

# CS480/680: Introduction to Machine Learning

## Lec 22: Differential Privacy

Yaoliang Yu



UNIVERSITY OF  
**WATERLOO**

FACULTY OF MATHEMATICS  
**DAVID R. CHERITON SCHOOL  
OF COMPUTER SCIENCE**

April 1, 2025

# The Netflix Challenge

		Inside Out	Good Will Hunting	Mean Girls	Terminator	Titanic	Warrior
							
Tina Fey		3	1	5	1	?	1
Helen Mirren		2	?	?	2	5	1
Sylvester Stallone		1	3	1	4	2	5
Tom Hanks		?	3	1	?	4	3
George Clooney		2	2	1	3	1	4

- $\langle \text{user}, \text{movie}, \text{date of rating}, \text{rating} \rangle$
- ~1M ratings, .5M users, 20k movies

# 1M Prize



# Video Privacy Protection Act of 1988



- Up to \$2,500 for then 2M users of Netflix

---

A. Narayanan and V. Shmatikov. "Robust De-anonymization of Large Sparse Datasets". In: *IEEE Symposium on Security and Privacy*. 2008, pp. 111–125.



# Linkage Attack

## ▸ Do you share voter information with other agencies or groups?

Yes. Elections Canada shares [voter information](#) from the National Register of Electors with all provincial and territorial electoral agencies and with some municipalities for election purposes only. Sharing voter registration information improves the accuracy of voters lists, making it easier to vote. It also reduces duplication, saving taxpayer money.

As required by the *Canada Elections Act*, we also provide voters lists (containing name, address and unique identifier number) to candidates, members of Parliament and registered and eligible political parties, who may use the information for specific, authorized purposes. Refer to the [Guidelines for Use of the Lists of Electors](#) to learn more.

Note that we do not share voter information with any other organizations, including social media platforms and media.

# Linkage Attack

## Do you share voter information with other agencies or groups?

Yes. Elections Canada shares [voter information](#) from the National Register of Electors with all provincial and territorial electoral agencies and with some municipalities for election purposes only. Sharing voter registration information improves the accuracy of voters lists, making it easier to vote. It also reduces duplication, saving taxpayer money.

As required by the *Canada Elections Act*, we also provide voters lists (containing name, address and unique identifier number) to candidates, members of Parliament and registered and eligible political parties, who may use the information for specific, authorized purposes. Refer to the [Guidelines for Use of the Lists of Electors](#) to learn more.

Note that we do not share voter information with any other organizations, including social media platforms and media.

[Subscribe](#)[Latest Issues](#)**SCIENTIFIC  
AMERICAN.**[Cart 0](#)[Sign In](#) | [Stay Informed](#)[THE SCIENCES](#)[MIND](#)[HEALTH](#)[TECH](#)[SUSTAINABILITY](#)[EDUCATION](#)[VIDEO](#)[PODCASTS](#)[BLOGS](#)[PUBLICATIONS](#)[Q](#)

THE SCIENCES

## Confirmed: The U.S. Census Bureau Gave Up Names of Japanese-Americans in WW II

Government documents show that the agency handed over names and addresses to the Secret Service

By IB Michel on March 30, 2007

# Anonymization is not Enough

<i>ZIP Code</i>	<i>Birth Date</i>	<i>Gender</i>	<i>Race</i>
33171	7/15/71	m	Caucasian
02657	2/18/73	f	Black
20612	3/12/75	m	Asian

**Table 2. Deidentified Data that Are Not Anonymous.**

# Anonymization is not Enough

<i>ZIP Code</i>	<i>Birth Date</i>	<i>Gender</i>	<i>Race</i>
33171	7/15/71	m	Caucasian
02657	2/18/73	f	Black
20612	3/12/75	m	Asian

