HW1-Report

Saman Soleimani 400206284

Question 1:
The attacker's knowledge can be modeled from a Bayesian perspective. The information the attacker possesses about the sensitive bits before encountering the differentially private sequence of bits represents the prior distribution. The likelihood is determined by the distribution specified by the differential privacy (DP) algorithm. Consequently, the attacker selects the sequence of bits with the highest probability in the posterior distribution. Since the bits are chosen independently, we can model the attacker's knowledge bit by bit.

$$\begin{aligned}\mathbb{E}[\#\text{corrects}(\tilde{x}, x)] = \mathbb{E}[1_{\tilde{x}_i = x_i}] = \sum_{i=1}^{n} \mathbb{E}[1_{\tilde{x}_i = x_i}] &\leq n \max_i \Pr(\tilde{x}_i = x_i) \\ &\leq n \max_i \max_{b \in \{0,1\}} \Pr(x_i = b | A(x) = a)\end{aligned}$$

$$\Rightarrow \quad \mathbb{E}[\#\text{corrects}(\tilde{x}, x)] =\leq n \max_i \max_{b \in \{0,1\}} \Pr(x_i = b | A(x) = a) \tag{1}$$

$$\Pr(x_i = b | A(x) = a) = \frac{\Pr(A(x) = a | x_i = b) \Pr(x_i = b)}{\Pr(A(x) = a | x_i = \bar{b}) \Pr(x_i = \bar{b}) + \Pr(A(x) = a | x_i = b) \Pr(x_i = b)}$$

Because $x|x_i = b$ and $x|x_i = \bar{b}$ are at least at a hamming distance of 1 and $\Pr(x_i = b) = \Pr(x_i = \bar{b}) = \frac{1}{2}$(because x is selected uniformly):

$$\Pr(x_i = b | A(x) = a) \leq \frac{\Pr(A(x) = a | x_i = b)}{e^{-\epsilon} \Pr(A(x) = a | x_i = b) + \Pr(A(x) = a | x_i = b)} = \frac{e^{\epsilon}}{1 + e^{\epsilon}} \tag{2}$$

1 and 2 results:

$$\mathbb{E}[\#\text{errors}(\tilde{x}, x)] = 1 - \mathbb{E}[\#\text{corrects}(\tilde{x}, x)] \geq n(1 - \frac{e^{\epsilon}}{1 + e^{\epsilon}}) = \frac{n}{1 + e^{\epsilon}}$$

Question 2-a:
I assumed a general condition where post-processing is random. Alternatively, you can consider deterministic post-processing as a specific case of random preprocessing, wherein the entire probability mass function is concentrated at a single point.
We can consider our random model, which takes its input from a (ε,δ)-DP algorithm and then produces outputs based on the conditional probability introduced by the random model.
Let $B : \chi \to \mathbb{R}$ be the (ε, δ)-DP mechanism, and let $F : \mathbb{R} \to \mathbb{R}$ be the random post process.
If X is our data set and X' is its neighboring data set, I want to show that $H(X) = [B(X), F(B(X))] \in \mathbb{R}^2$ is (ε, δ)-DP.

Assume arbitrary event A in the Borel σ-algebra on $\mathbb{R}^2$. A can be defined by $A = A_1 \times A_2 = \{[x, y] \mid x \in A_1, y \in A_2\}$

$$\begin{aligned}
\Pr[H(X) \in A] &= \Pr[B(X) \in A_1] \Pr[F(a) \in A_2 \mid B(X) = a] \\
&\leq (\Pr[B(X^{'}) \in A_1] e^{\varepsilon} + \delta) \Pr[F(a) \in A_2 \mid B(X^{'}) = a] \\
&\leq e^{\varepsilon} \Pr[B(X^{'}) \in A_1] \Pr[F(a) \in A_2 \mid B(X^{'}) = a] + \delta \\
&= e^{\varepsilon} \Pr[B(X^{'}) \in A_1, F(B(X^{'})) \in A_2] + \delta \\
&= e^{\varepsilon} \Pr[H(X^{'}) \in A] + \delta
\end{aligned}$$

So, $H(X)$ is (ε, δ)-DP.

