

Basics of Machine Learning

CMPUT 296

University of Alberta

Winter 2020

Martha White



UNIVERSITY OF ALBERTA

What is machine learning?

- “The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience.” -Mitchell
- “...the subfield of AI concerned with programs that learn from experience.” -Russell & Norvig
- “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 or other outcomes of interest.” -Murphy

What is this course about?

- The world is full of information and data
- Much of that data is noisy and has an element of uncertainty
 - We have incomplete knowledge of the environment (partial observability)
 - Actions of other actors not provided
- Machine learning algorithms help us analyze that data
- **Goal:** understand machine learning algorithms by deriving them from the beginning
 - our focus will be on **prediction** (on new data)

A starting example

- Our goal: obtain a function to predict house prices, given age
 - $f(\text{age}) = \text{price of house}$
- Imagine you have a dataset of previous house sales this year, with attribute information Age and target Price
 - $(\text{age}_1, \text{price}_1), (\text{age}_2, \text{price}_2), \dots, (\text{age}_9, \text{price}_9)$
- Presumably these previous houses give us some information about the function between age and price
- Idea: if we can learn a function to accurately recreate these (age, price) pairs, then it could provide good predictions

Formalizing the problem

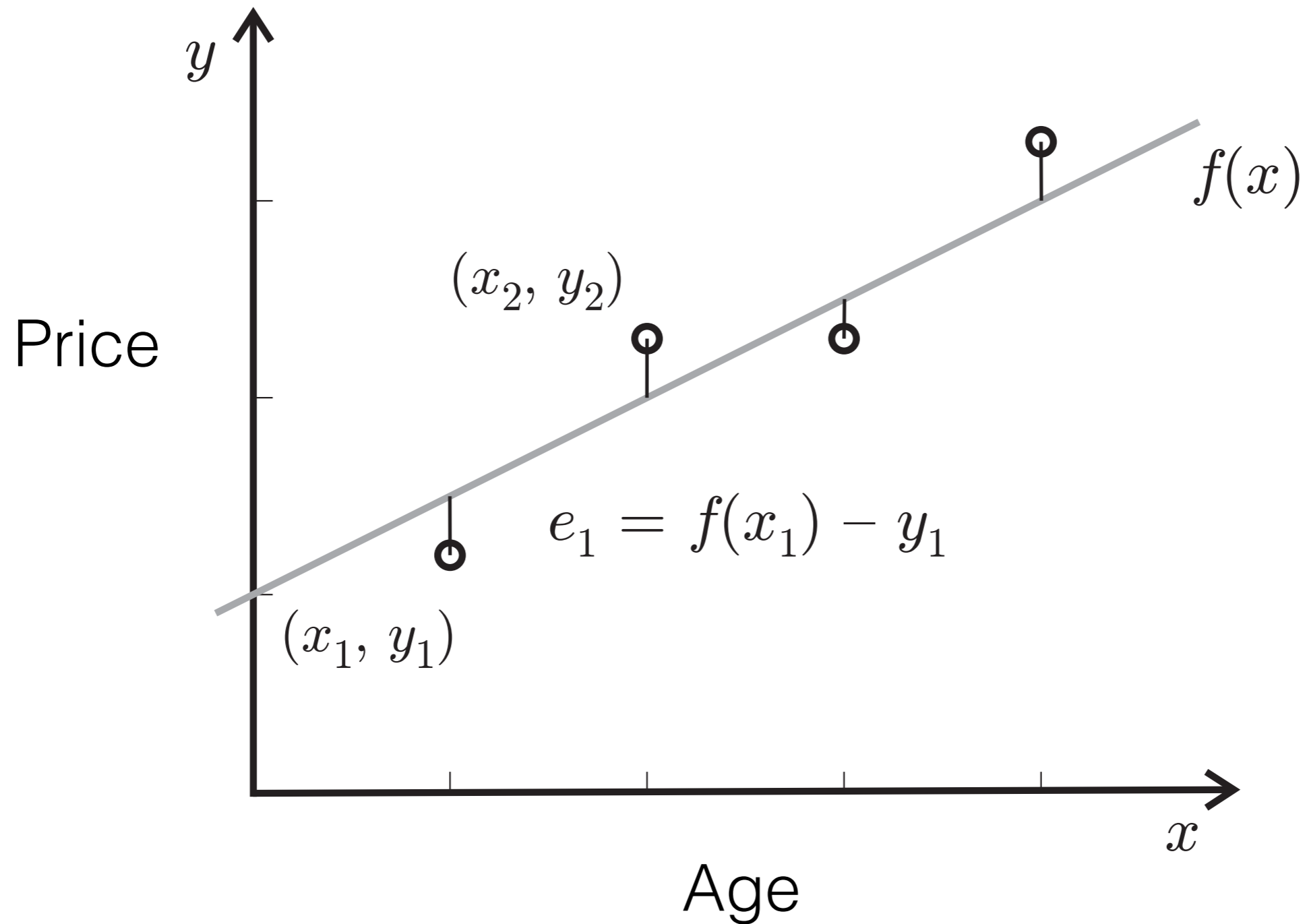
Let $x = \text{age}$ and $y = \text{price}$

