LECTURE 13. 05/09

Last Class	Today
1. PCA	1. Privacy in ML
2. Computing SVD	2. Why?
3. Application	3. What?
	4. How ... ?

Privacy means "Proof of Privacy".

Sensitive Datasets: → Medical Records

→ Genetic data

→ Search Logs

→

AOL: Released such a database.

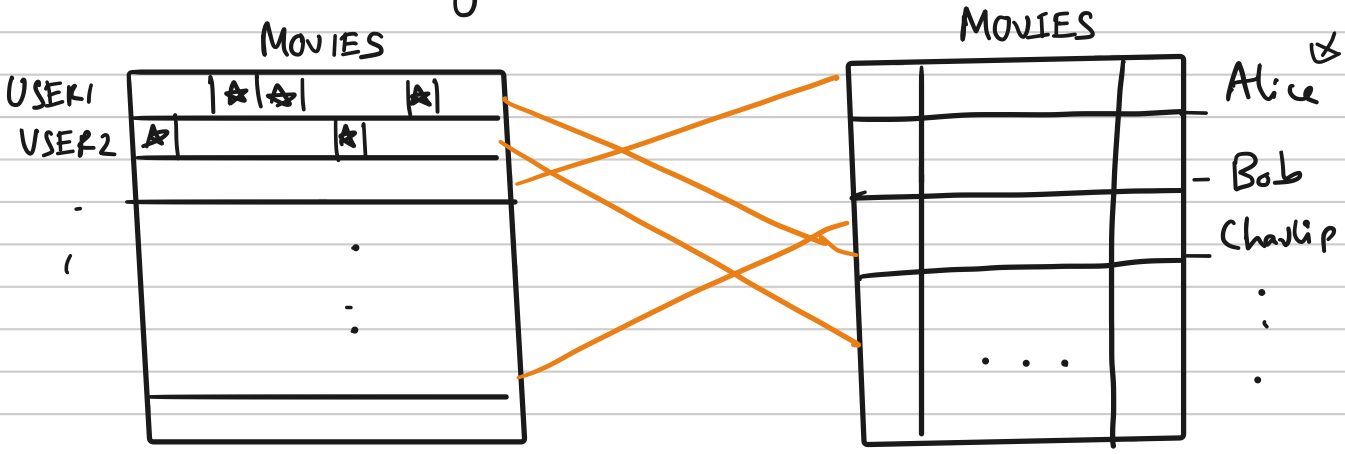
→ NY TaxiCab.

→ 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.

→ Netfix Challenge Dataset. IMDB Ratings Database



Narayanan & Shmatikov (2008)

→ Netfix Challenge 2: 10 Million $. ✗ Was Canceled because of a Lawsuit for violating privacy.

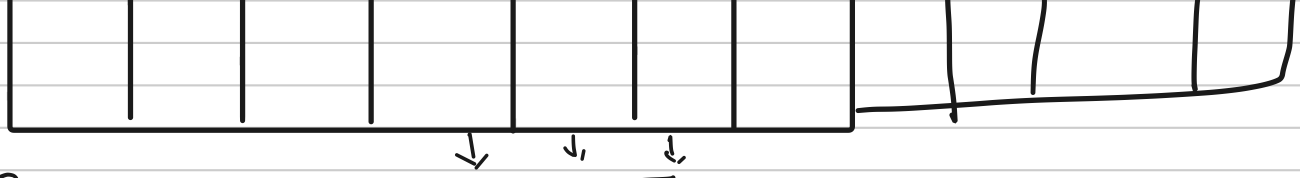
→ Example 3:

Mass: Group Insurance Commission

→ William Weld Governor

→ Every State employees hospital visit records are available. (but anonomized to preserve privacy).

Name	SSN	ID	SEX	AGE	ZIP	HEALTH	WT	CHO	. . .	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.

"You like Star Wars."

→ Person 1: $X_1 \in \{0,1\}$

Person 2: $X_2 \in \{0,1\}$

⋮ ⋮ ⋮

Person n: $X_n \in \{0,1\}$

→ What we want is to estimate the average $p = \frac{X_1 + X_2 + \cdots + X_n}{n}$.

Silliest: No privacy

X_1 ?

Full privacy:

Y_1 : "Noisy version of X_1"

X_2

$\vdots$

X_n

Released to public $\rightarrow$

$\downarrow$

Can Compute p exactly!

Y_2

$\vdots$

Y_n: "Noisy version X_n"

Warner 1965: "Randomised 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.

$\rightarrow$ $\tilde{p}$ is going to be a function of $Y_1, Y_2, \ldots, Y_n$.

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

$= \frac{1}{2} - \gamma + 2\gamma \cdot X_i$

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

Suggests: Given the noisy information $Y_1, \ldots, Y_n$,

Estimate

$$\bar{p} = \frac{\left(\frac{Y_1 + \cdots + Y_n}{n}\right) - \left(\frac{1}{2} - \gamma\right)}{2\gamma}.$$

Claim: $\Pr\left[\,|\bar{p} - p| > \delta\right] \leq \frac{\delta}{4\gamma \cdot \sqrt{n}}$.

