Sharif University of Technology

Differential Privacy

Instructor: Prof. M. H. Yassaee

Name: Khashayar Ghaffari **Student ID:** 98100215

1 Differential Privacy and Reconstruction Attacks

We have, $\mathbb{E}[E_i] = \mathbb{P}[\tilde{x}_i \neq x_i] = \mathbb{P}[x_i = 1 \mid A(x) = a]$.

$$\Rightarrow \mathbb{E}[||\tilde{x} - x||_1] = \mathbb{E}[\sum_i E_i] = \sum_i \mathbb{E}[E_i]$$

Now we can say,

$$\frac{\mathbb{P}[x_i = 1 \mid A(x) = a]}{\mathbb{P}[x_i = 0 \mid A(x) = a]} = \frac{\mathbb{P}[A(x) = a \mid x_i = 1]\mathbb{P}[A(x) = a]}{\mathbb{P}[A(x) = a \mid x_i = 0]\mathbb{P}[A(x) = a]}$$

$$= \frac{\mathbb{P}[A(x) = a]\mathbb{P}[x_i = 1]\mathbb{P}[A(x) = a]}{\mathbb{P}[A(\tilde{x}) = a]\mathbb{P}[x_i = 0]\mathbb{P}[A(x) = a]} = \frac{\mathbb{P}[A(x) = a]}{\mathbb{P}[A(\tilde{x'} = a)]} \leq e^{\epsilon}$$

Also we know that, $\mathbb{P}[x_i = 1 \mid A(x) = a] + \mathbb{P}[x_i = 0 \mid A(x) = a] = 1$, so we have,

$$\mathbb{P}[x_i = 1 \mid A(x) = a] \geq \frac{1}{1 + e^{\epsilon}} \Rightarrow \mathbb{E}[\#errors] \geq \frac{n}{1 + e^{\epsilon}}$$

2 Approximate Differential Privacy

1. Let $M : \mathcal{X}^n \rightarrow \mathcal{Y}$ be a (ϵ, δ)-differentially private, and let $F : \mathcal{Y} \rightarrow \mathcal{Z}$ be an arbitrary randomized mapping. Then we prove that $F \circ M$ is (ϵ, δ)-differentially private.
   Since F is a randomized function, we can consider it to be a distribution over deterministic functions f. The privacy proof follows for every neighbouring dataset X, X' and $T \subseteq \mathcal{Y}$:

$$\mathbb{P}[F(M(X)) \in T] = \mathbb{E}_{f\sim F}[\mathbb{P}[M(X) \in f^{-1}(T)]]$$

$$\leq \mathbb{E}_{f\sim F}[e^{\epsilon}\mathbb{P}[M(X') \in f^{-1}(T)] + \delta]$$
$$= e^{\epsilon}\mathbb{P}[F(M(X')) \in T] + \delta$$

2. Let $S = \mathcal{X}_1 \times \mathcal{X}_2$, then we have $\mathbb{P}[x \in S] = \mathbb{P}[x_1 \in \mathcal{X}_1]\mathbb{P}[x_2 \in \mathcal{X}_2 \mid x_1 \in \mathcal{X}_1]$.
   Since the algorithms are differentially private,

$$\mathbb{P}[x_1 \in \mathcal{X}_1]\mathbb{P}[x_2 \in \mathcal{X}_2 \mid x_1 \in \mathcal{X}_1] \leq (e^{\epsilon_1}\mathbb{P}[x_1' \in \mathcal{X}_1] + \delta_1)(e^{\epsilon_2}\mathbb{P}[x_2' \in \mathcal{X}_2 \mid x_1 \in \mathcal{X}_1] + \delta_2)$$

$$\Rightarrow \mathbb{P}[x \in S] \leq (e^{\epsilon_1}\mathbb{P}[x_1' \in \mathcal{X}_1] + \delta_1)(e^{\epsilon_2}\mathbb{P}[x_2' \in \mathcal{X}_2 \mid x_1 \in \mathcal{X}_1] + \delta_2)$$
$$= (e^{\epsilon_1}\mathbb{P}[x_1' \in \mathcal{X}_1]e^{\epsilon_2}\mathbb{P}[x_2' \in \mathcal{X}_2 \mid x_1 \in \mathcal{X}_1])$$
$$+\delta_1 \underbrace{(e^{\epsilon_2}\mathbb{P}[x_2' \in \mathcal{X}_2 \mid x_1 \in \mathcal{X}_1] + \delta_2)}_{\leq 1, \text{ (Approximate DP Theorem)}} + \delta_2 \underbrace{(e^{\epsilon_1}\mathbb{P}[x_1' \in \mathcal{X}_1] + \delta_1)}_{\leq 1} - \delta_1\delta_2$$
$$\leq (e^{\epsilon_1+\epsilon_2}\mathbb{P}[x_1' \in \mathcal{X}_1]\mathbb{P}[x_2' \in \mathcal{X}_2 \mid x_1 \in \mathcal{X}_1]) + \delta_1 + \delta_2 = e^{\epsilon_1+\epsilon_2}\mathbb{P}[x' \in S] + \delta_1 + \delta_2$$