Make difference between $f(x_i)$ and y_i small

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

Need to pick a form for f

Linear function f

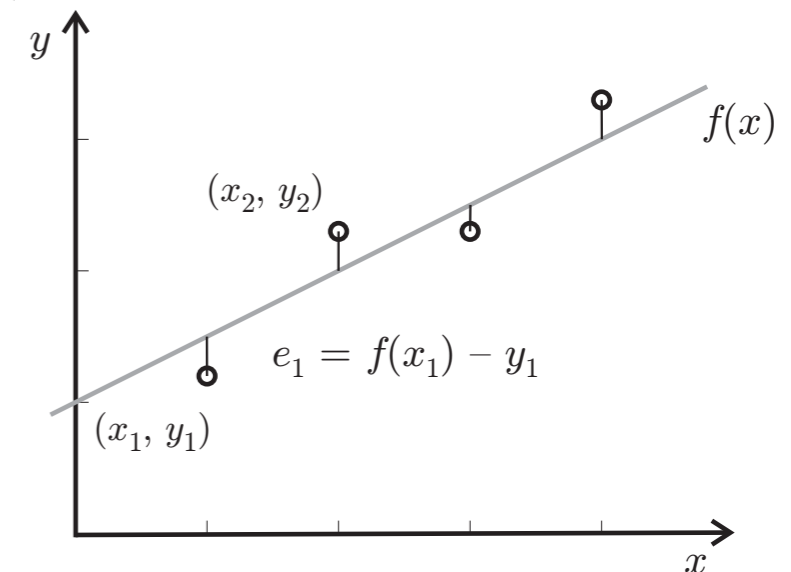


$$f(x) = w_0 + w_1 x$$

Solving for the optimal function

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

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



- We will see later how to solve this
- Questions:
 - Would you use this to predict the price of a house? Why or why not?
 - Will this predict well? How do we know?
 - If not, what is missing to make these assessments?

Probabilities to the rescue

- We can specify the problem using a **probabilistic approach**
- e.g., it is unlikely that there is deterministic function from age to price
 - many houses will have the same age but different prices
- Instead can specify the problem so there is a distribution over targets (price), given attributes about the item (age)
- Does this mean we think the world is stochastic, not deterministic?
 - stochasticity largely comes from partial observability
 - if knew age, size, number of rooms, if the queen lives there, etc., maybe the outcome is a deterministic price

Topics

- Background in Probability (Chapter 2)
- Introduction to Estimation, with Sample Averages (Chapter 3)
 - How to determine confidence in our estimates (concentration inequalities)
 - Sample Complexity and Convergence Rate
- Introduction to Optimization (Chapter 4)
- Parameter Estimation (Chapter 5)
 - MAP and Maximum Likelihood
 - Bayesian estimation
- Introduction to Prediction (Chapter 6)

Topics (cont...)

