Protecting Privacy in ML

Introduction to Differential Privacy

Jesse Cresswell, PhD
Senior Machine Learning Scientist
Layer 6 AI at TD

Oct. 13 2021 - Big Data & AI Toronto

layer 6
AI at TD

Agenda

- What do we mean by privacy?
- Attempts at privacy
- Privacy as a resource
- Differential Privacy
- Applications

Layer 6 - Who We Are

layer 6
AI at TD

Layer 6

ML in finance, healthcare

Active research groups:

CV, NLP, Deep Gen, RL, RecSys, Privacy

What do we mean by privacy?

What do we mean by privacy?

Protecting people's privacy is a critical priority in healthcare and banking.

This term is often used, but rarely defined.

Canada has several laws about protecting "personal information" - data about an individual that can **identify the person**, on its own or combined with other data.

These laws cover collection, use, and disclosure of personal information, but they may not capture the spirit of what privacy should mean in data analysis.

Is there a rigorous way to define privacy, relevant to how we use data in ML?

What privacy is not.

Cryptography is often brought up in discussions on privacy.

After all, cryptography allows us to send “private” messages to trusted parties.

The Enigma machine, cryptocurrencies, and the darkweb are secured by cryptography.

What privacy is not.

Cryptography is often brought up in discussions on privacy.

After all, cryptography allows us to send “private” messages to trusted parties.

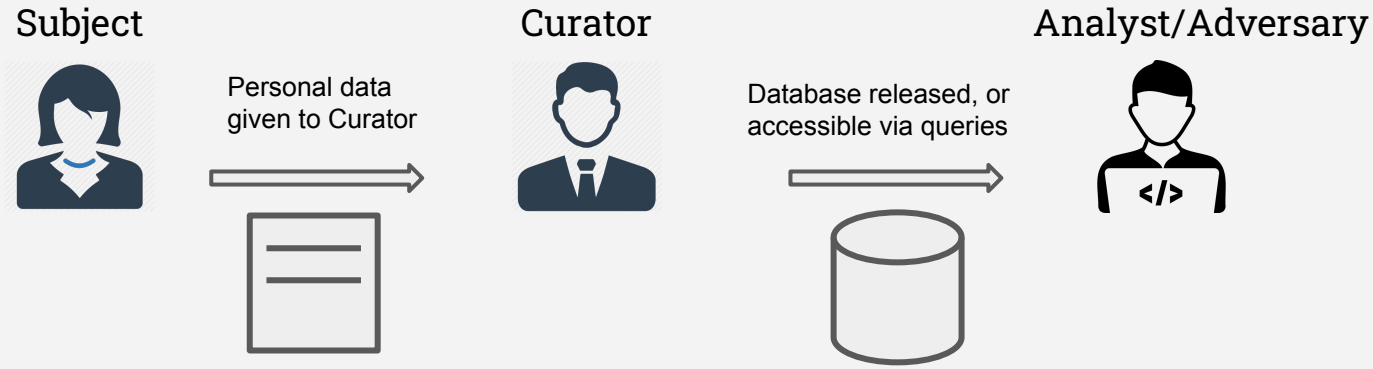
Cryptography is a **security** protocol. Our message is secure from prying eyes.
But the entire content of our message is revealed to the recipient.

Cryptography is **binary** - either all information is revealed, or none is.

Privacy is about protecting information belonging to individuals. In many cases we want to **reveal general information** publicly, but cannot expose personal data.

Setting for defining privacy

Subjects - individuals who give, or refuse to give, their personal data
Curator - trusted centralized party that holds data and controls the database
Analysts/Adversaries - access the database for good or for evil



Privacy is a promise from the curator to the subjects that **no harm** will come to them by including their data in the database.

layer 6
AI at TD

The promise of privacy

Curator's promise - *"You will not be affected, adversely or otherwise, by allowing your data to be used in the database, no matter what other data sets are available*."

Imagine an individual choosing to provide their data to a study on smoking.

Whether or not the individual gives their data, the study is likely to conclude that smoking causes health problems. Actions will be taken - insurers raise premiums on smokers, or our individual chooses to stop smoking.

However, it was **not the inclusion of the individual's data** that led to these outcomes, but the study as a whole. The promise was not broken.