**Table 2. Deidentified Data that Are Not Anonymous.**

The 1997 voting list for Cambridge, Massachusetts, contains demographics on 54,805 voters. Of these, birth date, which contains the month, day, and year of birth, alone can uniquely identify the name and address of 12 percent of the voters. One can identify 29 percent of the list by just birth date and gender, 69 percent with only a birth date and a 5-digit ZIP code, and 97 percent (53,033 vot-

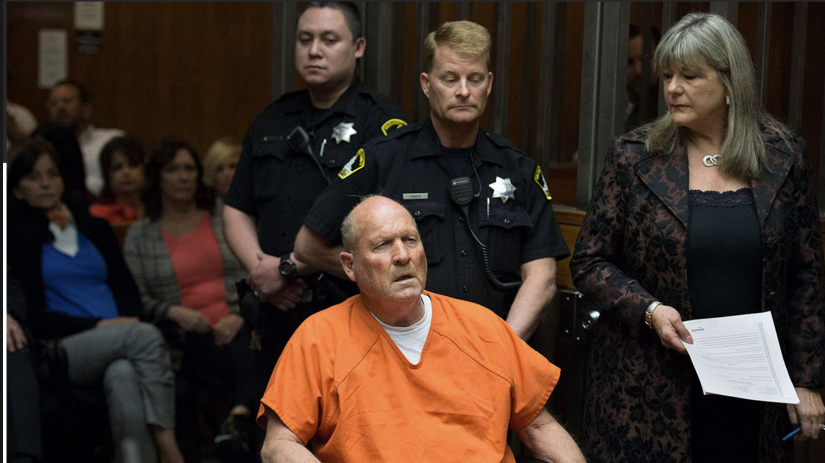
birth date alone	12%
birth date and gender	29%
birth date and 5-digit ZIP code	69%
birth date and full postal code	97%

**Table 3. Uniqueness of Demographic Fields in Cambridge, Massachusetts, Voter List.**

# We will find you: DNA search used to nab Golden State Killer can home in on about 60% of white Americans

Researchers call for limiting how ancestry databases can be used to protect privacy

11 OCT 2018 • BY JOCELYN KAISER



<https://www.science.org/content/article/we-will-find-you-dna-search-used-nab-golden-state-killer-can-home-about-60-white>

# Differencing Attack

```
Diff "C:/Qt/Qt5.6.0/Examp... | X | main.cpp"
C:/Qt/Qt5.6.0/Examples/Qt-5.6/qtwebengine/webeng...
1 /*****
2 **
3 ** Copyright (C) 2016 The Qt Company Ltd.
4 ** Contact: http://www.qt.io/licensing/
5 **
6 ** This file is part of the examples of the Qt Tool
7 **
8 ** $QT_BEGIN_LICENSE:BSD$
9 ** You may use this file under the terms of the BSD
10 **
11 ** "Redistribution and use in source and binary form, without
12 ** modification, are permitted provided that the file
13 **
14 ** Skipped 71 lines...

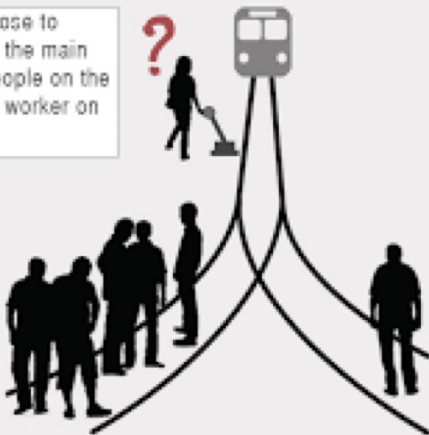
C:/Qt/Qt5.7.0-beta/Examples/Qt-5.7/qtwebengine...
1 /*****
2 **
3 ** Copyright (C) 2016 The Qt Company Ltd.
4 ** Contact: https://www.qt.io/licensing/
5 **
6 ** This file is part of the QtWebEngine module o
7 **
8 ** $QT_BEGIN_LICENSE:BSD$
9 ** Commercial License Usage
10 ** Licensees holding valid commercial Qt license
11 ** accordance with the commercial license agreem
12 ** Software or, alternatively, in accordance wit
13 ** a written agreement between you and The Qt Co
14 ** and conditions see https://www.qt.io/terms-co
15 ** information use the contact form at https://w
16 **
17 ** BSD License Usage
18 ** Alternatively, you may use this file under th
19 ** as follows:
20 **
21 ** "Redistribution and use in source and binary
22 ** modification, are permitted provided that the
23 **
24 ** Skipped 71 lines...
```