- Introduction to Prediction (Chapter 6)
 - Understanding our objective for prediction, formally
- Linear Regression (and Polynomial Regression) (Chapter 7)
- Generalization Error and Evaluating our Models (Chapter 8)
- Regularization and Constraining our Function Space (Chapter 9)
- Logistic Regression and Linear Classifiers (Chapter 10)
- Bayesian Linear Regression (Chapter 11)
- (Maybe) Missing Data and Semi-supervised Learning

Basic information

Class meets:

Time: TR 2:00pm – 3:20 pm

Place: CCIS 1 140

Instructor:

Martha White

Office: ATH 3-05

Email: whitem@ualberta.ca

Web: marthawhite.ca

Office Hours:

Time: T 3:20pm-5:00pm (is this a good time?) — No office hours this week

Place: ATH 3-05

Class Web Site:

GitHub pages: <https://marthawhite.github.io/mlbasics/>

Teaching Assistants

Archit Sakhadeo: sakhadeo@ualberta.ca

Ehsan Ahmadi: eahmadi@ualberta.ca

Hager Radi: radi@ualberta.ca

Homayoon Farrahi: farrahi@ualberta.ca

Kiarash Aghakasiri: aghakasi@ualberta.ca

Shibhansh Dohare: dohare@ualberta.ca

Vincent Liu: vliu1@ualberta.ca

Single lab on Wednesday, from 5-8. No Lab the first week

Lab/Office Hours:

- Labs will typically be question and answer sessions
- Might have some labs have tutorials, if needed/requested

Come to labs and ask questions!

Policy for Interacting with TAs

- Before meeting, ask a couple of clear questions over email. Sometimes that is enough to clarify the issue
- **Preferably, post your question on eClass.** Many others likely have the same questions.
 - TAs will answer questions within 24 hours
- If you do need to meet, you should meet with a TA for no more than 20 minutes. Come prepared so that you can use that time efficiently
 - I have instructed the TAs to turn you away if its longer than that
- No arguing for marks with TAs
 - should only ask: “Can you help me understand this question?”
 - Even if you get 10% improvement on an assignment, at best this translates into 1% for the final grade. It is a waste of your (and everyone’s) time, and is unfair.

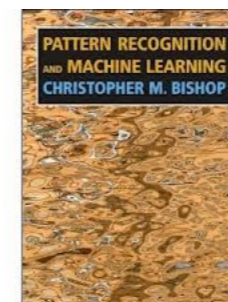
Textbook information

- **Main notes** provided on github course site
 - written by myself (about 100 pages), its a fast read (big margins)
- **[Optional] reference material/recommended readings:**
 - An Introduction to Statistical Learning, 2013 (accessible and concise)

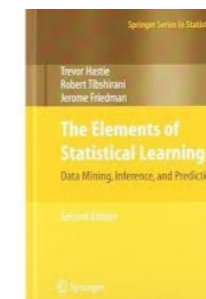
- Bayesian Reasoning and Machine Learning by D. Barber, Cambridge Press 2012.



- Pattern Recognition and Machine Learning by C. M. Bishop, Springer 2006.



- The Elements of Statistical Learning by T. Hastie, R. Tibshirani, and J. Friedman, 2009



My expectations

- Basic mathematical skills
 - some calculus
 - some probabilities
 - some linear algebra (mostly just vectors and dot products)
 - I will give crash courses in these along the way, so if you do not have this background, a willingness to learn these topics is a prerequisite
- You are hardworking and motivated to learn (machine learning)
- You are motivated to think beyond the material and ask open-ended questions

What to expect from this course

- I know you have expectations of me too
 - I will try to be transparent in marking and course choices
 - I am here to help you learn; I will treat you with respect and listen thoughtfully to your questions (I love to answer questions and give advice!)
 - Feel free to give me feedback (e.g., Miss Martha, you are talking too fast)
 - I have provided a mechanism for anonymous feedback; you can also talk to the TAs about any issues
- This course will be quite mathematical, with derivations of details
 - this is absolutely necessary, and will make you much more skilled in ML

“Why is there so much math?”

