

LECTURE 13. 05/09

Last Class

1. PCA
2. Computing SVD
3. Application

Today

1. Privacy in ML
2. Why?
3. What?
4. How...?

Privacy means "Proxy of Privacy".

Sensitive Datasets : → Medical Records

→ Genetic data

→ Search Logs

→

NAME	EMAIL	SEARCH SENTENCE	LOCATION	DATE
.	.			

AOL: Released such a database.

→ NY TaxiCab

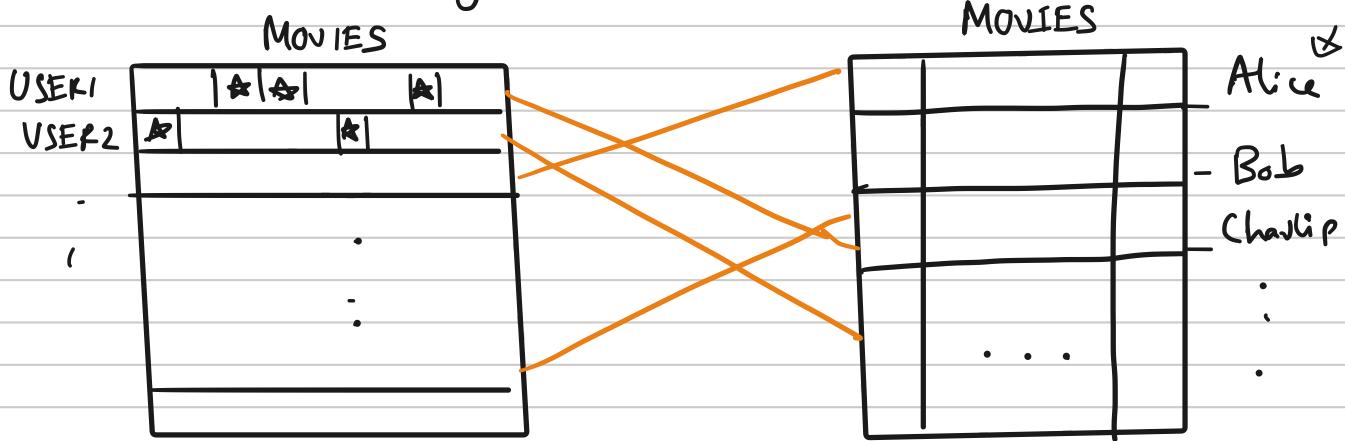
→ FOIL Request to get Taxi fare data

→ (medallion, license, vendor_id, rate_type, pickup_time,
drop-off time, pickup location, dropoff_location, fare.)

To calculate income earned by various cab drivers.

→ Netflix Challenge Dataset.

IMDB Ratings Database



Narayan & Shmatikov (2008)

→ Netflix Challenge 2: 10 Million \$. X Was canceled because of a Lawsuit for violating privacy.

→ Example 3:

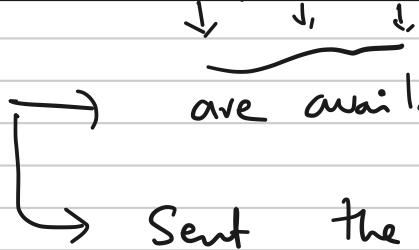
Mars: Group Insurance Commission

→ William Weld Governor

→ Every state employees hospital visit records are available. (but anonymized to preserve privacy).

Name	SSN	ID	SEX	AGE	ZIP	HEAM	WT	(Hb)	...	Admit
------	-----	----	-----	-----	-----	------	----	------	-----	-------

Sweeney:



are available by voter records database.

Sent the Governor his medical record information.

→ "Reconstruction Attacks":

How to get "guaranteed" Privacy?

→ First example: Simplest data analysis tool.

→ Person 1 : $X_1 \in \{0,1\}$ "You like Star Wars."
Person 2 : $X_2 \in \{0,1\}$
⋮ ⋮ ⋮
Person n : $X_n \in \{0,1\}$

→ What we want is to estimate

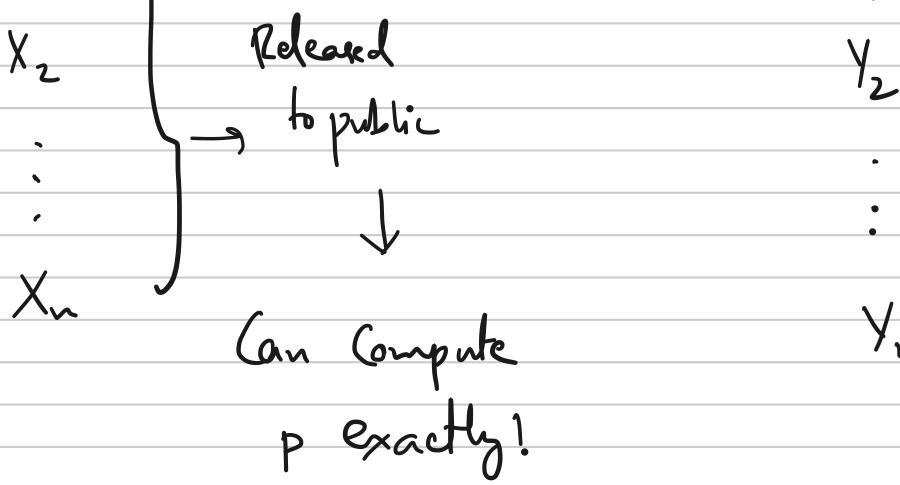
$$\text{the average } p = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Stuart: No privacy

$X_1 \quad ?$

Full privacy:

V. : "Noisy version of X "



Warren 1965: "Randomized Response". (RR)

Each user

$$Y_i = \begin{cases} X_i & \text{with prob } \frac{1}{2} + \gamma \\ 1 - X_i & \text{with prob } \frac{1}{2} - \gamma. \end{cases}$$

$0 < \gamma < \frac{1}{2}.$

$\gamma = 0$: "Full Privacy"

$\gamma = \frac{1}{2}$: "No privacy" at all.

→ \tilde{p} is going to be a function of Y_1, Y_2, \dots, Y_n .

$$\begin{aligned} \rightarrow \mathbb{E}[Y_i] &= \left(\frac{1}{2} + \gamma\right) X_i + \left(\frac{1}{2} - \gamma\right) (1 - X_i) \\ &= \frac{1}{2} - \gamma + 2\gamma \cdot X_i \end{aligned}$$

$$\Rightarrow X_i = \frac{\mathbb{E}[Y_i] - \left(\frac{1}{2} - \gamma\right)}{2\gamma}.$$

Suggests: Given the noisy information Y_1, \dots, Y_n ,

Estimate

$$\bar{P} = \frac{\left(\frac{Y_1 + \dots + Y_n}{n} \right) - \left(\frac{1}{2} - \delta \right)}{2\delta}.$$

Claim: $P_r [|\bar{P} - p| > \delta] \leq \frac{s}{4\delta \cdot \sqrt{n}}$.