(Comes from computing Variance of $\bar{p}$)

Claim: With 75% chance my estimate

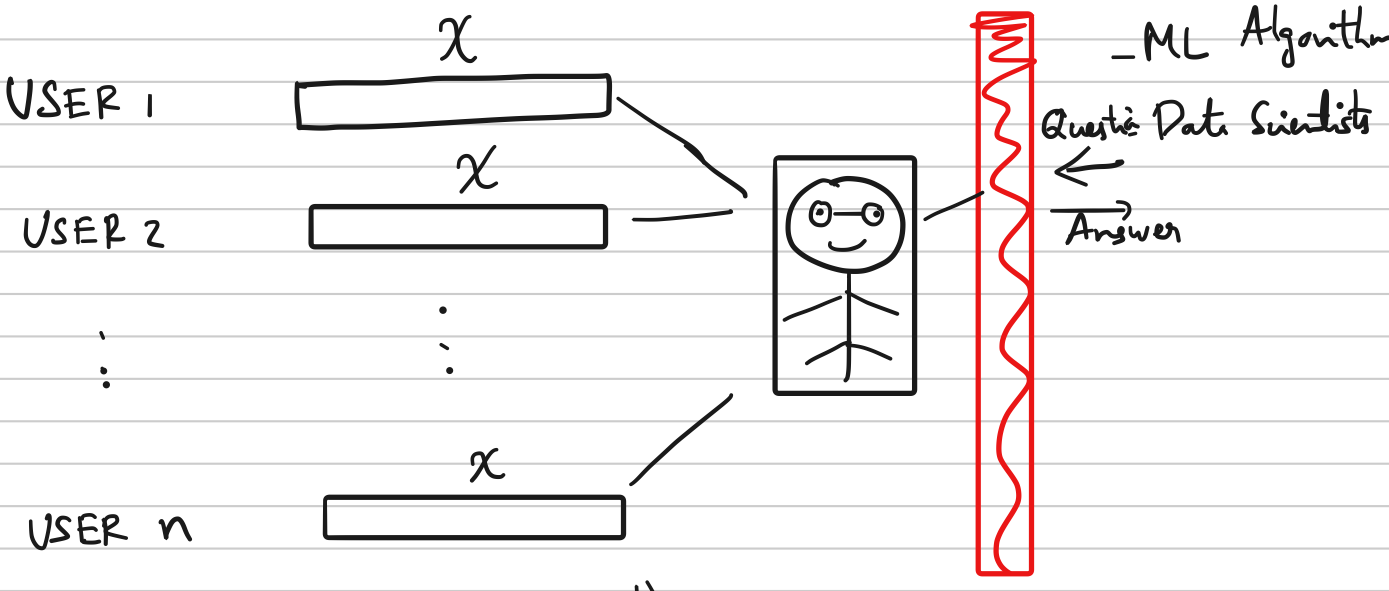
$$|\bar{p} - p| \leq \frac{1}{\gamma\sqrt{n}} = \alpha$$

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

DIFFERENTIAL PRIVACY

→ "CENTRAL DIFFERENTIAL PRIVACY"

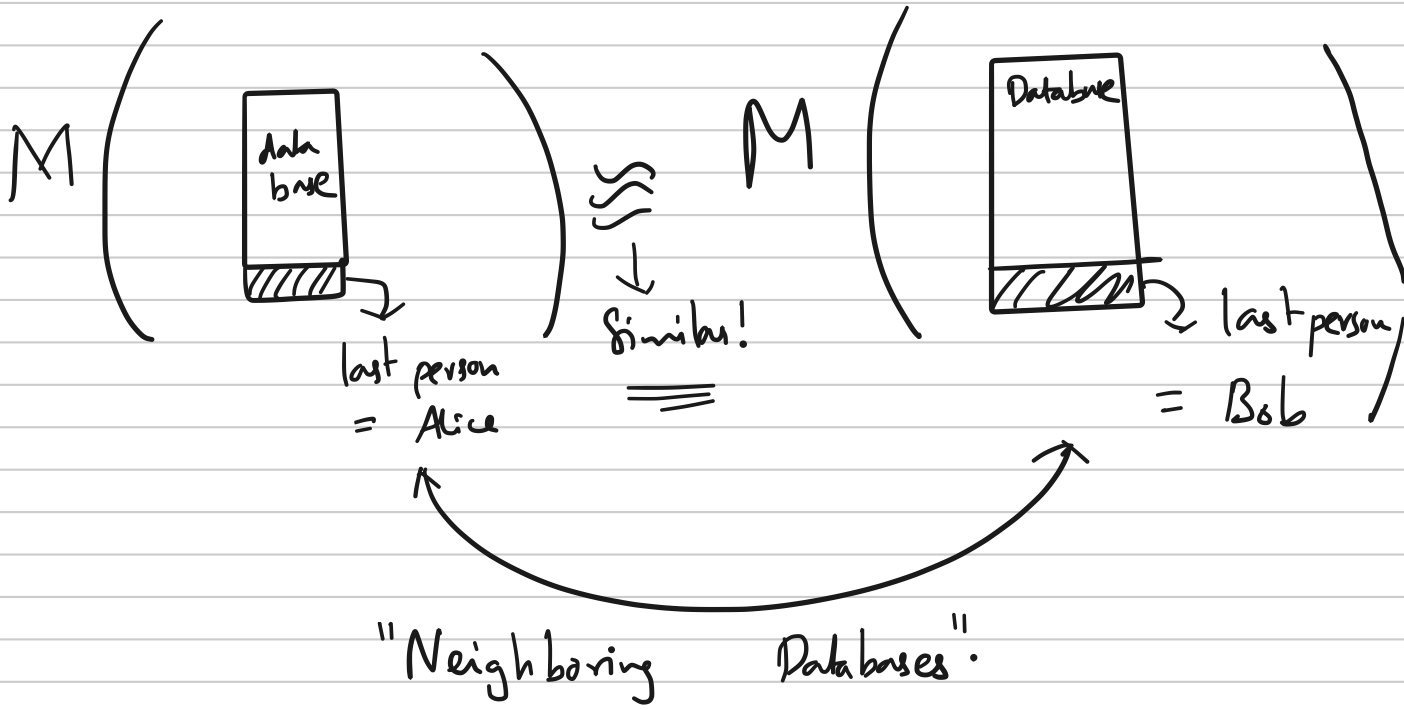
→ "TRUSTED CURATOR MODEL".



