Differential Privacy

A. M. Sadeghzadeh, Ph.D.

Sharif University of Technology
Computer Engineering Department (CE)
Data and Network Security Lab (DNSL)

December 29, 2024

Today's Agenda

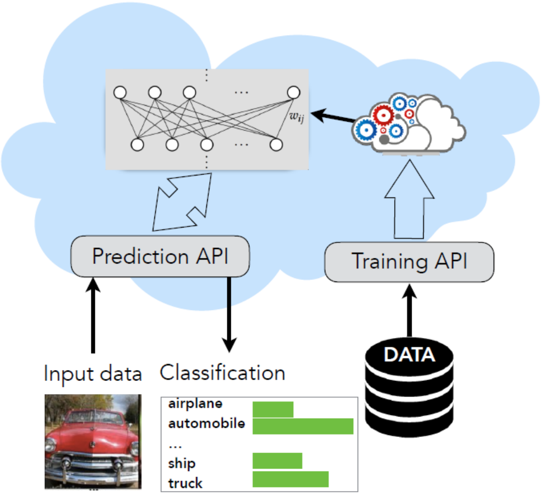
Recap

Machine Learning as a Service

Machine Learning as a Service

Google Cloud Platform

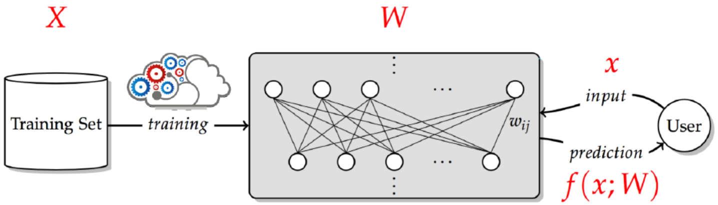
amazon web services™



(Shokri, 2020)

Privacy Risks in Machine Learning

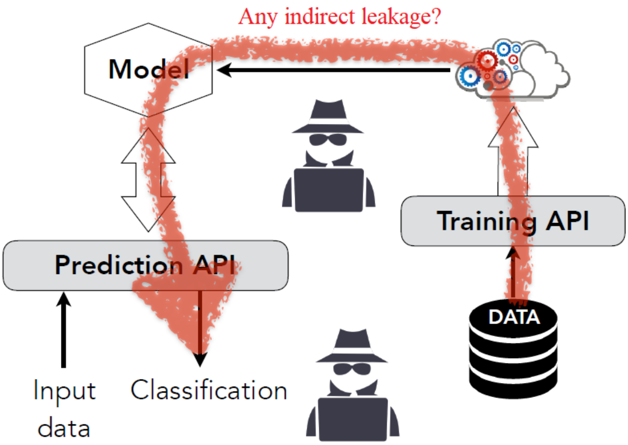
- What is training data leakage? Inferring information about members of X, beyond what can be learned about its underlying distribution.



(Shokri, 2020)

Membership Inference Attack

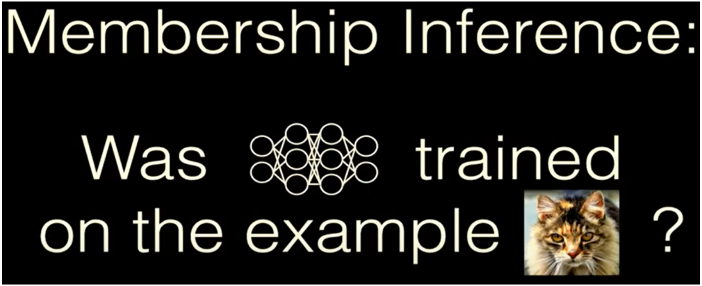
- Given a model, can an adversary infer whether data point x is part of its training set?



