Intro to Differential Privacy

CSE 484

Wilson Tang, wtang06@cs.washington.edu

Andrew Wei, nowei@cs.washington.edu

Dao Yi, daoyee@cs.washington.edu

Motivation

2003 - Dinur-Nissim attack

- With enough accurate data, attacker can reconstruct almost the entire underlying dataset

We want our data **anonymized** while still **useful**?

Reconstruction Attacks
- Census Reconstruction

Setup:

- (age, sex, race)
- Only know aggregated data
- (D)s are suppressed data to protect

Assume age is in [1,125], then for any three people, there are C(125,3)≈300k combinations.

		AGE		
STATISTIC	GROUP	COUNT	MEDIAN	MEAN
1A	total population	7	30	38
2A	female	4	30	33.5
2B	male	3	30	44
2C	black or African American	4	51	48.5
2D	white	3	24	24
3A	single adults	(D)	(D)	(D)
3B	married adults	4	51	54
4A	black or African American female	3	36	36.7
4B	black or African American male	(D)	(D)	(D)
4C	white male	(D)	(D)	(D)
4D	white female	(D)	(D)	(D)
5A	persons under 5 years	(D)	(D)	(D)
5B	persons under 18 years	(D)	(D)	(D)
5C	persons 64 years or over	(D)	(D)	(D)

Note: Married persons must be 15 or over

Reconstruction Attacks
- Census Reconstruction

Add constraints to male group:

- Count=3
- Median=40, must one male aged 30
- Mean=44
- Assume ages in [0,125]

WIth this there's 30 possibilities

A	B	C
1	30	101
2	30	100
3	30	99
4	30	98
5	30	97
…		
25	30	77
26	30	76
27	30	75
28	30	74
29	30	73
30	30	72

		AGE		
STATISTIC	GROUP	COUNT	MEDIAN	MEAN
1A	total population	7	30	38
2A	female	4	30	33.5
2B	male	3	30	44
2C	black or African American	4	51	48.5
2D	white	3	24	24
3A	single adults	(D)	(D)	(D)
3B	married adults	4	51	54
4A	black or African American female	3	36	36.7
4B	black or African American male	(D)	(D)	(D)
4C	white male	(D)	(D)	(D)
4D	white female	(D)	(D)	(D)
5A	persons under 5 years	(D)	(D)	(D)
5B	persons under 18 years	(D)	(D)	(D)
5C	persons 64 years or over	(D)	(D)	(D)

Note: Married persons must be 15 or over

Reconstruction Attacks
- Census Reconstruction

With more constraints, we can extract more information, eventually reconstruct dataset

A SAT problem (NP-Hard)

Anyone can perform such attack using a SAT-Solver

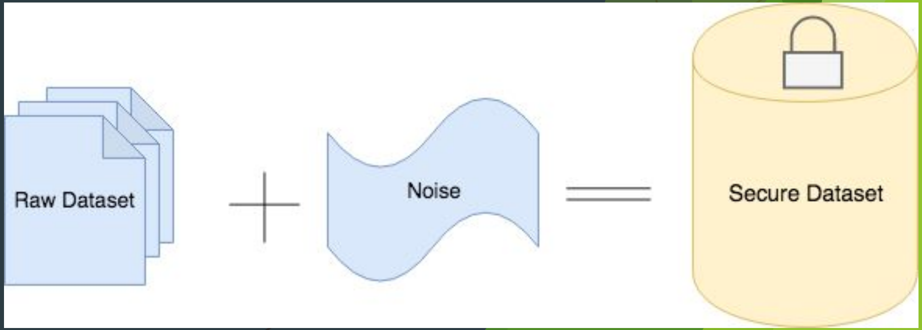
What is Differential Privacy?

We want our data **<u>anonymized</u>** while still **<u>useful</u>**.

- Anonymized: Cannot determine if an individual is in the dataset; Cannot reconstruct
- Useful: Data still represent information faithfully

What we do?

- Add noise to dataset in a way that
- An adversary cannot tell if any individual data were changed arbitrarily



ε-differential privacy [Dwork, McSherry, Nissim, Smith 2006]

Differential Privacy:

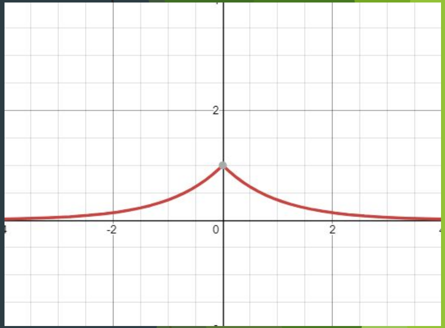
$$\Pr[A(D_1) = t] \leq \exp(\varepsilon) \cdot \Pr[A(D_2) = t]$$

