UNIVERSITY OF CAMBRIDGE

Differential Privacy

Talay M Cheema[1] and Ferenc Huszár[2]

[1]Department of Engineering, University of Cambridge
[2]Department of Computer Science and Technology, University of Cambridge

Outline

Differential Privacy in general

- Motivation and definitions
- The Laplace and exponential mechanisms
- δ-approximate DP and the Gaussian mechanism
- Zero-concentrated DP

Differential privacy in machine learning

- DP-SGD
- DP and generalisation

Why bother?

- Privacy is subjectively *important*

Why bother?

- Privacy is subjectively *important*
- Naive approaches are inadequate
  - **Anonymisation** foiled by using side-information
  - **Large queries** allow differencing attacks
  - **Benign facts** may not be benign...
  - **Query auditing** is hard, and non-answers are informative

Why bother?

- Privacy is subjectively *important*
- Naive approaches are inadequate
  - **Anonymisation** foiled by using side-information
  - **Large queries** allow differencing attacks
  - **Benign facts** may not be benign...
  - **Query auditing** is hard, and non-answers are informative
- Computational security and federated learning do different, *complementary* things

The setup

Users interact with a *trusted curator* of a database.

- Consider two databases x and x' which differ in one entry – x includes your data, x' doesn't.
- Users ask for some f to be computed on the database – e.g., number of PhD students in CBL; average age of students in CBL.
- The curator uses a noisy function ϕ instead.

The setup

Users interact with a *trusted curator* of a database.

- Consider two databases x and x' which differ in one entry – x includes your data, x' doesn't.
- Users ask for some f to be computed on the database – e.g., number of PhD students in CBL; average age of students in CBL.
- The curator uses a noisy function ϕ instead.

Your participation in the database should bring you no disadvantage

Privacy loss as a random variable

Privacy loss

If $\phi(x) \sim P, \phi(x') \sim P'$, then let the privacy loss be

$$\lambda(x||x') = \log \frac{P(r)}{P'(r)}, \qquad r \sim P.$$

- $\lambda(x||x')$ is the improvement of the Bayesian log odds in favour of x rather than x' *(in favour of you being in the database)*.

Privacy loss as a random variable

Privacy loss

If $\phi(x) \sim P, \phi(x') \sim P'$, then let the privacy loss be

$$\lambda(x||x') = \log \frac{P(r)}{P'(r)}, \qquad r \sim P.$$

- $\lambda(x||x')$ is the improvement of the Bayesian log odds in favour of x rather than x' *(in favour of you being in the database)*.
- This is a worst case assessment – an adversary may need a lot of side information to gain this much information.

ε differential privacy

Strict differential privacy

A function ϕ is ε differentially private if for *every* adjacent pair x, x'

$$\Pr[\lambda(x||x') \leq \varepsilon] = 1$$

Reflections

- Contrast with cryptographic methods – any user may be an adversary
- Contrast with information theory – worst case analysis rather than averages
- Privacy is guaranteed for *individuals* – privacy for arbitrary groups precludes learning

Outline

Differential Privacy in general

- Motivation and definitions
- The Laplace and exponential mechanisms
- δ-approximate DP and the Gaussian mechanism
- Zero-concentrated DP

Differential privacy in machine learning

- DP-SGD
- DP and generalisation

‘Just add noise’

$$\phi(x) = f(x) + \nu$$

A few issues...

- What if underestimates are much worse than overestimates?

'Just add noise'

$$\phi(x) = f(x) + \nu$$

A few issues...

- What if underestimates are much worse than overestimates?
- If ν has scale 1, but $f(x) - f(x') = 1000$...

'Just add noise'

$$\phi(x) = f(x) + \nu$$

A few issues...

- What if underestimates are much worse than overestimates?
- If ν has scale 1, but $f(x) - f(x') = 1000$...

Sensitivity

The ℓ_p sensitivity of a function f is

$$\Delta_p f = \sup_{x, x' \text{adjacent}} ||f(x) - f(x')||_p$$

The Laplace mechanism

The Laplace mechanism is ε-DP.

$$\phi(x) = f(x) + \nu, \qquad \nu \sim \text{Lap}\left(\frac{\Delta_1 f}{\varepsilon}\right)$$

The Laplace mechanism

The Laplace mechanism is ε-DP.

$$\phi(x) = f(x) + \nu, \qquad \nu \sim \text{Lap}\left(\frac{\Delta_1 f}{\varepsilon}\right)$$

Proof. $P(r) \propto \exp(-\frac{\varepsilon||f(x)-r||_1}{\Delta_1 f})$

The Laplace mechanism

The Laplace mechanism is ε-DP.

$$\phi(x) = f(x) + \nu, \qquad \nu \sim \text{Lap}\left(\frac{\Delta_1 f}{\varepsilon}\right)$$

Proof. $P(r) \propto \exp(-\frac{\varepsilon||f(x)-r||_1}{\Delta_1 f})$

$$\begin{aligned}\lambda(x||x') = \log\frac{P(r)}{P'(r)} &= \frac{\varepsilon||f(x')-r||_1}{\Delta_1 f} - \frac{\varepsilon||f(x)-r||_1}{\Delta_1 f} \\ &\le \frac{\varepsilon||f(x')-f(x)||_1}{\Delta_1 f} \\ &\le \varepsilon\end{aligned}$$

