6.S060
Lecture 24
Introduction to Differential Privacy
Outline

• Motivation

• Part I:
  – Differential Privacy (DP) Basics
  – DP pros and cons, deployment, challenges

• Part II:
  – DP for Statistics


Motivation
Data Privacy: The Problem

• Given a dataset with sensitive information, such as:
  – Census data
  – Health records
  – Social network activity
  – Telecommunications data

• How can we:
  – enable desirable uses of the data
  – while protecting the privacy of the data subjects?
Approach 1: Encrypt the Data

**Problems:** How to search over data or compute statistics? Who has the encryption key?

---

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Blood</th>
<th>...</th>
<th>HIV?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen</td>
<td>F</td>
<td>B</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Jones</td>
<td>M</td>
<td>A</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Smith</td>
<td>M</td>
<td>O</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Ross</td>
<td>M</td>
<td>O</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Lu</td>
<td>F</td>
<td>A</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Shah</td>
<td>M</td>
<td>B</td>
<td></td>
<td>Y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Blood</th>
<th>...</th>
<th>HIV?</th>
</tr>
</thead>
<tbody>
<tr>
<td>100101</td>
<td>001001</td>
<td>110101</td>
<td>...</td>
<td>110111</td>
</tr>
<tr>
<td>101010</td>
<td>111010</td>
<td>111111</td>
<td>...</td>
<td>001001</td>
</tr>
<tr>
<td>001010</td>
<td>100100</td>
<td>011001</td>
<td>...</td>
<td>110101</td>
</tr>
<tr>
<td>001110</td>
<td>010010</td>
<td>110101</td>
<td>...</td>
<td>100001</td>
</tr>
<tr>
<td>110101</td>
<td>000000</td>
<td>111001</td>
<td>...</td>
<td>010010</td>
</tr>
<tr>
<td>111110</td>
<td>110010</td>
<td>000101</td>
<td>...</td>
<td>110101</td>
</tr>
</tbody>
</table>
Approach 2: Anonymize the Data

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Blood</th>
<th>...</th>
<th>HIV?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen</td>
<td>F</td>
<td>B</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Jones</td>
<td>M</td>
<td>A</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Smith</td>
<td>M</td>
<td>O</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Ross</td>
<td>M</td>
<td>O</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Lu</td>
<td>F</td>
<td>A</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Shah</td>
<td>M</td>
<td>B</td>
<td></td>
<td>Y</td>
</tr>
</tbody>
</table>

Problems?
All it takes is a knowledge of a small number of attributes to identify/name the person!

Uniquely identify > 60% of the US population [Sweeney `00, Golle `06]
How many movies required on average to uniquely identify a user?

Four!
Narayanan-Shmatikov Set-Up

• Dataset: \( x \) = set of records (e.g., Netflix ratings)
• Adversary’s inputs:
  – \( x' \) = subset of records from \( x \), distorted slightly
  – \( aux \) = auxiliary information about a record \( r \in D \) (e.g., a particular identifiable user’s IMDB ratings)
• Adversary’s goal: output either
  – \( r' \in x' \) = record that is “close” to \( r \), or
  – \( \bot \) = failed to find a match
• For the $1m Netflix Challenge, a dataset of 5,00,000 subscribers’ ratings (less than 1/10 of all subscribers) was released (total of 100m ratings over 6 years).

• Out of 50 sampled IMDB users, two standouts were found, with eccentricities of 28 and 15.

• Reveals all movies watched from only those publicly rated on IMDB.

• Class action lawsuit, cancelling of Netflix Challenge II.

Message: Any attribute can be a “quasi-identifier”
Approach 3: Mediate Access

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Blood</th>
<th>HIV?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen</td>
<td>F</td>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td>Jones</td>
<td>M</td>
<td>A</td>
<td>N</td>
</tr>
<tr>
<td>Smith</td>
<td>M</td>
<td>O</td>
<td>N</td>
</tr>
<tr>
<td>Ross</td>
<td>M</td>
<td>O</td>
<td>Y</td>
</tr>
<tr>
<td>Lu</td>
<td>F</td>
<td>A</td>
<td>N</td>
</tr>
<tr>
<td>Shah</td>
<td>M</td>
<td>B</td>
<td>Y</td>
</tr>
</tbody>
</table>

Problems: Curator sees all the data. What queries are allowed? How much do they leak?
Part I
Differential privacy

- **Requirement**: effect of each individual should be “hidden”

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]
Differential privacy

- Requirement: Adversary should not be able to tell if any one person’s data were changed arbitrarily
### Differential privacy

- **Requirement:** Adversary should not be able to tell if any one person’s data were changed arbitrarily.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Blood</th>
<th>...</th>
<th>HIV?</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>B</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>M</td>
<td>O</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>M</td>
<td>O</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>F</td>
<td>A</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>M</td>
<td>B</td>
<td></td>
<td>Y</td>
</tr>
</tbody>
</table>
Differential privacy