- “How many people have disease X?”
- “How many people, not named YY, have disease X?”

# Just Sacrifice A Few?

## The trolley problem

The person can choose to divert the tram from the main track, saving five people on the track, but killing the worker on the other track.



# Restricted Access





# Example

---

- Consider a medical study about smoking and cancer
- Should a smoker participate?
- If yes, may lead to higher insurance premium
- But may also benefit from learning health risks
- Has the smoker's privacy been compromised?

Participate or not, impact on the smoker is likely the same

Have you cheated in any exam?

# Randomized Response

- Want to estimate the percentage of cheaters
- If ask bluntly, almost certainly will under-estimate
- Toss a coin: head, answer honestly; tail, answer randomly
  - cheaters: w.p.  $\frac{3}{4}$  say yes
  - non-cheaters: w.p.  $\frac{1}{4}$  say yes
  - $\frac{3}{4}p + \frac{1}{4}(1 - p) = \frac{1}{4} + \frac{1}{2}p =$  percentage of yes
- Plausible deniability for everyone
- What happens if we ask this question repeatedly?

---

S. L. Warner. "Randomised response: a survey technique for eliminating evasive answer bias". *Journal of the American Statistical Association*, vol. 60, no. 309 (1965), pp. 63–69.

# Differential Privacy

- Let  $M : \mathcal{D} \rightarrow \mathcal{Z}$  be a **randomized** mechanism
- **$(\epsilon, \delta)$ -DP** if for any  $D, D' \in \mathcal{D}$  differing by one data point, for any event  $E \subseteq \mathcal{Z}$ ,

$$\Pr[M(D) \in E] \leq \exp(\epsilon) \cdot \Pr[M(D') \in E] + \delta$$

- dataset  $D, D'$  fixed; randomness from the mechanism
- the smaller  $\epsilon$  or  $\delta$  is, the stricter the privacy requirement
- **$(\epsilon, 0)$ -DP** if  $\delta = 0$ , a.k.a.  **$\epsilon$ -DP**
- $\epsilon$  (roughly) bounds log odds ratio:  $\boxed{\epsilon \leq 1}$  often considered “good”
- $\delta$  allows rare, possibly catastrophic event (to trade utility): often,  $\boxed{\delta \ll 1/|\mathcal{D}|}$

# Randomized Response is $(\log 3, 0)$ -DP

$$\log \frac{\Pr[\mathbf{M}(D) \in E]}{\Pr[\mathbf{M}(D') \in E]} = \log \frac{\int_E p(\mathbf{x}) \, d\mathbf{x}}{\int_E q(\mathbf{y}) \, d\mathbf{y}}$$

# Randomized Response is $(\log 3, 0)$ -DP

$$\log \frac{\Pr[\mathbf{M}(D) \in E]}{\Pr[\mathbf{M}(D') \in E]} = \log \frac{\int_E p(\mathbf{x}) \, d\mathbf{x}}{\int_E q(\mathbf{y}) \, d\mathbf{y}} = \log \int_E \frac{p(\mathbf{x})}{q(\mathbf{x})} \cdot \frac{q(\mathbf{x})}{\int_E q(\mathbf{y}) \, d\mathbf{y}} \, d\mathbf{x}$$