(comes from Computing Variance of \bar{P})

Claim: With 75% chance my estimate

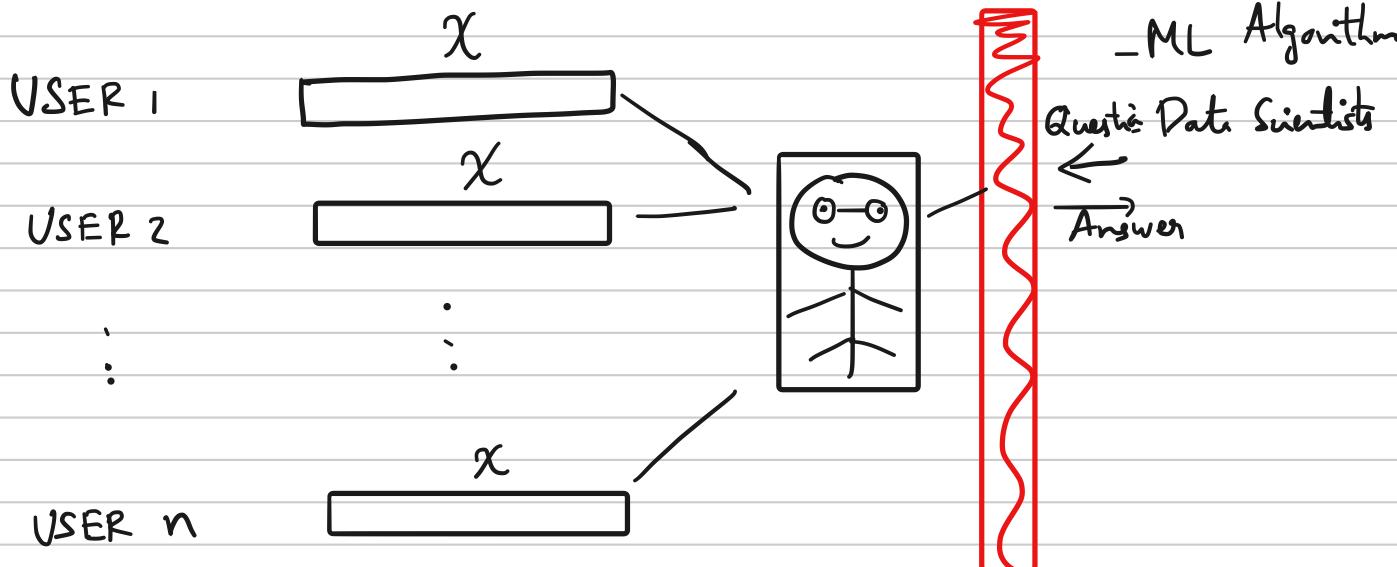
$$|\bar{P} - p| \leq \frac{1}{8\sqrt{n}} = \alpha$$

You need at least $n \geq \frac{1}{\delta^2 \cdot \alpha^2}$ people.

DIFFERENTIAL PRIVACY

→ "CENTRAL DIFFERENTIAL PRIVACY"

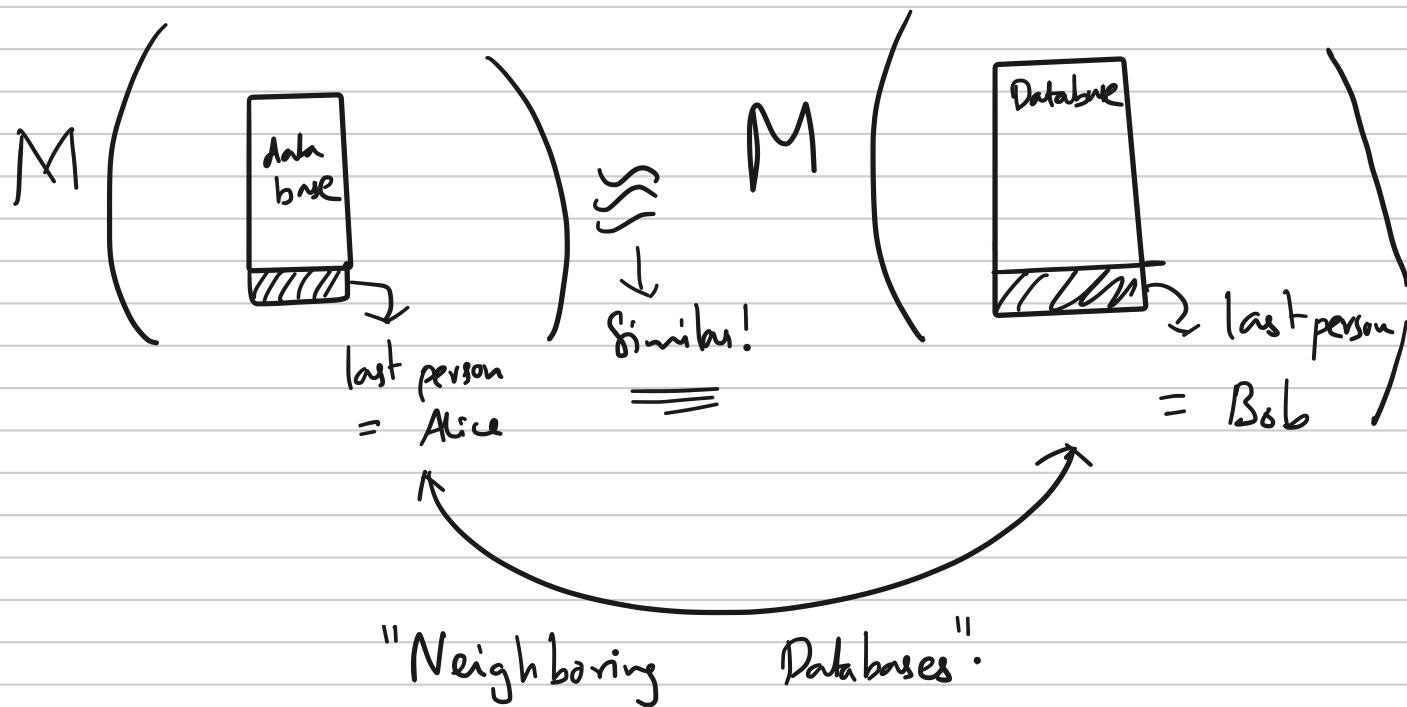
→ "TRUSTED CURATOR MODEL".



