HW2-Report

Saman Soleimani 400206284

Question 1:

a) Assume X and $X^{'}$ are two neighboring datasets that differs on ith data.

$$
\begin{aligned}
GS = \max_{X \sim X^{'}} \|\nabla_\theta L(\theta, X) - \nabla_\theta L(\theta, X^{'})\|_1 \\
&= \frac{1}{n} \max_{x_i \neq x_i^{'}} \| - \frac{\begin{bmatrix}1 & x_{i1} & \dots & x_{id}\end{bmatrix}^T y_i e^{-y_i f_\theta(x_i)}}{1 + e^{-y_i f_\theta(x_i)}} \\
&+ \frac{\begin{bmatrix}1 & x_{i1}^{'} & \dots & x_{id}^{'}\end{bmatrix}^T y_i^{'} e^{-y_i^{'} f_\theta(x_i^{'})}}{1 + e^{-y_i^{'} f_\theta(x_i^{'})}} \|_1 \\
&\leq \max_{x_i \in [0,1]^d} \frac{2}{n} \| \frac{\begin{bmatrix}1 & x_{i1} & \dots & x_{id}\end{bmatrix}^T y_i e^{-y_i f_\theta(x_i)}}{1 + e^{-y_i f_\theta(x_i)}} \|_1 \\
&= \frac{2}{n}(d+1)\frac{e}{1+e}
\end{aligned}
$$

If we want to use the basic composition of the Laplace mechanism, the privacy cost in each iteration should be $\frac{\varepsilon}{T}$. Thus, the Laplace noise $Lap\left(\frac{2T(d+1)\frac{e}{1+e}}{n\varepsilon}\right)$ should be added to the gradient of the loss function in each iteration. The variance of this noise is $\frac{8T^2(d+1)^2\frac{e^2}{(1+e)^2}}{n^2\varepsilon^2}$.

b) $\log(1+e^x)$ is convex and nondecreasing, and $yf_\theta(x)$ is convex with respect to θ. Therefore, the composition of these functions, which is the loss function on one sample, is convex. Consequently, the loss function on all samples, $L(\theta, X)$, is convex. So, if $\theta^* = \underset{\theta \in B_2(0,R)^d}{\operatorname{argmin}} L(\theta, X)$, $g^t = \nabla_\theta L(\theta^t, X)$ is the gradient, $\tilde{g}^t$ is an unbiased estimate of the gradient (similar to what we encounter in Noisy PGD), and η is the learning rate:

$$
\begin{aligned}
&\mathbb{E}[L(\theta^t, X) - L(\theta^*, X)] = \mathbb{E}[L(\theta^t, X)] - L(\theta^*, X) \\
&\leq \mathbb{E}\left[\frac{1}{\eta}\langle \eta\tilde{g}^t, \theta^t - \theta^* \rangle\right] \\
&\leq \mathbb{E}\left[\frac{1}{2\eta}\left(\|\eta\tilde{g}^t\|^2 + \|\theta^t - \theta^*\|^2 - \|\theta^t - \eta\tilde{g}^t - \theta^*\|^2\right)\right]
\end{aligned}
$$

Because $\mathcal{C}$, the domain set of θ is convex, if $\tilde{u}^t = \theta^t - \eta\tilde{g}^t$, then this property holds for the projection function $\pi_{\mathcal{C}}(.) : \|\pi_{\mathcal{C}}(\tilde{u}^t) - y\|^2 \leq \|\tilde{u}^t - y\|^2; \forall y \in \mathcal{C}$.

$$\begin{aligned}\mathbb{E}[L(\theta^t, X)] - L(\theta^*, X) &\leq \mathbb{E}\left[\frac{1}{2\eta}\left(\|\eta\tilde{g}^t\|^2 + \|\theta^t - \theta^*\|^2 - \|\pi_{\mathcal{C}}(\theta^t - \eta\tilde{g}^t) - \theta^*\|^2\right)\right] \\ &= \mathbb{E}\left[\frac{1}{2\eta}\left(\|\eta\tilde{g}^t\|^2 + \|\theta^t - \theta^*\|^2 - \|\theta^{t+1} - \theta^*\|^2\right)\right] \\ &= \frac{\eta}{2}\mathbb{E}[\|\tilde{g}^t\|^2] + \frac{1}{2\eta}\mathbb{E}[\|\theta^t - \theta^*\|^2 - \|\theta^{t+1} - \theta^*\|^2]\end{aligned}$$

$$\Rightarrow \quad \mathbb{E}[L(\theta^t, X)] - L(\theta^*, X) \leq \frac{\eta}{2}\mathbb{E}[\|\tilde{g}^t\|^2] + \frac{1}{2\eta}\mathbb{E}[\|\theta^t - \theta^*\|^2 - \|\theta^{t+1} - \theta^*\|^2] \quad (1)$$

