Data 102 Fall 2022 Lecture 1

Data, Inference, and Decisions:

what does that mean?

Your Instructors

Jacob Steinhardt

Ramesh Sridharan

This Course Has Two Big Ideas:

- Making Decisions Under Uncertainty

- Modeling in the Real World: Assumptions & Robustness

Big Ideas

- Making Decisions Under Uncertainty
- Modeling in the Real World: Assumptions & Robustness

Course Topics

- Repeated binary decision-making
- Causal Inference
- Bayesian & frequentist modeling
- Prediction: regression & nonparametrics
- Quantifying uncertainty (intervals and more)
- Interpretability
- Concentration inequalities
- Sequential decisions w/feedback
- Matching Markets
- Robustness
- Privacy

Logistics

- Everything you’ll need to know will be on the course website or Ed

data102.org/fa22

Problem setup: what are we trying to do?

1. We observe data: x, y
2. We want to understand hidden (unknown) state of the world: θ

Data: x	**Data: y**	**Unknown:** θ
-	Heights in a sample	Average population height
-	Video from a car camera/sensor	What objects/people are near the car?
Patient medical records	Patient health outcomes	Prediction formula for health outcomes
Phone usage (survey)	Happiness (survey)	How much does phone usage *cause* happiness to increase/decrease?

Assumptions: Bayesian/Frequentist and (Non)parametric

- Bayesian vs Frequentist
  - Frequentist: data (y) are random, unknowns (θ) are ***fixed***
  - Bayesian: data (y) are random*, unknowns (θ) are ***random***
  - Sounds simple, but has huge consequences!

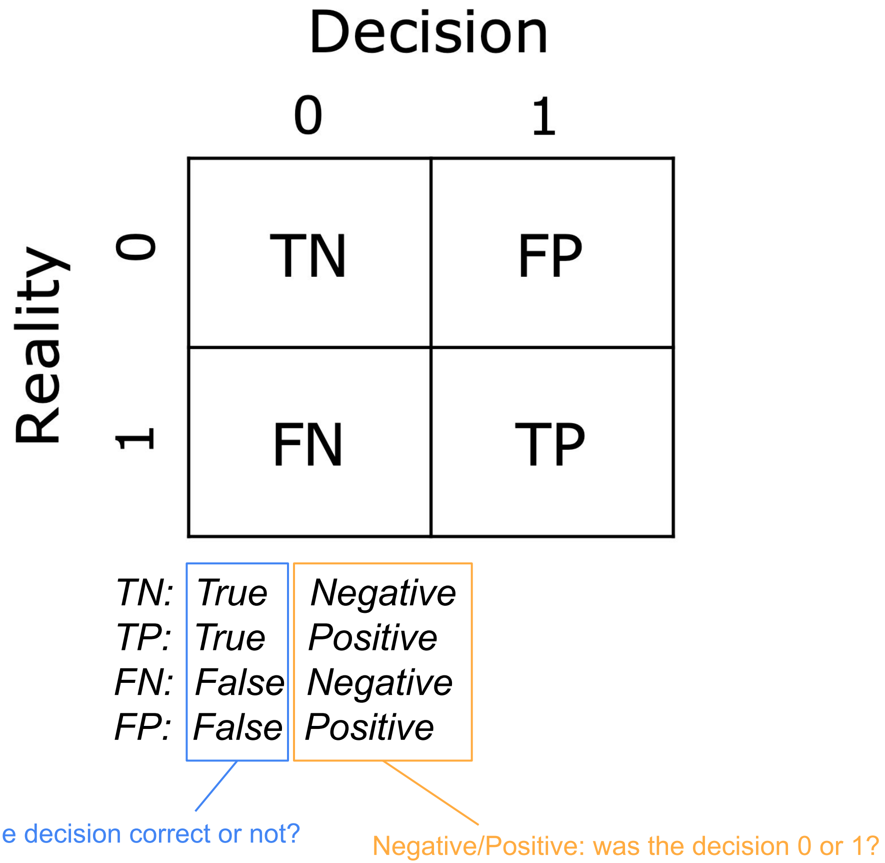
- Parametric vs Nonparametric
  - Parametric
    - Make assumptions about relationship between unknowns (θ) and data (y)
    - Use assumptions to find θ from y
  - Nonparametric:
    - Don’t bother with assumptions
    - Find any good function f so that θ = f(y)
    - (there’s another definition we’ll talk about later in the semester too)

Binary Decision Making

- The simplest kind of decision: yes or no (0 or 1)
- Setup
  - Reality is 0 or 1
  - We observe noisy data, and use that to make a decision (our best guess for reality)
  - Our decision is 0 or 1
- Examples
  - COVID testing
  - Fraud detection
  - Predicting recidivism (will someone commit another crime?)
  - Detecting underground oil wells
  - Movie/TV recommendations

Binary Decision Making

- Examples
  - COVID testing
  - Fraud detection
  - Predicting recidivism
  - Detecting an underground oil well
  - Movie/TV recommendations



Decision

Reality	0	1
0	TN	FP
1	FN	TP

TN: True Negative
TP: True Positive
FN: False Negative
FP: False Positive

True/False: was the decision correct or not?

Negative/Positive: was the decision 0 or 1?

Multiple Decisions

Single decision

		Decision	
		0	1
Reality	0	TN	FP
	1	FN	TP

Multiple decisions

		Decision	
		0	1
Reality	0	n_{00}	n_{01}
	1	n_{10}	n_{11}

Multiple Decisions

- We usually don’t know “Reality”

- In real-world scenarios, we also need to make more than one decision

- Next: strategies and theory around how to make those decisions intelligently

“Row-wise” rates: *what if we knew reality?*

- TNR: $$\frac{n_{00}}{n_{00} + n_{01}}$$ *specificity*
- FPR: $$\frac{n_{01}}{n_{00} + n_{01}}$$
- TPR: $$\frac{n_{11}}{n_{10} + n_{11}}$$ *sensitivity* *recall*
- FNR: $$\frac{n_{10}}{n_{10} + n_{11}}$$

		Decision	
		0	1
Reality	0	n_{00}	n_{01}
	1	n_{10}	n_{11}

Wikipedia: Sensitivity and Specificity

A column-wise rate: what if we made a “1” decision?

		Decision	
		0	1
Reality	0	n_{00}	n_{01}
	1	n_{10}	n_{11}

False discovery proportion (FDP):

$$\frac{n_{01}}{n_{01} + n_{11}}$$

P(R = 0 | D = 1)