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

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

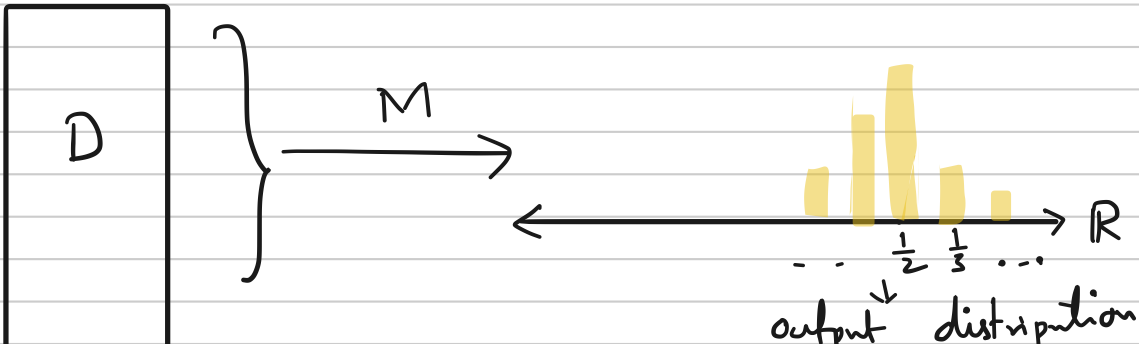
($\mathcal{X}^n$: database; $\mathcal{Y}$: answer for the "query").

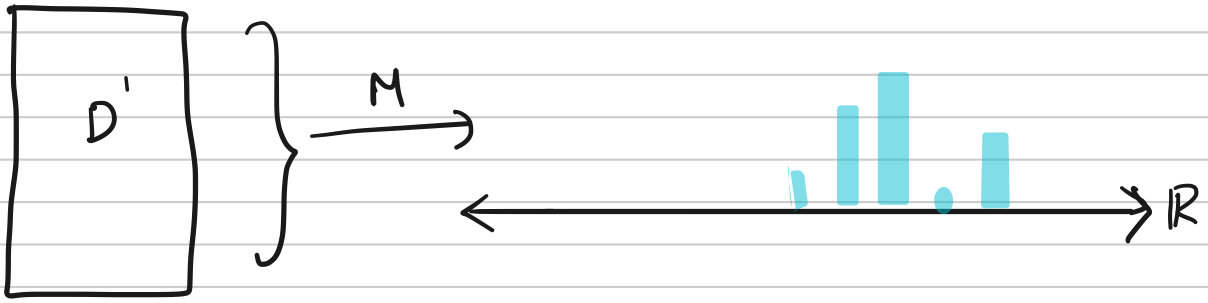


Defn: $D, D' \in \mathcal{X}^n$ are neighboring if they differ in exactly one row.

Differential privacy: $M: \mathcal{X}^n \to \mathcal{Y}$ [DMNS06] is ε-Differentially private if $\forall$ neighboring databases D, D', $\forall\ y \in \mathcal{Y}$

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





Interpretation: $e^{-\varepsilon} \Pr[M(D')=y] \leq \Pr[M(D)=y]$

$\lesssim 1-\varepsilon$

$\leq e^{\varepsilon} \Pr[M(D')=y].$

$\lesssim 1+\varepsilon$

DMNS06: Apple

Google

Microsoft

US Census Bureau 2020

→ Diff. Privacy is quantitative.
Small ε corresponds to better privacy.

→ ε should be thought 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.

11:15 AM Mon May 9
55%
nytimes.com
TECHNOLOGY
The New York Times
SUBSCRIBE FOR $1/WEEK
LOG IN

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 gwinnett

Special offer. Subscribe for ~~$4.25~~ $1 a week.
EXPAND

LECTURE 14 05/11

LAST CLASS:

- → Pitfalls of privacy
- → Differential privacy
- → Randomized Response

TODAY

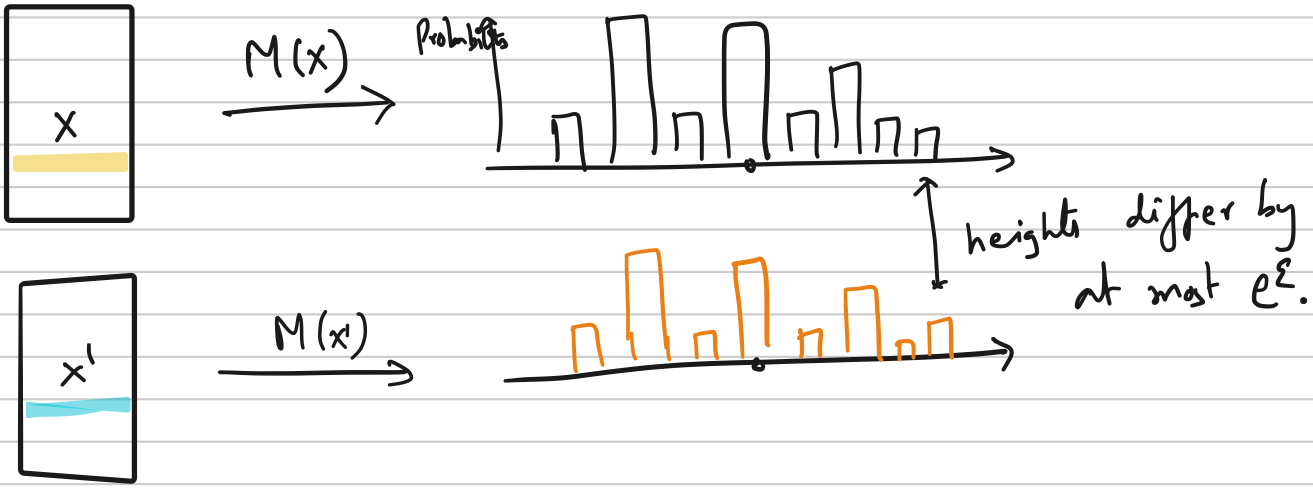
- → DP of Rand. Response
- → Laplacian Mechanism
- → Global sensitivity
- → Group privacy

HW3 Due today @ 9:59PM.

