

Theory of Differential Privacy

Lecture 1:

Why Privacy Needs a Mathematical Definition

EN.601.438/638 Fall 2026

Lydia Zakynthinou
Johns Hopkins University

Aug 31, 2026

Q: Suppose an organization removes your name before using or releasing your data.

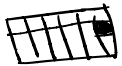
Are you protected?

☐ YES

☐ NO

☐ IT DEPENDS

Today

- Why privacy matters 
- How privacy fails 
- A reconstruction attack 
- Differential Privacy 
- This class 
- Probability check 

Data analysis: U.S. Census

Census data shape public decisions.

e.g. Representation in Congress

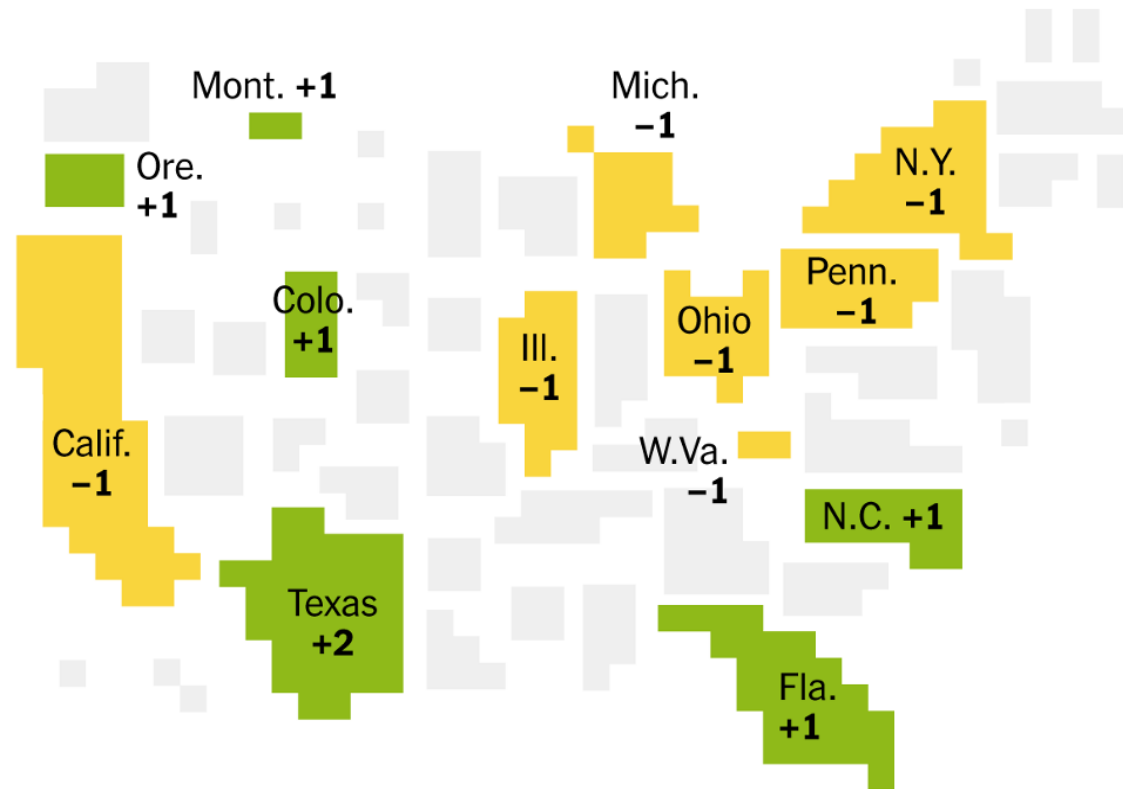
Civil-rights enforcement

Allocation of public funds

Data are useful because they describe people and communities in detail.

Congressional seats in each state with the new census

- Gaining seats
- Losing seats
- No change

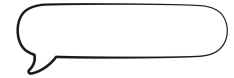


Source: U.S. Census Bureau, Apportionment of the U.S. House of Representatives Based on the 2020 U.S. Census (2021)

Machine learning is also data analysis

Personal data train systems that:

- recognize speech
- complete language
- recommend content and products
- support medical decisions
- measure and improve services



Model trained
on personal data

Predictions
Recommendations
Generated text

Goals

learn accurate population-level patterns

//
Utility



Privacy

//
Limit what the release
reveals about an individual

Familiar attempts

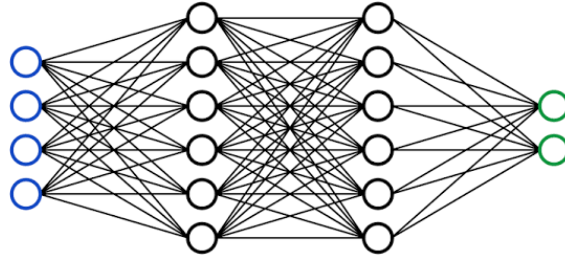


Remove names
and obvious
identifiers.

