# Introduction

CMPUT 296: Basics of Machine Learning

Chapter 1

# Don't Come to Campus

- **All** of Computing Science's courses are online-only this semester
- CSC and Athabasca Hall are **closed**
  - You can only come if you are explicitly required to by an instructor
  - Even in that case, the Chair and/or Dean need to sign off

# 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?

You need to either construct rules by hand, or derive 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 of 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 recreates these pairs could also provide good predictions

# 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**

$$\text{minimize} \sum_{i=1}^{9} (f(x_i) - y_i)^2$$

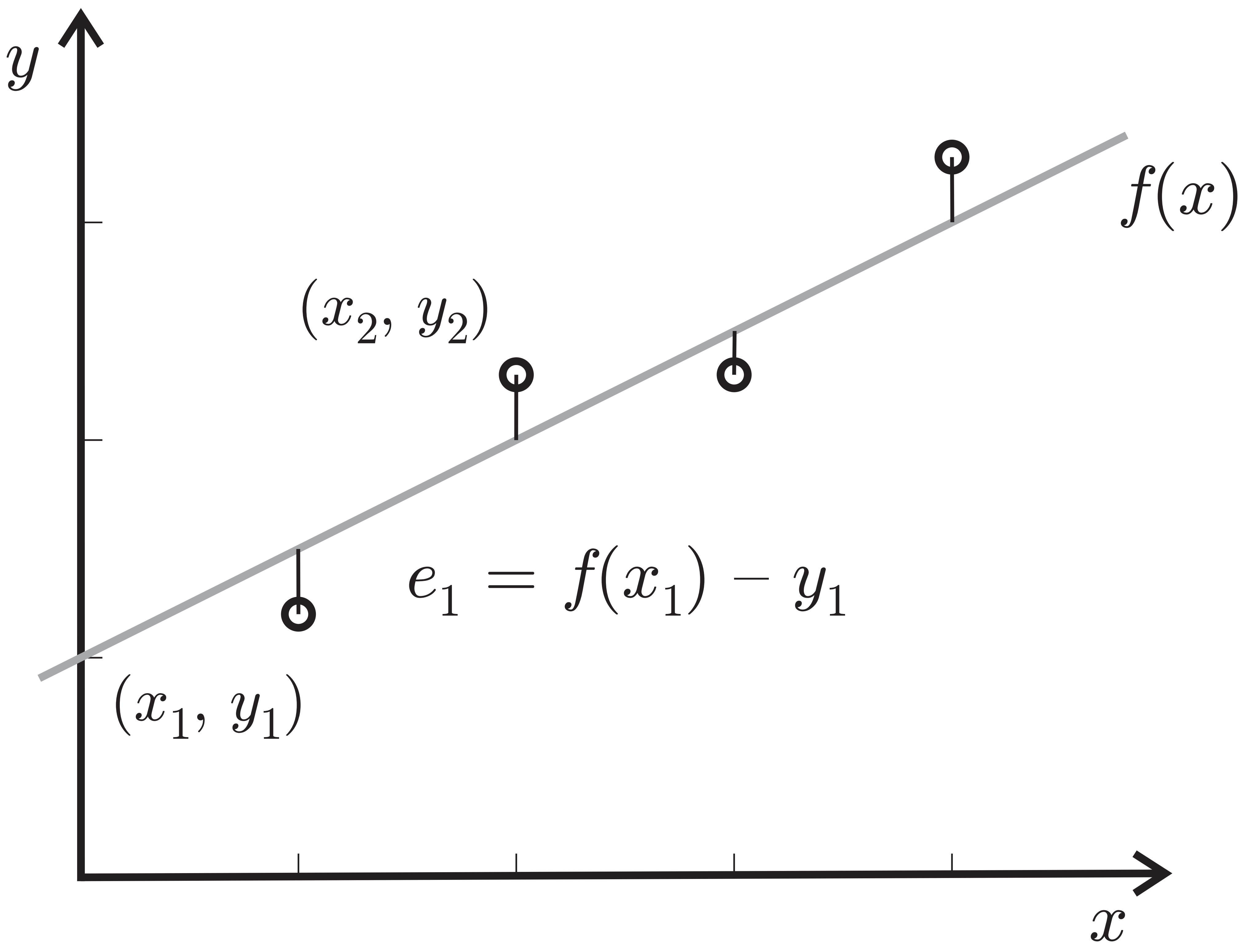
**Questions:**

1. If $f$ can be literally **any** function, then what is the solution?
   - Is that desirable?
2. What could we do instead?
3. Why are we **squaring** the difference?