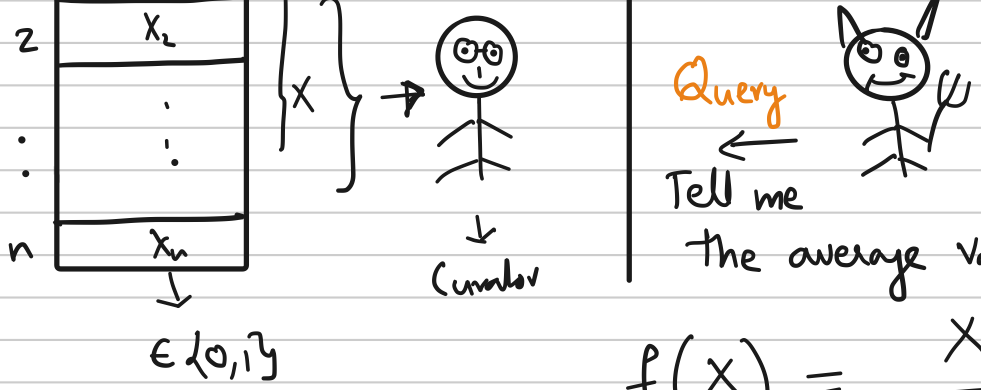
Differential Privacy

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

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



Computing the Mean Differentially Privately



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

Q: What should curator do?

Q: What guarantee does Randomized Response provide?

Randomized Response:

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

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

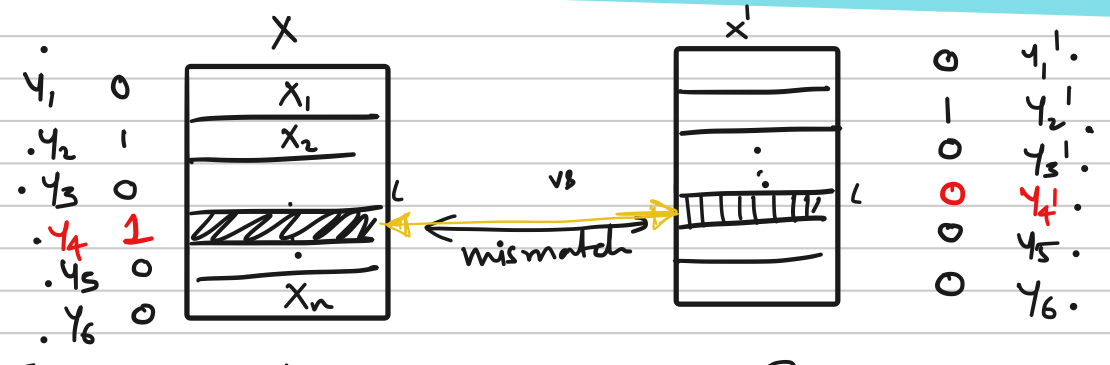
"Intuitively": $\gamma = \frac{1}{2}$: Bad privacy

$\gamma = 0$: Good privacy.

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

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

Proof:



Fix a possible answers $a \in R$

$$\Pr\left[RR_\gamma(x) = a\right] \quad \text{vs} \quad \Pr\left[RR_\gamma(x') = a\right].$$

RR_γ ↓

Randomized-Response with rate γ

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

$$\frac{Pr\left[RR_\gamma(x) = a\right]}{Pr\left[RR_\gamma(x') = a\right]} \leq \frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma} = \frac{1+2\gamma}{1-2\gamma} = e^{O(\gamma)}.$$

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

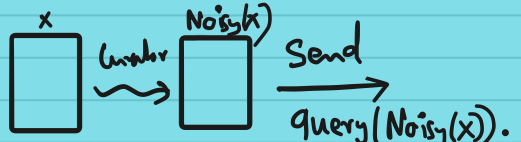
Summary: Randomised Response with noise rate γ is $O(\gamma)$- DP and achieves error $\frac{1}{\gamma\sqrt{n}}$ for the mean.

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

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

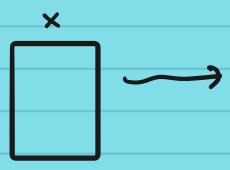
TWO FUNDAMENTAL WAYS TO GET PRIVACY

"INPUT PERTURBATION



Ex: Randomized Response

"OUTPUT PERTURBATION



1. Curator computes Query(x)
2. Release Query(x) + "Noise".

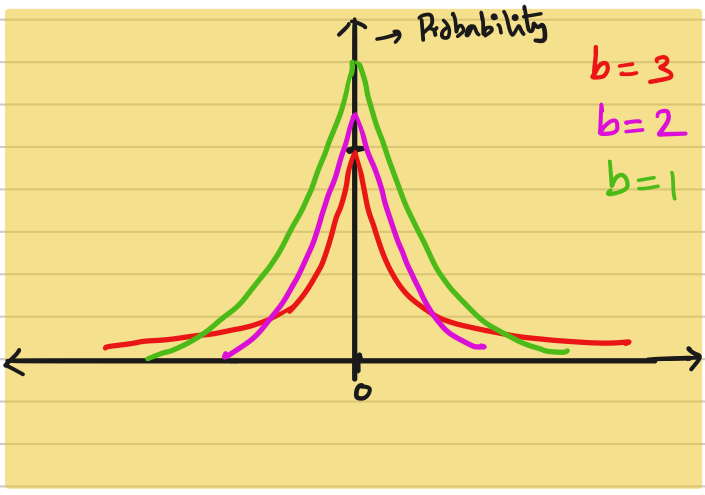
Ex: "Laplacian Mechanism".

LAPLACIAN MECHANISM FOR RELEASING MEAN.