3 Differentially Private LSR Problem

Let $\rho_{\epsilon,\delta} = \frac{\epsilon^2}{ln\frac{1}{\delta}}$, and let $C = max\ \mathbb{E}$ of l_2-sensitivity, then we can say,

$$C = \max_t \ \mathbb{E}[||\nabla L(\theta, X) - \nabla L(\theta^* < X)||_2^2] = \max_t \ \mathbb{E}[||\nabla L(\theta, X)||_2^2]$$

$$= \max_t \ \mathbb{E}[||\theta_t^T 2(y - \theta_t^T x)||_2] \leq R^2.16R^2 d = 16R^4$$

We know that, $\mathbb{E}L(\theta^{priv}, X) - \min_{\theta \in B_2(0,R)^d} L(\theta, X) \leq \frac{RC}{\sqrt{T}} + \frac{2RC\sqrt{d}}{n\sqrt{\rho_{\epsilon,\delta}}}$.

Now it is enough to set, $T \geq \frac{4R^2C^2}{\alpha^2}$ and $n_0 \geq \frac{4RC\sqrt{d}}{\alpha\sqrt{\rho_{\epsilon,\delta}}}$.

Then for each $n \geq n_0$, we can see that $\frac{RC}{\sqrt{T}} + \frac{2RC\sqrt{d}}{n\sqrt{\rho_{\epsilon,\delta}}} \leq \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$. So we have the following:

$$\mathbb{E}L(\theta^{priv}, X) - \min_{\theta \in B_2(0,R)^d} L(\theta, X) \leq \alpha.$$

4 An Unknown Private Algorithm

Employ the exponential mechanism with an output range $\mathcal{X}$ and a utility function $u : \mathcal{X}^n \times \mathcal{X}$ defined as

$$u(X, y) = \min \left\{ |\{i : x_i \leq y\}|, |\{i : x_i \geq y\}| \right\}.$$

Note that y lies between $\min_{i=1}^n x_i$ and $\max_{i=1}^n x_i$ when $u(X, y) \geq 1$. Also, $\mathrm{OPT}(X) = \max_{y \in X} u(X, y)$ satisfies $\mathrm{OPT}(X) \geq \frac{n}{2}$ by choosing y as the $\lfloor \frac{n}{2} \rfloor$-th element in the sorted order of X.
The sensitivity Δu of u is at most 1, evident from the fact that for neighboring datasets X and X_0,

$$|\{i : x_i \leq y\} - \{i : x_{0i} \leq y\}| \leq 1, \quad |\{i : x_i \geq y\} - \{i : x_{0i} \geq y\}| \leq 1$$

Therefore, $|u(X, y) - u(X_0, y)| \leq 1$.
By the utility guarantee for the exponential mechanism, for the random output y,

$$\mathbb{P}\left(u(X, y) \geq \mathrm{OPT}(X) - \frac{2}{\varepsilon} \ln\left(\frac{|N|}{\beta}\right)\right) \geq 1 - \beta.$$

Given $\mathrm{OPT}(X) \geq \frac{n}{2}$, if $n \geq \frac{4}{\varepsilon} \ln\left(\frac{|N|}{\beta}\right) + 3$, then $\mathrm{OPT}(X) - \frac{2}{\varepsilon} \ln\left(\frac{|N|}{\beta}\right) \geq 1$, leading to

$$\mathbb{P}(u(X, y) \geq 1) \geq 1 - \beta.$$

This implies that $u(X, y) \geq 1$ ensures y is between $\min_{i=1}^n x_i$ and $\max_{i=1}^n x_i$, validating the algorithm's desired property. The ε-differential privacy of the algorithm follows from the privacy analysis of the exponential mechanism.