(Shokri, 2020)

Membership Inference Attack

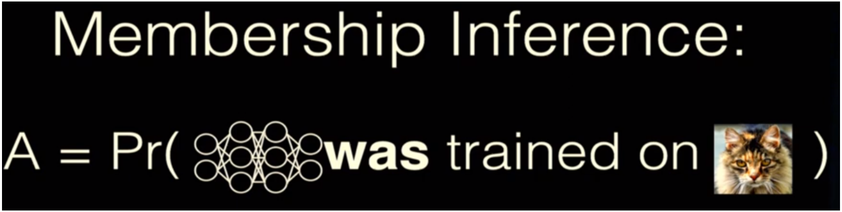
- Given a model, can an adversary infer whether data point x is part of its training set?



(Carlini, 2022)

Membership Inference Attack

- Given a model, can an adversary infer whether data point x is part of its training set?



(Carlini, 2022)

Differential Privacy

Differential Privacy

source: http://www.recode.net/2016/6/15/11940908/mossberg-apple-is-still-a-world-of-its-own

Differential Privacy

Differential privacy describes a promise, made by a data holder, or curator, to a data subject:

C. Dwork and A. Roth

You will **not be affected**, adversely or otherwise, by **allowing your data to be used in any study or analysis**, no matter what other studies, data sets, or information sources, are available.

Differential Privacy

Differential privacy describes a promise, made by a data holder, or curator, to a data subject:

C. Dwork and A. Roth

You will **not be affected**, adversely or otherwise, by **allowing your data to be used in any study or analysis**, no matter what other studies, data sets, or information sources, are available.

Differential privacy addresses the paradox of **learning nothing about an individual** while **learning useful information about a population**.

Privacy

A medical database may teach us that smoking causes cancer. Has an smoker been harmed by the analysis?

- Perhaps
  - affecting an insurance company's view of a smoker's long-term medical costs.
  - He may also be helped — learning of his health risks

Privacy

A medical database may teach us that smoking causes cancer. Has an smoker been harmed by the analysis?

- Perhaps
  - affecting an insurance company's view of a smoker's long-term medical costs.
  - He may also be helped — learning of his health risks

It is certainly the case that **more is known about him after the study** than was known before, but was his information **leaked**?

- Differential privacy will take the view that **it was not.**

Privacy

A medical database may teach us that smoking causes cancer. Has an smoker been harmed by the analysis?

- Perhaps
  - affecting an insurance company's view of a smoker's long-term medical costs.
  - He may also be helped — learning of his health risks

It is certainly the case that **more is known about him after the study** than was known before, but was his information **leaked**?

- Differential privacy will take the view that **it was not.**

Differential privacy: the impact on the smoker is the same **independent** of whether or not he was in the study.

- Differential privacy promises that the probability of harm/benefit **was not significantly increased** by their choice to participate.

Defining Private Data Analysis

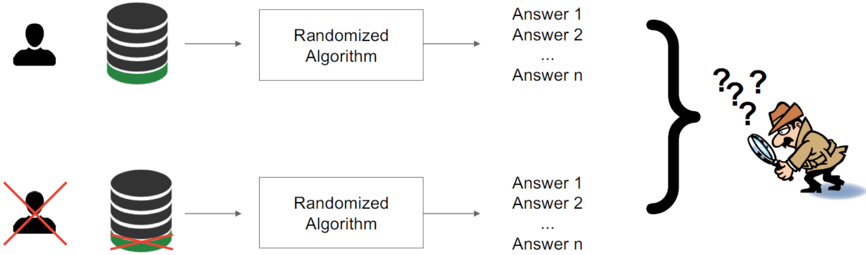
The analyst **knows no more about any individual** in the data set **after the analysis** is completed than she knew before the analysis was begun.

- The adversary's **prior and posterior views about an individual** (i.e., before and after having access to the database) **shouldn't be too different**"
  - Access to the database shouldn't change the adversary's views about any individual **too much**.

Differential Privacy

Differential Privacy ensures that any sequence of outputs (responses to queries) is **essentially equally likely** to occur, **independent** of the presence or absence of **any individual**.

- If **nothing is learned about an individual** then the individual **cannot be harmed by the analysis**.