# Randomized Response is $(\log 3, 0)$ -DP

$$\begin{aligned} \log \frac{\Pr[\mathbf{M}(D) \in E]}{\Pr[\mathbf{M}(D') \in E]} &= \log \frac{\int_E p(\mathbf{x}) \, d\mathbf{x}}{\int_E q(\mathbf{y}) \, d\mathbf{y}} = \log \int_E \frac{p(\mathbf{x})}{q(\mathbf{x})} \cdot \frac{q(\mathbf{x})}{\int_E q(\mathbf{y}) \, d\mathbf{y}} \, d\mathbf{x} \\ &\quad (\text{Jensen's inequality}) \leq \int_E \log \left( \frac{p(\mathbf{x})}{q(\mathbf{x})} \right) \cdot \frac{q(\mathbf{x})}{\int_E q(\mathbf{y}) \, d\mathbf{y}} \, d\mathbf{x} \end{aligned}$$

# Randomized Response is $(\log 3, 0)$ -DP

$$\begin{aligned}\log \frac{\Pr[M(D) \in E]}{\Pr[M(D') \in E]} &= \log \frac{\int_E p(\mathbf{x}) \, d\mathbf{x}}{\int_E q(\mathbf{y}) \, d\mathbf{y}} = \log \int_E \frac{p(\mathbf{x})}{q(\mathbf{x})} \cdot \frac{q(\mathbf{x})}{\int_E q(\mathbf{y}) \, d\mathbf{y}} \, d\mathbf{x} \\ &\quad (\text{Jensen's inequality}) \leq \int_E \log \left( \frac{p(\mathbf{x})}{q(\mathbf{x})} \right) \cdot \frac{q(\mathbf{x})}{\int_E q(\mathbf{y}) \, d\mathbf{y}} \, d\mathbf{x} \\ &\quad (\text{mean} \leq \text{max}) \leq \max_{\mathbf{x}} \log \frac{p(\mathbf{x})}{q(\mathbf{x})} \leq \epsilon\end{aligned}$$



# Randomized Response is $(\log 3, 0)$ -DP

$$\begin{aligned}\log \frac{\Pr[M(D) \in E]}{\Pr[M(D') \in E]} &= \log \frac{\int_E p(\mathbf{x}) \, d\mathbf{x}}{\int_E q(\mathbf{y}) \, d\mathbf{y}} = \log \int_E \frac{p(\mathbf{x})}{q(\mathbf{x})} \cdot \frac{q(\mathbf{x})}{\int_E q(\mathbf{y}) \, d\mathbf{y}} \, d\mathbf{x} \\ &\quad (\text{Jensen's inequality}) \leq \int_E \log \left( \frac{p(\mathbf{x})}{q(\mathbf{x})} \right) \cdot \frac{q(\mathbf{x})}{\int_E q(\mathbf{y}) \, d\mathbf{y}} \, d\mathbf{x} \\ &\quad (\text{mean} \leq \text{max}) \leq \max_{\mathbf{x}} \log \frac{p(\mathbf{x})}{q(\mathbf{x})} \leq \epsilon\end{aligned}$$

- Consider when  $D$  has a cheater and  $D'$  has a non-cheater:

$$\begin{aligned}- \log \frac{\Pr[M(D)=\text{Yes}]}{\Pr[M(D')=\text{Yes}]} &= \log \frac{3/4}{1/4} = \log 3 \\ - \log \frac{\Pr[M(D)=\text{No}]}{\Pr[M(D')=\text{No}]} &= \log \frac{1/4}{3/4} = -\log 3\end{aligned}$$

# DP in Practice

- Apple: reportedly  $\epsilon = 6$  in MacOS,  $\epsilon = 14$  in iOS10 and  $\epsilon = 2$  for health types
- Facebook: e.g.,  $\epsilon = 1.453$  and  $\delta = 1e - 5$
- Google: e.g.,  $\epsilon$  up to 9
- LinkedIn: each query uses  $\epsilon = 0.15$  and  $\delta = 1e - 10$
- Microsoft: e.g.,  $\epsilon = 12$  and  $\delta = 5.8e - 6$
- US Census Bureau: e.g.,  $\epsilon = 13.64$  and  $\delta = 1e - 5$