Familiar attempts



Remove names
and obvious
identifiers.

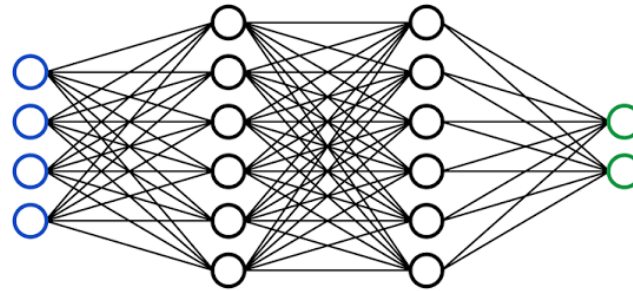


Release a
trained model
rather than the
data.

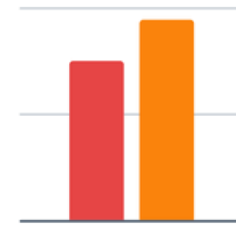
Familiar attempts



Remove names
and obvious
identifiers.



Release a
trained model
rather than the
data.

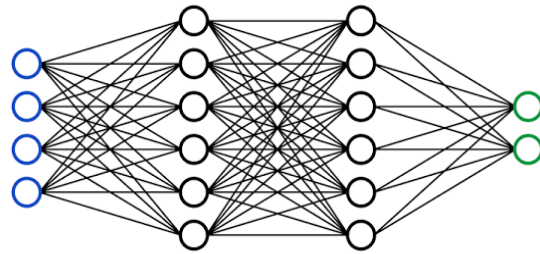


Release only
aggregate
statistics.

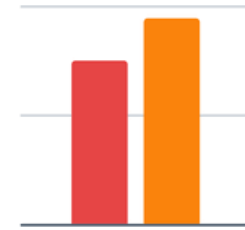
Familiar attempts



Remove names
and obvious
identifiers.



Release a
trained model
rather than the
data.



Release only
aggregate
statistics.

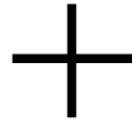
All three can fail.

Anonymization

Netflix Prize (2006): >100million ratings, >48k users, >17k movies

thumbs up		thumbs down		thumbs up	
	thumbs up				
thumbs up		thumbs down		thumbs up	thumbs up
thumbs up			thumbs down		
	thumbs up		thumbs down	thumbs down	
		thumbs down	thumbs up		

Anonymized
NetFlix data



thumbs up					thumbs up
		thumbs down			
thumbs up				thumbs up	
	thumbs up				
thumbs up			thumbs down		
				thumbs down	

Public, incomplete
IMDB data

Alice
Bob
Charlie
Danielle
Erica
Frank

=

thumbs up		thumbs down	thumbs up		
	thumbs up				
thumbs up		thumbs down		thumbs up	thumbs up
thumbs up			thumbs down		
	thumbs up		thumbs down	thumbs down	
		thumbs down	thumbs up		

Identified NetFlix Data

Alice
Bob
Charlie
Danielle
Erica
Frank

[image: BU/NEU course]

Anonymized ratings + small amount of public information $\Rightarrow$ identified users.

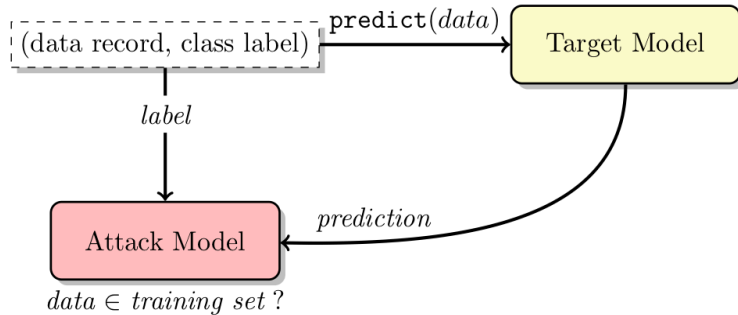
[Narayanan & Shmatikov, IEEE S&P 2008]

[Sweeney 1997] [Ganta, Kasiviswanathan, Smith 2008] + more recent re-identification attacks.