(Papernot , 2019)

Randomized Response

A technique developed in the social sciences to collect **statistical information about embarrassing or illegal behavior**

Randomized Response

A technique developed in the social sciences to collect **statistical information about embarrassing or illegal behavior**

Study participants are told to report whether or not they have property P as follows:

1. Flip a coin.
2. If **tails**, then respond truthfully.
3. If **heads**, then flip a second coin and respond “Yes” if heads and “No” if tails.

Randomized Response

A technique developed in the social sciences to collect **statistical information about embarrassing or illegal behavior**

Study participants are told to report whether or not they have property P as follows:

1. Flip a coin.
2. If **tails**, then respond truthfully.
3. If **heads**, then flip a second coin and respond “Yes” if heads and “No” if tails.

Privacy comes from the **plausible deniability** of any outcome.

Randomized Response

A technique developed in the social sciences to collect **statistical information about embarrassing or illegal behavior**

Study participants are told to report whether or not they have property P as follows:

1. Flip a coin.
2. If **tails**, then respond truthfully.
3. If **heads**, then flip a second coin and respond "Yes" if heads and "No" if tails.

Privacy comes from the **plausible deniability** of any outcome.

Let p is the true fraction of participants having property P the expected number of "Yes" answers is

$$\mathbb{E}(\text{"}yes\text{"}|p) = \frac{1}{4}(1-p) + \frac{3}{4}p = \frac{1}{4} + \frac{p}{2} \rightarrow p = 2\mathbb{E}(\text{"}yes\text{"}|p) - \frac{1}{2}$$

Randomization Is Essential

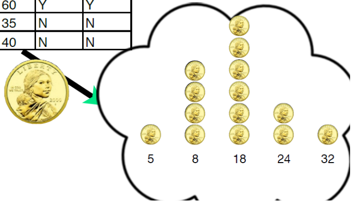
Suppose, for the sake of contradiction, that we have a non-trivial deterministic algorithm.

- Non-triviality says that there exists a query and two databases that yield different outputs under this query.

Changing one row at a time we see there exists a **pair of databases differing only in the value** of a single row, on which the **same query yields different outputs**.

- An adversary knowing that the database is one of these two almost identical databases learns the value of the data in the unknown row.

name	DOB	sex	weight	smoker	lung cancer
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/3/46	F	140	N	N
Ellen Jones	4/24/59	F	160	Y	Y
Jennifer Kim	3/1/70	F	135	N	N
Rachel Waters	9/5/43	F	140	N	N



(Katrina Ligett, 2017)

A Randomized Algorithm

In general, a randomized algorithm with domain A and (discrete) range B will be associated with a mapping from A to the probability simplex over B, denoted $\Delta(B)$

A Randomized Algorithm

In general, a randomized algorithm with domain A and (discrete) range B will be associated with a mapping from A to the probability simplex over B, denoted $\Delta(B)$

Definition 2.1 (Probability Simplex)**.** Given a discrete set B, the *probability simplex* over B, denoted $\Delta(B)$ is defined to be:

$$\Delta(B) = \left\{ x \in \mathbb{R}^{|B|} : x_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{|B|} x_i = 1 \right\}$$

A Randomized Algorithm

In general, a randomized algorithm with domain A and (discrete) range B will be associated with a mapping from A to the probability simplex over B, denoted $\Delta(B)$

Definition 2.1 (Probability Simplex). Given a discrete set B, the *probability simplex* over B, denoted $\Delta(B)$ is defined to be:

$$\Delta(B) = \left\{ x \in \mathbb{R}^{|B|} : x_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{|B|} x_i = 1 \right\}$$

Definition 2.2 (Randomized Algorithm). A randomized algorithm $\mathcal{M}$ with domain A and discrete range B is associated with a mapping $M : A \to \Delta(B)$. On input $a \in A$, the algorithm $\mathcal{M}$ outputs $\mathcal{M}(a) = b$ with probability $(M(a))_b$ for each $b \in B$. The probability space is over the coin flips of the algorithm $\mathcal{M}$.