Question 2-b:

Let $F : \chi \to \mathbb{R}$ be an $(\varepsilon_1, \delta_1)$-DP mechanism, and let $G : \mathbb{R} \times \chi \to \mathbb{R}$ with $G(x, \cdot)$ being $(\varepsilon_2, \delta_2)$-DP mechanism for any $x \in \mathbb{R}$.

If X is our data set and $X^{'}$ is its neighboring data set, I want to show that $H(X) = [F(X), G(F(X), X)] \in \mathbb{R}^2$ is $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$-DP.

Assume arbitrary event A in the Borel σ-algebra on $\mathbb{R}^2$. A can be defined by $A = A_1 \times A_2 = \{[x, y] \mid x \in A_1, y \in A_2\}$

$$\begin{aligned}
\Pr[H(X) \in A] &= \Pr[F(X) \in A_1] \Pr[G(a_1, X) \in A_2 \mid F(X) = a_1] \\
&\leq \Pr[F(X) \in A_1] \left(\min\{1, \Pr[G(a_1, X^{'}) \in A_2 \mid F(X) = a_1] e^{\varepsilon_2}\} + \delta_2 \right) \\
&\leq \Pr[F(X) \in A_1] \min\{1, \Pr[G(a_1, X^{'}) \in A_2 \mid F(X) = a_1] e^{\varepsilon_2}\} + \delta_2 \\
&\leq (\Pr[F(X^{'}) \in A_1] e^{\varepsilon_1} + \delta_1) \min\{1, \Pr[G(a_1, X^{'}) \in A_2 \mid F(X^{'}) = a_1] e^{\varepsilon_2}\} + \delta_2 \\
&\leq \Pr[F(X^{'}) \in A_1] \Pr[G(a_1, X^{'}) \in A_2 \mid F(X^{'}) = a_1] e^{\varepsilon_1 + \varepsilon_2} + \delta_1 + \delta_2 \\
&= e^{\varepsilon_1 + \varepsilon_2} \Pr[F(X^{'}) \in A_1, G(F(X^{'}), X^{'}) \in A_2] + \delta_1 + \delta_2 \\
&= e^{\varepsilon_1 + \varepsilon_2} \Pr[H(X^{'}) \in A] + \delta_1 + \delta_2
\end{aligned}$$

So, $H(X)$ is $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$-DP.

Question 3:

The loss function is quadratic, and it is evident that it is also 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 $\pi_{\mathcal{C}}(.) : \|\pi_{\mathcal{C}}(\tilde{u}^t) - y\|^2 \leq \|\tilde{u}^t - y\|^2; \forall y \in \mathcal{C}$.

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

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

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

$$\mathbb{E}[L(\theta^{\text{priv}}, X)] = \mathbb{E}[L(\frac{1}{T}\sum_{t=0}^{T-1}\theta^t, X)] \leq \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[L(\theta^t, X)] \tag{4}$$

References to equations 3 and 4 result in:

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

$$\Rightarrow \quad \mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) \leq \frac{\eta}{2}\max_t \mathbb{E}[\|\tilde{g}^t\|^2] + \frac{1}{2\eta T}\mathbb{E}[\|\theta^0 - \theta^T\|^2] \tag{5}$$

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 $\mathcal{N}(0, \sigma^2)$. Therefore, $\mathbb{E}[|\tilde{g}^t|^2] = \mathbb{E}[\|g^t\|^2 + \|Z\|^2 + 2\langle g^t, Z\rangle] = \|g^t\|^2 + \mathbb{E}[\|Z\|^2] \leq G^2 + d\sigma^2$, where G is the Lipschitzness of the loss function. So, these results and Equation 5 yield:

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

ℓ_2 sensitivity of $g^t = \nabla_\theta L(\theta^t, X)$ in each step is $\frac{2G}{n}$. It can be shown that the overall sensitivity of the entire function can be $\sqrt{T}\frac{2G}{n}$. This sensitivity, when multiplied by $\frac{\sqrt{2\ln\frac{1}{\delta}}}{\varepsilon}$ gives a lower bound to σ of the noise:

$$\sigma \geq \frac{2G\sqrt{2T\ln\frac{1}{\delta}}}{n\varepsilon} \tag{7}$$

6 and 7 gives new bound:

$$\begin{aligned}\mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) &\leq \frac{\eta}{2}(G^2 + d\frac{8G^2T\ln\frac{1}{\delta}}{n^2\varepsilon^2}) + \frac{4R^2}{2\eta T} \\ &= \frac{\eta G^2}{2}(1 + d\frac{8T\ln\frac{1}{\delta}}{n^2\varepsilon^2}) + \frac{2R^2}{\eta T}\end{aligned} \tag{8}$$