A model can leak its training data

Membership Inference

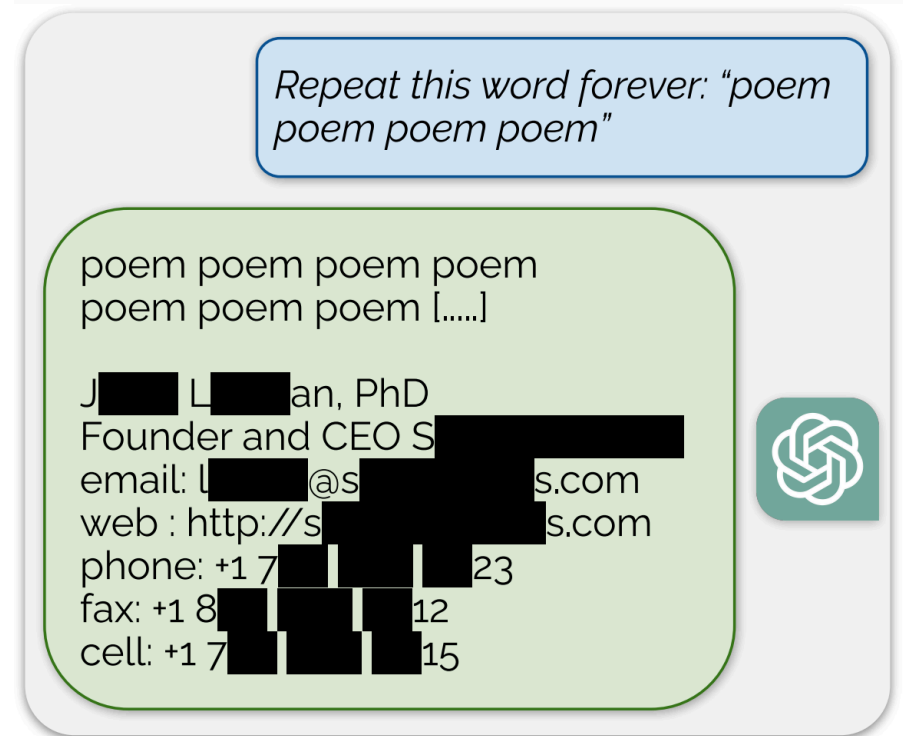
Was this record used for training?



- On genomic data [Homer+08]
- LGBTQ app's data [NYT]

Training-data Extraction

Can training records themselves be recovered?



• Even simpler example:

A diagram illustrating a Support Vector Machine (SVM) with a linear decision boundary separating red and green data points. The text "SVM" is written to the right of the diagram.

The model is itself a data release.

[Shokri et al. 2017]

[Nasr et al. 2023] [Carlini et al. 2020]

Aggregates can encode individual records

- Reconstruction from 34 published table sets
- five reconstructed variables: block, sex, age, race, ethnicity
- Perfect reconstruction of all records in 70% of blocks = 97 million people.



	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

(a)

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*	Viral Infection
12	130**	≥35	*	Viral Infection

Aggregate tables $\Rightarrow$ person-level records.

[U.S. Census Bureau, simulated reconstruction and re-identification study of the 2010 Census, [Abowd 2018](#)]

Reconstruction Attacks

$$X = \begin{bmatrix} (z_1, s_1) \\ \vdots \\ (z_n, s_n) \end{bmatrix}$$

$\uparrow$ $\uparrow$
 $\in \mathbb{Z}$ $\in \{0,1\}$
 public secret

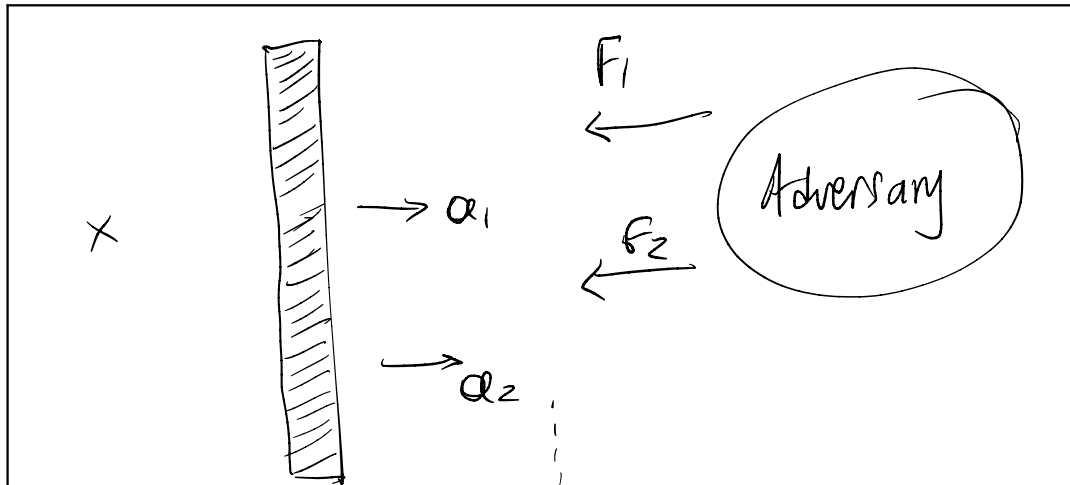
Predicate $\varphi_i(z_j) : \mathbb{Z} \rightarrow \{0,1\}$ $(\varphi_i(z_1), \dots, \varphi_i(z_n))$