Database

We will think of databases x as being collections of records from a universe $\mathcal{X}$. It will often be convenient to represent databases by their histograms: $x \in \mathbb{N}^{|\mathcal{X}|}$, in which each entry x_i represents the number of elements in the database x of type $i \in \mathcal{X}$.

name	DOB	sex	weight	smoker	lung cancer
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/3/46	F	140	N	N
Ellen Jones	4/24/59	F	160	Y	Y
Jennifer Kim	3/1/70	F	135	N	N
Rachel Waters	9/5/43	F	140	N	N

(Katrina Ligett, 2017)

Distance Between Databases

Definition 2.3 (Distance Between Databases)**.** The ℓ_1 norm of a database x is denoted $\|x\|_1$ and is defined to be:

$$\|x\|_1 = \sum_{i=1}^{|\mathcal{X}|} |x_i| \, .$$

The ℓ_1 distance between two databases x and y is $\|x - y\|_1$

Note that $\|x\|_1$ is a measure of the *size* of a database x (i.e., the number of records it contains), and $\|x - y\|_1$ is a measure of how many records *differ* between x and y.

Differential Privacy

Differential Privacy: A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ,δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \approx_\epsilon \Pr[\mathcal{M}(y) \in \mathcal{S}]$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$.

Differential Privacy

Differential Privacy: A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ,δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\text{Pr}[\mathcal{M}(x) \in \mathcal{S}] \leq (1 + \epsilon)\text{Pr}[\mathcal{M}(y) \in \mathcal{S}]$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$.

Differential Privacy

Differential Privacy: A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ,δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\epsilon)\Pr[\mathcal{M}(y) \in \mathcal{S}]$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$.

Differential Privacy

Differential Privacy: A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ,δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\epsilon)\Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$.

Differential Privacy

Differential Privacy: A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ,δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\epsilon)\Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$.

- If $\delta = 0$, we say that $\mathcal{M}$ is ϵ-differentially private.

Differential Privacy

Differential Privacy: A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ,δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\epsilon)\Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$.

- If $\delta = 0$, we say that $\mathcal{M}$ is ϵ-differentially private.
- δ is a negligible function.
  - δ that are less than the inverse of any polynomial in the size of the database.

Differential Privacy

Differential Privacy: A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ,δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\text{Pr}[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\epsilon)\text{Pr}[\mathcal{M}(y) \in \mathcal{S}] + \delta$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$.

- If $\delta = 0$, we say that $\mathcal{M}$ is ϵ-differentially private.
- δ is a negligible function.
  - δ that are less than the inverse of any polynomial in the size of the database.
- The definition is symmetric.

Differential Privacy

Differential Privacy: A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ,δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\epsilon)\Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$.

- If $\delta = 0$, we say that $\mathcal{M}$ is ϵ-differentially private.
- δ is a negligible function.
  - δ that are less than the inverse of any polynomial in the size of the database.
- The definition is symmetric.
- Differential privacy **is a definition, not an algorithm**.
  - For a given computational task T and a given value of ϵ **there will be many differentially private algorithms** for achieving T in an ϵ-differentially private manner. **Some will have better accuracy than others.**

Differential Privacy

Differential Privacy: A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ,δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

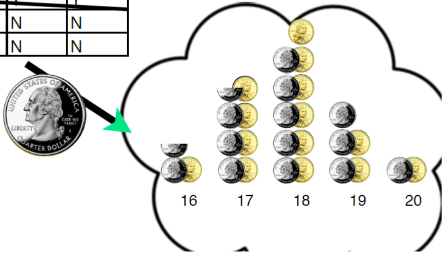
$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\epsilon)\Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$.