Privacy vs utility

- number of PhD students in CBL – $\Delta_1 f = ?$
- average age of students in CBL – $\Delta_1 f \approx ?$

Privacy vs utility

- number of PhD students in CBL – $\Delta_1 f = 1$
- average age of students in CBL – $\Delta_1 f \approx ?$

Privacy vs utility

- number of PhD students in CBL – $\Delta_1 f = 1$
- average age of students in CBL – $\Delta_1 f \approx a_{\max}/n$

Privacy vs utility

- number of PhD students in CBL – $\Delta_1 f = 1$
- average age of students in CBL – $\Delta_1 f \approx a_{\max}/n$
- Accuracy is compromised if noise is high...

Privacy vs utility

- number of PhD students in CBL – $\Delta_1 f = 1$
- average age of students in CBL – $\Delta_1 f \approx a_{\max}/n$
- Accuracy is compromised if noise is high...

The exponential mechanism

Let the utility of $f(x) = r$ be $u(x, r)$. Then for ε-DP, output r with distribution

$$p(r) \propto \exp\left(\frac{\varepsilon u(x, r)}{2 \max_r \Delta_1 u(\cdot, r)}\right).$$

This has strong utility guarantees, and the Laplace mechanism is a special case.

Outline

Composition

The total privacy loss of k ε-DP functions is $k\varepsilon$. To do better we need a relaxation.

Composition

The total privacy loss of k ε-DP functions is $k\varepsilon$. To do better we need a relaxation.

δ-approximate differential privacy

A function ϕ is δ-approximately ε differentially private (or (ε, δ)-DP) if for *every* adjacent pair x, x'

$$\Pr[\lambda(x||x') \leq \varepsilon] \geq 1 - \delta$$

A reasonable *worst case privacy loss.*

Advanced composition

The advanced composition theorem

For any δ', the composition of k (ε, δ)-DP mechanisms is $(\varepsilon', k\delta + \delta')$-DP with

$$\varepsilon' = \varepsilon\sqrt{2k \log \frac{1}{\delta'}} + \frac{1}{2}k\varepsilon^2$$

$\varepsilon' \approx \sqrt{k}\varepsilon$ for $k << \varepsilon^2$ if we allow a moderate leakage δ'.

The Gaussian mechanism

Gaussian mechanism version 1

For any $\varepsilon \in (0, 1), \delta > 0, c^2 = 2 \log \frac{1.25}{\delta}$, for (ε, δ)-DP

$$\phi(x) = f(x) + \nu \qquad \nu \sim \mathcal{N}(0, \sigma^2) \qquad \sigma = \frac{c\Delta_2 f}{\varepsilon}$$

Gaussian mechanism version 2

For any $\varepsilon > 0, \delta \in (0, 0.5), c^2 = 2 \log \frac{2}{\sqrt{16\delta+1}-1}$, for (ε, δ)-DP

$$\phi(x) = f(x) + \nu \qquad \nu \sim \mathcal{N}(0, \sigma^2) \qquad \sigma = \frac{(c + \sqrt{c^2 + \varepsilon})\Delta_2 f}{\varepsilon\sqrt{2}}$$

Outline

Differential Privacy in general

- Motivation and definitions
- The Laplace and exponential mechanisms
- δ-approximate DP and the Gaussian mechanism
- Zero-concentrated DP

Differential privacy in machine learning

- DP-SGD
- DP and generalisation

Towards a relaxation

Rényi divergence

The divergence of order $\alpha \in (1, \infty)$ is

$$D_\alpha(P||P') = \frac{1}{\alpha - 1} \log \int \left(\frac{P(r)}{P'(r)} \right) dP(r)$$
$$= \frac{1}{\alpha - 1} \log \mathbb{E}[e^{(\alpha-1)\lambda(x||x')}]$$

- $D_1(P||P') = D_{KL}(P||P') = \mathbb{E}[\lambda(x||x')]$
- $D_\infty(P||P') = \sup_r \lambda(x||x')$
- $D_\alpha(P||P')$ is increasing in α

Strict ε-DP: $D_\infty(P||P') \leq \varepsilon$ for every x, x' adjacent.

Strict ε-DP: $D_\infty(P||P') \leq \varepsilon$ for every x, x' adjacent.

Zero concentrated DP

ϕ is (ξ, ρ)-zCDP if for every adjacent x, x', and every $\alpha \in (1, \infty)$

$$D_\alpha(P||P') \leq \xi + \rho\alpha$$

- Clearly, $(\varepsilon, 0)$-zCDP $\iff$ ε-DP

Strict ε-DP: $D_\infty(P||P') \leq \varepsilon$ for every x, x' adjacent.

Zero concentrated DP

ϕ is (ξ, ρ)-zCDP if for every adjacent x, x', and every $\alpha \in (1, \infty)$

$$D_\alpha(P||P') \leq \xi + \rho\alpha$$

- Clearly, $(\varepsilon, 0)$-zCDP $\iff$ ε-DP
- More generally, zCDP characterises the decay of λ
- There are conversions between the two forms
- zCDP yields nice analyses of the Gaussian mechanism and group privacy

Outline