(Think $X = \mathbb{R}^d$)

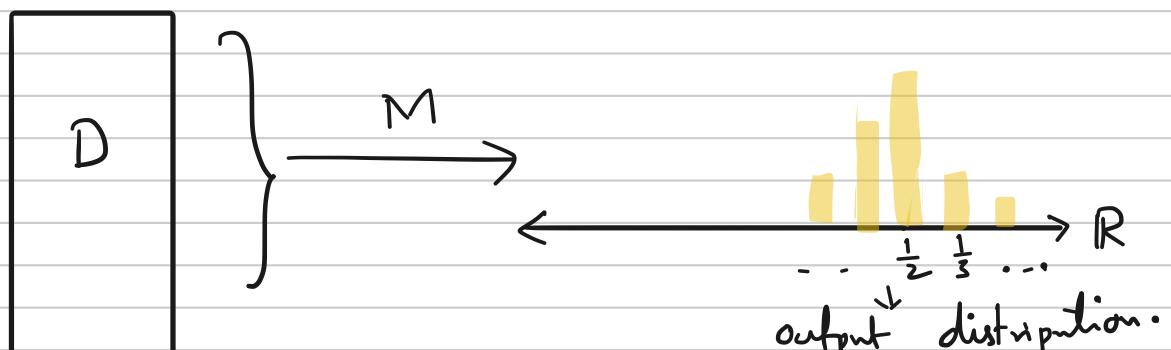
(curator) $M: X^n \rightarrow Y$ (\mathbb{R}).
(database)

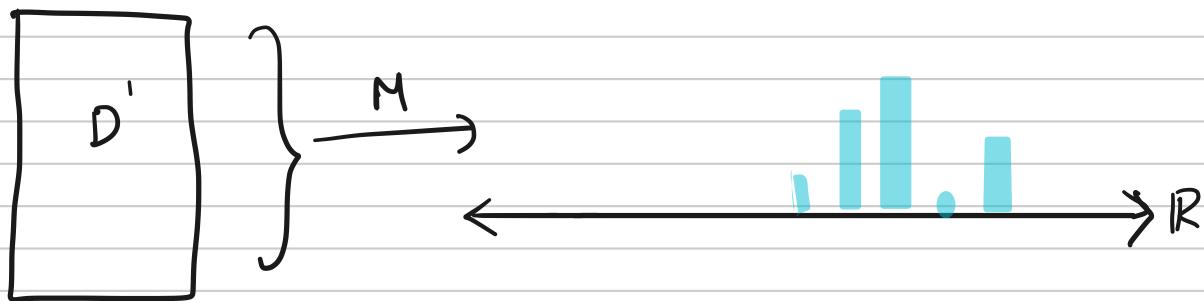
(answer for the "query").



Defn: $D, D' \in X^n$ are neighboring if they differ in exactly one row.

Differential privacy: $M: X^n \rightarrow Y$ [DMNS06] is ϵ -Differentially private if \forall neighboring databases D, D' , $\forall y \in Y$
 $e^{-\epsilon} \cdot \Pr[M(D)=y] \leq \Pr[M(D')=y] \leq e^{\epsilon} \Pr[M(D')=y]$.





Interpretation:

$$\begin{aligned}
 e^{-\varepsilon} \Pr[M(D') = y] &\leq \Pr[M(D) = y] \\
 &\leq e^{\varepsilon} \Pr[M(D) = y].
 \end{aligned}$$

≈ 1 - ε ↓ ≈ 1 + ε

DMNSes: Apple

Google

Microsoft

US Census Bureau 2020

- Diff. Privacy is quantitative.
- Small ε corresponds to better privacy.
- ε should be thought of as $\varepsilon = 0.01$.
- This is a worst-case guarantee on the databases.
- Probabilities are close multiplicatively.
- e^ε vs $1 \pm \varepsilon$ is just a convenience.

A Face Is Exposed for AOL Searcher No. 4417749

[Give this article](#)

By Michael Barbaro and Tom Zeller Jr.

Aug. 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision swinnett

Special offer. Subscribe for \$4.25 \$1 a week.

[EXPAND](#)

LAST CLASS:

- Pitfalls of privacy
- Differential privacy
- Randomized Response

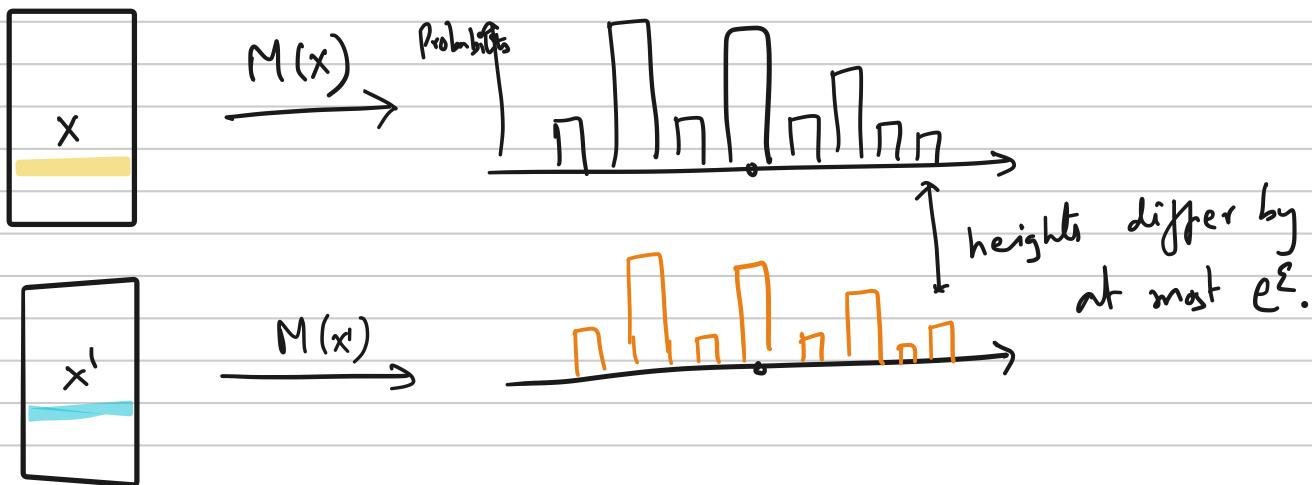
TODAY

- DP of Rand. Response
- Laplacian Mechanism
- Global sensitivity
- Group privacy

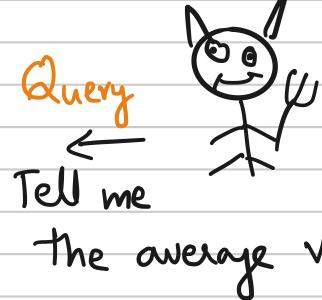
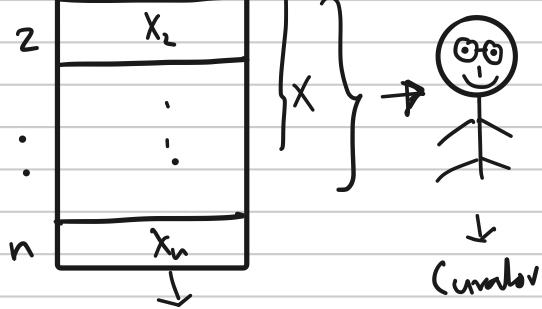
HW3 Due today @ 9:59PM.

Differential Privacy

A mechanism $M: X^n \rightarrow Y$ is ϵ -Differentially Private if for any two adjacent databases X, X' , and any $y \in Y$

$$e^{-\epsilon} \cdot \Pr[M(X)=y] \leq \Pr[M(X)=y] \leq e^{\epsilon} \cdot \Pr[M(X)=y]$$


Computing the Mean Differentially Privately



ϵ do, i.y

$$f(X) = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Q: What should curator do?

Q: What guarantee does Randomized Response provide?