- If $\delta = 0$, we say that $\mathcal{M}$ is ϵ-differentially private.
- δ is a negligible function.
  - δ that are less than the inverse of any polynomial in the size of the database.
- The definition is symmetric.
- Differential privacy **is a definition, not an algorithm**.
  - For a given computational task T and a given value of ϵ **there will be many differentially private algorithms** for achieving T in an ϵ-differentially private manner. **Some will have better accuracy than others.**
- $\epsilon \approx 1$ and $\delta \ll \frac{1}{N}$ (generally speaking, one digit ϵ is good) where N is the size of dataset.

Differential Privacy

name	DOB	sex	weight	smoker	lung cancer
John Doe	12/1/51	M	185	Y	N
Jane Smith	3/3/46	F	140	N	N
~~Ellen Jones~~	~~4/24/59~~	~~F~~	~~160~~	~~Y~~	~~Y~~
Jennifer Kim	3/1/70	F	135	N	N
Rachel Waters	9/5/43	F	140	N	N

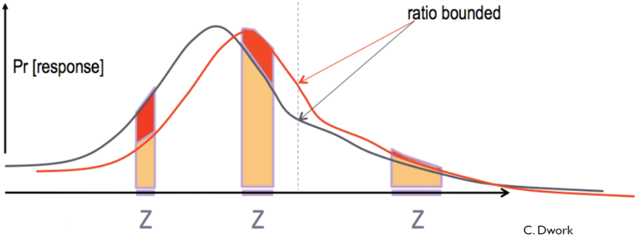


(Katrina Ligett, 2017)

Differential Privacy

ϵ-differential privacy

$$\Pr[M(x_1) \in S] \leq e^{\varepsilon} \Pr[M(x_2) \in S]$$

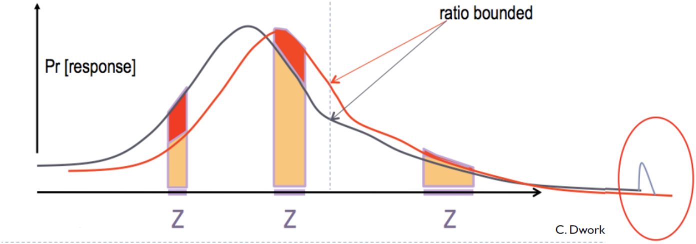


(Katrina Ligett, 2017)

Differential Privacy

(ϵ, δ)-differential privacy

$$\Pr[M(x_1) \in S] \leq e^{\varepsilon} \Pr[M(x_2) \in S] + \delta$$



(Katrina Ligett, 2017)

Privacy Loss

(ϵ, δ)-differential privacy ensures that for all adjacent x, y, the **absolute value of the privacy loss will be bounded by** ϵ with probability at least $1 - \delta$.

The quantity

$$\mathcal{L}^{(\xi)}_{\mathcal{M}(x)\|\mathcal{M}(y)} = \ln\left(\frac{\Pr[\mathcal{M}(x) = \xi]}{\Pr[\mathcal{M}(y) = \xi]}\right)$$

is important to us; we refer to it as the *privacy loss* incurred by observing ξ. This loss might be positive (when an event is more likely under x than under y) or it might be negative (when an event is more likely under y than under x).

What is the privacy loss of Randomized response mechanism?

Claim 3.5. The version of randomized response described above is $(\ln 3, 0)$-differentially private.

What is the privacy loss of Randomized response mechanism?

Claim 3.5. The version of randomized response described above is $(\ln 3, 0)$-differentially private.

Proof. Fix a respondent. A case analysis shows that $\Pr[\text{Response} = \text{Yes}|\text{Truth} = \text{Yes}] = 3/4$. Specifically, when the truth is "Yes" the outcome will be "Yes" if the first coin comes up tails (probability $1/2$) or the first and second come up heads (probability $1/4$)), while $\Pr[\text{Response} = \text{Yes}|\text{Truth} = \text{No}] = 1/4$ (first comes up heads and second comes up tails; probability $1/4$). Applying similar reasoning to the case of a "No" answer, we obtain:

$$\frac{\Pr[\text{Response} = \text{Yes}|\text{Truth} = \text{Yes}]}{\Pr[\text{Response} = \text{Yes}|\text{Truth} = \text{No}]}$$

$$= \frac{3/4}{1/4} = \frac{\Pr[\text{Response} = \text{No}|\text{Truth} = \text{No}]}{\Pr[\text{Response} = \text{No}|\text{Truth} = \text{Yes}]} = 3.$$