> How can we learn something about a population,
> while not affecting individuals?

Attempts at privacy

Attempts at privacy - Anonymization

Some approaches are commonly used, but don't preserve the promise.

The curator may remove personal information, and publish the rest.

Banks and credit bureaus use anonymization to share customer information. Hospitals release data on medical images and diagnoses without personal information.

Name	Age	Job	Salary
[redacted]	23	Clerk	50,000
[redacted]	45	Driver	60,000
[redacted]	61	Lawyer	100,000

Attempts at privacy - Anonymization

Linkage Attacks

It can be possible to de-anonymize records using outside information.

Netflix released data on sparse user movie ratings - de-anonymized by comparing to public ratings on IMDB. [Narayanan & Shmatikov 2008]

Name	Age	Job	Salary
[redacted]	23	Clerk	50,000
[redacted]	45	Driver	60,000
[redacted]	61	Lawyer	100,000

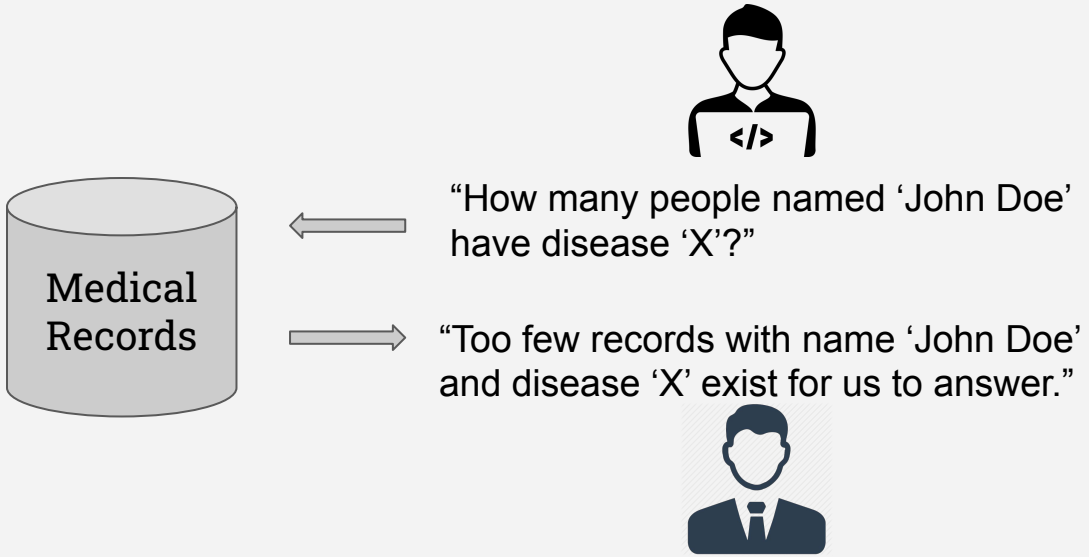
Name	Age	Job	Favorite Sport
Joe	[redacted]	Clerk	Squash
Jun	[redacted]	Driver	Soccer
Jaya	[redacted]	Lawyer	Hockey

layer 6
AI at TD

Attempts at privacy - Restrict Queries

The curator does not release the database, but will answer certain queries.

To avoid harm to individuals, queries will only be answered if they involve a large enough group.



layer 6
AI at TD

Attempts at privacy - Restrict Queries

Differencing Attacks

If an individual is known to be in the database then 2 queries over large sets will reveal info about the individual:

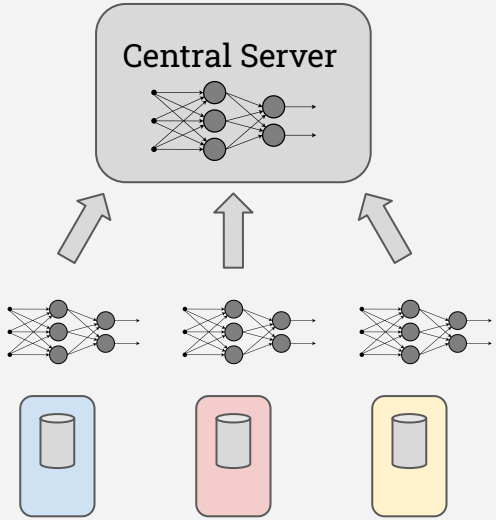
layer 6
AI at TD

Attempts at privacy - Federated Learning

Federated Learning is a distributed ML approach where data is not stored on a centralized server.