Randomized Response:

$$Y_i = \begin{cases} x_i & \text{with prob } \frac{1}{2} + \gamma \\ 1 - x_i & \text{" .. } \end{cases}$$

$$\text{Output } \tilde{f}(x) = \frac{y_1 + \dots + y_n}{n}.$$

"Intuitively": $\gamma = \frac{1}{2}$: Bad privacy

$\gamma = 0$: Good privacy.

$$\Pr \left[|\tilde{f}(x) - f(x)| \geq \frac{1}{\delta \sqrt{n}} \right] \leq \frac{1}{4}.$$

Claim: Randomized Response as above
is $O(\delta)$ - Differentially Private.

Proof:

		x
y_1	0	x_1
y_2	1	x_2
y_3	0	.
y_4	1	x_n
y_5	0	.
y_6	0	.

	x'
0	y'_1
1	y'_2
0	y'_3
0	y'_4
0	y'_5
0	y'_6

Fix a possible answers $a \in \mathbb{R}$

$$\Pr_{x \sim X} [RR(x) = a]$$

$$\Pr_{x' \sim X'} [RR(x') = a].$$

Randomized-Response with rate γ

All the coordinates in x, x' except one are same.

$$\frac{P_r [RR_y(x) = a]}{P_r [RR_y(x') = a]} \leq \frac{\frac{1}{2} + \gamma}{\frac{1}{2} - \gamma} = \frac{1+2\gamma}{1-2\gamma} = e^{O(\gamma)}$$

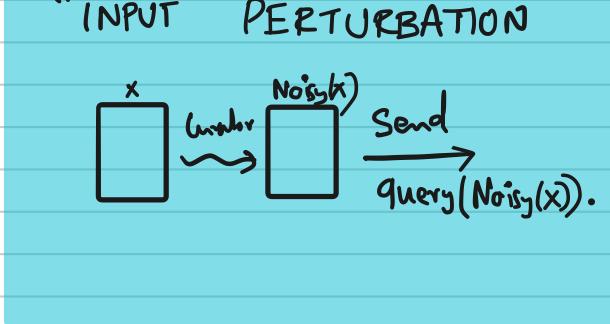
(as changing one entry can only change the probability of 0 or 1 between these two values).

Summary: Randomized Response with noise rate γ is $O(\gamma)$ -DP and achieves error $\frac{1}{\gamma\sqrt{n}}$ for the mean.

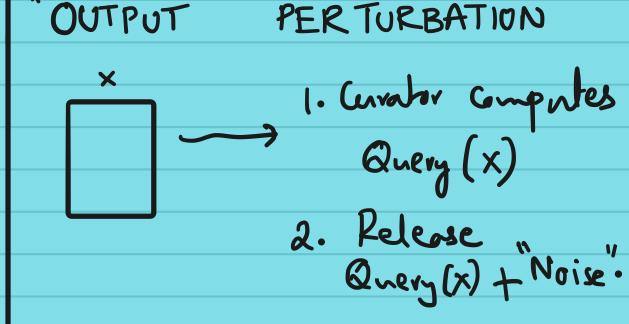
Remark: If we want ϵ -Diff. Privacy and error $\leq \alpha$, then we need $n \geq \frac{C}{\epsilon^2 \alpha^2}$ (for some constant C). if using RR.

Q: Can we achieve better trade-off between privacy, accuracy, size of database? (ϵ) (α) (n)

TWO FUNDAMENTAL WAYS TO GET PRIVACY



Ex: Randomized Response



Ex: "Laplacian Mechanism".

LAPLACIAN MECHANISM FOR RELEASING MEAN.

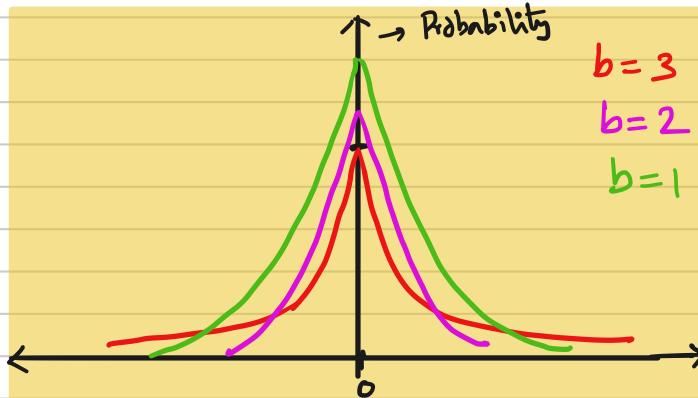
1. Compute $f(x) = \frac{x_1 + x_2 + \dots + x_n}{n}$

2. Release $f(x) + z$

$$z \sim \text{Laplacian}(b)$$

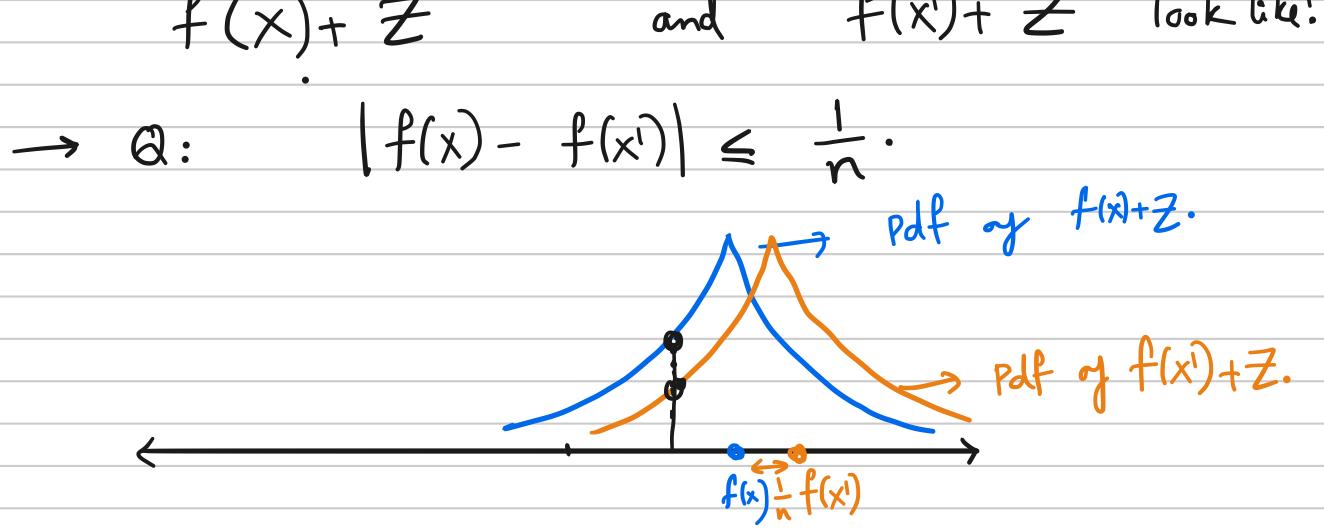
Laplacian with mean 0 and variance b is the distribution whose probability density function

$$P(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right).$$



Claim: Laplacian Mechanism with $b = \frac{1}{\epsilon n}$ satisfies ϵ -Differential privacy for releasing the mean.

