SEPT 8 LECTURE

INTRO TO COURSE

> Goals for the Course →

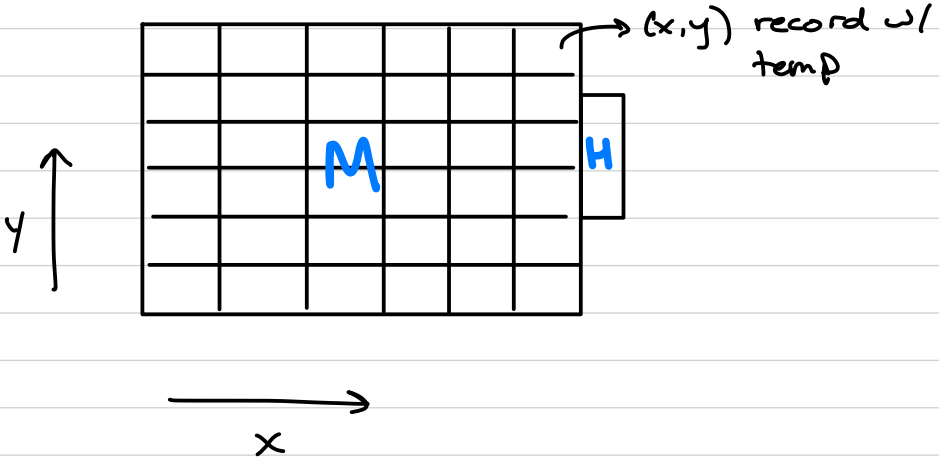
- learn ML + data science tools
- improve our programming skills and workflow
- learn about ethical data

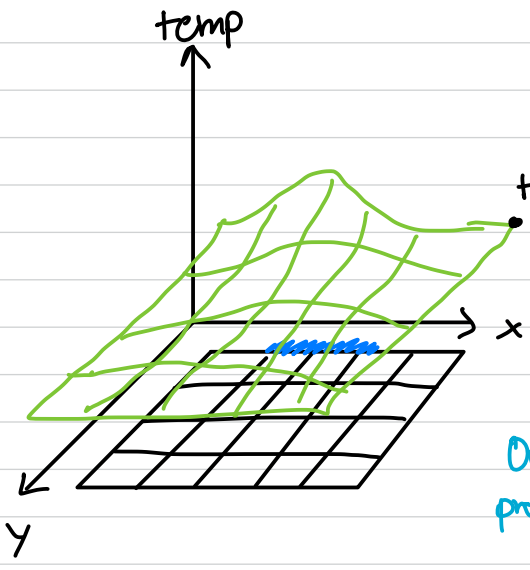
> Syllabus →

- All assignments posted already
- Group project topics provided through Spark

> What is Data Science? →

A plate w/ a hot plate on the side





$f(x, y, t) \Rightarrow$ temperature

Consider how we can use the model to predict the temperature.

Our model describes how the heat propagated through the plate.

If we are equally good at explaining every outcome, then we have no knowledge. That will not help predict anything.

The Game

(a, b, c)
(2, 4, 6)

Try to uncover the rule by submitting examples and the ruler setter will truthfully if the example fits the rule.

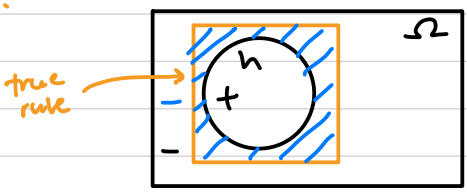
Participant A

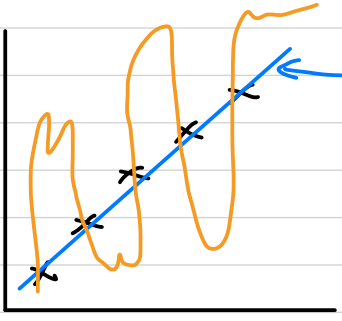
(2, 4, 3)	N
(10, 12, 14)	Y
(5, 7, 9)	Y

rule : $(a < b < c)$

Not all examples give the same amount of knowledge.

TRY NEGATIVE EXAMPLES



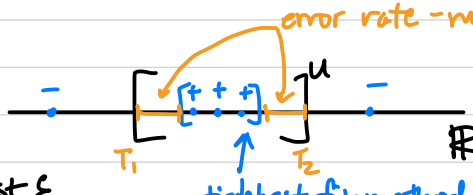


Okam razor philosophy?
Simpler models > complex

Try to minimize confirmation/positive bias

» Comp Learning Theory →

- What is learnable?
- How many examples does it take to learn something



Define $T_1 + T_2$ s.t. being in $T_1 + T_2$ at most ε

$P(T_1 \cup T_2) < \varepsilon$

To be reasonably close
u^- has an error rate of at most ε w/ prob. p

GIT

- GH is the interface/browser that lets us upload to Git repositories online and have version control.

- Git is the version control sys.

> Why Use It →

1. Progress loss minimized by uploading to the cloud
2. Ease of iterating through diff versions of code
3. Collab is productive