1. Compute $f(x) = \frac{X_1 + X_2 + \cdots + X_n}{n}$
2. Release $f(x) + Z$, where $Z \sim \text{Laplacian}(b)$

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

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

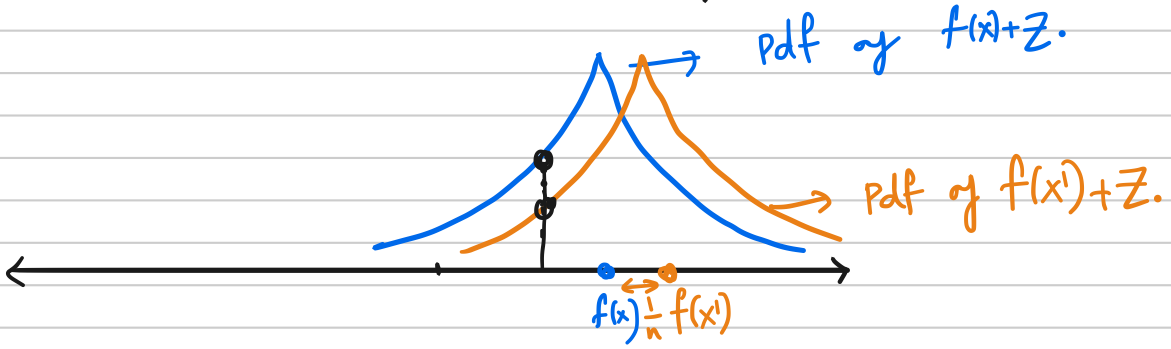


Claim: Laplacian Mechanism with $b = \frac{1}{\varepsilon n}$ satisfies ε- Differential privacy for releasing the mean.

Proof: $Z \equiv \text{Laplacian}(b)$.

$f(x)+Z$ and $f(x')+Z$ look like:

→ Q: $|f(x) - f(x')| \leq \frac{1}{n}$.



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

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

$$= \frac{1}{2b} \cdot e^{-\frac{|a-f(x)|}{b}} \cdot$$ (definition of Laplacian)

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

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

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

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

$$= e^{-\frac{1}{b}\left(|a-f(x)| - |a-f(x')|\right)}$$

$|a-f(x)| - |a-f(x')|$ is at most $\frac{1}{n}$ in

$\frac{1}{n \cdot b}$ magnitude.

$\leq e$

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

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

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 $b = \frac{1}{n \cdot \varepsilon}$ achieves ε-DP, and $Pr[|Answer - f(x)| \geq \frac{2}{n\varepsilon}] \leq \frac{1}{4}$.

Remark: To achieve ε-DP, and α-accuracy, we need $n \geq \frac{2}{\varepsilon \cdot \alpha}$.

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

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

Sensitivity: $f: X^n \to \mathbb{R}$

$$S_1(f) = \max_{x, x'} |f(x) - f(x')|.$$

x, x': two neighboring databases

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

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

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

$f: X^n \to \mathbb{R}^k$. $S_1(f) = \max_{x, x'} \sum_{i=1}^{k} |f(x)_i - f(x')_i|$.

x, x': two adjacent databases

Laplacian Mechanism:

1. Compute $f(x)$
2. Compute k independent Laplacian noise variables $Z_1, Z_2, \ldots, Z_k \sim \text{Laplacian}(b)$
3. Output $f(x) + (Z_1, Z_2, \ldots, Z_k)$.

$f: X^n \to \mathbb{R}^k$ has sensitivity $S_1(f) \leq S$, Then Laplacian Mechanism with noise $b = \frac{S}{\varepsilon}$ achieves ε-DP.

Proof: $f: \mathcal{X} \to \mathbb{R}^k$.

X, X' are two neighboring databases.

$$M(x) = f(x) + (Z_1, Z_2, \ldots, Z_k) \quad ; \quad M(x') = f(x') + (Z_1, \ldots, Z_k)$$

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

$$\cdot \left(\frac{1}{2b}\right) \cdot \exp\left(-\frac{|a_2 - f(x)_2|}{b}\right)$$

$$\vdots$$

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

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

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

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

$$\leq \exp\left(\frac{S_1(f)}{b}\right)$$

Summary:

→ Defined DP

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

→ "Global Sensitivity" : Laplacian Mechanism

Post-Processing Differential Privacy.

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

Proof: $\Pr[f \circ M(x) = z]$ — x, x' neighboring databases

$$= \sum_{y: f(y)=z} \Pr[M(x) = y]$$ (let's say f is deterministic)

$$\leq \sum_{y: f(y)=z} e^{\varepsilon} \Pr[M(x') = y]$$

$$= e^{\varepsilon} \cdot \Pr[f \circ M(x') = z].$$

GROUP PRIVACY

What if we have databases x, x' that differ in at most k rows.
$M: \mathcal{X}^n \to \mathcal{Y}$ is ε-Differentially private.

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

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

$x^{(1)} =$ [illegible]

$x^{(2)}$ } k steps form a chain of databases
$\vdots$ } where one entry is changed
$x^{(k)} = x'$ } each time.

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

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

$$\vdots$$

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

LECTURE 15 05/16

Last Class	Today
- Diff Privacy of RR	- Counting Queries / Histogram Queries
- Global Sensitivity approach	- Answering "Max" Queries
- Laplacian Mechanism	- Exponential Mechanism
- Post Processing / Group Privacy	- Approx. Diff Privacy

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

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

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

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

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

Counting Queries: I make k counting queries.

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

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

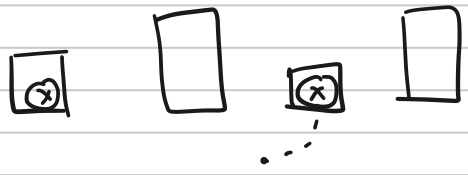
$\Rightarrow$ Adding $\text{Lap}\left(\frac{k}{\varepsilon}\right)$ gives us ε-DP.

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

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, ...

