L21: Privacy

Jeff M. Phillips

April 3, 2019

Think if you were the data point.

Should I release a public data set?

Health Data: lots regulation

public data

age, height, zip code, name | private data

? cancer

Example: Heath Records

Story Time:

Example: Heath Records

Story Time:

- In 2000, Massachusetts released all state employee's medical records in an effort for researchers to be able to study them.

Example: Heath Records

Story Time:

- In 2000, Massachusetts released all stated employee's medical records in an effort for researchers to be able to study them.
- They wiped all ids, but kept zip codes, birthday, gender. Was declared anonymized by the government.

Example: Heath Records

Story Time:

- In 2000, Massachusetts released all stated employee's medical records in an effort for researchers to be able to study them.
- They wiped all ids, but kept zip codes, birthday, gender. Was declared anonymized by the government.
- In Massachusetts, it was possible to buy voter data for $20. It has names, birthday, zip codes, and birthdays of all voters.

Example: Heath Records

Story Time:

- In 2000, Massachusetts released all stated employee's medical records in an effort for researchers to be able to study them.
- They wiped all ids, but kept zip codes, birthday, gender. Was declared anonymized by the government.
- In Massachusetts, it was possible to buy voter data for $20. It has names, birthday, zip codes, and birthdays of all voters.
- A grad student, Latanya Sweeney combined the two to identify the governor of Massachusetts. Story is, she mailed him his own health records!

Example: Heath Records

Story Time:

- In 2000, Massachusetts released all stated employee's medical records in an effort for researchers to be able to study them.
- They wiped all ids, but kept zip codes, birthday, gender. Was declared anonymized by the government.
- In Massachusetts, it was possible to buy voter data for $20. It has names, birthday, zip codes, and birthdays of all voters.
- A grad student, Latanya Sweeney combined the two to identify the governor of Massachusetts. Story is, she mailed him his own health records!
- Dr. Sweeney now teaches at Harvard.

Modify dataset $D \xrightarrow{?} D'$

- Hopefully analysis on D' meaningful
- but can't identify individuals

$D \to D'$ remove data

k-anonymity: For each set of unique public traits in D' there are at least k data points.

Could still be $1/k$ chance correct

Maybe all k have cancer?

ℓ-diversity: starts w/ k-anonymity

ℓ-classes ($\ell = 2$: cancer / no cancer)

Each set (of $2k$) must have at least one in each class.

↳ "plausible deniability"

What if $\frac{k-1}{k}$ have cancer?

t-closeness: starts w/ k-anonymity + ℓ-diversity

Distribution of private traits must be t-close to all data

How tall is Sylvester Stallone?

- "Sly Stallone is height average NJ man."
- Outside: Average NJ man 5' 8";

Example: Netflix Prize

Story Time:

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  And another set $D_2 = \{\langle \text{user-id, movie, date of grade} \rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  And another set $D_2 = \{\langle \text{user-id, movie, date of grade} \rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle \text{user-id, movie, date of grade, grade} \rangle\}$.
  And another set $D_2 = \{\langle \text{user-id, movie, date of grade} \rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets
  $D_1 = \{\langle$user-id, movie, date of grade, grade$\rangle\}$.
  And another set $D_2 = \{\langle$user-id, movie, date of grade$\rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!
- Raters of movies also rate on IMDB (with user id, time stamp)

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle$user-id, movie, date of grade, grade$\rangle\}$.
  And another set $D_2 = \{\langle$user-id, movie, date of grade$\rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!
- Raters of movies also rate on IMDB (with user id, time stamp)
- Researchers showed that by linking who rated similar sets of movies, with similar scores and times, they could identify many people.

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle$user-id, movie, date of grade, grade$\rangle\}$.
  And another set $D_2 = \{\langle$user-id, movie, date of grade$\rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!
- Raters of movies also rate on IMDB (with user id, time stamp)
- Researchers showed that by linking who rated similar sets of movies, with similar scores and times, they could identify many people.
- (maybe watched embarrassing films on Netflix, not listed on IMDB)

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle$user-id, movie, date of grade, grade$\rangle\}$.
  And another set $D_2 = \{\langle$user-id, movie, date of grade$\rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!
- Raters of movies also rate on IMDB (with user id, time stamp)
- Researchers showed that by linking who rated similar sets of movies, with similar scores and times, they could identify many people.
- (maybe watched embarrassing films on Netflix, not listed on IMDB)
- Class action lawsuit filed (lated dropped) against Netflix.