Prog: $z \sim \text{Laplacian}(b)$.



$$\Pr[M(x) = a] = \Pr[f(x) + Z = a]$$

$$= \Pr[Z = a - f(x)]$$

$$= \frac{1}{2b} \cdot e^{-\frac{|a-f(x)|}{b}}. \quad \begin{matrix} \text{(definition} \\ \text{of Laplacian)} \end{matrix}$$

$$\Pr[M(x') = a] = \Pr[f(x') + Z = a]$$

$$= \Pr[Z = a - f(x')]$$

$$= \frac{1}{2b} \cdot e^{-\frac{|a-f(x')|}{b}}.$$

$$\frac{\Pr[M(x) = a]}{\Pr[M(x') = a]} = \frac{e^{-\frac{|a-f(x)|}{b}}}{e^{-\frac{|a-f(x')|}{b}}} = e^{-\frac{1}{b}(|a-f(x)| - |a-f(x')|)}$$

is at most $\frac{1}{n}$ in

$$\leq e^{\frac{1}{n \cdot b}}$$

magnitude.

$b = \frac{1}{\epsilon n}$ then, we get

$$\frac{\Pr[M(x) = a]}{\Pr[M(x) = a]} \leq e^\epsilon.$$

Claim: $\Pr[|Z| > t \cdot b] \leq \exp(-t)$.

Laplacian distribution has exponentially decaying tails.

Claim: $\Pr[|Z| > 2b] \leq \exp(-2) \leq \frac{1}{4}$.

Claim: Laplacian Mechanism with $b = \frac{1}{n \cdot \epsilon}$
 achieves ϵ -DP, and $\Pr[|\text{Answer} - f(x)| \geq \frac{2}{n \epsilon}] \leq \frac{1}{4}$.

Remark: To achieve ϵ -DP, and α -accuracy,

$$\text{we need } n \geq \frac{2}{\epsilon \cdot \alpha}.$$

(In comparison, RR to achieve ϵ -DP & α -accuracy
 needed $n \geq \frac{1}{\epsilon^2 \cdot \alpha^2}$.)

Queries: $f: X^n \rightarrow Y$.

Sensitivity: $f: \mathcal{X}^n \rightarrow \mathbb{R}$ $S_1(f) = \max_{\substack{x, x' \\ \text{two neighboring databases}}} |f(x) - f(x')|$.

$$S_1(\text{Mean}) = \frac{1}{n}.$$

Thm: If $f: \mathcal{X}^n \rightarrow \mathbb{R}$ that has $S_1(f) = S$,
 Then, Laplacian Mechanism with noise $b = \frac{S}{\epsilon}$
 achieves ϵ -DP.

Proof: Everything stays same as for the mean,
 just use that $|f(x) - f(x')| \leq S$.

$$f: \mathcal{X}^n \rightarrow \mathbb{R}^k . \quad S_1(f) = \max_{\substack{x, x' \\ \text{two adjacent databases}}} \sum_{i=1}^k |f(x)_i - f(x')_i| .$$

Laplacian Mechanism:

1. Compute $f(x)$

2. Compute k independent Laplacian noise variables

$z_1, z_2, \dots, z_k \sim \text{Laplacian}(b)$

3. Output $f(x) + (z_1, z_2, \dots, z_k)$.

$f: \mathcal{X}^n \rightarrow \mathbb{R}^k$ has sensitivity $S_1(f) \leq S$,

Then Laplacian Mechanism with noise $b = \frac{S}{\epsilon}$ achieves ϵ -DP.

Proof:

$f: X \rightarrow \mathbb{R}$.

x, x' are two neighboring databases.

$$M(x) = f(x) + (z_1, z_2, \dots, z_k) ; M(x') = f(x') + (z_1, \dots, z_k)$$

$$\Pr[M(x) = (a_1, a_2, \dots, a_k)] = \left(\frac{1}{2b}\right) \cdot \exp\left(-\frac{|a_1 - f(x)_1|}{b}\right) \cdot \left(\frac{1}{2b}\right) \cdot \exp\left(-\frac{|a_2 - f(x)_2|}{b}\right)$$

⋮

$$\cdot \left(\frac{1}{2b}\right) \cdot \exp\left(-\frac{|a_k - f(x)_k|}{b}\right)$$

$$= \left(\frac{1}{2b}\right)^k \cdot \exp\left(-\frac{1}{b} \left(\sum_{i=1}^k |a_i - f(x)_i|\right)\right).$$

$$\Pr[M(x') = a] = \left(\frac{1}{2b}\right)^k \cdot \exp\left(-\frac{1}{b} \left(\sum_{i=1}^k |a_i - f(x')_i|\right)\right).$$

$$\Rightarrow \frac{\Pr[M(x) = a]}{\Pr[M(x') = a]} \leq \exp\left(-\frac{1}{b} \cdot \sum_{i=1}^k |f(x)_i - f(x')_i|\right)$$

$$\leq \exp\left(-\frac{\sum_i S_i(f)}{b}\right)$$

Summary:

→ Defined DP

→ Randomized Response (to get ϵ -DP for releasing mean)

→ "Global Sensitivity": Laplacian Mechanism

Post- Processing Differential Privacy.

Thm: Let $M: \mathcal{X}^n \rightarrow \mathcal{Y}$ is ϵ -differentially private.
 $f: \mathcal{Y} \rightarrow \mathcal{Z}$. Then $F \circ M: \mathcal{X}^n \rightarrow \mathcal{Z}$ is also
 ϵ -Differentially Private.

Proof: $\Pr_{x \sim \mathcal{X}} [F \circ M(x) = z]$

x, x'
neighboring
databases

$$= \sum_{y: F(y)=z} \Pr_x [M(x) = y] \quad (\text{let's say } F \text{ is deterministic})$$

$$\leq \sum_{y: F(y)=z} e^\epsilon \Pr_x [M(x') = y]$$

$$y: F(y)=z$$

$$= e^\epsilon \cdot \Pr_x [F \circ M(x') = z].$$

GROUP PRIVACY

What if we have databases \underline{x}, x' that differ in at most k rows.

$M: \mathcal{X}^n \rightarrow \mathcal{Y}$ is ϵ -Differentially private.

$$+ y \quad \Pr_x [M(x) = y] \leq e^{k\epsilon} \cdot \Pr_x [M(x') = y].$$

Proof: $X^{(0)} = X$

$$X^{(1)} = \{ \dots \}$$

$x^{(0)}, \dots, x^{(k)}$ form a chain of databases
 where one entry is changed
 each time.

$$\Pr[M(x^{(0)}) = y] \leq e^\varepsilon \cdot \Pr[M(x^{(1)}) = y]$$

$$\leq e^\varepsilon \cdot e^\varepsilon \cdot \Pr[M(x^{(2)}) = y]$$

⋮
⋮
⋮

$$\leq e^{k\varepsilon} \cdot \Pr[M(x^{(k)}) = y].$$

Last Class

- Difff Privacy by RR
- Global Sensitivity approach
- Laplacian Mechanism
- Post Processing / Group Privacy

Today

- Counting Queries / Histogram Queries
- Answering "Max" Queries
- Exponential Mechanism
- Approx. Difff Privacy

Recall: Laplacian Mechanism gave us ϵ -DP from Sensitivity.

