Lecture 10

Today

- Finish up PTR
- DP for modern ML
- Swap vs add/remove
- Composition & PLRV

Propose-Test-Release, Again

Example 1: releasing the mode

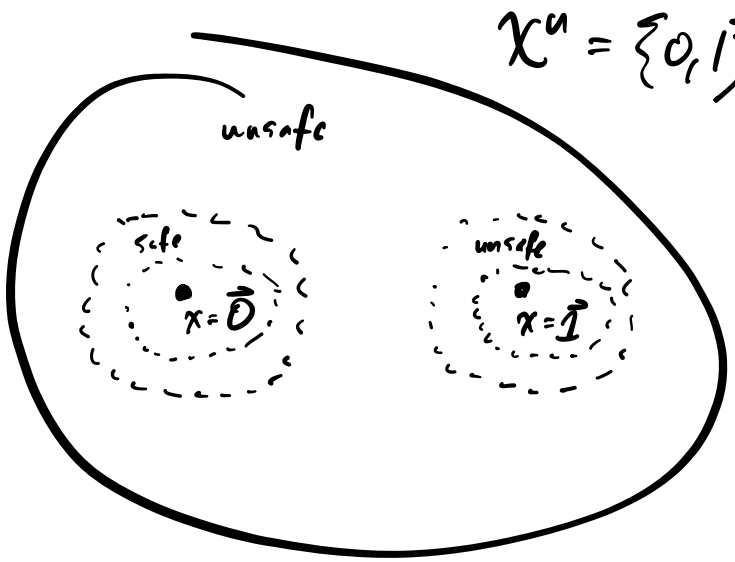
$x_1, \ldots, x_n \in \{0, 1\}$

base algo: $A(x) = \text{mode}(x)$

not DP

recall $SAFE^A_{\varepsilon,\delta} = \{x \in \mathcal{X}^n : \forall x' \sim x,\ A(x) \approx_{\varepsilon,\delta} A(x')\}$

Q: which datasets are safe?



PTR allows us to:

on $x \in$ unsafe, return $\perp$

on x far from unsafe, return mode(x)

inbetween, do one or the other

Exercises

1. Can approximate mode from a noisy histogram. When is PTR approach better or worse?
2. How do both approaches work if $\mathcal{X} = \mathbb{R}$?

In notes: mean estimation

base alg: $\frac{1}{n}\sum_{i=1}^{n} x_i + \mathcal{N}(0, \mathbb{I})$

(kind of)

safe datasets $\Longleftrightarrow$ datasets which are tightly clustered

DP for modern machine learning

Need stochastic first-order methods on large models, non-convex losses

Goal

Thousands of iterations

High dimensions

Stochastic mini-batches

Distributed training

Privacy Tools

New definitions: CDP, RDP, f-DP

DP-GD & dimension-independence

Privacy amplification by subsampling

DP-FTPL & correlated noise

New definitions for better composition & beyond

Def A is differentially private if for all **adjacent** x and x' we have

$$A(x) \approx A(x')$$

Def (swap) Datasets $x = (x_1, \ldots, x_n)$ and $x' = (x'_1, \ldots, x'_n)$ are tuples, swap-adjacent if same except in one entry.

Def (add/remove) Datasets $x = \{x_1, \ldots, x_n\}$ and $x' = \{x'_1, \ldots, x'_m\}$ are add/remove adjacent if x is x' plus one more element (or vice versa).

Theorists usually prefer swap (can assume n known)

Practitioners usually prefer add/remove.

Exercise How do these relate to each other?

Composition & the PLRV

Def For alg $A: \mathcal{X}^n \to \mathcal{Y}$ and datasets x, x', output y, the **privacy loss** is

$$L_A^{x \to x'}(y) = \log\left(\frac{\Pr[A(x)=y]}{\Pr[A(x')=y]}\right)$$

the **privacy loss random variable** (PLRV) is $L_A^{x \to x'}(Y)$ for $Y \sim A(x)$.

PLRV for Gaussians ↘

Likelihood ratio, measures evidence for hypothesis

$$H_0 : Y \sim A(x)$$

$$H_1 : Y \sim A(x')$$

Let $A(x) = (A_1(x), A_2(x), \ldots, A_k(x))$, each independent / not adaptive.

Claim (PLRV for Gaussians)

Let $A(x) \sim \mathcal{N}(\mu, \sigma^2)$ and $A(x') \sim \mathcal{N}(\mu', \sigma^2)$.

Then $L_A^{x \to x'}(Y) \sim \mathcal{N}(\rho, 2\rho^2)$ for $\frac{(\mu-\mu')^2}{2\sigma^2}$

Pf For fixed y,

$$L_A^{x \to x'}(y) = \log\left(\frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]}{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu')^2}{2\sigma^2}\right)}\right)$$

$$= \cdots$$

$$= \frac{(\mu-\mu')(2y-\mu-\mu')}{2\sigma^2}$$

Affine function of Y!