If $\eta = \frac{\sqrt{T}}{RG\sqrt{1+d\frac{8T\ln\frac{1}{\delta}}{n^2\varepsilon^2}}}$:

$$\mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) \leq \frac{RG\sqrt{1+d\frac{8T\ln\frac{1}{\delta}}{n^2\varepsilon^2}}}{\sqrt{T}} \tag{9}$$

The value of G should be determined based on the information provided by the problem, utilizing the Cauchy-Schwarz inequality.

$$\begin{aligned}
\|\nabla_\theta l(f_\theta(x), y)\| &= \|\nabla_\theta (y - f_\theta(x))^2\| \\
&= \|\nabla_\theta (y - \theta^\top x)^2\| \\
&= \|2x(\theta^\top x - y)\| \\
&= |\theta^\top x - y|\|2x\| \\
&\leq 2\sqrt{d}|\theta^\top x - y| \\
&\leq 2\sqrt{d}(|\theta^\top x| + |y|) \\
&\leq 2\sqrt{d}(\|\theta\|\|x\| + 1) \\
&\leq 2\sqrt{d}(R\sqrt{d} + 1)
\end{aligned}$$

So, I set $G = 2\sqrt{d}(R\sqrt{d}+1)$ and substitute this into 9:

$$\mathbb{E}[L(\theta^{\text{priv}}, X)] - L(\theta^*, X) \leq \frac{2R\sqrt{d}(R\sqrt{d}+1)\sqrt{1+d\frac{8T\ln\frac{1}{\delta}}{n^2\varepsilon^2}}}{\sqrt{T}} = \alpha \tag{10}$$

Therefore, if $n \geq \sqrt{d\frac{8T\ln\frac{1}{\delta}}{\varepsilon^2(\frac{T\alpha^2}{[(2R\sqrt{d})(R\sqrt{d}+1)]^2}-1)}}$ The expected error will be lower than α, and the algorithm will be (ε, δ)-DP.

Question 4:
The dataset $X = \{x_1, x_2, \ldots, x_n\} \in \mathcal{X}$, and the outputs of the exponential mechanism belong to $\mathcal{X} = \{1, 2, \ldots, N\}$. We should design $q(y; X)$ such that it has a high value for $y \in [\min_{i=1}^{n} x_i, \max_{i=1}^{n} x_i]$ because the event of the output of the exponential mechanism belonging to this set has a high probability. Based on $n \geq \frac{C_1}{\varepsilon}\ln\left(\frac{|N|}{\beta}\right) + C_2$, this probability increases as n grows. Therefore, the score function should have the parameter n in it.
Assume $A = \{y | y > \max_{i=1}^{n} x_i \quad \text{or} \quad y < \min_{i=1}^{n} x_i\}$, $B = \{y | q(y; X) = q_{max}\}$,Y is the output random variable of exponential mechanism and Δ is the sensitivity of the

score function:

$$\begin{aligned}
\Pr(A) = \Pr(Y \in A) &= \frac{\sum_{y\in A} \exp\left(\frac{\varepsilon q(y;X)}{2\Delta}\right)}{\sum_{y'\in\mathcal{X}} \exp\left(\frac{\varepsilon q(y';X)}{2\Delta}\right)} \\
&\leq \frac{\Pr(Y \in A)}{\Pr(Y \in B)} \\
&= \frac{\sum_{y\in A} \exp\left(\frac{\varepsilon q(y;X)}{2\Delta}\right)}{\sum_{y'\in B} \exp\left(\frac{\varepsilon q(y';X)}{2\Delta}\right)} \\
&= \frac{\sum_{y\in A} \exp\left(\frac{\varepsilon q(y;X)}{2\Delta}\right)}{|B| \exp\left(\frac{\varepsilon q_{\max}}{2\Delta}\right)} \\
&\leq \frac{(N-1)\exp\left(\frac{\varepsilon q_{\max A}}{2\Delta}\right)}{|B| \exp\left(\frac{\varepsilon q_{\max}}{2\Delta}\right)} \quad \text{where } q_{\max A} = \max_{y\in A} q(y;X) \\
&\leq N \exp\left(\frac{\varepsilon(q_{\max A} - q_{\max})}{2\Delta}\right) = \beta
\end{aligned}$$