Example: Netflix Prize

Story Time:

- In 2006, Netflix (e.g., DVDs) released awesome data sets $D_1 = \{\langle$user-id, movie, date of grade, grade$\rangle\}$.
  And another set $D_2 = \{\langle$user-id, movie, date of grade$\rangle\}$.
  Wants researchers to predict grade on D_2.
  (Had another similar private data D_3 to evaluate grades : cross validation.)
- If certain improvement over Netflix's algorithm, get \$1 million!
- Led to lots of cool research!
- Raters of movies also rate on IMDB (with user id, time stamp)
- Researchers showed that by linking who rated similar sets of movies, with similar scores and times, they could identify many people.
- (maybe watched embarrassing films on Netflix, not listed on IMDB)
- Class action lawsuit filed (lated dropped) against Netflix.
- Netflix Prize had proposed sequel, dropped in 2010 for more privacy concerns.

Differential Privacy

Trait of two data sets D_1, D_2

Goal: D_1 and D_2 similar statistical properties

Rule: for any $x \in D_1$ cannot know its value from D_2

---

allowed access to D_1, D_2 w/ query q.

$$\forall_{q \in Q} \quad \frac{P_r[q(D_1) \in E]}{P_r[q(D_2) \in E]} \leq e^{\varepsilon} \approx 1 + \varepsilon$$

broader $Q, E \rightarrow$ the more I learn about D_1

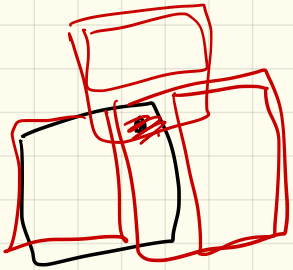
event

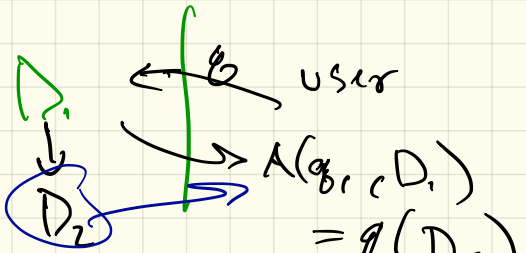
error param

smaller ε starts $\rightarrow$ more similar

Two Main types of mechanisms

- **Interactive** D_1 exists
  - ↳ I can ask queries to oracle
  - ↳ return answers.





- **Non-Interactive** $D_1 \rightarrow$ add Noise $\rightarrow D_2 \rightarrow$ release D_2

Laplacian Mechanism

$D_1 = \{x_1, x_2, \ldots x_n\}$ $\quad x_i \in \mathbb{R}$

ex. height.

$D_2 = \{x_1', x_2', \ldots x_n'\}$

$x_i' = x_i + Lap(\varepsilon)$

$= c\, e^{-\varepsilon}$

$$\frac{P_r[x_1 \in D_1 \geq 70]}{P_r[x_1' \in D_2 \geq 70]} = \frac{c\, e^{-4\varepsilon}}{c\, e^{-3\varepsilon}} = e^{(-\varepsilon)} \approx 1 - \varepsilon$$

66

67

Interactive Database

$D_1 = [0\ 1\ 0\ 1\ 1\ \$\ 0\ 0\ 1]$

$D_2 = [0\ 1\ 0\ 1\ 1\ \$\ P\ P\ 1]$

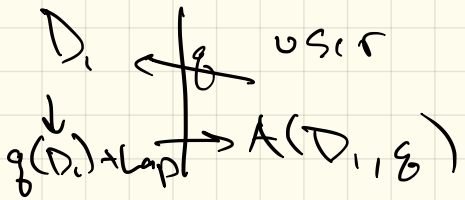
$1 = \text{Hamm}(D_1, D_2)$

$q\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 0$

Ask queries $q \in Q = \{ q = 1^D \text{ xor } q \}$

4 bit vector

q = how many people in some range have cancer?



$$\text{Ham}(D_1, D_2) = 1$$

$$\frac{\Pr[A(D_1, q) = x]}{\Pr[A(D_2, q) = x]} = \frac{\text{Lap}(|x - g(D_1)|)}{\text{Lap}(|x - g(D_2)|)}$$

observation

$$= e^{(\varepsilon(|x - g(D_1)| - |x - g(D_2)|))}$$

$$\leq e^{(\varepsilon \underbrace{|g(D_1) - g(D_2)|}_{\leq 1})}$$

$$\leq e^{\varepsilon} \approx 1 + \varepsilon$$