If $f(y) = ay + b$ and $Y \sim \mathcal{N}(\mu, \sigma^2)$,

then $f(Y) \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Rest as exercise ☒

Then to composition ↗

Then PLRV

$$L_A^{x \to x'}(Y) = \log\left(\frac{\Pr[A_1(x) = Y_1] \cdots \Pr[A_k(x) = Y_k]}{\Pr[A_1(x') = Y_1] \cdots \Pr[A_k(x') = Y_k]}\right)$$

$$= \sum_{j=1}^{k} \log\left(\frac{\Pr[A_j(x) = Y_j]}{\Pr[A_j(x') = Y_j]}\right)$$

$$= \sum_{j=1}^{k} L_{A_j}^{x \to x'}(Y_j)$$

Sum of independent random variables!

Central Limit Theorem says: $\xrightarrow{k \to \infty}$ Gaussian

<u>Claim</u> Suppose A_1 is ε-DP.

Let $L_1 \triangleq L_{A_1}^{x \to x'}(Y_1)$. Then $\mathbb{E}[L_1] \le \frac{1}{2}\varepsilon^2$

$\mathrm{Var}[L_1] \le \varepsilon^2$

<u>Proof sketch</u> . . . □

So $\mathbb{E}\left[L_A^{x \to x'}(Y)\right] = \sum_{j=1}^{k} \mathbb{E}[Z_j] \le \frac{1}{2}\varepsilon^2 k$

$\mathrm{Var}\left[L_A^{x \to x'}(Y)\right] = \sum_{j=1}^{k} \mathrm{Var}[Z_j] \le \varepsilon^2 k.$

CLT says: $L_A^{x \to x'}(Y) \approx \mathcal{N}\left(\frac{1}{2}\varepsilon^2 k, \varepsilon^2 k\right)$

Are we done? When PLRV is exactly Gaussian, have privacy analysis

How to make rigorous?

Standard tool: Berry-Esséen Theorem converges to normal at $O\left(\frac{1}{\sqrt{k}}\right)$ rate, too slow. Must take $\delta \geq \frac{1}{\sqrt{k}}$.

Our approach: look at MGF.

Moment Generating Functions

Def The moment generating function of a RV X is $M_X(\lambda) \triangleq \mathbb{E}\left[e^{\lambda X}\right]$.

Under some assumptions, $\mathbb{E}[X^n] = M_X^{(n)}(0)$ ← n-th derivative of $M_X(t)$

(moments)

Prove Hoeffding bound. State lemma.

use $\lambda = \frac{4n\varepsilon}{(b-a)^2}$ use $\mu = 0$

Markov's inequality $\Pr[X \geq t] \leq \frac{\mathbb{E}[Y]}{t}$

b/c $e^{\lambda X}$ is nonneg

Def A RV X with $\mathbb{E}X = 0$ is subgaussian if $\exists$ constant k such that $\forall \lambda \geq 0$, $\mathbb{E}[e^{\lambda X}] \leq e^{k\lambda^2/2}$.

Back to Differential Privacy

Def Alg A satisfies ρ-zero-Concentrated differential privacy if $\forall x, x'$ the PLRV

$$L = L_A^{x \to x'}(Y) \quad \text{satisfies}$$

$$\forall \lambda \geq 0 \quad \mathbb{E}\left[e^{\lambda L}\right] \leq \exp\left(\lambda(\lambda+1)\rho\right)$$

1) ε-DP is $\frac{1}{2}\varepsilon^2$-zCDP.

2) Composing k $\frac{1}{2}\varepsilon^2$-zCDP algorithms gives $\left(\frac{1}{2}\varepsilon^2 k\right)$-zCDP

3) $\left(\frac{1}{2}\varepsilon^2 k\right)$-zCDP gives (ε', δ)-DP for any $\delta > 0$ and $\varepsilon' = \varepsilon\sqrt{2k\log(1/\delta)} + \frac{1}{2}\varepsilon^2 k$.

Def A satisfies (α, ε)-Rényi differential privacy, (α,ε)-RDP, if for all $x \sim x'$

$$L = L_A^{x \to x'}(Y)$$

satisfies

$$\mathbb{E}\left[e^{(\alpha-1)Z}\right] \le \exp\left((\alpha-1)\cdot\varepsilon\right)$$

Claim If A satisfies $(\alpha, \alpha\rho)$-RDP for all $\alpha \in (1,\infty)$, then it satisfies ρ-zCDP.

Equivalently,

Def The Rényi divergence of P,Q at order α is

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1}\log \mathbb{E}_{x\sim Q}\left(\frac{P(x)}{Q(x)}\right)^\alpha$$

Def A satisfies (α,ε)-RDP if $\forall x \sim x'$ we have $D_\alpha(A(x)\|A(x')) \le \varepsilon$.