Intro

Three goals for this course:

- Develop data science and machine learning toolkit - process and intuition behind how these tools came into being
- Develop programming process. The difference between a good programmer and a great programmer is really in their workflow - great programmers have a better workflow that enables them to produce better code.
- Learn to think critically about ethical challenges with data.

FATE: fairness, accountability, transparency, and ethics.

*Review of the syllabus

What does it mean to actually learn something? From a scientific perspective, what does it mean to have knowledge? What we'll see is that knowledge takes the form of anticipatable predictions.

Imagine that you're presented with a metal plate. On the right side of the plate is a heater, which is heating up the metal plate. You notice overtime that the metal plate is warming up, and we'd like to understand how the metal plate heats up over time.

To do this, we split up the metal plate into grid sections. At each (x, y) coordinate, we'll record the temperature at each time t.

If we plot this, we'll have a 3D plot, with the (x, y) grid on the x and y dimensions, and the height z is the temperature. Each plot is at a single timestamp, for example, t_0. However, ideally, we'd like a 4D plot, so that we can view time as part of this as well. We collect all of this data, and we then want to model how the temp fluctuates over time and as a position on the grid.

Effectively, we are looking for a function $f(x, y, t) => temperature$ that models the behavior in the metal plate. From a scientific perspective, what is the knowledge contained in this model? Well, it's the anticipation of how the heat will propagate through the metal plate.

Another way of thinking of this is that the model explains in hindsight how the heat propagated. Well, that's a bit ambiguous - since it's in hindsight, how can we differentiate the model from anything else?

Let's suppose that instead we are again presented with another metal plate, and we still have the heater on the right side. However, now assume that the heat propagates from the left side first, not the right. Clearly, that doesn't make sense with the model we developed previoulsly. Ok, so then what now? This new manner of heat propagation falsifies our model in many ways.

Well, the scientific method is to examine the set of events that led to the current situation - our model expected a certain set of events, and if those events don't occur, then our model is in trouble. How then, can we have a deterministic prediction of the outcome?

This is the scientific approach - need to consider the sequence of events to develop a sense of anticipation. If, however, we are equally good at explaining every possible event, then we don't really have "knowledge." That would be magic.

In order for a theory to contain knowledge, it necessarily needs to be falsifiable. And therefore, in order to try and confirm a theory, you need to be willing to try and falsify it.

Let's consider a different example: suppose we try to play a game. The professor has a rule that governs triples of integers. The goal is for the students to guess what the rule is. The only way that they can get insight to the rule, is by submitting examples or guesses. The professor will then respond truthfully indicating whether or not the guess is correct.

to start, the professor gives us an initial data point: (2,4,6).

- Participant one submits: (2,4,3) -> No
- (6,8,10) -> yes
- (10, 12, 14) -> yes

Then, participant 1 says I know the rule, writes it down, and submits it. Professor created a poll where we had to submit responses we would have liked to test. Majority response was (-6, -4, -2)

One possible hypothsis for this is $h = (x, x+2, x+4)$

This is kind of what data science is - we only have a certain number of examples, and we are trying to discover the rule from just those guesses. We can only do as much with the data that we have been given - there are some inherit limitations, which we'll address in more detail later in the course.

Some limitations from a high level:

1. Not all examples contribute equally to learning.
2. A set of examples are not always representative of the rule.
3. There are possibly infinitely many rules that can go through a finite number of examples.
    1. To address this issue, we usually rely on occam's razor: the simplest explanation is probably the right one - how can we choose amongst an infinite set of functions that pass through our data points? Well, the simpler the better, and that should generally drive our data science workload.

Two definitions:

- Given hypothesis h, positive examples are examples that expect a positive result.
- Given hypothesis h, negative examples are examples that expect a negative result.

Another poll presented on which of the above examples are positive and negative. Example 1 was negative, 2&3 positive.

There are many other possible examples, such as $(x, 2x, 3x)$ In this case, our positive negative examples change!

Imagine that the set of examples is a rectangle, the true rule is an area inside that, and our hypothesis is a circle inside the true rule. In this case, there is an entire area where our examples may not indicate that our hypothesis is not equal to the true rule. If we never hit that set, we just confirm our bias.

The rule that the Professor was actually thinking is $(a < b < c)$.

This phenomenon is called confirmation bias: that we often fail to find negative examples that negate our own hypothesis. Something like only 20% of people successfully manage to find negative examples that remove them from confirmation bias.

Phrase to remember: garbage in, garbage out.

Data science is a very powerful set of techniques, that generally work extremely well, that lead to all sorts of philosophical questions. We'll have to navigate these questions as future data scientists.