<https://desfontain.es/blog/real-world-differential-privacy.html>

# A Hypothesis Testing View

- Consider null hypothesis  $H_0 : D$  and alternative hypothesis  $H_1 : D'$
- Or simply two classes  $Y = 0$  vs.  $Y = 1$
- Treat  $\hat{Y} := \llbracket M(\cdot) \in E \rrbracket$ 
  - $\Pr(M(D) \in E) = \Pr(\hat{Y} = 1 \mid Y = 0)$ : false positive rate; type-1 error
  - $\Pr(M(D') \in E) = \Pr(\hat{Y} = 1 \mid Y = 1)$ : true positive rate; power
- DP:  $\text{FPR} \leq \exp(\epsilon) \cdot \text{TPR} + \delta$

$$\mathbb{D}_\alpha(M(D) \| M(D')) := \frac{1}{\alpha - 1} \log \mathbb{E}_{\mathbf{X} \sim q} \left( \frac{p(\mathbf{X})}{q(\mathbf{X})} \right)^\alpha \leq \epsilon$$

$$\text{equivalently, } \mathbb{E}_{\mathbf{X} \sim p} e^{(\alpha-1)r(\mathbf{X})} \leq e^{(\alpha-1)\epsilon}$$

- $p$  and  $q$  are the densities of  $M(D)$  and  $M(D')$ , respectively
- **Log odds ratio**:  $r = \log \frac{p}{q}$ ; a.k.a. privacy loss
- $\mathbb{D}_\alpha = \log \left[ \mathbb{E}_{\mathbf{X} \sim p} (r(\mathbf{X}))^{\alpha-1} \right] \frac{1}{\alpha-1}$  increasing w.r.t.  $\alpha \geq 1$ , in particular
  - $\alpha \downarrow 1 \implies \mathbb{D}_\alpha \rightarrow \text{KL}$
  - $\alpha \rightarrow \infty \implies \mathbb{D}_\alpha \rightarrow \max_{\mathbf{x}} \log \frac{p(\mathbf{x})}{q(\mathbf{x})}$ , used in  $(\epsilon, 0)$ -DP (see slide 14)

# The Many Shades of DP

- $\epsilon$ -DP: log odds ratio  $r$  uniformly bounded by  $\epsilon$
- $(\epsilon, \delta)$ -DP: roughly, with probability  $1 - \delta$ , we have  $r \leq \epsilon$ 
  - anything can happen for the remaining  $\delta$  probability
  - sacrificing some  $\delta$  proportion for (much?) better utility
  - the smaller  $\epsilon$  or  $\delta$  is, the stronger the privacy guarantee
- $\alpha$ -DP: bounds the exponential moment of  $r$ 
  - smoother transition than  $(\epsilon, \delta)$ -DP
  - implies  $(\epsilon, \delta)$ -DP by e.g. [Markov's inequality](#)
  - the **bigger**  $\alpha$  or the smaller  $\epsilon$  is, the stronger the privacy guarantee

# Calculus for DP

- Post-processing: If  $M$  is DP, so is  $T \circ M$  for any  $T$
- Parallel composition:  $D = \cup_k D_k$ , each  $M_k$  is DP, then  $M(D) := (M_1(D_1), \dots, M_K(D_K))$  is DP
- Sequential composition:  $(M(D), N(D, M(D)))$  is  $(\alpha, \epsilon_N + \epsilon_M)$ -RDP
  - cannot ask too many questions or run ML algorithms for too many epochs!
  - often been heavily abused in practice
- Differ by a group of  $k$ :  $(k\epsilon, 0)$ -DP
- Subsampling

# Gaussian Mechanism

$$M(D) := f(D) + \xi, \quad \text{where } \xi \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

