SHARIF UNIVERSITY OF TECHNOLOGY

Differential Privacy

Instructor: Prof. M. H. Yassaee

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

1 Differentially Private Logistic Regression

(a) We know that:

$$L(\theta, X) = \frac{1}{n}\sum_{i=1}^{n} log(1 + e^{-y_i \sum_{k=1}^{d} \theta_k x_k + \theta_0}) \Rightarrow \frac{\partial L}{\partial \theta_j} = \frac{x_j}{n}\sum_{i=1}^{n}\frac{-y_i e^{-y_i f_\theta(x_i)}}{1 + e^{-y_i f_\theta(x_i)}}$$

Now we want upper bound on the ℓ_1-sensitivity of $\nabla L(\theta, X)$, so assume that X and X' are neighbours.

$$||\nabla L(\theta, X) - \nabla L(\theta, X')||_1 = |\frac{\partial L(\theta, X)}{\partial \theta_j} - \frac{\partial L(\theta, X')}{\partial \theta_j}| = |\frac{x_j - x'_j}{n}\sum_{i=1}^{n}\frac{-y_i e^{-y_i f_\theta(x_i)}}{1 + e^{-y_i f_\theta(x_i)}}|$$

$$|\frac{(x_j - x'_j)}{n}\sum_{i=1}^{n} y_i \frac{e^{-y_i f_\theta(x_i)}}{1 + e^{-y_i f_\theta(x_i)}}| \leq |\frac{(x_j - x'_j)}{n}\sum_{i=1}^{n} y_i| = |\frac{x_j - x'_j}{n}||\sum_{i=1}^{n} y_i| \leq \frac{1}{n}\sum_{i=1}^{n}|y_i| \leq 1$$

2 Graph Privacy and Different Types of Sensitivity

(a) General Sensitivity:

$$GS_q = \max_{G,G':G\sim G'} |q(G) - q(G')| \geq \max |q(G)| - \min |q(G')| \geq n - 0 = n$$

$$\Rightarrow GS_q \geq n$$

Now, let G be a graph with n vertices and zero edges, and G' be a graph with n vertices and edge set $E' = \{(1,2),(1,3),...,(1,n)\}$. We have $G \sim G'$, because G and G' are different just in edges of vertex 1. Also we have,

$$|q(G) - q(G')| = n - 0 = n$$
$$\Rightarrow GS_q \leq n$$

So we claim that, $GS_q = n$.

(b) Local Sensitivity:

$$LS_q(G) = \max_{G':G\sim G'} |q(G) - q(G')|$$

For find the graph G' which $LS_q(G) = |q(G) - q(G')|$, notice that we should choose one vertex like v from G and do one the following to construct G':

- Connect v to all other vertices.
- Remove all of the edges which connected to v.

In the first case we have $q(G') = 0$, so $|q(G) - q(G')| = q(G)$.
In the second case we have $q(G') = q(G) + 1 +$ [#Vertices connected to v and has degree 1], so we have, $|q(G) - q(G')| = 1 +$ [#Vertices connected to v and has degree 1].

$$\Rightarrow LS_q(G) = \max_{v\in V(G)} \{q(G),\ 1 + [\text{\#Vertices connected to } v \text{ and has degree } 1]\} \geq 1.$$

Let K_n be a complete graph of order $n \geq 3$, then we have, $LS_q(K_n) = 1$, so for all $n \geq 3$ we have $\min_G\ LS_q(G) = 1$.
If $n = 2$, then it is simple to see that $\min_G\ LS_q(G) = 2$.

(c) Local Sensitivity Over $\mathcal{H}$:

$$\max_{G\in\mathcal{H}} LS_q(G) = \max_{G\in\mathcal{H}} \max_{G'\in\mathcal{G}:G\sim G'} |q(G) - q(G')| \leq GS_q = n$$

Let G be a graph with n vertices and zero edges, then by definition we have, $G \in \mathcal{H}$. And also let G' be a graph with n vertices and edge set $E' = \{(1,2),(1,3),...,(1,n)\}$. We have, $|q(G) - q(G')| = n$. (G and G' are neighbours.) So we have $\max_{G\in\mathcal{H}}\ LS_q(G) \geq n$. By first inequality we claim that, $\max_{G\in\mathcal{H}}\ LS_q(G) = n$.

(d) Restricted Sensitivity on $\mathcal{H}$: By our explanations in part (b) we can say that, $RS_q^{\mathcal{H}} \leq d + 1$. Also let G be a graph and assume that one of its vertices like v is connected to d vertices which has degree 1. So we can construct G' (Similar to part (b)) which has this property:

$$|q(G) - q(G')| = 1 + [\text{\#Vertices connected to } v \text{ and has degree } 1] = 1 + d$$

$$\Rightarrow RS_q^{\mathcal{H}} = d + 1$$

3 Lipschitz Extensions

(a)
$GS_f = \infty$ because for all $t > 0$ we can find two neighbour vectors such that, $|f(x) - f(x')| > t$.
$LS_f(x) = \max_{x':x\sim x'} |f(x) - f(x')| = \max_{x':x\sim x'} \frac{1}{n}|x_j - x'_j| = \infty$
$RS_f^{\mathcal{H}} = \max_{x\sim x'} \frac{1}{n}|\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x'_i| = \max_{x,x'} \frac{1}{n}|x_j - x'_j| = \frac{b-a}{n}$
$g(x) = \frac{1}{n}\sum_{i=1}^{n} \max\{a,\ \min\{b,\ x_i\}\}$ is Lipschitz Extension.

(b)
$GS_f = \infty$ similar to part (a)
$\min_{x\in\mathcal{G}} LS_f(x) = \min_{x\in\mathcal{G}} \max_{x':x\sim x'} |f(x) - f(x')| = \min_{x\in\mathcal{G}} \max_{x':x\sim x'} |median(x) - median(x')|$ if we put $x_1 = x_2 = ... = x_n$, then we have $median(x) = x_1 = median(x')$. (for all $n \geq 3$)
$\Rightarrow \min_{x\in\mathcal{G}} LS_f(x) = 0$
$RS_f^{\mathcal{H}} = \max_{x\sim x'} |median(x) - median(x')| \leq b - a$
If $x = (a, a, ..., a, a, b, b, ..., b)$ and x'=(a,a,...,a,b,b,b,...,b), then we have, $|median(x) - median(x')| = b - a$.
So we claim that, $RS_f^{\mathcal{H}} = b - a$.

(c)
Question 2!