# Linear function space

**Definition:**

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

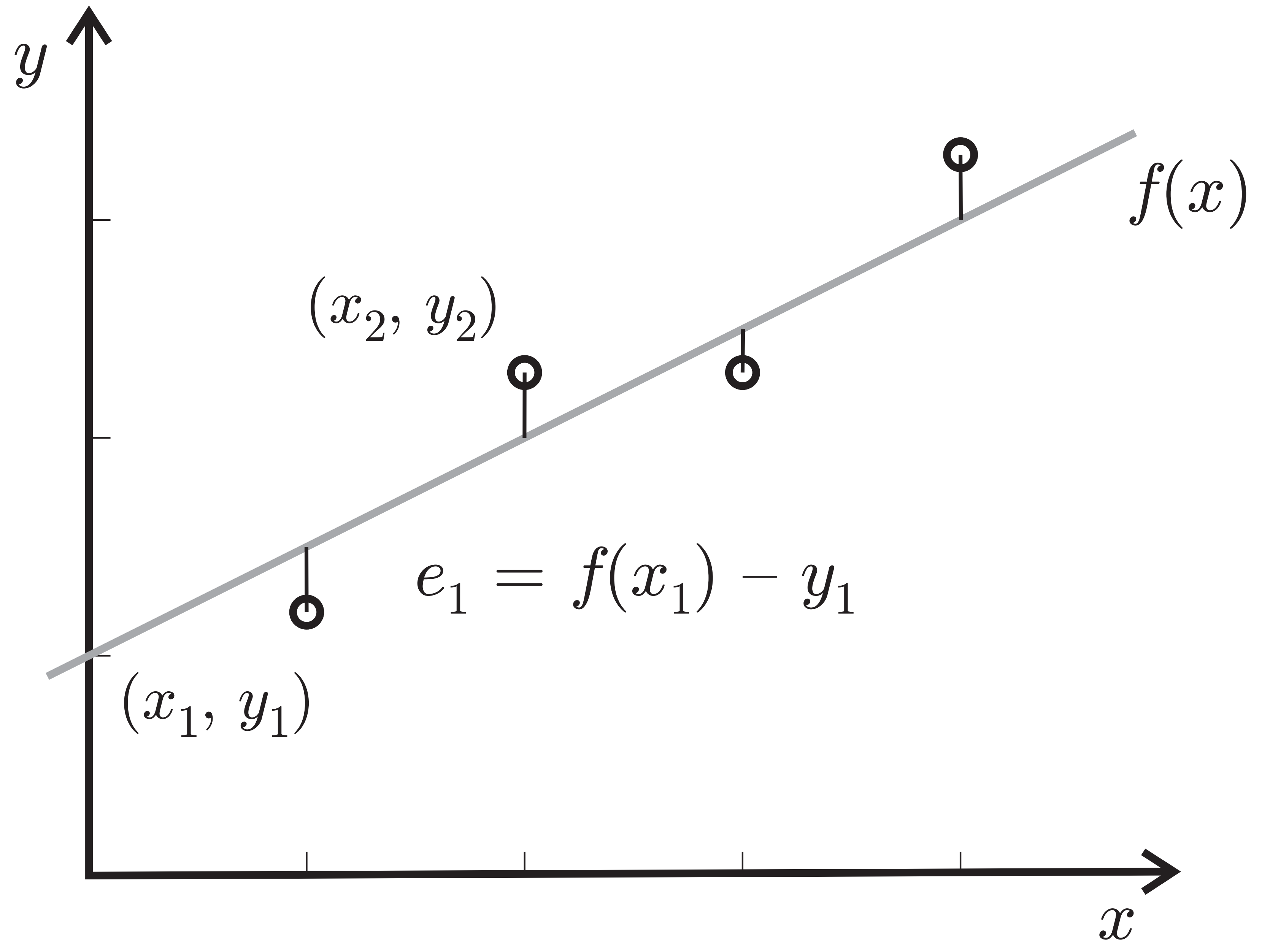


# Solving for the optimal function

**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 ...

# 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:** jrwright.info/mlbasics
  - This is the main source of information about the class
  - Slides, readings
- **Access-controlled course information:** eClass
  - Assignments, forum, video recordings, link to lecture meetings
  - **Discussion forum** for **public** questions about assignments, lecture material, etc.
- **Email:** james.wright@ualberta.ca for **private** questions (health problems, grades, etc.)
- **Lectures:** Tuesdays and Thursdays, 12:30-1:50pm on Google Meet
  - Lectures will be recorded and posted on eClass
- **Office hours:** immediately after lecture

# Teaching Assistants

Liam Peetpare: peetpare@ualberta.ca

Ehsan Ahmadi: eahmadi@ualberta.ca

- **Office hours:** twice per week (see eClass for times and Meet link)
    - Typically question/answer sessions
    - Please no arguing for marks
    - Sometimes pre-scheduled **tutorials**
- **No TA office hours this week**

# Readings

- Readings from Basics of ML textbook
  - Available on course site
  - It's a fast read
- See jrwright.info/mlbasics/schedule.htm for sections
- Optional readings listed on website also

# 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

# "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

# Grading

- 30%: Assignments
  - Mixture of mathematical problem sets and programming exercises
- 5%: Quiz on **October 8**
- 20%: Midterm exam on **October 29**
- 35%: Final exam (date TBD)
- 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
  - You may start the exam at any time during a **24 hour period**