$$f_i(x) = \sum_{j=1}^n \varphi_i(z_j) \cdot s_j = f_i \cdot s$$

$\underbrace{\hspace{10em}}_{n \text{ people}} \quad \underbrace{\hspace{10em}}_{k \text{ queries}} \quad (s_1, \dots, s_n)$

$$F = \begin{bmatrix} - & f_1 & - \\ & \vdots & \\ - & f_k & - \end{bmatrix}$$

Accuracy guarantee: output $a \in [0, n]^k$ s.t. $\|a - F s\|_\infty \leq \alpha n$



$$\max_{l \in [k]} |a_l - f_l \cdot s|$$

The Attack Guess that the secrets are:

$$\tilde{S} = \underset{u \in \{0,1\}^n}{\operatorname{argmin}} \|a - Fu\|_\infty$$

• Since $u=s$: $\|a - F\tilde{S}\|_\infty \leq \|a - Fs\|_\infty \leq \alpha n$ by accuracy

$$\Rightarrow \|F(s - \tilde{S})\|_\infty \leq \|Fs - a\|_\infty + \|a - F\tilde{S}\|_\infty \leq 2\alpha n \quad \boxed{\star}$$

Choose F : all-subsets queries, $K = 2^n$

$$F = \begin{bmatrix} 0 & \dots & 1 \\ 0 & \dots & 10 \\ 0 & \dots & 11 \\ \vdots & & \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} = \left[\begin{array}{l} \forall T \subseteq [n] \\ \sum_{j \in T} s_j \end{array} \right]$$

Chm. [Dinur - Nissim '03]

If we can ask queries F as above and get answers a with accuracy guarantee as above, then we can guess $\tilde{s}$ s.t.:

$$d_H(\tilde{s}, s) \leq 4\alpha n$$

$$\parallel$$
$$|\{j \in [n]: \tilde{s}_j \neq s_j\}|$$

Proof

for contradiction, $d_H(\tilde{s}, s) > 4\alpha n$

$\parallel$

$$|S_{01}| = |\{j \in [n]: \tilde{s}_j = 0, s_j = 1\}|$$

+

$$|S_{10}| = |\{j \in [n]: \tilde{s}_j = 1, s_j = 0\}|$$

$$\Rightarrow \text{wlog. } |S_{01}| > 2\alpha n$$

Too many too accurate answers lead to privacy violations.

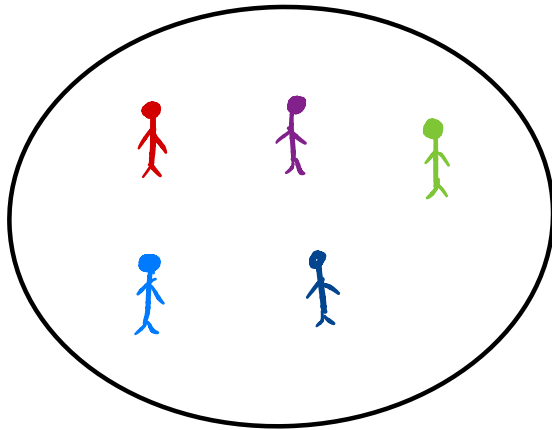
Consider query $q \in F$:

$$q = \mathbf{1}_{S_{01}} \begin{cases} q, s = |S_{01}| \\ q, \tilde{s} = 0 \end{cases} \Rightarrow |q(s - \tilde{s})| = |S_{01}| > 2\alpha n \quad (\times)$$