- “...I just want to use Machine Learning to solve Problem X”
- **Point 1:** Application of algorithms is simpler when you understand their development and underlying assumptions
 - and so you will be better at solving Problem X
- **Point 2:** Becoming comfortable with Math and Formalizing Problems is an important part of the CS program

CS is not about becoming an expert programmer

- CS is about problem solving, through the medium of computing
- You will learn somewhat to program, but our primary goal is to turn you into problem solvers. This means carefully designing solutions, which means:
 - Formalizing the problem
 - Understanding different potential approaches, to make informed choices
 - Evaluating your solution, in terms of correctness, efficiency, maintenance, etc.
- A part of being good at problem solving is being comfortable with mathematical concepts AND a logical thinker (clarity)

This course focuses on understanding, correctness and clarity

- Four Thought Questions, where you have to communicate your understanding clearly
- Exams are all short answer, with a mix of conceptual questions and derivation questions
 - You need to clearly communicate what you know
- I will grade you on understanding (regurgitation is largely useless)

Marks

- 30%: Assignments (3), a mixture of mathematical and programming exercises (code provided in python)
 - Must complete assignments independently
- 5%: Quiz, February 13, in class
- 20%: Midterm exam, March 12, in class
- 35%: Final exam, April 22, 9 a.m.
- 10%: Thought questions (4), show that you're reading and thinking about the material

Assignment grading

- Coarse binned grading:
 - 80 - 100 gets rounded up to 100 (yah!)
 - 60 - 80 gets rounded up to 80 (:D)
 - 40 - 60 gets rounded up to 60
 - **0 - 40 gets rounded down to 0**
- See policy listed in Assignment 1 (released on website)

More about assignments

- **All assignments are individual:**
 - The assignments are feasible on your own; you will learn much more and get more out of the course when you accomplish this
 - In-your-head-rule: you can discuss with classmates to some extent, but no writing down solutions on a piece of paper, or copying code
- Acknowledge sources (websites, books) in your documents
 - **DO NOT copy code** from the internet as your solution
 - If you find code, understand it and then write code yourself
 - Surprising fact: code online is not necessarily good
- Write legibly and upload scanned document (or use LaTeX)
- We are very serious about academic honesty; assignments are systematically tested for plagiarism and cheating

Three Exams

- Giving **Clear Answers** to Short Answer questions takes practice
- The first Quiz is only worth 5%, to give you a chance to make some mistakes and still improve
- Midterm already worth quite a lot (20%)
- And the Final even more (35%)
- I will also give practice questions
 - and go over an example of a wrong (but seemingly reasonable) answer

Goal of Assignments and Exams is different

- Assignments are for you to learn
- Exams are to evaluate you
 - which of course helps you learn
- Consequence: doing well on the assignments does not mean you will do well in the course
 - You need to do the assignments to learn the material, so not doing well on them likely means you will not do well
 - If you do poorly on the Quiz and Midterm, it is an important indicator for final performance in the course

Thought Questions

- University is about thinking and asking questions
 - a question does not have to have an answer, but it can be thought-provoking or make you think about a topic differently
- “Thought questions” correspond to (short) readings in the notes, and should demonstrate you’ve read and thought about the topics
- There are no stupid questions (so ask any questions in class/office hours/email), but “thought questions” are to demonstrate insight and so I have some requirements

General format for Thought Questions

- (1) First show/explain how you understand a concept
- (2) Given this context, propose a follow-up question
- (3) [Optional] Propose an answer to the question, or how you might find it
- **Suggestion:** framing a coherent, concise thought is a skill. When writing your thought question, ask yourself: is this clear? Read it again after writing it, with a critical eye

Example of a “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.

Examples of “Bad” Thought Questions

- I don't understand linear regression. Could you explain it again?
 - i.e. a request for me to explain something, 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 easily be answered from reading the definitions in the notes
 - But the following modification would be good: I understand that probability mass functions are for discrete random variables and probability density functions 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 types of functions?

More on grading

- This course is curved; I decide on the letter thresholds at the end
 - Note that I don't enjoy having to assign you a grade, but it helps you learn
- Constructive criticism is important: my marking will not be easy, because then you cannot learn where you are deficient
 - In the future, you might have a hard time finding someone to tell you when something is not good enough.
 - Your undergrad should be a time when you are given that gift (:D)
- A good indicator for final performance is performance on the exams, which are a large percentage of the grade. If you fail the midterm and final (less than 50% on both), then you will likely get an F in the course.