Due to the convexity of the loss landscape and the application of Jensen's inequality, we can infer:

$$\mathbb{E}[L(\theta^{\text{priv}}, X)] = \mathbb{E}[L(\frac{1}{T}\sum_{t=0}^{T-1}\theta^t, X)] \leq \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[L(\theta^t, X)] \quad (2)$$

References to equations 1 and 2 result in:

$$\begin{aligned}\mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) &\leq \frac{\eta}{2T}\sum_{t=0}^{T-1}\mathbb{E}[\|\tilde{g}^t\|^2] + \frac{1}{2\eta T}\sum_{t=0}^{T-1}\mathbb{E}[\|\theta^t - \theta^*\|^2 - \|\theta^{t+1} - \theta^*\|^2] \\ &\leq \frac{\eta}{2}\max_t \mathbb{E}[\|\tilde{g}^t\|^2] + \frac{1}{2\eta T}\mathbb{E}[\|\theta^0 - \theta^*\|^2 - \|\theta^T - \theta^*\|^2] \\ &\leq \frac{\eta}{2}\max_t \mathbb{E}[\|\tilde{g}^t\|^2] + \frac{1}{2\eta T}\mathbb{E}[\|\theta^0 - \theta^T\|^2]\end{aligned}$$

$$\Rightarrow \quad \mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) \leq \frac{\eta}{2}\max_t \mathbb{E}[\|\tilde{g}^t\|^2] + \frac{1}{2\eta T}\mathbb{E}[\|\theta^0 - \theta^T\|^2] \quad (3)$$

Because $\theta \in B_2(0, R)^d$, it follows that $\|\theta^0 - \theta^T\|^2 \leq 4R^2$. Let $\tilde{g}^t = g^t + Z$, where Z is assumed to be a vector with entries sampled from $Lap\left(\frac{2T(d+1)\frac{e}{1+e}}{n\varepsilon}\right)$. Therefore, $\mathbb{E}[|\tilde{g}^t|^2] = \mathbb{E}[\|g^t\|^2 + \|Z\|^2 + 2\langle g^t, Z\rangle] = \|g^t\|^2 + \mathbb{E}[\|Z\|^2] \leq G^2 + (d+1)\frac{8T^2(d+1)^2\frac{e^2}{(1+e)^2}}{n^2\varepsilon^2}$, where G is the Lipschitzness of the loss function. So, these results and Equation 3 yield:

$$\mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) \leq \frac{\eta}{2}(G^2 + \frac{8T^2(d+1)^3\frac{e^2}{(1+e)^2}}{n^2\varepsilon^2}) + \frac{4R^2}{2\eta T} \quad (4)$$

The value of G should be determined based on the information provided by the problem.

$$\|\nabla_\theta L(\theta, X)\| \leq \frac{1}{n}\sum_{i=1}^{n}\max_{x_i \neq x_i'}\|-\frac{\begin{bmatrix}1 & x_{i1} & \dots & x_{id}\end{bmatrix}^T y_i e^{-y_i f_\theta(x_i)}}{1 + e^{-y_i f_\theta(x_i)}}\| \leq \frac{e}{1+e}\sqrt{d+1}$$

So, I set $G = \frac{e}{1+e}\sqrt{d+1}$ and substitute this into 4:

$$\mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) \leq \frac{\eta}{2}((d+1)\frac{e^2}{(1+e)^2} + \frac{8T^2(d+1)^3\frac{e^2}{(1+e)^2}}{n^2\varepsilon^2}) + \frac{4R^2}{2\eta T} \quad (5)$$

Therefore, if we set $T = \frac{n\varepsilon}{\sqrt{8}(d+1)}$, 5 gives:

$$\mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) \leq \eta((d+1)\frac{e^2}{(1+e)^2}) + \frac{4R^2}{2\eta T} \tag{6}$$

I set $\eta = \frac{\sqrt{2}\frac{1+e}{e}R}{\sqrt{(d+1)T}}$, 6 gives:

$$\mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) \leq \sqrt{\frac{2e^2R^2(d+1)}{(1+e)^2T}} = \sqrt{\frac{2\sqrt{8}e^2R^2(d+1)^2}{(1+e)^2n\varepsilon}}$$

Because $\mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) \leq \alpha$, so the underlying inequality should be true:

$$\sqrt{\frac{2\sqrt{8}e^2R^2(d+1)^2}{(1+e)^2n\varepsilon}} \leq \alpha$$
$$\rightarrow \frac{2\sqrt{8}e^2R^2(d+1)^2}{(1+e)^2\alpha^2\varepsilon} \leq n$$

So, K should be greater than $\frac{2\sqrt{8}e^2}{(1+e)^2}$.