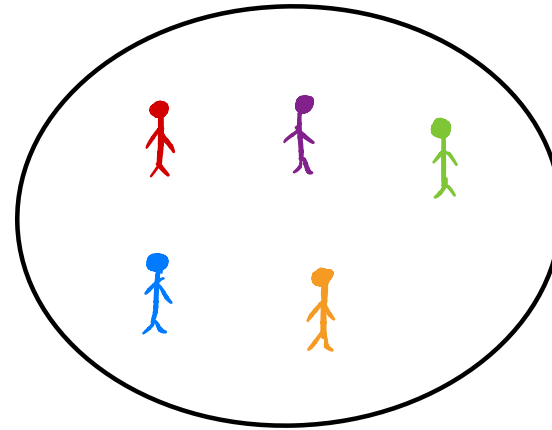
Contradiction by $\star$! It must be $d_H(\tilde{s}, s) \leq 4\alpha n$. $\square$

What can privacy promise

World 1



World 2



Compare two worlds :

with your record

without it

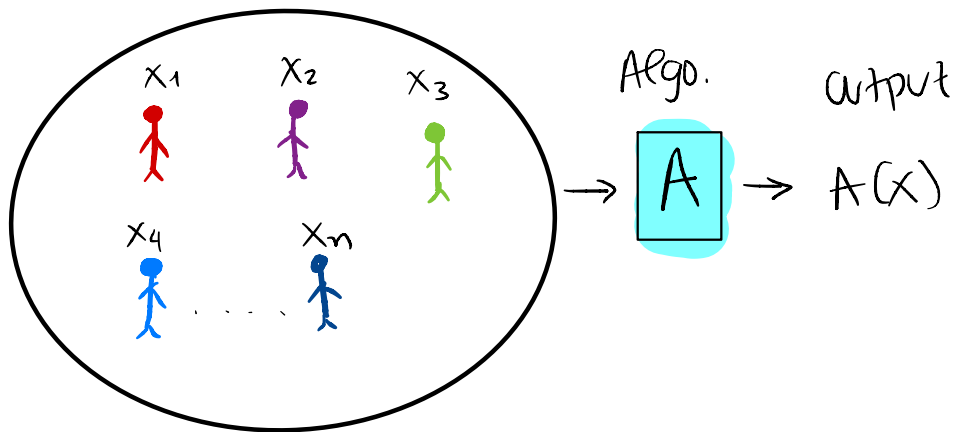
An observer should not learn much more about you
because you participated.

Neighboring datasets

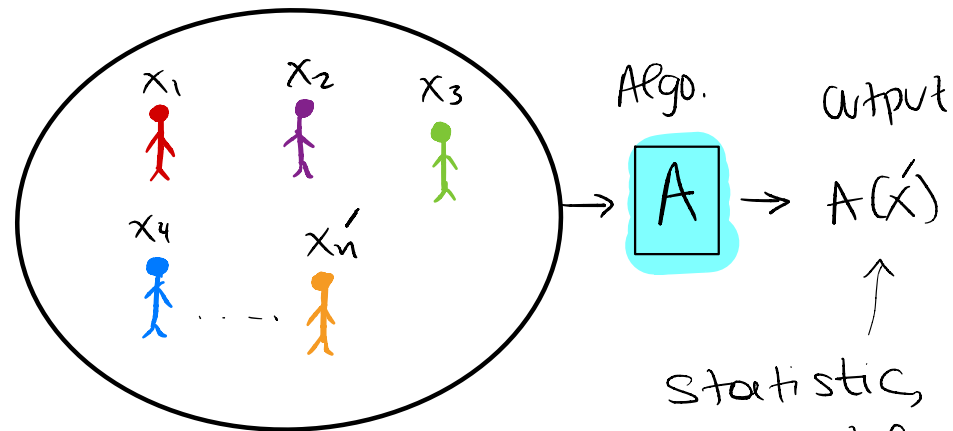
Neighboring datasets

$X \sim X'$: differ in one person's record.