$\square$

Laplace Mechanism

Sensitivity

The ℓ_1 sensitivity of a function f captures the magnitude by which a single individual's data can change the function f in the worst case

- Intuitively, the uncertainty in the response that we must introduce in order to hide the participation of a single individual.

The sensitivity of a function gives an upper bound on how much we must perturb its output to preserve privacy.

Definition 3.1 (ℓ_1-sensitivity). The ℓ_1-sensitivity of a function $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$ is:

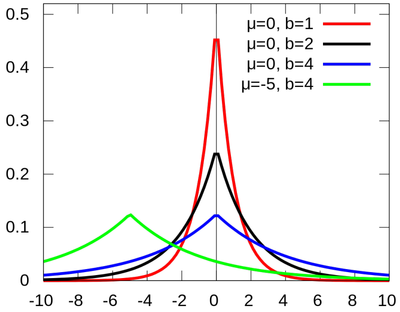
$$\Delta f = \max_{\substack{x,y \in \mathbb{N}^{|\mathcal{X}|} \\ \|x-y\|_1 = 1}} \|f(x) - f(y)\|_1.$$

The Laplace Distribution

Definition 3.2 (The Laplace Distribution). The Laplace Distribution (centered at 0) with scale b is the distribution with probability density function:

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

The variance of this distribution is $\sigma^2 = 2b^2$. We will sometimes write $\text{Lap}(b)$ to denote the Laplace distribution with scale b.



The Laplace Mechanism

Definition 3.3 (The Laplace Mechanism). Given any function $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$, the Laplace mechanism is defined as:

$$\mathcal{M}_L(x, f(\cdot), \varepsilon) = f(x) + (Y_1, \ldots, Y_k)$$

where Y_i are i.i.d. random variables drawn from $\mathrm{Lap}(\Delta f/\varepsilon)$.

The Laplace Mechanism

Theorem 3.6. The Laplace mechanism preserves $(\varepsilon, 0)$-differential privacy.

The Laplace Mechanism

Theorem 3.6. The Laplace mechanism preserves $(\varepsilon, 0)$-differential privacy.

Proof. Let $x \in \mathbb{N}^{|\mathcal{X}|}$ and $y \in \mathbb{N}^{|\mathcal{X}|}$ be such that $\|x - y\|_1 \leq 1$, and let $f(\cdot)$ be some function $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$. Let p_x denote the probability density function of $\mathcal{M}_L(x, f, \varepsilon)$, and let p_y denote the probability density function of $\mathcal{M}_L(y, f, \varepsilon)$. We compare the two at some arbitrary point $z \in \mathbb{R}^k$

$$\frac{p_x(z)}{p_y(z)} = \prod_{i=1}^{k} \left(\frac{\exp(-\frac{\varepsilon|f(x)_i - z_i|}{\Delta f})}{\exp(-\frac{\varepsilon|f(y)_i - z_i|}{\Delta f})} \right) = \prod_{i=1}^{k} \exp\left(\frac{\varepsilon(|f(y)_i - z_i| - |f(x)_i - z_i|)}{\Delta f} \right)$$

$$\leq \prod_{i=1}^{k} \exp\left(\frac{\varepsilon|f(x)_i - f(y)_i|}{\Delta f} \right) = \exp\left(\frac{\varepsilon \cdot \|f(x) - f(y)\|_1}{\Delta f} \right) \leq \exp(\varepsilon),$$

where the first inequality follows from the triangle inequality, and the last follows from the definition of sensitivity and the fact that $\|x - y\|_1 \leq 1$. That $\frac{p_x(z)}{p_y(z)} \geq \exp(-\varepsilon)$ follows by symmetry. $\square$

Example: Counting Queries

Counting queries are queries of the form "**How many elements in the database satisfy Property** P?"

