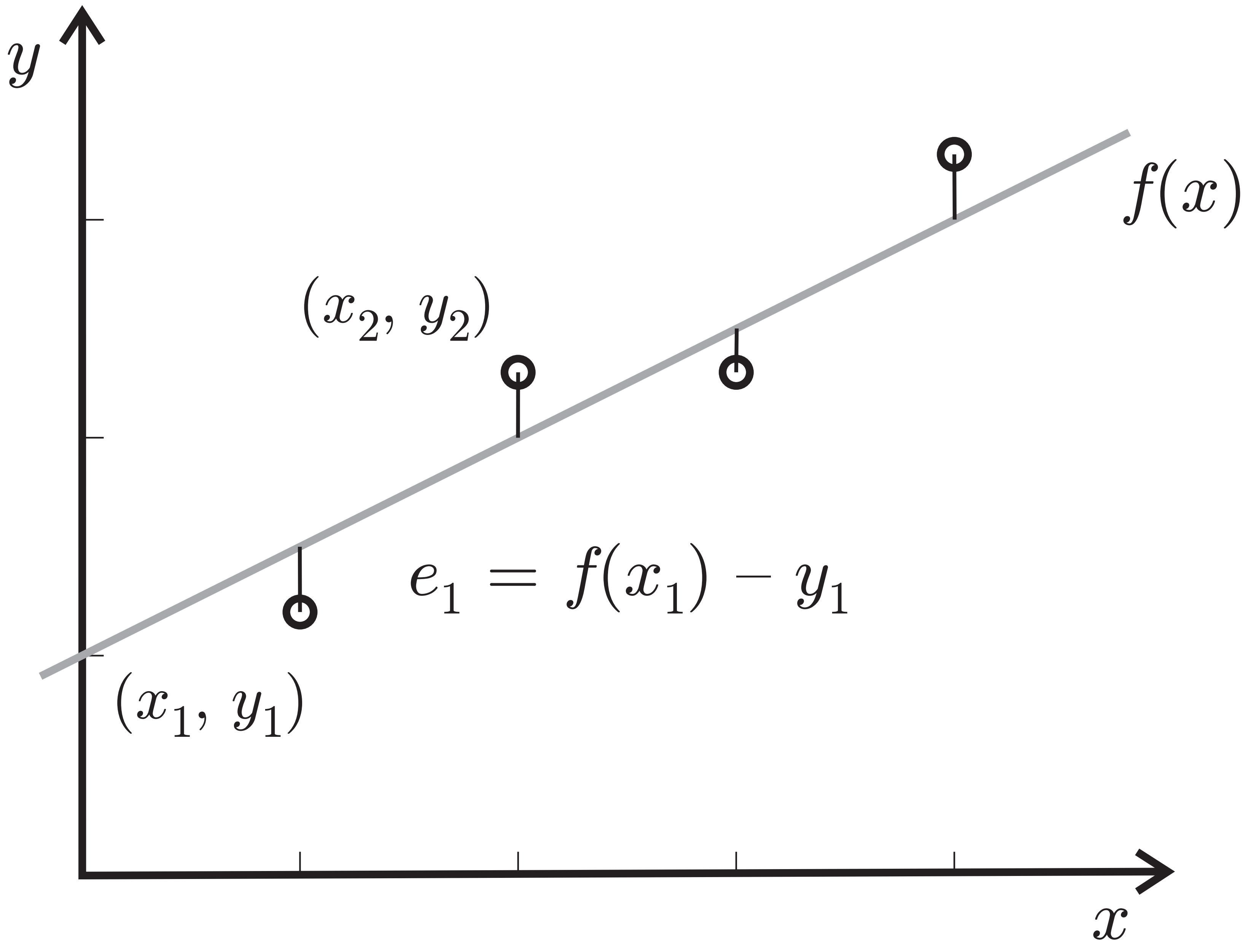


Solving for the optimal function

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:

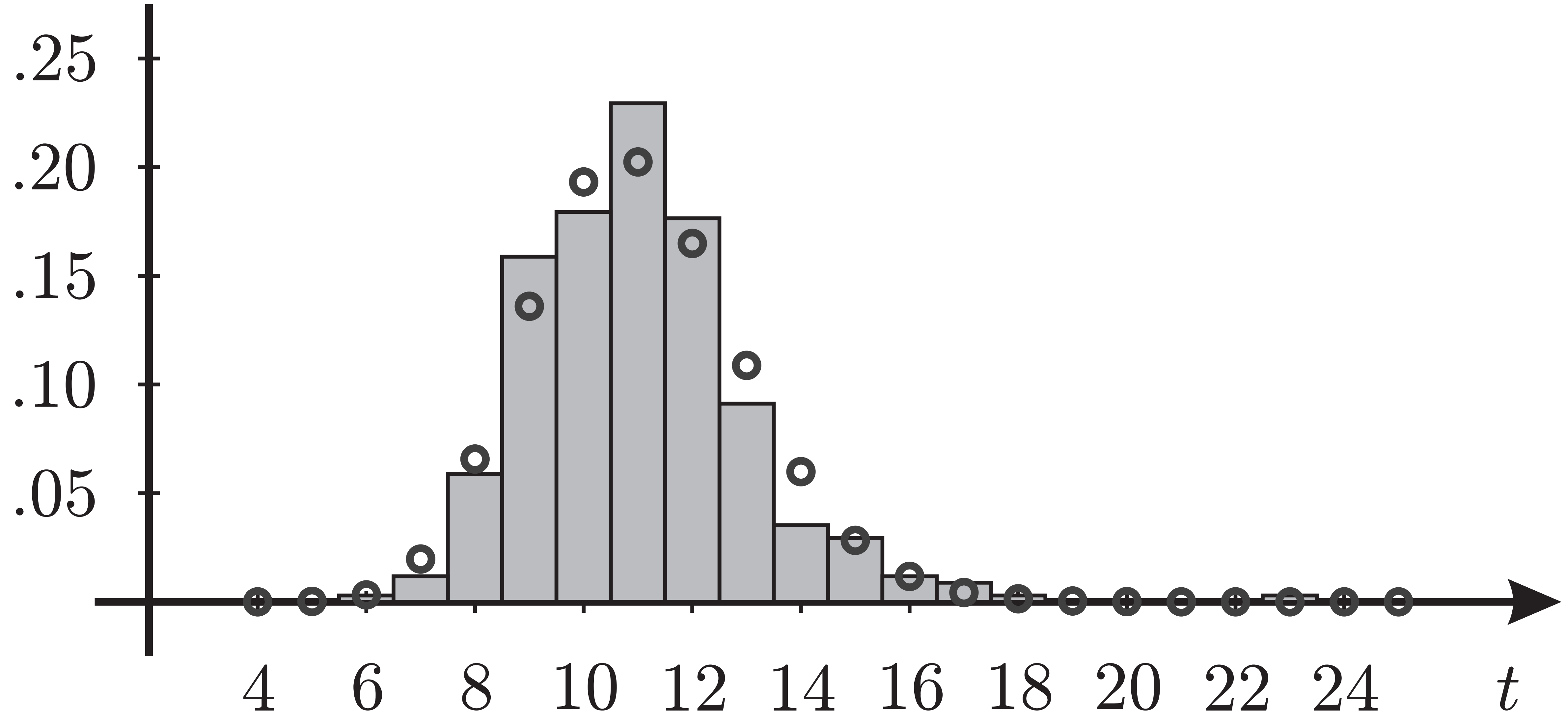
1. Would you use this to predict the value of a house? Why/why not?
2. Will this predict well? How do we know?
3. What is missing to make these assessments?

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

- 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
- 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

Course topics #2

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

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:** Tuesdays and Thursdays, 2:00-3:20pm on Zoom
  - Lectures will be recorded and posted on eClass
- **Office hours:** immediately after lecture on Tuesday

Teaching Assistants

Chunlok Lo
Erfan Miahi
Gábor Mihucz
Kiarash Aghakasiri
Maryam Hashemzadeh
Rohan Nuttall

- **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, but we will have a python tutorial**

Lab

- Lab primarily used as a time to work on assignments and ask TAs questions
- Sometimes, there will be planned tutorials (this week on python)
- Usually, two TAs will come for the first hour and do the following:
    - Set up multiple breakout rooms where you can discuss assignments in smaller groups and potentially identify any points of confusion
    - One TA will stay in the main room and the other will be in a breakout room
    - You can go to either TA to ask questions
    - The TAs will maintain the order by typing your name in the chat
- For the second hour, only one TA will be present

Readings

- Readings from Basics of ML textbook
  - Available on course site
  - It's a fast read
- See the schedule for sections and for reading deadlines
- Readings have an associated marked component called Thought Questions

Prerequisites

- Basic mathematics
    - Some calculus
    - Some probabilities
    - Some linear algebra (vectors and dot products mostly)
    - Crash courses/refreshers along the way
- Motivation to learn
- Motivation to think **beyond the material**
    - This is what thought questions are meant to practice
- I welcome feedback, both during and outside of lecture

Let’s do our first poll!

- I’d like to get a sense of your background
- This will help me pace appropriately, and design lab tutorials to fill in gaps

"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
- If you want to go into **graduate studies**, you need good marks
- 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)

Grading

- 30%: Assignments
  - Mixture of mathematical problem sets and programming exercises
- 5%: Quiz on **February 25**
- 20%: Midterm exam on **March 18**
- 35%: Final exam **April 28**
- 10%: Thought questions

Assignments

- Three 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 be on eClass
    - Exam format detailed in FAQ

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

Lectures

- Lectures will mostly involve me writing on my iPad (like a whiteboard)
- I highly encourage you to ask any question
  - You can type questions in chat
  - You can raise your hand and then ask outloud
  - We will use slido 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 conceptual problem
  - sometimes jumping jacks or sit ups
- 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**

Summary

- 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?”
- All exams will be on eClass
- Readings from a free textbook (written by me)

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