Sensitivity:

$$\Delta f = \max \| f(D_1) - f(D_2) \|_1$$

To get ε-differential privacy, we can add noise, y, according to:

$$y \sim \mathrm{Lap}(0, \Delta f/\varepsilon) \qquad \Pr[y] \propto \exp(-\varepsilon |y| / \Delta f)$$



Lap(0, 1)

A is randomized algorithm that takes a dataset as input

f is a function

D_1, D_2 are datasets that differ in one element

t is the result of a query to the statistical database

ε is the privacy loss parameter/the privacy budget, a positive real number

Lap is the Laplace distribution

Example

Suppose that $\mathbf{x} \in \{-1, 0, 1, \ldots, 59, 60\}^n$

Note that $|\mathbf{x}| = n$

$f(\mathbf{x}) = \mathrm{sum}(\mathbf{x}) = \Sigma_i(x_i)$

$\Delta f = \max \| f(D_1) - f(D_2) \|_1 = 61$

Let us choose $\varepsilon = 2$.

Then $A(\mathbf{x}) = f(\mathbf{x}) + Y$, where $Y \sim \mathrm{Lap}(0, 61/2)$

So $A(x)$ is 2-differentially private

We get that $f(\mathbf{x}) = 71$

x
1
3
4
4
3
55
1

Possible results $A(\mathbf{x})$
82.4386
66.1222
79.5883
66.4617
51.2372
101.7916
64.7038
…

Deep Learning with Differential Privacy

DL models train on a LOT of data, sometimes sensitive

DL models can “memorize” training data

Differential privacy adds noise to the model so it won’t learn the exact data

Abadi et. al (2016) implements differentially private SGD

Algorithm 1 Differentially private SGD (Outline)

Input: Examples $\{x_1, \ldots, x_N\}$, loss function $\mathcal{L}(\theta) = \frac{1}{N}\sum_i \mathcal{L}(\theta, x_i)$. Parameters: learning rate η_t, noise scale σ, group size L, gradient norm bound C.

Initialize θ_0 randomly

for $t \in [T]$ **do**

Take a random sample L_t with sampling probability L/N

Compute gradient

For each $i \in L_t$, compute $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i)$

Clip gradient

$\bar{\mathbf{g}}_t(x_i) \leftarrow \mathbf{g}_t(x_i) / \max\left(1, \frac{\|\mathbf{g}_t(x_i)\|_2}{C}\right)$

Add noise

$\tilde{\mathbf{g}}_t \leftarrow \frac{1}{L}\left(\sum_i \bar{\mathbf{g}}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I})\right)$

Descent

$\theta_{t+1} \leftarrow \theta_t - \eta_t \tilde{\mathbf{g}}_t$

Output θ_T and compute the overall privacy cost (ε, δ) using a privacy accounting method.

Future work +
Problems that still exist

Repeated queries will reveal the underlying data

Higher privacy budget can mitigate this problem, but it still exists

Groups of attackers can “plan” these attacks

Value	Occurrences
1	18
2	23
3	35
4	28
5	20

Legal/Ethics: Responsibility

If a company implements differential privacy on a dataset, but the dataset still becomes de-anonymized, should the company take responsibility for it?

- Legally, can they be sued?
- Ethically, is it their responsibility to protect the people?

AOL 2006 data breach - $5,000+ compensation

Manifest-no refusal 6 - “resist the market-driven force to commodify the human experience”

Legal/Ethics:
Learning on anonymized data

Even if data is anonymized, is it ethical to train ML models on personal data?

- Will people have a say as to how their data is used?
- Are people comfortable with their data being used that way?
- Can anonymized users remove their data?

Conclusion

Differential Privacy:

- Provides a mathematical definition of privacy loss
- Helps us keep data private
- Lets us learn from data in a privacy-preserving manner

References

- Irit Dinur and Kobbi Nissim.(2003) Revealing information while preserving privacy. In Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems ,PODS '03,. ACM, New York, NY, USA, 202–210
- Dwork C., McSherry F., Nissim K., Smith A. (2006) Calibrating Noise to Sensitivity in Private Data Analysis. In: Halevi S., Rabin T. (eds) Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science, vol 3876. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11681878_14
- Adabi M., Chu A., Goodfellow I., McMahan H., Mironov I., Talwar K., Zhang L. (2016) Deep Learning with Differential Privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security pp. 308-318, 2016. https://arxiv.org/abs/1607.00133
- Simson Garfinkel, John M. Abowd, and Christian Martindale. 2018. Understanding Database Reconstruction Attacks on Public Data: These attacks on statistical databases are no longer a theoretical danger. Queue 16, 5, Pages 50 (September-October 2018), 26 pages. https://doi.org/10.1145/3291276.3295691
- https://www.nytimes.com/2006/08/09/technology/09aol.html
- https://en.wikipedia.org/wiki/Differential_privacy
- https://www.manifestno.com/home