Instead, the model is trained on the device where data is collected (e.g. smartphone), and updates are aggregated by the central server.

Intuitively this seems “more private”, since the data is kept on the personal device where it is generated.



Clients with local, private data send model updates only.

layer 6
AI at TD

Attempts at privacy - Federated Learning

Memorization

Except, there are many examples of neural networks memorizing individual training examples which can be recovered by analyzing the model. [Fredrikson, Jha, Ristenpart 2015]

Image seen by model during training

Reconstructed image inferred from trained model

Privacy as a resource

Fundamental Law of Information Recovery

The earliest definitions of privacy insisted that “nothing should be learned about an individual when data is released”.

This is almost comical, as if we cannot learn *anything* about *anyone*, then there is no information to be gained at all!

An example of general information we would like to extract is “20% of the data population are diabetic”.

But now we have learned information about individuals - each individual belongs to a population where the prevalence of diabetes is 20%.

Name	Diabetic
Joe	No
Jun	No
Jaya	No
Janet	No
Jaro	Yes

layer 6
AI at TD

Fundamental Law of Information Recovery

There is a fundamental tradeoff - learning anything from data analysis necessitates learning something about the underlying data.

Privacy is **not binary** - we can learn something about individuals without learning everything.

Instead, privacy may be thought of as a **resource**. The more information we gain from a data source, the more **privacy is used up**.

Differential privacy gives a formal definition of this resource.

Differential Privacy

Randomized Response

Survey on sensitive subject - Have you committed tax fraud?
Nobody would truthfully respond to this survey and incriminate themselves.

Ask subjects to flip a coin so that they have plausible deniability

Answer
Truthfully

Flip again
If Heads: “Yes”
If Tails: “No”

Analyst gets an estimate of the true answer by understanding random process.

Randomness in **responses to queries** is the key to solving issues discussed above.

Randomized Queries

Adding randomness is a good way to achieve privacy.

Differential Privacy works with **randomized queries** that return answers from a **database**.

Instead of returning the true answer to a query, each possible answer (including the true one) has some probability of being returned.

We think of a database as a collection of records.

Each record contains various points of information about one Subject.

layer 6
AI at TD

Differential Privacy [Dwork et al. 2006]

Consider two databases $\mathcal{D}$ and $\mathcal{D}'$ which differ by one record, and a randomized query $\mathcal{M}[\mathcal{D}]$ which acts on databases to give a result.

The **query is differentially private** if the **results are almost indistinguishable** for all databases that differ by one record.

A query is (ϵ,δ)-differentially private if for all subsets of the output $\mathcal{S} \subseteq \text{Range}[\mathcal{M}]$ and all pairs of databases that differ by one record

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

Intuitively, the likelihood of any result from the query must be almost unchanged when one datapoint is added or removed.

layer 6
AI at TD

Differential Privacy

A query is (ϵ,δ)-differentially private if for all subsets of the output $\mathcal{S} \subseteq \text{Range}[\mathcal{M}]$ and all pairs of databases that differ by one record

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

Counterexample:

Query that returns the exact count of records that satisfy property P.

On neighbouring databases

$$\Pr(\mathcal{M}[\mathcal{D}] = n) = 1 \quad \text{and} \quad \Pr(\mathcal{M}[\mathcal{D}'] = n) = 0$$

so this exact count query cannot be DP unless ϵ=∞ or δ=1.

layer 6
AI at TD

Differential Privacy

A query is (ϵ,δ)-differentially private if for all subsets of the output $\mathcal{S} \subseteq \text{Range}[\mathcal{M}]$ and all pairs of databases that differ by one record

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

- Randomized - Any DP-query must be randomized
- Quantitative - ϵ and δ are numerical, where lower means better privacy
- Worst Case - We assume nothing about what the input datasets are like

layer 6
AI at TD

Differential Privacy

A query is (***ϵ,δ***)-differentially private if for all subsets of the output $\mathcal{S} \subseteq \text{Range}[\mathcal{M}]$ and all pairs of databases that differ by one record

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

Curator's promise - *"You will not be affected, adversely or otherwise, by allowing your data to be used in the database, no matter what other data sets are available."*

By adding one Subject's data, differential privacy makes the promise that the likelihood of any result will change only by a small factor.

layer 6
AI at TD

Guarantees

After accessing the database with a differentially private query, can an attacker **bring in outside info** to weaken the privacy guarantee?