Question 2:

a)For global sensitivity, a graph on $\mathcal{G}$ should be found. If we change the connectivity of edges incident to one node, the number of isolated nodes should be maximum. A star graph is a good choice because if we remove the edges incident to the center node, all n nodes will be isolated. So, $GS_q = n$.

b) For minimum local sensitivity, $\min_{G\in\mathcal{G}} \max_{G\sim G'} \|q(G) - q(G')\|$, a graph should be found such that for all its neighbors, the difference between the number of isolated nodes of this graph and all its neighbors should be minimum. A complete graph should be a good choice because, if we remove even all edges incident to a node, only one node will be isolated. So, $\min_{G\in\mathcal{G}} LS_q = 1$.

c) For the maximum local sensitivity on $\mathcal{H}$, $\max_{G\in\mathcal{H}} \max_{G\sim G'} \|q(G) - q(G')\|$, I need to find a graph in the hypothesis set such that the difference between the number of isolated nodes in this graph and all its neighbors is maximized. The hypothesis set consists of graphs with a maximum vertex degree at most d. Therefore, let's consider an n-node empty graph for G with n-node isolated, and G' as a star graph with n nodes and 0 nodes isolated. Note that G' may not be in the hypothesis set. Thus, $\max_{G\in\mathcal{H}} LS_q = n$.

d) The restricted sensitivity, $\max_{G,G'\in\mathcal{H},G\sim G'} |q(G) - q(G')|$, is similar to the maximum local sensitivity on $\mathcal{H}$; the difference is that the neighboring graph should be in $\mathcal{H}$. Therefore, assume an n-node empty graph for G with n-node isolated, and G' consists of two subgraphs: the first is a star graph with $d+1$ nodes, and the second is an empty graph with $n-(d+1)$ nodes. This configuration results in $n-(d+1)$ isolated nodes. Hence, $RS_q^{\mathcal{H}} = d+1$.

Question 3:

a)

i) The global sensitivity is infinite because one can change some finite data to infinite data in a neighboring dataset.

ii) If we consider every dataset, we can replace one data point with a very large or very small value, leading to infinite local sensitivity. Therefore, the minimum local sensitivity is infinite.

iii) For the restricted sensitivity, if we replace one data point equal to a with b, the restricted sensitivity becomes $\frac{b-a}{n}$.

iv) Function f from $\mathcal{H}$ has restricted sensitivity $\frac{b-a}{n}$; thus, it is $\frac{b-a}{n}$ Lipschitz with respect to the Hamming norm. An explicit Lipschitz extension from $\mathcal{H}$ to $\mathcal{G}$ is given by:

$$\widetilde{f}(x) = \frac{1}{n}\sum_{i=1}^{n}(\text{truncate } x_i \text{ to } [a,b])$$

Therefore, this function is in $\mathcal{G}$ and is Lipschitz with $\frac{b-a}{n}$.

b)

i) For global sensitivity, because the dataset can include everything in real numbers, the gap between the middle of the ordered dataset and their adjacent data can be infinite. So, the global sensitivity is infinite.

ii) For the minimum local sensitivity, if all data in the dataset has unique quantity, choosing any arbitrary data and changing it to any real number will not affect the median. The new data cannot be placed between the gap of the middle of the ordered dataset and their adjacent data because this gap is zero. Therefore, the minimum local sensitivity is zero.

iii) For the restricted sensitivity, the gap between the middle of the ordered dataset and their adjacent data can be at most $b-a$.

c)

i)For global sensitivity, a graph on $\mathcal{G}$ should be found. If we change the connectivity of edges incident to one node, the number of isolated nodes should be maximum. A star graph is a good choice because if we remove the edges incident to the center node, all n nodes will be isolated. So, $GS_q = n$.

ii) For the minimum local sensitivity, $\min\limits_{G\in\mathcal{G}} \max\limits_{G\sim G'} \|q(G)-q(G')\|$, a graph should be found such that for all its neighbors, the difference between the number of isolated nodes of this graph and all its neighbors should be minimum. A complete graph should be a good choice because, if we remove even all edges incident to a node, only one node will be isolated. So, $\min\limits_{G\in\mathcal{G}} LS_q = 1$.

iii) The restricted sensitivity, $\max\limits_{G,G'\in\mathcal{H},G\sim G'} |q(G)-q(G')|$, is similar to the maximum local sensitivity on $\mathcal{H}$; the difference is that the neighboring graph should be in $\mathcal{H}$. Therefore, assume an n-node empty graph for G with n-node isolated, and G' consists of two subgraphs: the first is a star graph with $d+1$ nodes, and the second is an empty graph with $n-(d+1)$ nodes. This configuration results in $n-(d+1)$ isolated nodes. Hence, $RS_q^{\mathcal{H}} = d+1$.