CS291A Introduction to Differential Privacy **Fall 2021**

Lecture 1: Course Overview / Privacy Challenges (September 27)

Lecturer: Yu-Xiang Wang *Scribes: Dan Qiao*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

1.1 A simple mathematic model

Consider a simple mathematical model, the dataset consists of n individuals. Each person has one secret bit of information from $\{0,1\}$, i.e. $x = [1,0,0,1,1,0,1,0,\cdots,1] \in \{0,1\}^n$. The adversary can use normalized linear query, i.e. he chooses $q \in \{0,1\}^n$ to get $\frac{1}{n}\sum_{i=1}^n q_i x_i = \frac{1}{n}q^T x = \frac{1}{n}\langle q, x\rangle$ or an approximate value of this term. We say an algorithm is blatantly non-private if one can reconstruct 90% of the dataset (secret bit vector) using its output.

Definition 1.1. *(blatant non-privacy, due to Dinur and Nissim). A mechanism $M : X^n \to Y$ is called blatantly non-private if for every $x \in X^n$, one can use M(x) to compute an $x' \in X^n$, such that x' and x differ in at most n/10 coordinates (with high probability over the randomness of M).*

The reconstruction attack is to find a dataset that is consistent with the observations. We have k linear queries $q_1, q_2, \cdots, q_k \in \{0,1\}^n$, $Q = [q_1, \cdots, q_k]^T \in \{0,1\}^{k\times n}$. An algorithm returns answers that are $\alpha-accurate$, i.e. it returns $y_1, y_2, \cdots, y_k \in [0,1]$, such that for any $i \in [k]$, $|y_i - \frac{1}{n}q_i^T x| \leq \alpha$. The reconstruction attack chooses $x' = argmin_{\tilde{x}\in\{0,1\}^n} max_{i\in[k]} |y_i - \frac{1}{n}q_i^T \tilde{x}|$. Because the true dataset $x \in \{0,1\}^n$, $max_{i\in[k]}|y_i - \frac{1}{n}q_i^T x'| \leq \alpha$.

1.2 All linear queries with constant error

Any algorithm that answers all 2^n linear queries with constant error implies blatant non-privacy.

Theorem 1.2. *(reconstruction from many queries with large error). Let $x \in \{0,1\}^n$. If we are given, for each $q \in \{0,1\}^n$, a value $y_q \in \mathbb{R}$ such that*

$$|y_q - \frac{\langle q, x\rangle}{n}| \leq \alpha.$$

Then one can use the y_q's to compute $x' \in \{0,1\}^n$ such that x and x' differ in at most 4α fraction of coordinates.

Proof. Show that for any $\tilde{x}$, such that $|y_q - \frac{q^T\tilde{x}}{n}| \leq \alpha$ for any $q \in \{0,1\}^n$, $\|\tilde{x} - x\|_1 \leq 4\alpha$.

$$\begin{aligned}\frac{1}{n}\|\tilde{x} - x\|_1 &= \frac{1}{n}\sum_{i=1}^n |\tilde{x}_i - x_i| \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{I}(\tilde{x}_i - x_i > 0)(\tilde{x}_i - x_i) + \frac{1}{n}\sum_{i=1}^n \mathbb{I}(\tilde{x}_i - x_i < 0)(x_i - \tilde{x}_i).\end{aligned}$$

Let $q \in \{0,1\}^n$ such that $q_i = 1$ iff $\tilde{x}_i - x_i > 0$.
Let $\tilde{q} \in \{0,1\}^n$ such that $\tilde{q}_i = 1$ iff $\tilde{x}_i - x_i < 0$.

$$\begin{aligned}\frac{1}{n}\|\tilde{x} - x\|_1 &= \frac{1}{n}|q^T(\tilde{x} - x)| + \frac{1}{n}|\tilde{q}^T(\tilde{x} - x)| \\ &\leq |\frac{q^T\tilde{x}}{n} - y_q| + |\frac{q^T x}{n} - y_q| + |\frac{\tilde{q}^T x}{n} - y_{\tilde{q}}| + |\frac{\tilde{q}^T\tilde{x}}{n} - y_{\tilde{q}}| \\ &\leq 4\alpha\end{aligned}$$

The first equation is because of the definition of q and $\tilde{q}$. The second inequality is because of triangular inequality. The third inequality is because each of the four terms is upper bounded by α. $\square$

1.3 O(n) linear queries with $O(\frac{1}{\sqrt{n}})$ error

Theorem 1.3. *(reconstruction from few queries with small error). There exists $c > 0$ and $q_1, \cdots, q_n \in \{0,1\}^n$ such that any mechanism that answers the normalized inner-product queries specified by $q_1, q_2, \cdots, q_n$ to within error at most $c/\sqrt{n}$ is blatantly non-private.*