No, differentially private queries are **immune to post-processing**.

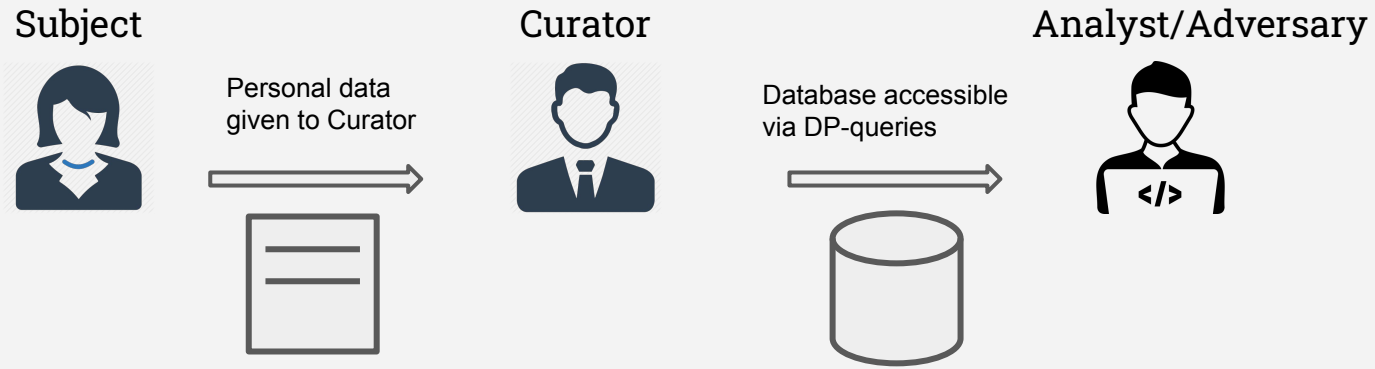
Privacy is a resource - it gets used up as the database is accessed more.

Differentially private queries quantify how much privacy is consumed.

Applying the same query several times is still differentially private,
but uses up more privacy.

Applications

Differential Privacy in Practice



The private database is accessible only through differentially private queries. Each access consumes some of the privacy budget.

We need to design *useful* queries that are differentially private.

Deterministic queries on a database like counts, sums, and averages will not maintain privacy. We can modify any query by **adding random noise to the result**.

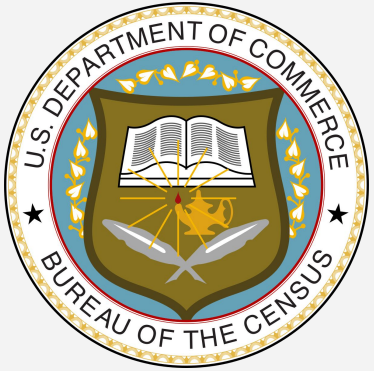
layer 6
AI at TD

Releasing general information from surveys

Differential privacy is useful for extracting general information from datasets while protecting the privacy of individuals.

In 2020 the U.S. Census Bureau began using differential privacy to analyse and release demographic information.

By comparison, the 2010 census used “swapping” - some attributes were exchanged between records - but this is ineffective against reconstruction attacks.