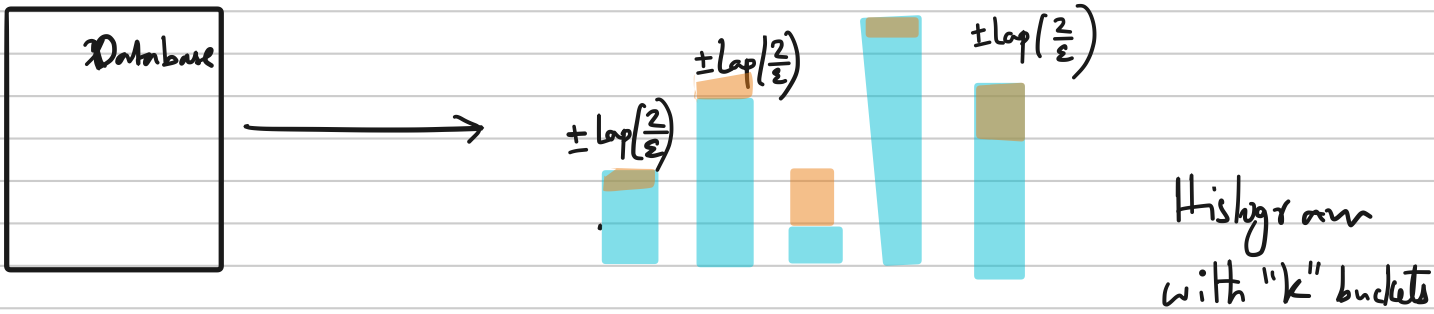
$S_1(\text{Histogram Query}) \leq 2$.



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

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

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



Histogram with "k" buckets

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

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

Remark: "Union Bound"

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

$\Pr[\mathcal{E}_1 \vee \mathcal{E}_2 \vee \dots \vee \mathcal{E}_k]$ =

$\mathcal{E}_1$: event 1, $\mathcal{E}_2$: event 2

$\leq \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2] + \dots + \Pr[\mathcal{E}_k]$.

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

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

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.

→ $n = 300,000,000$ people

→ Histogram query with $k = 10,000$.

→ $\ln(10,000) \approx 9.2$.

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

Differentially Private Selection

Example (Most Common Medical Condition):

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

Approach 1: $f: \mathcal{X}^n \to \mathbb{R}^k$

Noise needed grows with k.
- Release the counts of each disease privately
- Compute max of the released counts.

$\text{Lap}\left(\frac{k}{\varepsilon}\right)$.

Approach 2: **Report "Noisy Max"**

$f: \mathcal{X}^n \to \mathbb{R}^k$ $\quad (0, 0 \ldots 0, 1, 0, \ldots 0)$

- Add noise $\text{Lap}\left(\frac{1}{\varepsilon}\right)$ to each 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"

 → "Painting"

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" $\mathcal{R}$.

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

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

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

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

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

Exponential Mechanism

Given setup as above,

$M_E(X, u, \mathcal{R})$ selects and outputs an element $r \in \mathcal{R}$ with probability proportional to $\exp\left(\frac{\varepsilon \cdot u(X, r)}{\Delta u}\right)$.

$$\Delta u = \max_{r \in \mathcal{R}} \max_{X, X' \text{ two neighboring databases}} \left| u(X, r) - u(X', r) \right|.$$

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

Proof: Take two neighboring databases X, X'.

$$\frac{Pr[M_E(X) = r]}{Pr[M_E(X') = r]}$$

$$Pr[M_E(X) = r] \propto \exp\left(\frac{\varepsilon u(X, r)}{\Delta u}\right)$$

$$Pr[M_E(X) = r] = \frac{\exp\left(\frac{\varepsilon u(X, r)}{\Delta u}\right)}{\sum_{s \in R} \exp\left(\frac{\varepsilon u(X, s)}{\Delta u}\right)}$$

$$Pr[M_E(X') = r] = \frac{\exp\left(\frac{\varepsilon u(X', r)}{u}\right)}{\sum_{s \in R} \exp\left(\frac{\varepsilon u(X', s)}{\Delta u}\right)}$$

By definition for any $s \in R$

$$|u(X, s) - u(X', s)| \leq \Delta u.$$

$$\Rightarrow \quad \exp\left(\frac{\varepsilon u(x',s)}{\Delta u}\right) \le e^{\left(\frac{\varepsilon\cdot(u(x,s)+\Delta u)}{\Delta u}\right)}$$

$$= e^{\varepsilon} \cdot e^{\frac{\varepsilon u(x,s)}{\Delta u}}.$$

$$\frac{Pr\left[M_E(x)=r\right]}{Pr\left[M_E(x')=r\right]} = \frac{\exp\left(\frac{\varepsilon u(x,r)}{\Delta u}\right)}{\exp\left(\frac{\varepsilon u(x',r)}{\Delta u}\right)} \cdot \frac{\sum_s \exp\left(\frac{\varepsilon u(x',s)}{\Delta u}\right)}{\sum_s \exp\left(\varepsilon \frac{u(x,s)}{\Delta u}\right)}$$

$$\le e^{\varepsilon} \cdot \exp\left(\frac{\varepsilon(u(x,r)-u(x',r))}{\Delta u}\right)$$

$$\le e^{2\varepsilon}.$$

Thm: $Pr\left[u\left(M_E(x,u,R)\right) \le OPT_u(x) - \frac{\Delta u}{\varepsilon}\cdot(\ln|R| + t)\right] \le e^{-t}.$

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

Rand. Response, Laplacian Mechanism, Exponential Mechanism

ε - Differential Privacy

"Pure" Differential Privacy.

Approximate Differential Privacy

ε-DP : $\forall$ neighboring databases

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

As an example if $\Pr[M(x) = t]$ is 2^{-100},
Then we need $\Pr[M(x') = t] \geq 2^{-100} \cdot e^{-\varepsilon}$.