Counting Queries: Number of people/items in database that have certain property.

$$f: X^n \rightarrow \mathbb{R}$$

$$\text{Sensitivity } S_1(\text{Counting Query}) = 1.$$

\Rightarrow Adding Laplacian noise $\text{Lap}\left(\frac{1}{\epsilon}\right)$ gives us ϵ -DP.

Counting Queries: I make k counting queries.

$$f: X^n \rightarrow \mathbb{R}^k$$

$$S_1(f) = \text{Sensitivity} \leq k.$$

\Rightarrow Adding $\text{Lap}\left(\frac{k}{\epsilon}\right)$ gives us ϵ -DP.

Histogram Queries: $f: X^n \rightarrow \mathbb{R}^k$

each person is in one of k categories and we want counts of how many people of each category.

Example: Histogram of ages 1-10, 10-20, 20-30, ...

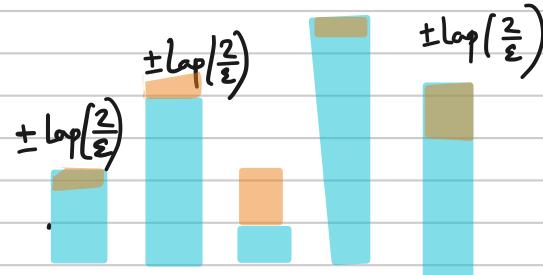
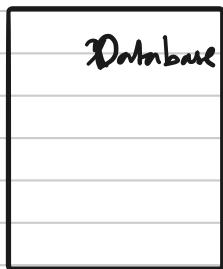
$S_1(\text{Histogram Query}) \leq 2$. 

\Rightarrow Adding $\text{Lap}\left(\frac{2}{\epsilon}\right)$ gives us ϵ -DP.

Remark:

If $Y \sim \text{Lap}(b)$, then

$$\Pr[|Y| > b \cdot t] \leq e^{-t}.$$



Histogram
with "k" buckets

$\Pr[\text{Count in bucket } i \text{ is off by more than } \frac{t}{\epsilon}] \leq e^{-t}$.

$\Pr[\text{there is some bucket whose count is off by more than } \frac{t}{\epsilon}] \leq k \cdot e^{-t}$.

Remark: "Union Bound"

$$k \cdot e^{-10 \cdot \log k}$$

$$= e^{-10}.$$

$$\Pr[\Sigma_1 \vee \Sigma_2 \vee \dots \vee \Sigma_k] = \\ \downarrow \text{event 1} \quad \downarrow \text{event 2} \\ \leq \Pr[\Sigma_1] + \Pr[\Sigma_2] + \dots + \Pr[\Sigma_k].$$

If $t > 10 + \log k$, then

$$\Pr[\text{there is some bucket whose count is off by more than } \frac{10 + \log k}{\epsilon}] \leq e^{-10}.$$

Example (First Names): Suppose we wish to calculate which first names from a list of 10,000 names is most common among participants in 2010 census.

$$\rightarrow n = 300,000,000 \text{ people}$$

$$\rightarrow \text{Histogram query with } k = 10,000.$$

$$\rightarrow \ln(10,000) \approx 9.2.$$

$$\rightarrow \text{If we want privacy with } \epsilon \approx 0.1, \text{ then taking } t = \frac{10 + \ln(10,000)}{\epsilon} = \frac{20}{0.1} = 200.$$

Differentially Private Selection

Example (Most Common Medical Condition):

I have 10 diseases and want to know which is most common.

Approach 1:

$$f: \mathcal{X}^n \rightarrow \mathbb{R}^k$$

- Noise needed
grows with k .
 $\text{Lap}\left(\frac{k}{\epsilon}\right)$.
- Release the counts of each disease privately
 - Compute max of the released counts.

Approach 2:

Report "Noisy Max"

- $f: \mathcal{X}^n \rightarrow \mathbb{R}^k$
- Add noise $\text{Lap}\left(\frac{1}{\epsilon}\right)$ to each $(0, 0, \dots, 0, 1, 0, \dots, 0)$ disease's count
 - Compute the max of these noisy counts
 - Release the id of the disease with the noisy max.

Theorem: Noisy Max is ϵ -Differentially Private.
(see [DR14] claim 3.9)

Exponential

Mechanism

→ Digital Goods Auction: "Digital Good"



User 1 → 1

User 2 → 1

User 3 → 1

User 4 → 3.01

If Price is 3.02 → Revenue is zero

If Price is 1 → Revenue is 4

If Price is 3.01 → Revenue is 3.01.

We have some "Range" R .

We have a "Utility" function $u: \mathcal{X} \times R \rightarrow \mathbb{R}$

Goal: Compute/Release $\arg \max_{r \in R} u(x, r)$.

Example: If $R = \text{Names of diseases}$

$u(x, \text{id}) = \# \text{ people with disease}$

$\Rightarrow \arg \max_{r \in R} u(x, r) = \text{Most common disease.}$

Exponential Mechanism

Given setup as above,

$M_E(x, u, R)$ selects and outputs an element
 $r \in R$ with probability proportional to $\exp\left(\frac{\sum_i u(x, i)}{\Delta u}\right)$.

$$\Delta u = \max_{r \in R} \max_{\substack{x, x' \text{ two} \\ \text{neighboring databases}}} |u(x, r) - u(x', r)|.$$

Claim: Exponential Mechanism is (2ϵ) Differentially Private.

Proof: Take two neighboring databases X, X' .

$$\frac{\Pr[M_E(x) = \delta]}{\Pr[M_E(x') = \delta]}$$

$$\rightarrow \Pr[M_E(x) = \delta] \propto \exp\left(\frac{\epsilon u(x, \delta)}{\Delta u}\right)$$

$$\Pr[M_E(x) = \delta] = \frac{\exp\left(\frac{\epsilon u(x, \delta)}{\Delta u}\right)}{\sum_{\delta' \in R} \exp\left(\frac{\epsilon u(x, \delta')}{\Delta u}\right)}$$

$$\Pr[M_E(x') = \delta] = \frac{\exp\left(\frac{\epsilon u(x', \delta)}{\Delta u}\right)}{\sum_{\delta' \in R} \exp\left(\frac{\epsilon u(x', \delta')}{\Delta u}\right)}$$

By definition for any $\delta \in R$

$$|u(x, \delta) - u(x', \delta)| \leq \Delta u.$$

$$(c - c_1) \quad (c_1 + c_2(x, \delta) + \Delta u)$$

$$\Rightarrow \exp\left(\frac{\sum u(x_i, s)}{\Delta u}\right) \leq e^{\left(\frac{\sum (u(x_i, s) + \Delta u)}{\Delta u}\right)} \\ = e^{\sum \frac{u(x_i, s)}{\Delta u}}.$$

$$\frac{\Pr_{\mathcal{E}}[M_E(x) = s]}{\Pr[M_E(x') = s]} = \frac{\exp\left(\frac{\sum u(x_i, s)}{\Delta u}\right)}{\exp\left(\frac{\sum u(x'_i, s)}{\Delta u}\right)} \cdot \frac{\sum \exp\left(\frac{\sum u(x_i, s)}{\Delta u}\right)}{\sum \exp\left(\frac{\sum u(x_i, s)}{\Delta u}\right)} \\ \leq e^{\sum \frac{u(x_i, s) - u(x'_i, s)}{\Delta u}} \\ \leq e^{2\sum \frac{u(x_i, s) - u(x'_i, s)}{\Delta u}}$$

Thm: $\Pr[u(M_E(x, u, R)) \leq \text{OPT}_u(x) - \frac{\Delta u}{\varepsilon} \cdot (\ln |R| + t)] \leq e^{-t}$.