  - Once you start you will have 2 hours (for midterm) or 6 hours (final)
  - Lecture will be cancelled on midterm and quiz dates

# 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

# Spot checks

- I won't be using a proctoring service for exams
- Instead, we will use **spot checks**
  - After every exam, some students will be selected to **verbally explain** their answers to a TA
  - If you can't explain how you got your answer, you may not get credit for the question

**Getting chosen for a spot check is not an accusation of cheating**

# Lectures

- Lectures take place on Google Meet
  - It's the **same URL** every time
  - URL is available on eClass
- Lectures will be **recorded**
  - Recordings on eClass
  - I won't make them public, because they will contain attendees' names
- Questions are encouraged!
  - In the text chat if you prefer

# 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, without any insight
- "Derive the maximum likelihood approach for a Gaussian."
  - i.e., an exercise question from a textbook
- "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?"

# Summary

- **Don't come to campus!**
- Course details at jrwright.info/mlbasics/ and on eClass: https://eclass.srv.ualberta.ca/course/view.php?id=64044
- This class is about **understanding** machine learning techniques by understanding their basic **mathematical underpinnings**
- Exams will be **spot checked** but not proctored
- Readings in free textbook, with associated thought questions
- No TA office hours this week

# AI Seminar

**What:** Great talks on cutting-edge AI research
External (e.g., DeepMind, IBM) and internal speakers

**When:** Fridays at noon

~~But come at 11:45 for free pizza / good seats~~

**Where:** ~~CSC 3-33~~ Online Zoom meeting
**Calendar:** www.cs.ualberta.ca/~ai/cal/

**Announcements:** Sign up for **ai-seminar**
www.mailman.srv.ualberta.ca/