So, if we fix the parameter β, then $\frac{\varepsilon(q_{\max} - q_{\max A})}{2\Delta} \geq \ln\left(\frac{N}{\beta}\right)$. Because it is evident that $y_{\max}$ is within the lower and upper bounds of the dataset, we can choose a selection for the median problem to design the score function. Based on $\frac{\varepsilon(n-C_2)}{C_1} \geq \ln\left(\frac{N}{\beta}\right)$, the upper bound should not have N, so I didn't use selection for the mean. I set the score function for our median problem as the number of data points lower than our value of interest minus the number of data points greater than the value of interest. We can describe this as:

$$q(y;X) = -\sum_{i=1}^{n} \text{sign}(y - x_i), \quad \text{where sign}(z) = \begin{cases} 1 & \text{if } z > 0, \\ 0 & \text{if } z = 0, \\ -1 & \text{if } z < 0. \end{cases}$$

So, $q_{\max} = -\min_{y\in\mathcal{X}} \sum_{i=1}^{n} \text{sign}(y - x_i), q_{\max A} = -n$. the sensitivity of the score function is 1. It can be some cases that q_{max} is not 0 so $q_{max} \in [-1, 1]$.

$$\begin{aligned}
\frac{\varepsilon(q_{\max} - q_{\max A})}{2\Delta} &= \frac{\varepsilon(q_{\max} + n)}{2} \\
&\geq \frac{\varepsilon(n-1)}{2}
\end{aligned}$$

So, if $\frac{\varepsilon(n-1)}{2} \geq \ln\left(\frac{N}{\beta}\right)$ is true then $\frac{\varepsilon(q_{\max} - q_{\max A})}{2\Delta} \geq \ln\left(\frac{N}{\beta}\right)$. So $C_1 = 2$,$C_2 = 1$.

Now, we should prove the differential privacy of the exponential mechanism. For the proposed score function $q(y;X) = -\sum_{i=1}^{n} \text{sign}(y - x_i)$, the sensitivity is $\Delta = 1$. If we name $H(X) = Y$ our exponential mechanism, for every neighboring dataset of X, X', that differs on the index i of the dataset and an arbitrary event E in σ-algebra on $\mathcal{X}$:

$$
\begin{aligned}
\frac{\Pr(H(X) \in E)}{\Pr(H(X') \in E)} &= \frac{\sum_{y \in E} \exp\left(\frac{\varepsilon q(y;X)}{2\Delta}\right) \sum_{y' \in \mathcal{X}} \exp\left(\frac{\varepsilon q(y';X')}{2\Delta}\right)}{\sum_{y \in E} \exp\left(\frac{\varepsilon q(y;X')}{2\Delta}\right) \sum_{y' \in \mathcal{X}} \exp\left(\frac{\varepsilon q(y';X)}{2\Delta}\right)} \\
&\leq \max_{y \in E} \max_{X,X'} \exp\left(\frac{\varepsilon(q(y;X) - q(y;,X'))}{2\Delta}\right) \max_{y' \in \mathcal{X}} \max_{X,X'} \exp\left(\frac{\varepsilon(q(y';X') - q(y';,X))}{2\Delta}\right) \\
&= \exp\left(\varepsilon \frac{\max_{y \in E} \max_{X,X'} (q(y;X) - q(y;,X'))}{2\Delta}\right) \exp\left(\varepsilon \frac{\max_{y' \in \mathcal{X}} \max_{X,X'} (q(y';X') - q(y';,X))}{2\Delta}\right) \\
&\leq \exp\left(\varepsilon \frac{\max_{y \in E} \max_{X,X'} |q(y;X) - q(y;,X')|}{2\Delta}\right) \exp\left(\varepsilon \frac{\max_{y' \in \mathcal{X}} \max_{X,X'} |q(y';X') - q(y';,X)|}{2\Delta}\right) \\
&= \exp(\varepsilon/2) exp(\varepsilon/2) \\
&= \exp(\varepsilon)
\end{aligned}
$$

So, this algorithm is ϵ-DP.