Proof. Recall that the attack is $x' = argmin_{\tilde{x} \in \{0,1\}^n} max_{i \in [k]} |y_i - \frac{1}{n}q_i^T\tilde{x}|$ and $\|y - \frac{1}{n}Qx'\|_\infty \leq \frac{c}{\sqrt{n}}$, so $\frac{1}{n}\|Qx' - Qx\|_\infty \leq \frac{2c}{\sqrt{n}}$.
The choice of $q_1, \cdots, q_n$ is i.i.d. at random, q_i satisfies that $q_{ij} \sim Ber(0.5)$, i.e. $q_{ij} = 1$ with probability 0.5, $q_{ij} = 0$ with probability 0.5.
It suffices to show that for any $\tilde{x}$ such that $\|\tilde{x} - x\|_1 \geq 0.1n$, with high probability $\tilde{x}$ doesn't satisfy $\frac{1}{n}\|Q\tilde{x} - Qx\|_\infty \leq \frac{2c}{\sqrt{n}}$. As long as there exists a single query out of $i \in [n]$ such that $\frac{1}{n}|q_i^T(\tilde{x} - x)| > \frac{2c}{\sqrt{n}}$, then $\tilde{x}$ can not possibly be x'.
Let $z = \tilde{x} - x$, then $z_i \in \{-1, 0, 1\}$. $q^T z = \sum_{j=1}^n q_j z_j = 0.5\#(z_i = 1) - 0.5\#(z_i = -1) + \sum_{t=1}^{\#(z_i=1)+\#(z_i=-1)} f_t$, where f_t's are i.i.d and $f_t = 0.5$ with probability 0.5, $f_t = -0.5$ with probability 0.5.
As n goes to infinity, $\#(z_i = 1) + \#(z_i = -1)$ also goes to infinity, because of C.L.T. (lemma 1.4), the difference between $P[\frac{1}{\sqrt{\#(z_i=1)+\#(z_i=-1)}}\sum_{t=1}^{\#(z_i=1)+\#(z_i=-1)} f_t \leq x]$ and $\phi(4x)$ will converge to 0 uniformly, which means the difference between $P[\frac{1}{\sqrt{n}}q^T z \leq x]$ and $P[N(\frac{0.5\#(z_i=1)-0.5\#(z_i=-1)}{\sqrt{n}}, \frac{\#(z_i=1)+\#(z_i=-1)}{4n}) \leq x]$ will converge to 0 uniformly. So for n large enough,

$$P[\frac{1}{\sqrt{n}}|q^T z| \leq 2c] \leq P[|N(\frac{0.5\#(z_i = 1) - 0.5\#(z_i = -1)}{\sqrt{n}}, \frac{\#(z_i = 1) + \#(z_i = -1)}{4n})| \leq 2c] + \frac{1}{8}$$

Assume this normal distribution to be $N(\mu, \sigma^2)$, then the density function is smaller than $\frac{1}{\sqrt{2\pi}\sigma}$ at any point, which means the density function of this normal distribution is smaller than $\sqrt{\frac{20}{\pi}}$ at any point (here $\sigma^2 = \frac{1}{n}Var[q^T z] \geq \frac{1}{40}$). So

$$\begin{aligned}P[\frac{1}{n}|q^T z| \leq \frac{2c}{\sqrt{n}}] &= P[\frac{1}{\sqrt{n}}|q^T z| \leq 2c] \\ &\leq \sqrt{\frac{20}{\pi}} \times 4c + \frac{1}{8}.\end{aligned}$$

We can choose c sufficiently small such that $P[\frac{1}{n}|q^T z| \leq \frac{2c}{\sqrt{n}}] \leq 0.25$ for n large enough.

In this way,

$$\begin{aligned} P(\tilde{x}\ is\ selected) &\leq P(max_i|\frac{1}{n}q_i^T\tilde{x} - y_i| \leq \frac{c}{\sqrt{n}}) \\ &\leq P(max_i|\frac{1}{n}q_i^T(\tilde{x} - x)| \leq \frac{2c}{\sqrt{n}}) \\ &\leq 0.25^n \\ &= 2^{-2n}. \end{aligned}$$

The third inequality holds because q_i's are independent.
So

$$\begin{aligned} P(any\ \tilde{x}\ is\ selected\ such\ that\ \|\tilde{x} - x\|_1 > 0.1n) &\leq |\{\tilde{x} : \|\tilde{x} - x\|_1 > 0.1n\}| \times P(\tilde{x}\ is\ selected) \\ &\leq 2^n \times 2^{-2n} \\ &= 2^{-n} \end{aligned}$$

The first inequality is because of union bound. The second inequality is because there are at most 2^n possible $\tilde{x}$'s.
This means with high probability, the attack will output x' such that $\|x' - x\|_1 \leq 0.1n$. $\square$

Lemma 1.4. *(Lindeberg–Lévy C.L.T., Berry-Esseen C.L.T.) Suppose* $\{X_1, \ldots, X_n\}$ *is a sequence of i.i.d. random variables with* $\mathbb{E}[X_i] = \mu$ *and* $\operatorname{Var}[X_i] = \sigma^2 < \infty$*. Then as* n *approaches infinity, the random variables* $\sqrt{n}(\bar{X}_n - \mu)$ *converge in distribution to a normal* $\mathcal{N}(0, \sigma^2)$*:* $\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$.
In addition, the convergence is uniform in z *in the sense that*

$$\lim_{n\to\infty} \sup_{z\in\mathbb{R}} \left|\mathbb{P}\left[\sqrt{n}(\bar{X}_n - \mu) \leq z\right] - \Phi\left(\frac{z}{\sigma}\right)\right| = 0 \ .$$

1.4 Conclusion

Any algorithm that answers too many questions too accurately will result in a blatant reconstruction of the dataset.
The attacks are not computationally efficient, but efficient attacks exist, via a linear programming relaxation.

$$x' = argmin_{\tilde{x}\in[0,1]^n} max_{i\in[k]}|y_i - \frac{1}{n}q_i^T\tilde{x}|.$$