A mechanism $M: X^n \to Y$ is (ε, δ) - Diff. Private if $\forall\ t \in Y$, x, x' neighboring databases,

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

Ex: $\varepsilon = 1$. $\varepsilon = 1$,

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

Example:

For each person → $1-\delta$: Replace record with junk

For each person → δ: Keep record

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

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

Last Class	Today
→ Examples of DP for practically relevant queries → Exponential Mechanism	→ Gaussian Mechanism (Approx. DP) → Private ERM.

1. Assignment 3 grades are up.
2. Assignment 4 will be up today by 6PM.

Approx DP (ε, δ)-DP if $M: \mathcal{X}^n \to \mathcal{Y}$

$\forall$ neighboring databases x, x' and all $t \in \mathcal{Y}$,

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

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

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

Gaussian Mechanism.

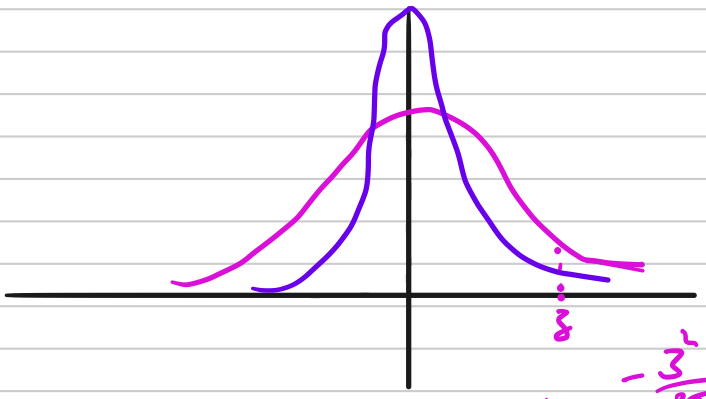
→ X

→ $f: \mathcal{X}^n \to \mathbb{R}^k$

→ Sample a noise vector $Z_1, \ldots, Z_k$ that are i.i.d $N(0, \sigma^2)$

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

(Recall: Gaussian

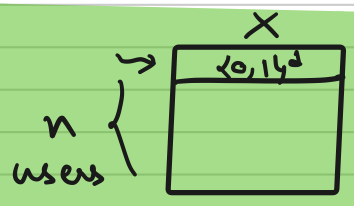


$\Pr[z=\delta] = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{z^2}{2\sigma^2}}$.

Thm: Let $f: X^n \to \mathbb{R}^k$. Then Gaussian mechanism with $\sigma = \sqrt{2\ln\left(\frac{1.25}{\delta}\right)} \cdot \frac{S_2(f)}{\varepsilon}$ ensures (ε, δ)-DP.

Really think of this as $\sigma \approx c \cdot \sqrt{\ln\left(\frac{1}{\delta}\right)} \cdot \frac{S_2(f)}{\varepsilon}$.

Example:



$f(x) = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (\in \mathbb{R}^d)$.

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

$S_2(f) = \max_{x, x' \text{ neighboring}} \|f(x) - f(x')\|_2 = \frac{\sqrt{d}}{n}$.

$\left(\left\|\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)\right\|_2\right)$

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

Laplacian	"Mean Release"	Gaussian
→ Noise $\approx \frac{d}{n}$		Noise $\approx \frac{\sqrt{d}}{n}$
↝ Error: $\frac{d}{n}$ per coordinate		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

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

(Recall for Pure DP, we had

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

Recall: Suppose $M_1, \ldots, M_k$ are ε-Differentially Private. Then, their Composition $\underbrace{M_1 \circ M_2 \circ \cdots \circ M_k}$ is $k \cdot \varepsilon$-DP.

"Adaptive Queries".

Basic Composition Theorem for Approx DP:
If $M_1, M_2, \ldots M_k$ are (ε, δ)- DP adaptive queries, then the composition is $(k\varepsilon, k\delta)$- DP.

Advanced Composition Theorem:
If $M_1, \ldots M_k$ are ε- DP mechanisms for answering adaptive queries. Then, the composition is $\left(\varepsilon\sqrt{2k\ln\left(\frac{1}{\delta}\right)} + k\varepsilon^2, \delta\right)$- DP
(as long as $\varepsilon < 1$).

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

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

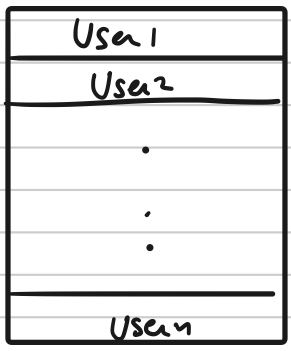
"A Reliable Leaderboard for Machine Learning Competitions".

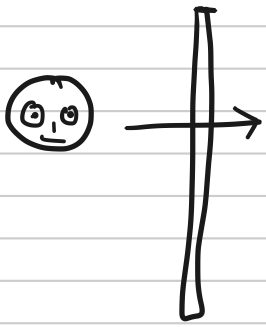
Advanced Composition Theorem:
If $M_1, \ldots M_k$ are (ε, δ') DP mechanisms for

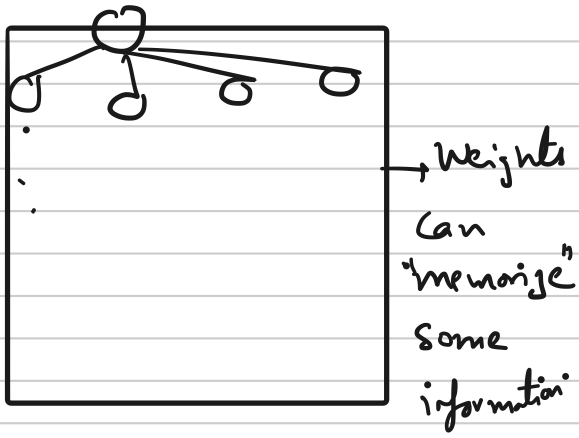
answering adaptive queries. Then, the Composition is $\left(\varepsilon\sqrt{2k\ln\left(\frac{1}{\delta}\right)} + k\varepsilon^2, k\delta' + \delta\right)$ – DP

(as long as $\varepsilon < 1$).