Example: Counting Queries

Counting queries are queries of the form "**How many elements in the database satisfy Property** P?"

- The sensitivity of a counting query is 1

Example: Counting Queries

Counting queries are queries of the form "**How many elements in the database satisfy Property** P?"

- The sensitivity of a counting query is 1
- $(\epsilon, 0)$-differential privacy can be achieved for counting queries by the addition of noise scaled to $1/\epsilon$, that is, by adding noise drawn from $\text{Lap}(1/\epsilon)$.

Example: Counting Queries

Counting queries are queries of the form "**How many elements in the database satisfy Property** P?"

- The sensitivity of a counting query is 1
- $(\epsilon, 0)$-differential privacy can be achieved for counting queries by the addition of noise scaled to $1/\epsilon$, that is, by adding noise drawn from $\mathrm{Lap}(1/\epsilon)$.

A fixed but arbitrary list of m **counting queries** can be viewed as a vector-valued query.

- The sensitivity is m

Example: Counting Queries

Counting queries are queries of the form "**How many elements in the database satisfy Property** P?"

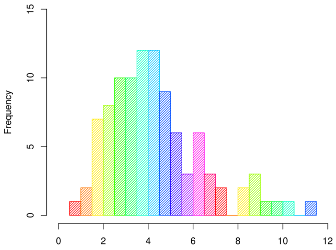
- The sensitivity of a counting query is 1
- $(\epsilon, 0)$-differential privacy can be achieved for counting queries by the addition of noise scaled to $1/\epsilon$, that is, by adding noise drawn from $\text{Lap}(1/\epsilon)$.

A fixed but arbitrary list of m **counting queries** can be viewed as a vector-valued query.

- The sensitivity is m
- $(\epsilon, 0)$-differential privacy can be achieved by adding noise scaled to m/ϵ to the true answer to each query.

Histogram Queries

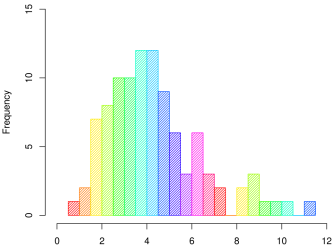
In this type of query the universe $\mathbb{N}^{|\mathcal{X}|}$ is **partitioned into cells**, and the query asks how many database elements lie in each of the cells.



Histogram Queries

In this type of query the universe $\mathbb{N}^{|\mathcal{X}|}$ is **partitioned into cells**, and the query asks how many database elements lie in each of the cells.

- Because the cells are disjoint, the addition or removal of a single database element can affect the count in exactly one cell. Hence the **sensitivity is 1**.
- $(\epsilon, 0)$-differential privacy can be achieved by adding noise scaled to $1/\epsilon$, that is, by adding noise drawn from $\mathrm{Lap}(1/\epsilon)$ to the true count in each cell.



Report Noisy Max

Report Noisy Max. Consider the following simple algorithm to determine which of m counting queries has the highest value: Add independently generated Laplace noise $\text{Lap}(1/\varepsilon)$ to each count and return the index of the largest noisy count (we ignore the possibility of a tie). Call this algorithm Report Noisy Max.

Claim 3.9. The Report Noisy Max algorithm is $(\varepsilon, 0)$-differentially private.