- Sensitivity:  $\Delta_2 f := \sup_{D \sim D'} \|f(D) - f(D')\|_{\Sigma^{-1}}^2$
- $(\alpha, \epsilon)$ -RDP with  $\epsilon = \frac{\alpha}{2} \Delta_2 f$
- $(\alpha, \epsilon)$ -RDP  $\implies (\epsilon + \frac{1}{\alpha-1} \log \frac{1}{\delta}, \frac{\delta}{\alpha})$ -DP
  - note  $\alpha \rightarrow \infty \implies \mathbb{D}_\alpha \rightarrow \max_{\mathbf{x}} \log \frac{p(\mathbf{x})}{q(\mathbf{x})} \implies (\epsilon, 0)$ -DP
  - to achieve  $\alpha \rightarrow \infty$  with Gaussian mechanism:  $\epsilon = \frac{\alpha}{2} \Delta_2 f \rightarrow \infty$

---

**Algorithm 1:** Differentially private stochastic gradient descent

---

**Input:** model  $\mathbf{w}$ ; data  $\mathbf{x}_1, \dots, \mathbf{x}_n$ ; noise  $\sigma$ , gradient bound  $C$ , batch size  $b$

```
1 for  $t = 0, 1, \dots$  do
2   sample a random batch  $B_t$  with size  $b$ 
3   for  $i \in B_t$  do
4      $\mathbf{g}_i \leftarrow \nabla_{\mathbf{w}} \ell(\mathbf{x}_i; \mathbf{w})$  // compute grad
5      $\mathbf{g}_i \leftarrow \mathbf{g}_i / \max\{1, \|\mathbf{g}_i\|_2 / C\}$  // grad clipping
6    $\mathbf{g} \leftarrow [\frac{1}{b} \sum_{i \in B_t} \mathbf{g}_i] + \sigma C \xi$  // adding noise
7    $\mathbf{w} \leftarrow \mathbf{w} - \eta \cdot \mathbf{g}$  // grad descent
8    $\mathbf{w} \leftarrow \mathbf{P}(\mathbf{w})$  // projection
```

---

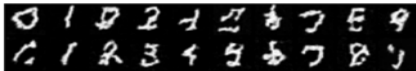




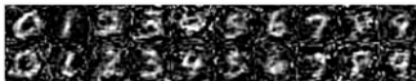
# Application in Generative Models

- Modern generative models are powerful, e.g., ChatGPT, DALLÉ-2
  - We can release the generative model as a proxy of releasing data
  - We can conduct data analysis / ML downstream tasks using generated data
- How to protect privacy when sensitive data (medical records, face images) are used in training?
- One solution: Differentially Private Generative Models - equip generative models with DP guarantees

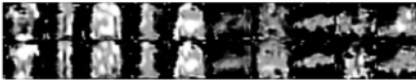
**DPDM**



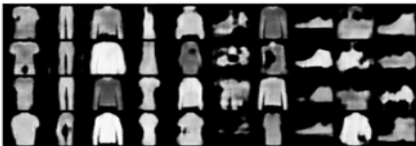
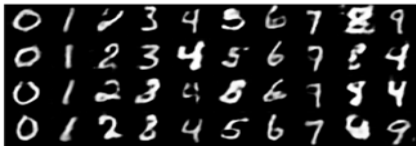
**DP-MERF**



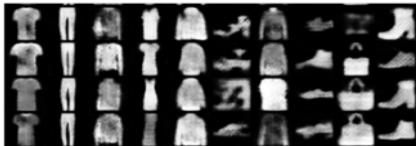
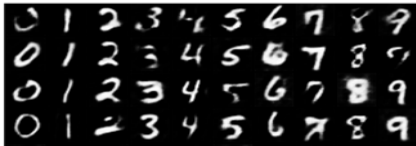
**PEARL**



**Ours (conditional)**



**Ours (parallel)**



**Figure 2: Qualitative comparison under  $(0.2, 10^{-5})$ -DP on MNIST and Fashion MNIST**