Remark: In the digital goods auction setting, we can take $|R| = \# \text{ users}$ (price is one of the bid values..)

Rand. Response, Laplacian Mechanism, Exponential Mechanism

$\underbrace{\hspace{10em}}$

ε -Differential Privacy

"Pure" Differential Privacy.

Approximate Differential Privacy

ϵ -DP : \forall neighboring databases

$$\frac{\Pr[M(x) = t]}{\Pr[M(x') = t]} \leq e^\epsilon.$$

As an example if $\Pr[M(x) = t]$ is 2^{-100} ,
then we need $\Pr[M(x) = t] \geq 2^{-100} \cdot e^{-\epsilon}$.

A mechanism $M: X^n \rightarrow Y$ is (ϵ, δ) -Diff. Private

if $\forall t \in Y$, x, x' neighboring databases,

$$\Pr[M(x) = t] \leq e^\epsilon \cdot \Pr[M(x') = t] + \delta$$

Ex: $\epsilon = 1$. $\delta = 1$,

In Pure-DP then

$$\Pr[M(x) = t] \leq e \cdot \Pr[M(x') = t]$$

In (ϵ, δ) -DP,

$$\Pr[M(x) = t] \leq e \cdot \Pr[M(x') = t] + \delta.$$

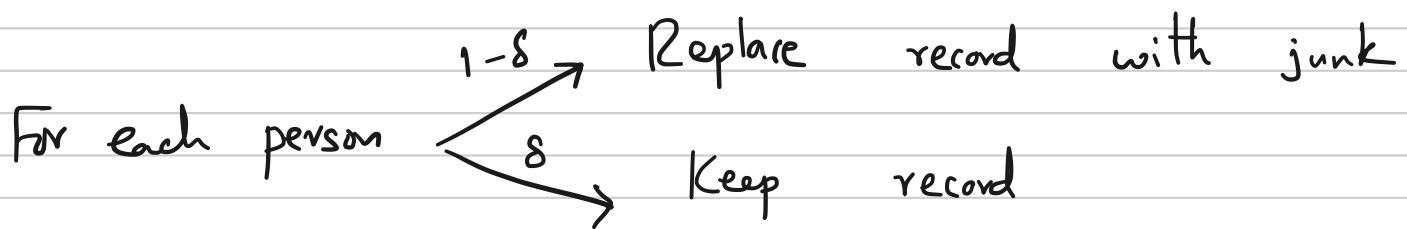
Remark: "We have privacy except with probability δ ".

Example:

(ϵ, δ) -DP:

Satisfies } M is with probability $1-\delta$ output junk
 (ϵ, δ) -DP! with probability δ output entire database.

Example:



\Rightarrow On average $\approx n \cdot \delta$ people's records are released as is

Always think of $\delta \ll \frac{1}{n}$.

Last class

- Examples of DP for practically relevant queries
- Exponential Mechanism

Today

- Gaussian Mechanism (Approx. DP)
- Private ERM.

1. Assignment 3 grades are up.

2. Assignment 4 will be up today by 6PM.

ApproxDP (ϵ, δ) -DP if $M: \mathcal{X}^n \rightarrow \mathcal{Y}$
if neighboring databases x, x' and all $t \in \mathcal{Y}$,
 $P_r[M(x) = t] \leq e^\epsilon \cdot P_r[M(x') = t] + \delta.$

L_2 -Sensitivity: $f: \mathcal{X}^n \rightarrow \mathbb{R}^k$

$$S_2(f) = \max_{\substack{x, x' \\ \text{neighboring databases}}} \|f(x) - f(x')\|_2.$$

Gaussian Mechanism.

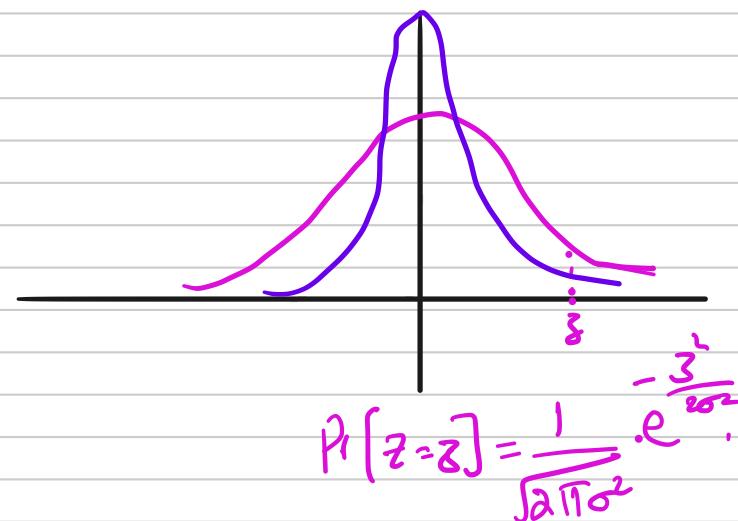
$$\rightarrow X$$
$$\rightarrow f: \mathcal{X}^n \rightarrow \mathbb{R}^k$$

→ Sample a noise vector z_1, \dots, z_k that are

$$\text{i.i.d } N(0, \frac{\sigma^2}{k})$$

→ Return $f(x) + (z_1, z_2, \dots, z_k)$.

(Recall: Gaussian)



Thm: Let $f: X^n \rightarrow \mathbb{R}^k$. Then Gaussian mechanism with $\sigma = \sqrt{2 \ln(\frac{1.25}{\delta})} \cdot \frac{S_1(f)}{\epsilon}$ ensures (ϵ, δ) -DP.

Really think of this as $\sigma \approx C \cdot \sqrt{\ln(\frac{1}{\delta})} \cdot \frac{S_1(f)}{\epsilon}$.

Example: n users $\xrightarrow{\quad} \boxed{x_0, x_1, \dots}$ $f(x) = \frac{1}{n} \sum_{i=1}^n x_i \quad (\in \mathbb{R}^d)$.

$$S_1(f) = \max_{\substack{x, x' \\ \text{neighboring}}} \|f(x) - f(x')\|_1 = \frac{d}{n}.$$

$$S_2(f) = \max_{\substack{x, x' \\ \text{neighboring}}} \|f(x) - f(x')\|_2 = \frac{\sqrt{d}}{n}. \quad \left(\left\| \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n} \right) \right\|_2 \right)$$

To use Laplacian mechanism, we would add much more noise than to use Gaussian mechanism.