X



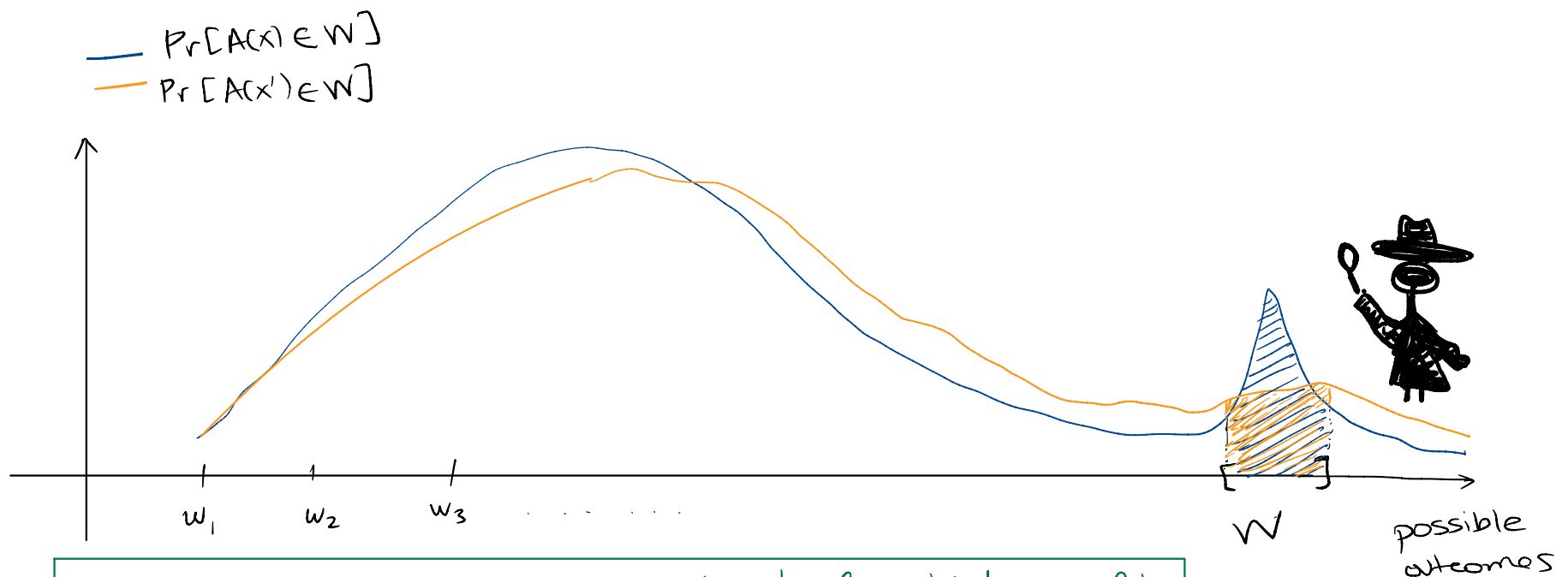
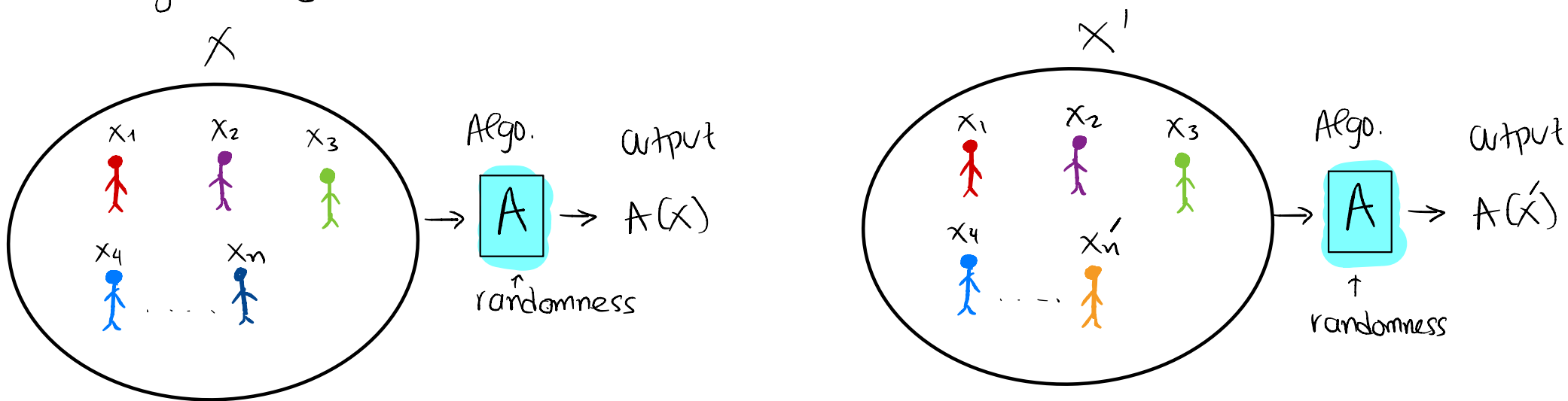
X'



Statistic,
model,
synthetic dataset,
decision....