Appealing Marks

- **Policy: No arguing for marks**
- We try to mark consistently and fairly
- If you feel like you deserved partial marks, then there are others that feel that way too. We marked everyone with the same standard
- Giving marks to those who ask means we introduce unfairness
 - it is always possible to second-guess ourselves for subjective questions, and your mark might not feel fair, but its as fair as we can make it
- **No matter what, I will not adjust marks**
 - except for a silly error like the numbers do not add up
- I will reconsider letter grade boundaries at the end of the course

In-Class Time

- I will mainly be writing on transparencies (or on the whiteboard)
- Classes will be mathematical, primarily focused on derivations
- The Plan for each 80 minutes:
 - 25 minutes of me writing
 - 3 minutes where you discuss a question with a neighbour
 - 25 minutes of me writing
 - 6 minutes Think-Pair-Share, 1 minute to get a volunteered answer
 - 20 minutes of me writing

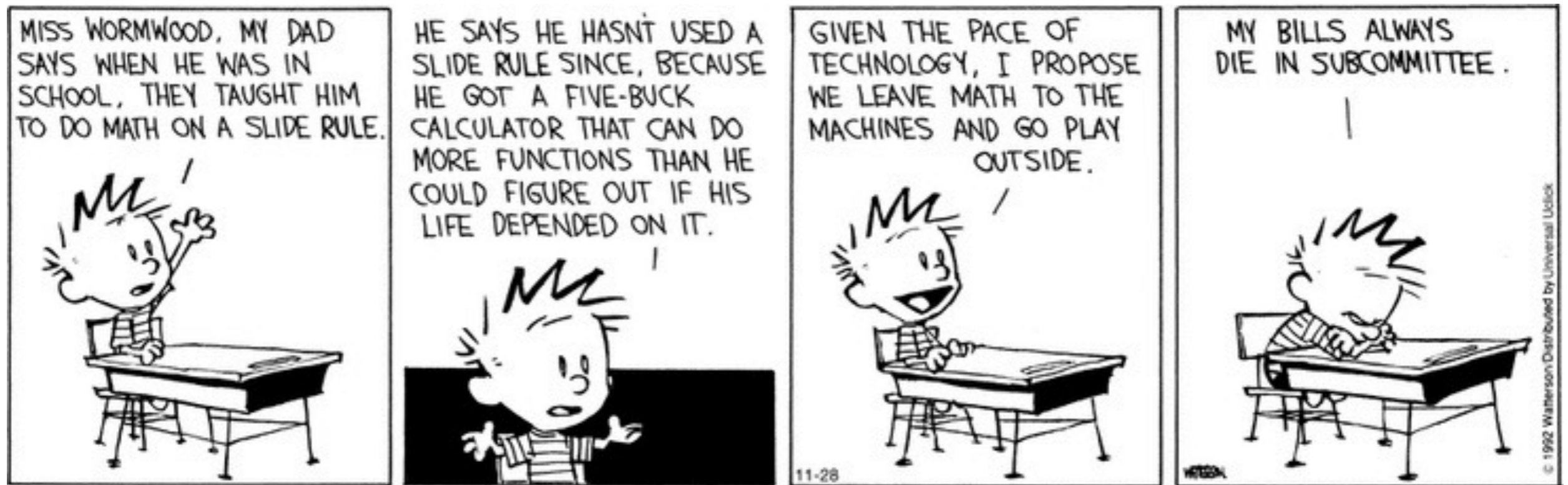
Asking Questions in Class Helps You Learn

- I want to encourage this as much as possible
- You can get 1% Bonus Marks (i.e., 1% to your final percentage) by asking **One** Question in class (what an easy mark!)
 - Email Myself (whitem@ualberta.ca) and Vincent Liu (vliu1@ualberta.ca) the day you asked the question, so we can add it to eClass
- You can ask a question raised by your neighbour: “I just discussed with Jill, and she had an interesting question. The question is...”
- **Note:** For any who choose not to discuss with their neighbour, I will call on you to share a question with the class

Think-Pair-Share

- Think of a problem where you think Machine Learning could be really useful
 - If you can't think of anything, try to think of a setting where prediction is useful
- Additionally think about:
 - What makes ML suitable for it?
 - What might be difficult about using ML for that problem?
- After 3 minutes, I will ask you to discuss with your neighbour
 - Take turns telling the other about your setting

Next: Crash Course in Probability



- Probabilities underly much of machine learning
 - enable precise modeling of uncertainty