- **Requirement**: Adversary should not be able to tell if any one person’s data were changed arbitrarily
Simple approach: random noise

- Very little noise needed to hide each person as $n \to \infty$
- This is just for one query
DP for one query/release

- **Requirement**: for all D, D’ differing on one row, and all q
  
  \[
  \text{Distribution of } M(D,q) \approx_\varepsilon \text{ Distribution of } M(D',q)
  \]
• **Requirement**: M is $\varepsilon$-DP if for all $D$, $D'$ differing on one row, and all $q$

\[
\forall \text{ sets } T, \Pr[M(D, q) \in T] \leq e^\varepsilon \cdot \Pr[M(D', q) \in T]
\]
The Laplace Mechanism

[Dwork-McSherry-Nissim-Smith ’06]

- Very little noise needed to hide each person as $n \to \infty$
- Theorem: The Laplace Mechanism is Differentially Private
Differential Privacy: Pros and Cons

+ Whatever an adversary learns about me, it could have learned from everyone else’s data
+ Mechanism cannot leak “individual-specific” information
+ Above interpretations hold regardless of adversary’s auxiliary information
+ Composes: k repetitions is kε differentially private
  - No protection for information that is not localized to a few rows.
  - No guarantee that subjects won’t be “harmed” by results of analysis
Differential Privacy Deployed

mostly focused on count and average statistics
Challenges for DP in Practice

- Accuracy for “small data” (small n)
- Modeling and managing privacy loss over time
- Analysts are used to working with raw data, not querying (slightly) noisy data
- Matching guarantees with privacy law and regulation
- ...

23
Part II
Setting

Private data set $D$ flows to the Privacy-preserving sanitizer, which produces a synthetic dataset, a summary statistic, and an ML model. These outputs are categorized as non-public and public data, separated by a privacy barrier.
Property of Sanitizer

- Aggregate information computable
- Individual information protected
Differentially Private Algorithm Design

• Global Sensitivity Method: statistics
• Exponential Method: optimization

• Problem:
  • Given function f, sensitive dataset D
  Find a differentially private approximation to f(D)

• Example: f(D) = mean of data points in D
The Global Sensitivity Method

Given: A function $f$, sensitive dataset $D$

Define: $\text{dist}(D, D') = \# \text{records that } D, D' \text{ differ by 1}$

Add or remove a record from $D$ to get $D'$

Global Sensitivity of $f$:

$$S(f) = \max_{\text{dist}(D, D') = 1} | f(D) - f(D')|$$
The Laplace Mechanism

Global Sensitivity of $f$ is $S(f) = \max_{\text{dist}(D, D') = 1} |f(D) - f(D')|$

Output $f(D) + Z$, where

$$Z \sim \frac{S(f)}{\epsilon} \text{Lap}(0, 1)$$

$\epsilon$-differentially private

Laplace distribution:

$$p(z|\mu, b) = \frac{1}{2b} \exp\left(-\frac{|z-\mu|}{b}\right)$$
Example: Mean

\[ f(D) = \text{Mean}(D), \text{ where each record is a scalar in } [0, 1] \]

Global Sensitivity of \( f = 1/n \)

**Laplace Mechanism:**

Output \( f(D) + Z \), where \( Z \sim \frac{1}{n\epsilon} \text{Lap}(0, 1) \)
Exponential Mechanism

Problem:
Given function $f(w, D)$, Sensitive Data $D$
Find differentially private approximation to

$$w^* = \arg\max_w f(w, D)$$

Example: $f(w, D) =$ accuracy of classifier $w$ on dataset $D$
Exponential Mechanism

Suppose for any $w$,\[ |f(w, D) - f(w, D')| \leq S\]
when $D$ and $D'$ differ in 1 record. Sample $w$ from:\[ p(w) \propto e^{\epsilon f(w, D)/2S}\]
for $\epsilon$-differential privacy.

\[
\begin{align*}
\text{argmax } f(w, D) \\
\end{align*}
\]
Example: Parameter Tuning

Given validation data $D$, $k$ classifiers $w_1, \ldots, w_k$ (privately) find the classifier with highest accuracy on $D$

Here, $f(w, D) =$ classification accuracy of $w$ on $D$

For any $w$, any $D$ and $D'$ that differ by one record,

$$|f(w, D) - f(w, D')| \leq \frac{1}{|D|}$$

So, the exponential mechanism outputs $w_i$ with prob:

$$\Pr(w_i) \propto e^{\epsilon |D| f(w_i, D) / 2}$$
Conclusion

• Differential privacy can help companies to learn more about a group of users without compromising the privacy of an individual within that group.

• Many of the world’s governments now have strict policies about how tech companies collect and share user data.
  – Companies need users’ data to provide high-quality services that benefit users, such as personalized recommendations.
  – Companies may face charges if they collect too much user data.