Privacy

Note Title 4/4/2016

- History Lesson
  - definitions
  - story
- Differential Privacy
- Ethics + Empathy

- If you were the data point.

Example: Health Records

Hospital: zip code, cancer / no cancer

↳ STORY

In 2000, Mass

↳ released med records of all state employees

↳ wiped ids

kept: zip code, birthday, gender

↳ In Mass: buy voter data $20

names, birthdays, zip code,

↳ grad student. Latanya Sweeney

→ ID ed governor of Mass.
mailed health records.

- K-anonymity: public data → narrow down to at least k people

(nothing to release)

- l-diversity: (k-anonymity) + at least l-distinct private traits

- t-closeness: (l-diversity) + distribution of private traits is t-close (in EMD) to the entire data.

(nothing to mine)

Sly Stallone 5' 8"

Netflix Challenge 2006

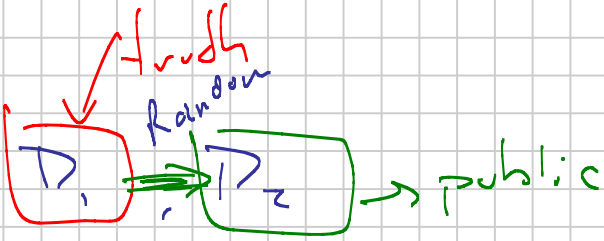
public D_1 = {<user-id, movie, date & grade, grade>}

D_2 = {<user-id, movie, date & grade, ——>}

evaluate D_3 = {< id, m, date, grade>} private

IMDB user-id, rating, time stamp, movie

Differential Privacy



Two Data Sets

- similar statistical analysis
- change 1 data point $D_1 \to D_2$

$D_1, D_2 \in \{0,1\}^n$

- $\text{Ham}(D_1, D_2) = 1$ (change 1 bit)
- $\dfrac{Pr[\mathcal{E}(D_1) \in R]}{Pr[\mathcal{E}(D_2) \in R]} \leq \exp(\varepsilon) \approx 1 + \varepsilon$

D_1 1 0 1 1 0 0 1 1
$\mathcal{E}$ 0 0 1 1 0 0 1 0

$|D_1 \text{ XOR } \mathcal{E}| \to |1\,0\,0\,0\,0\,0\,0\,1| = 2$

$R = [0, 3]$

1. Interactive Approach — $D_2 \leftarrow (D_1 + \text{noise})$ is private. But ask questions.

2. Non-Interactive Approach

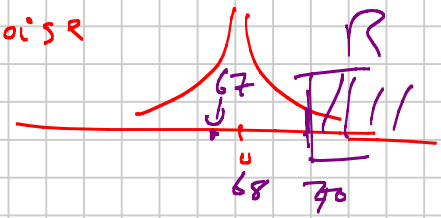
Release D_2

$D_1 \leftarrow$ True height of Sly Stalone : 68

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

$\text{Lap}(\varepsilon)$

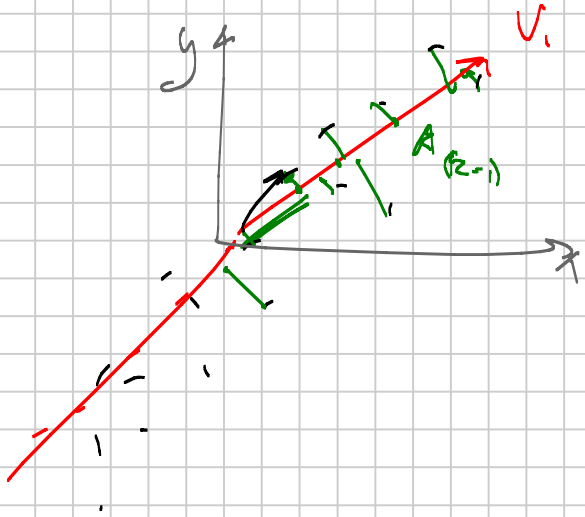
$\exp(-\varepsilon)$



$Pr[D_1 \geq 70] \approx \exp(-2\varepsilon)$ (68)

$Pr[D_2 \geq 70] \approx \exp(-3\varepsilon)$ (67)

$\dfrac{Pr[D_1 \geq 70]}{Pr[D_2 \geq 70]} = \dfrac{\exp(-2\varepsilon)}{\exp(-3\varepsilon)} = \exp(\varepsilon) \approx 1 + \varepsilon$


y
x
v_1
$A_{(k-1)}$