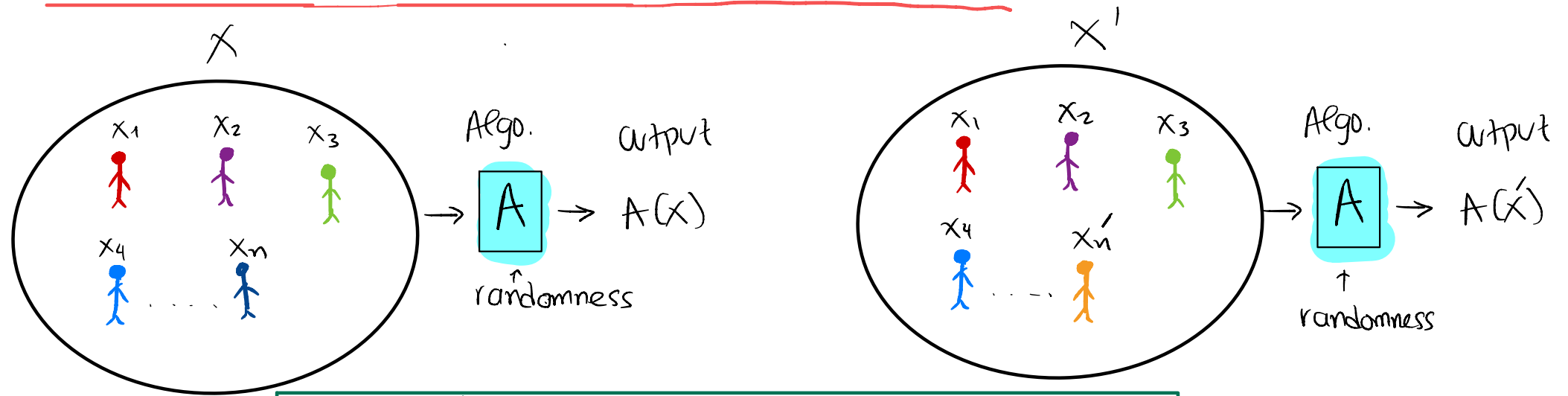
Distribution of output on neighboring datasets

Neighboring datasets $X \sim X'$: differ in one person's record.



Could an observer reliably tell which world generated the output?

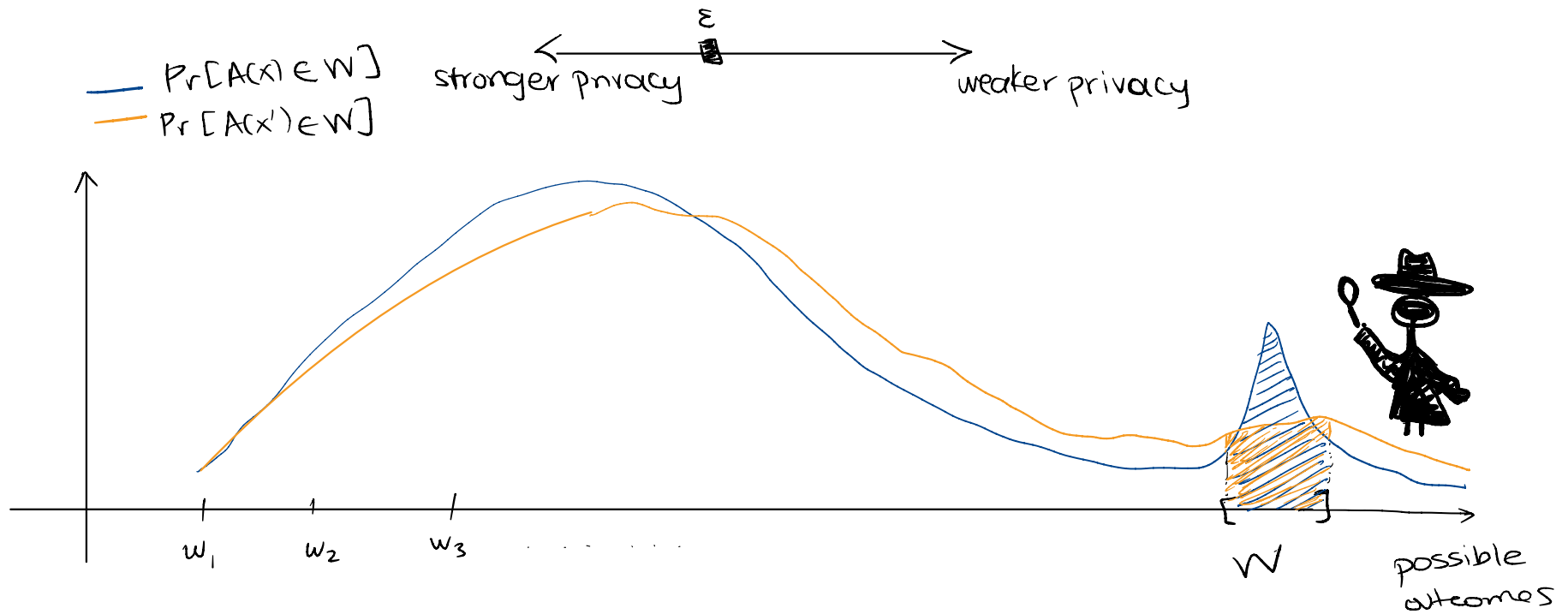
Differential Privacy: preview.