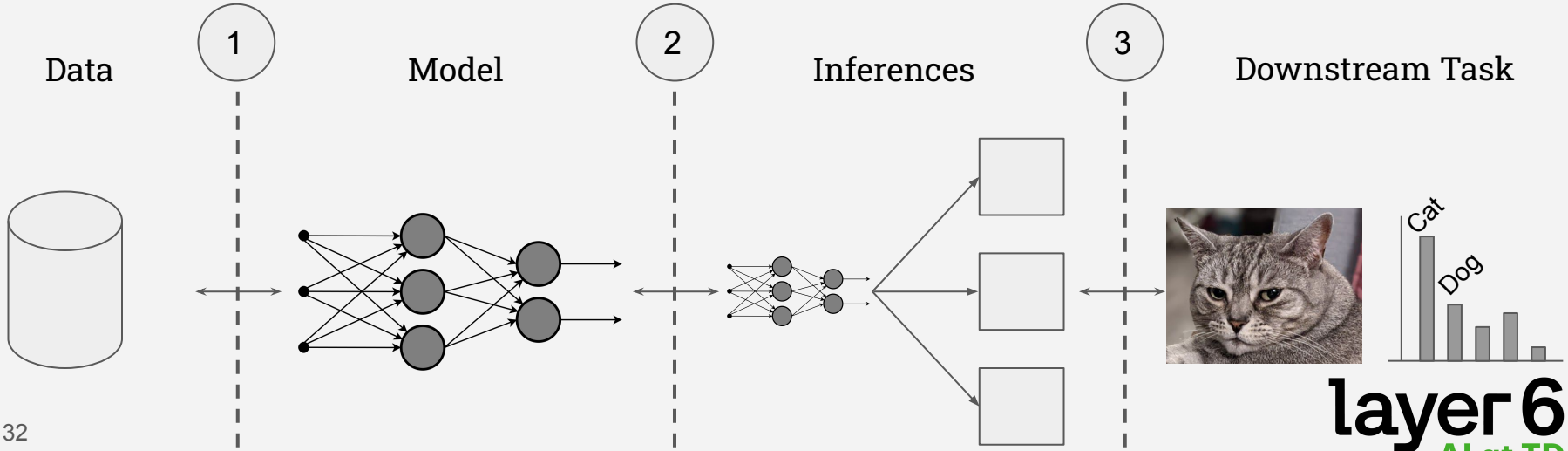


layer 6
AI at TD

Training ML models

Machine learning poses several challenges from a differential privacy perspective.

1. Many accesses to the database are needed (training steps)
2. Memorization of individual examples
3. Inference needs to be run many times

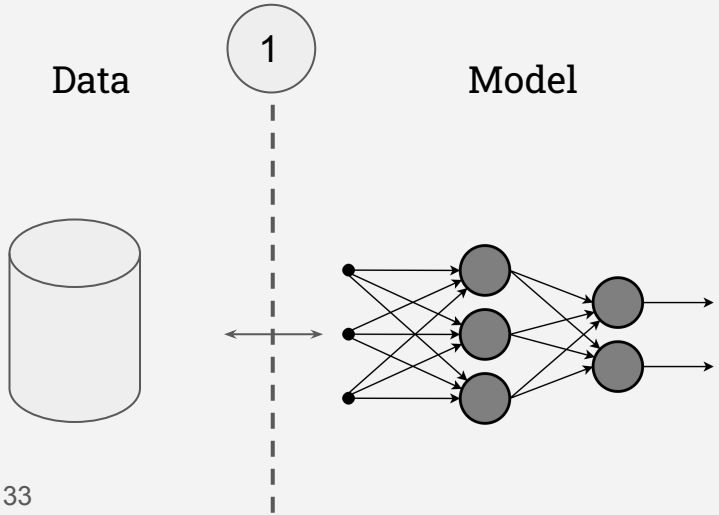


layer 6
AI at TD

Training ML models

Control privacy during training via gradient updates

- Gradients have unbounded sensitivity - **clip** to provide hard bound
- **Add noise** to the deterministic gradients
- Aggregate noisy gradients and take a gradient descent step



Gradient clipping and noising are standard regularization techniques.

Once model is trained, unlimited inference can be done due to the post-processing guarantee.

layer 6
AI at TD

Conclusions

Conclusions

Privacy is not binary. We cannot have perfect privacy for every individual, while providing useful information.

Privacy is a resource that is used up with every query.

Randomized results are essential for protecting privacy.

Differential privacy aims to quantify how much privacy a query uses.

Our job is to develop queries that minimize privacy use, while maximizing the utility/accuracy of results.

layer 6
AI at TD

Thank you!

jesse@layer6.ai

jesse.cresswell@td.com