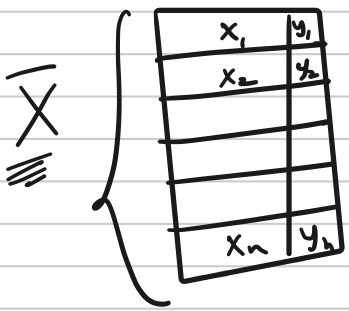
PRIVATE ML







PRIVATE EMPERICAL RISK MINIMIZATION



Parametric family of predictors $h_\theta : X \to Y$.

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

The algorithm gets

The loss that ... on predicting y_i from x_i.

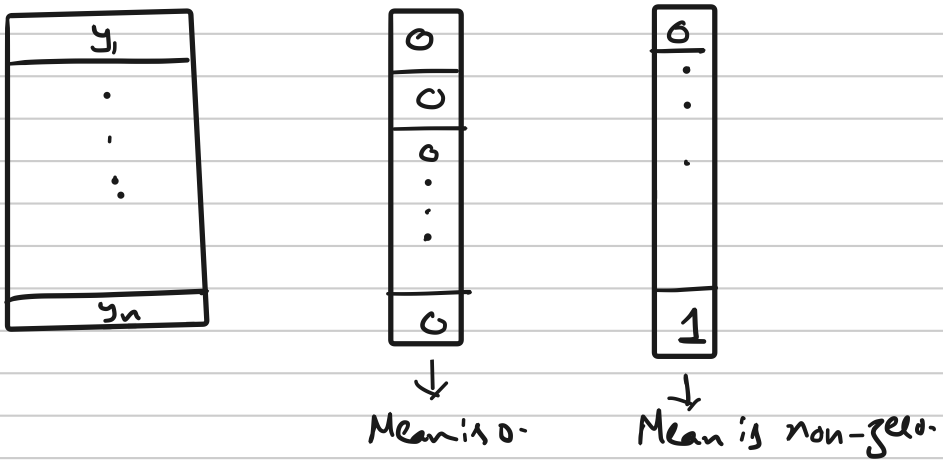
$$\theta^* = \arg\min_{\theta} L(\theta). \quad = f(\bar{X}).$$

We have to answer query f privately.

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

Problem: Sensitivity of θ^* is very high.

Example:



Prediction family is just constants.

loss is squared loss:

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

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

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

Three approaches for Private ERM

(a) . Output Perturbation → . Compute θ^* . Add noise.

(b) "Objective" Perturbation → . $\tilde{L}(\theta) = L(\theta) + \langle b, \theta \rangle$ ("Some noise vector").

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

"Privacy guarantees depend on exact solutions.

(c) "Gradient" Perturbation

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

Private Stochastic Gradient Descent.

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

SGD

→ Pick a start θ_0

→ For $t = 1, \ldots, T$:

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

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

(η_t: "Step-size")

DP-SGD

DP-SGD

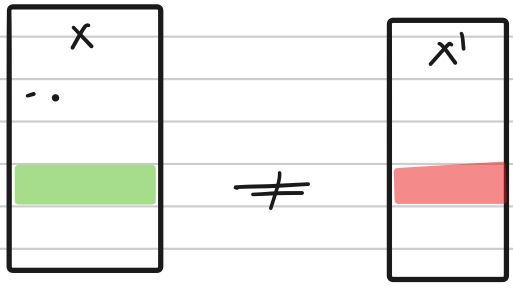
→ Pick a start θ_0

→ For $t = 1, \cdots, T$:

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

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

G_t typically Gaussian

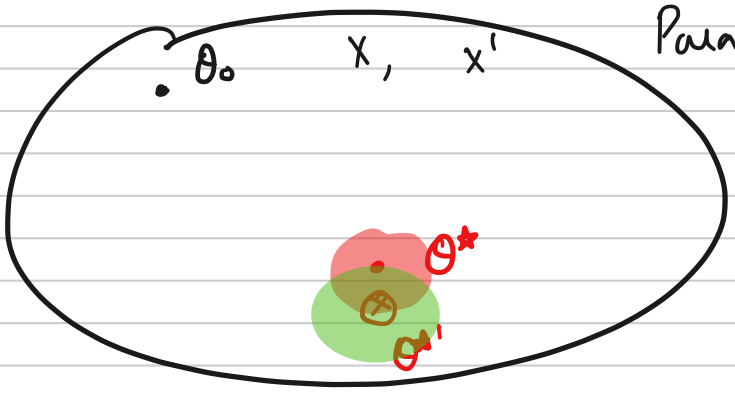


θ_0 θ_0

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.



Idea: Since SGD was "randomized" to begin with adding a bit of noise does not hurt too much.

1. How much noise needs to be added to ensure (ε, δ)-DP.

   Depends on how many iterations of SGD you are going to use?

   Depends on "sensitivity" of gradient function".

Advanced Composition is really helpful.

Suppose we were looking to solve

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

SGD for L convex, 1-Lipschitz and $\|\theta_0 - \theta^*\| \leq 1$

→ Then running "true SGD" for $T = O(1/\alpha^2)$ iterations gives $L(\cdot) - L(\theta^*) \leq \alpha$

↳ accuracy

THEOREM:

We can use Gaussian Mechanism + DP-SGD to get (ε, δ)-DP and accuracy α if $n > \frac{\sqrt{d}\sqrt{\log(1/\delta)}}{\varepsilon \cdot \alpha^2}$.

n: size of the database

REMARKS:

1. Privacy is important and ad-hoc solutions don't work.
2. We need to quantify privacy: DP is one of the best ways to do so.
3. Companies now use DP in aggregating data.
4. Many topics we did not cover...