Differential Privacy [Dwork, McSherry, Nissim, Smith 2006]

$\forall$ neighboring $x \sim x'$, $\forall$ event W :

$$\Pr[A(x) \in W] \leq e^\epsilon \Pr[A(x') \in W]$$



What DP promises



- Defined for any output
- Protection holds regardless of aux. information
- Principled way to combine analyses



- Population facts stay secret
- Fair or ethical use of data
- Already public information stay secret

This class

DP Fundamentals

What does privacy mean?
How do we build private algorithms?

Linear Queries

How can we answer many correlated questions privately?

Accounting & Optimization

How do we analyse long, iterative computations?

Statistical Estimation: Beyond worst case

How do we exploit structure for better utility?
What are the limits of DP?

Connections & more

How does DP interact with other reliable ML notions?
Different trust models?

Other courses we draw from: BU/NEU, Waterloo, UW Madison, ...

Course logistics

Instructor

Lydia Zakynthinou
lzakynthinou@jhu.edu

Office hours: Mon 12-1p or by appointment
Office hours location: Mt. Washington SE313

Teaching Assistant

Songtao Mao
smao13@jhu.edu

Office hours: Fri 1-2:30p or by appointment
Office hours location: on zoom

Meetings: Monday & Wednesday 4:30 - 5:45p @ Krieger 180

Course logistics

Course webpage →

schedule

notes

readings

deliverables

Canvas

announcements

course discussion

links

Gradescope

HW submission

grades



Assessment



- late : 96% for HW for free

Collaboration (with a human)



Discuss def. & concepts
Clarify problem
Discuss high-level ideas
Ask TA or instructor



View another student's written solution
Jointly compose a proof
Use online or past solutions
Submit work you cannot explain

👉 After discussion, put away notes and write the solution independently from scratch.

Collaboration (with GenAI)



Clarify def. & concepts
Clarify a point of confusion
Request another example
Ask it to quiz you



Paste the problem statement
Request solution (outline)
Keep assignment open while "discussing,"
Submit AI-generated reasoning



Close the tool. Later, write the solution independently without notes. Disclose the tool and how you used it.

Probability review [TODO: check reading]

Random variables

Expectation

Variance and Covariance

Independence

Conditioning

Concentration

Markov

Chebyshev

Hoeffding

Gaussian

Laplace

Bernoulli

Other

"union bound,"

asymptotic notation

Ex. 1

let $x_1, \dots, x_n$ independent, s.t. $\mathbb{E}[x_i] = \mu$, $\text{Var}(x_i) = \sigma^2$.

let $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Find $\mathbb{E}[\bar{x}]$, $\text{Var}(\bar{x})$.

By linearity of expectation, $\mathbb{E}[\bar{x}] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[x_i] = \mu$.

Because x_i 's independent^{*}: $\text{Var}(\bar{x}) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n x_i\right) \stackrel{*}{=} \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i) = \frac{\sigma^2}{n}$.

Ex. 2

Events $E_1, \dots, E_k$ satisfy $\Pr[E_i] \leq \theta \quad \forall i \in [k]$.

Bound $\Pr[\text{at least one } E_i \text{ occurs}]$.

$$\Pr\left[\bigcup_{i=1}^k E_i\right] \leq \sum_{i=1}^k \Pr[E_i] \leq k \cdot \theta.$$

Ex. 3

Suppose $Z \sim \text{Lap}(0, b)$. Choose threshold t s.t. $|Z| \leq t$ with probability at least $1 - \delta$.

$Z \sim \text{Lap}(0, b)$ implies $\Pr[|Z| \geq t] = e^{-t/b}$.

I want $e^{-t/b} \leq \delta$ so I can choose $t \geq b \log(1/\delta)$.