The Accuracy of the Laplace Mechanism

Fact 3.7. If $Y \sim \text{Lap}(b)$, then:

$$\Pr[|Y| \geq t \cdot b] = \exp(-t).$$

Theorem 3.8. Let $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$, and let $y = \mathcal{M}_L(x, f(\cdot), \varepsilon)$. Then $\forall \delta \in (0, 1]$:

$$\Pr\left[\|f(x) - y\|_\infty \geq \ln\left(\frac{k}{\delta}\right) \cdot \left(\frac{\Delta f}{\varepsilon}\right)\right] \leq \delta$$

Proof. We have:

$$\begin{aligned}
\Pr\left[\|f(x) - y\|_\infty \geq \ln\left(\frac{k}{\delta}\right) \cdot \left(\frac{\Delta f}{\varepsilon}\right)\right] &= \Pr\left[\max_{i \in [k]} |Y_i| \geq \ln\left(\frac{k}{\delta}\right) \cdot \left(\frac{\Delta f}{\varepsilon}\right)\right] \\
&\leq k \cdot \Pr\left[|Y_i| \geq \ln\left(\frac{k}{\delta}\right) \cdot \left(\frac{\Delta f}{\varepsilon}\right)\right] \\
&= k \cdot \left(\frac{\delta}{k}\right) \\
&= \delta
\end{aligned}$$

where the second to last inequality follows from the fact that each $Y_i \sim \text{Lap}(\Delta f / \varepsilon)$ and Fact 3.7. $\square$

Example: First Names

Suppose we wish to calculate which first names, from a list of 10,000 potential names, were the most common among participants of the 2010 census.

- This question can be represented as a query $f : \mathbb{N}^{|\mathcal{X}|} \rightarrow \mathbb{R}^{10000}$

Example: First Names

Suppose we wish to calculate which first names, from a list of 10,000 potential names, were the most common among participants of the 2010 census.

- This question can be represented as a query $f : \mathbb{N}^{|\mathcal{X}|} \rightarrow \mathbb{R}^{10000}$
- This is a histogram query, and so has sensitivity $\Delta f = 1$

Example: First Names

Suppose we wish to calculate which first names, from a list of 10,000 potential names, were the most common among participants of the 2010 census.

- This question can be represented as a query $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^{10000}$
- This is a histogram query, and so has sensitivity $\Delta f = 1$

We can simultaneously calculate the frequency of all 10000 names with $(\epsilon = 1, \delta = 0)$-differential privacy $(y = \mathcal{M}(\mathbb{N}^{|\mathcal{X}|}, f))$, and with probability 95% $(\delta = 0.05)$.

Example: First Names

Suppose we wish to calculate which first names, from a list of 10,000 potential names, were the most common among participants of the 2010 census.

- This question can be represented as a query $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^{10000}$
- This is a histogram query, and so has sensitivity $\Delta f = 1$

We can simultaneously calculate the frequency of all 10000 names with $(\epsilon = 1, \delta = 0)$-differential privacy $(y = \mathcal{M}(\mathbb{N}^{|\mathcal{X}|}, f))$, and with probability 95% $(\delta = 0.05)$.

$$Pr[\|f(x) - y\|_\infty \geq \ln(\frac{k}{\delta}).(\frac{\Delta f}{\epsilon})] \leq \delta$$

$$Pr[\|f(x) - y\|_\infty \geq \ln(\frac{10000}{0.05}).(\frac{1}{1})] \leq 0.05$$

$$Pr[\|f(x) - y\|_\infty \geq 12.2] \leq 0.05$$

Hence, no estimate will be off by more than an additive error of $\ln(10000/.05) \approx 12.2$, with probability 95%.

- That's pretty low error for a nation of more than 300,000,000 people!

The Gaussian Mechanism

Let $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^d$ be an arbitrary d-dimensional function, and define its ℓ_2 sensitivity to be $\Delta_2 f = \max_{\text{adjacent}x,y} \|f(x) - f(y)\|_2$. The *Gaussian Mechanism with parameter* σ adds noise scaled to $\mathcal{N}(0, \sigma^2)$ to each of the d components of the output.

Theorem A.1. Let $\varepsilon \in (0, 1)$ be arbitrary. For $c^2 > 2\ln(1.25/\delta)$, the Gaussian Mechanism with parameter $\sigma \geq c\Delta_2 f/\varepsilon$ is (ε, δ)-differentially private.

References

- Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, (Ch. 2).
- Gautam Kamath, CS 860, Fall 2020, University of Waterloo