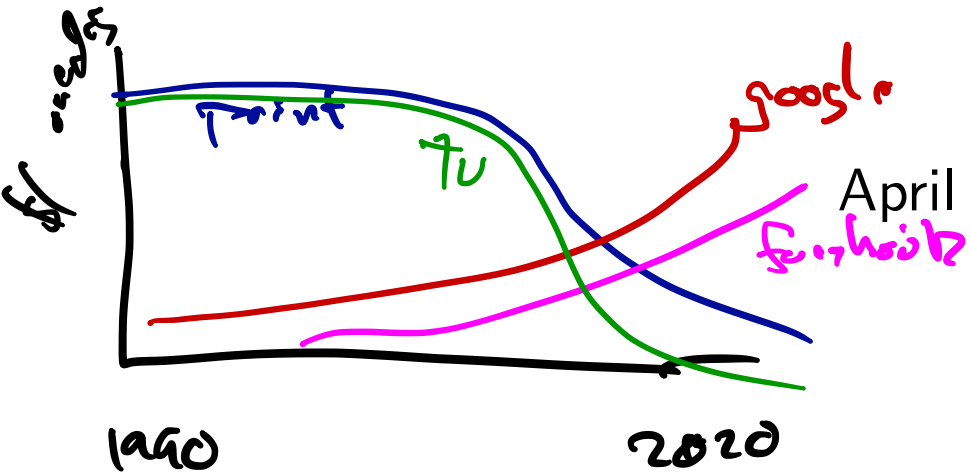
L21
~~L20~~: Privacy

Jeff M. Phillips



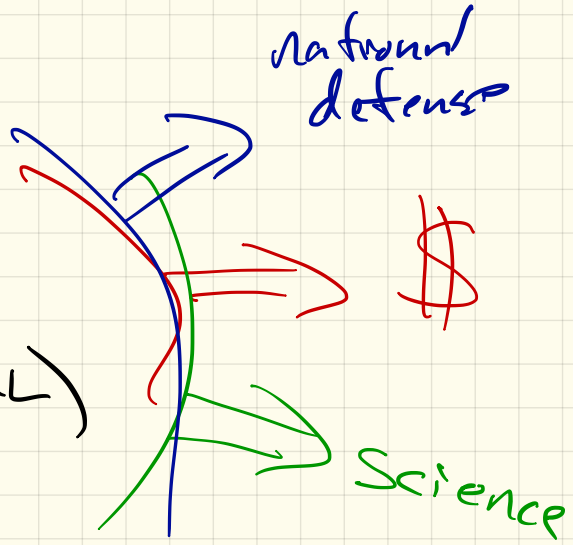
April 4, 2018

X is a data set

↳ clusters

↳ regression

↳ classify (ML)



$x \in X$

↑ this is good!

Example: Heath Records

Story Time:

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.

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~~ gender 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.

Define Anonymity

full data set D → map → D'

D': detailed as possible + not identifying any one $x \in D$

- **k-anonymity**: categorical data

  no group $S \subset D'$ will be only with specific traits and $|S| \geq k$.

- **l-diversity**: "l-well separated" hidden traits.

  each group S, has at least l traits

- **t-closeness**: of hidden traits must be t-close to distribution of full popula

- Figure out how Sylvester Stallone has average height of men in NJ.
- Survey in NJ average height 5', 8"

Example: Netflix Prize

Story Time:

Example: Netflix Prize

Story Time:

- In 2006, Netflix 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 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!

Example: Netflix Prize

Story Time:

- In 2006, Netflix 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!

Example: Netflix Prize

Story Time:

- In 2006, Netflix 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 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 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 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 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

D_1 ← true data

D_2 ← released (given access)

and somehow D_1, D_2 are close.

also protect individual data points.

---

queries $q \in Q$ such of D_2 or D_1

property R

Example $q(D_1)$ → ages of class

R → # older 30.

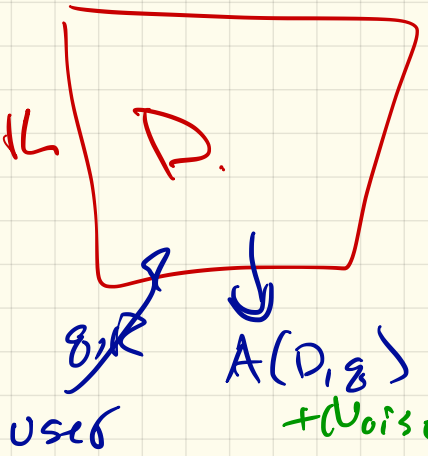
$$\frac{Pr[q(D_1) \in R]}{Pr[q(D_2) \in R]} \leq \exp(\varepsilon) \approx 1 + \varepsilon$$

e^{ε}

⇒ D_1, D_2 ε-differentially private.

Interactive Approach

Trusted Entity with D



Non-Interactive

store $D_2 = D_1 + \text{Noise}$

↑ Laplacian Noise

$\text{Lap}(\varepsilon) = c \exp(-\varepsilon)$

D ← true dataset Sly Stallones' height in inches

$+Lap(\varepsilon)$ ↙ ↘ $+Lap(\varepsilon)$

ξ ξ

D_1 68 D_2 67

.

$$\frac{Pr[D \geq 70 \mid D_1]}{Pr[D > 70 \mid D_2]} \approx \frac{e^{-2\varepsilon}}{e^{-3\varepsilon}} \quad e^{\varepsilon} \approx 1+\varepsilon$$