Laplacian

$$\rightarrow \text{Noise} \approx \frac{d}{n}$$

\approx Error: $\frac{d}{n}$ per coordinate

"Mem Release"

Gaussian

$$\text{Noise} \approx \frac{\sqrt{d}}{n}$$

Error $\approx \frac{\sqrt{d}}{n}$ per coordinate.

Group Privacy of Approximate DP:

Suppose X and x' are two databases that differ in k rows.

If M is (ϵ, δ) -DP

$$\Pr[M(x) = y] \leq e^{k\epsilon} \Pr[M(x') = y] + k \cdot e^{\epsilon} \cdot \delta.$$

(Recall for pure DP, we had

$$\Pr[M(x) = y] \leq e^{k\epsilon} \cdot \Pr[M(x') = y].$$

Recall: Suppose M_1, \dots, M_k are (ϵ, δ) -Differentially Private. Then, their composition $\underbrace{M_1 \circ M_2 \circ \dots \circ M_k}_1$ is $k \cdot \epsilon$ - DP.

↓ "Adaptive Queries".

Basic Composition Theorem for Approx DP:

If M_1, M_2, \dots, M_n are (ϵ, δ) -DP adaptive queries, then the composition is $(k\epsilon, k\delta)$ -DP.

Advanced Composition Theorem:

If M_1, \dots, M_k are ϵ -DP mechanisms for answering adaptive queries. Then, the composition is $(\epsilon \sqrt{2k \ln(\frac{1}{\delta})} + k\epsilon^2, \delta)$ -DP
(as long as $\epsilon < 1$).

Remark: Main win for advanced composition is that you get non-trivial privacy for the composition when $\epsilon \ll \frac{1}{\sqrt{k}}$.

Remark: The above was actually used to make Kaggle "leaderboard" better.

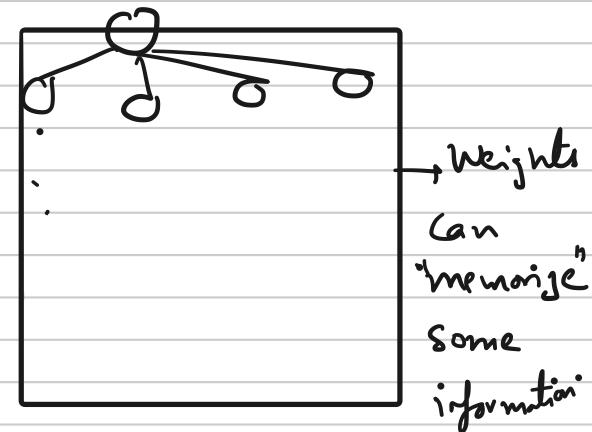
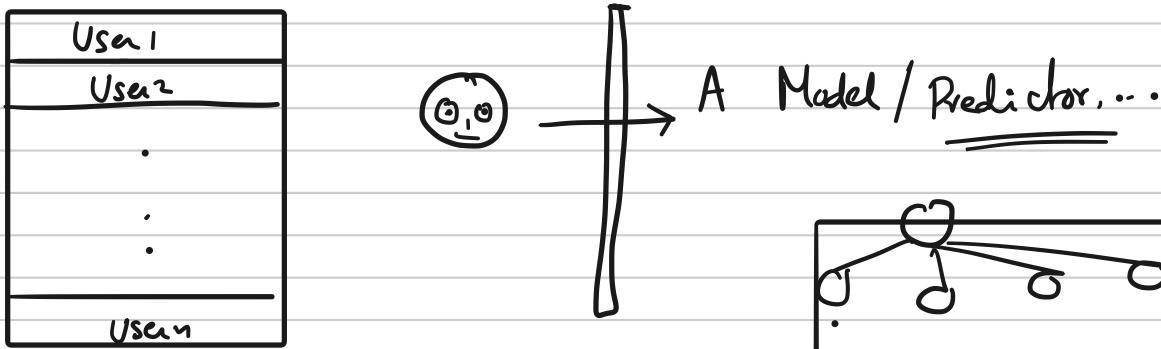
"A Reliable Leaderboard for Machine Learning Competitions".

Advanced Composition Theorem:

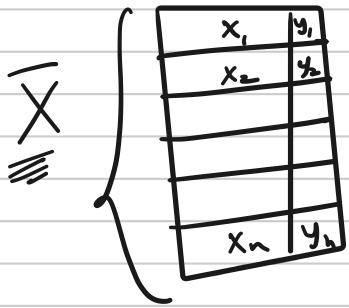
If M_1, \dots, M_k are (ϵ, δ) -DP mechanisms for

answering adaptive queries. Then, the composition
is $(\epsilon \sqrt{2k \ln(\frac{1}{\delta})} + k\epsilon^2, k\delta^{\frac{1}{2}} + \delta) - DP$
(as long as $\epsilon < 1$).

PRIVATE ML



PRIVATE EMPIRICAL RISK MINIMIZATION



Parametric family of predictors

$$h_{\theta}: x \rightarrow y.$$

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n l(\theta; x_i, y_i).$$

The loss function gets

The loss function is based on predicting y_i from x_i .

$$\hat{\theta}^* = \arg \min_{\theta} L(\theta) \cdot = f(\bar{x})$$

We have to answer query f privately.

Approach 1: Add input noise or output noise ...

Problem: Sensitivity of $\hat{\theta}^*$ is very high.

Example:

y_1
.
.
:
y_n

0
0
0
:
0

Mean is 0.

0
1
0
0
1

Mean is non-zero.

Predictor family is just constants.

loss is squared loss:

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - \theta)^2$$

$$\hat{\theta}^* = \arg \min L(\theta) = \frac{1}{n} (y_1 + \dots + y_n)$$

Recall that for DP, we need to have closeness in distributional outcomes.

Three approaches for Private ERM

(a) Output Perturbation

(b) "Objective" Perturbation

(c) "Gradient" Perturbation

Compute θ^*

Add noise.

$$\tilde{L}(\theta) = L(\theta) + \langle b, \theta \rangle$$

"Some noise vector".

Output $\arg \min \tilde{L}(\theta)$.

"Privacy guarantees depend on exact Solutions."

Intuitively Add additional noise to each gradient- step for Solving ERM.

Private Stochastic Gradient Descent.

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n l(\theta; x_i, y_i)$$

SGD

→ Pick a start θ_0

→ For $t = 1, \dots, T$:

ⓐ Pick index $i \in [n]$ uniformly at random

$$\theta_{t+1} = \theta_t - \eta_t \nabla_{\theta} l(\theta; x_i, y_i).$$

"Step-size"

→ Pick a start θ_0

→ For $t=1, \dots, T$:

① Pick index $i \in [n]$ uniformly at random

$$\theta_{t+1} = \theta_t - \eta_t \left(\nabla_{\theta} l(\theta; x_i, y_i) + \text{Noise}_t \right)$$

G_t typically Gaussian



θ_0

θ_0

{ Pick index. }
at random { }
 θ_1 θ_1

as long as picked index is
not the one entry where the
two differ, we have same state.

