Advanced Topics in Machine Learning (2024)
Assignment 3

October 3, 2024

Question 1

Algorithm Design

Denote that the voting preference of voter i is $x_i \in \{0,1\}^n$, $i \in \{1,2,\ldots,m\}$, that $x_{i,j} \in \mathbb{R}$ is the voting result of voter i for candidate j where $j \in \{1,2,\ldots,n\}$ and $\tilde{c}_i$ is the i-th element of C_ϵ. Define $f_k : \mathbb{R}^{m\times n} \to \mathbb{R}$, $f_k(X) = \sum_{i=1}^m x_{i,k}$, where $X = (x_1, x_2, \ldots, x_m)$, $k \in \{1,2,\ldots,n\}$. The algorithm is designed as follow:

Algorithm 1 Differential Privacy Voting Mechanism

Require: X, Privacy parameter: ϵ

1: **for** each $i = 1,2,\ldots,n$ **do**

2:     Sample noise $\Gamma_i \sim \text{Laplace}\left(\frac{n}{\epsilon}\right)$

3:     Compute the votes with noise $\tilde{c}_i = f_i(X) + \Gamma_i$

4: **end for**

5: **return** $C_\epsilon = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n)$

Proof of the privacy

For the function $f_k, k \in \{1,2,\ldots,n\}$, its sensitivity is evaluated as follow:

$$\Delta_{f_k} = \max_{X,X'} |\sum_{i=1}^m x_{i,k} - \sum_{i=1}^m x'_{i,k}|_1 = 1 \tag{1}$$

where $X' = (x'_1, x'_2, \ldots, x'_m)$.

Denote the whole algorith as $A : \mathbb{R}^{m\times n} \to \mathbb{R}^n$.

$$\begin{aligned}
\frac{Pr[A(X) = C_\epsilon]}{Pr[A(X') = C_\epsilon]} &= \frac{\prod_{i=1}^n \frac{\epsilon}{2n} \exp\left(-\frac{\epsilon|\tilde{c}_i - f_i(X)|}{n}\right)}{\prod_{i=1}^n \frac{\epsilon}{2n} \exp\left(-\frac{\epsilon|\tilde{c}_i - f_i(X')|}{n}\right)} \\
&= \prod_{i=1}^n \exp(-\frac{\epsilon}{n}(|\tilde{c}_i - f_i(X)| - |\tilde{c}_i - f_i(X')|)) \\
&\leq \exp(\frac{\epsilon}{n} \sum_{i=1}^n |f_i(X') - f_i(X)|) \quad \text{(Triangle inequality)} \\
&\leq \exp(\frac{\epsilon}{n} \sum_{i=1}^n 1) \quad \text{(Proved above)} \\
&= \exp(\epsilon)
\end{aligned} \tag{2}$$

Compute $\mathbb{E}[||C_\epsilon - C||_1]$

Given that $c_i = f_i(X), i \in \{1, 2, \ldots, n\}$, where $C = (c_1, c_2, \ldots, c_n)$, $\tilde{c}_i - c_i = \Gamma_i \sim Laplace(\frac{n}{\epsilon})$. Hence, $\mathbb{E}[\tilde{c}_i - c_i] = 0$. Then

$$\begin{aligned}
\mathbb{E}[||C_\epsilon - C||_1] &= \mathbb{E}[\sum_{i=1}^{n} \tilde{c}_i - c_i] \\
&= \sum_{i=1}^{n} \mathbb{E}[||\tilde{c}_i - c_i||_1] \\
&= nE[Lap(\frac{n}{\epsilon})] \\
&= n \cdot \frac{\epsilon}{2n} \int_{-\infty}^{\infty} \exp(-\frac{\epsilon|x|}{n})|x|dx \\
&= n \cdot \frac{\epsilon}{n} \int_{0}^{\infty} \exp(-\frac{\epsilon x}{n})xdx
\end{aligned} \tag{3}$$

Let $\sigma = -\frac{\epsilon x}{n}$, $dx = -\frac{n}{\epsilon}d\sigma$,

$$\begin{aligned}
\mathbb{E}[||C_\epsilon - C||_1] &= n \cdot \frac{n}{\epsilon} \int_{0}^{-\infty} e^\sigma \sigma d\sigma \\
&= \frac{n^2}{\epsilon} \left(\sigma e^\sigma - e^x\right)\Big|_0^{-\infty} \\
&= \frac{n^2}{\epsilon}
\end{aligned} \tag{4}$$

Question 2

Prove Privacy guarantee

Fix the elements of that vector for all but the argmax location for both neighbouring datasets and denote the event above is A. Denote the algorithm as F and neighboring datasets as S and S', according to the Laplace mechanism where $||c_i - c_i'||_1 \leq 1, i \in \{1, 2, \ldots, n\}$. For any outputs $u \in \mathbb{R}$:

$$\frac{Pr[F(S) = i^*|A]}{Pr[F(S') = i^*|A]} = \frac{Pr[\tilde{c}_{i^*} = u]}{Pr[\tilde{c}'_{i^*} = u]} \leq e^{\frac{\epsilon}{2}} \tag{5}$$

Then according to the Law of total probability,

$$\frac{Pr[F(S) = i^*]}{Pr[F(S') = i^*]} = \int_A \frac{Pr[F(S) = i^*|A]}{Pr[F(S') = i^*|A]} \tag{6}$$

Given that Laplace noise is sampled independently and randomly, for any subset K,

$$\begin{aligned}
\frac{Pr[F(S) \in K]}{Pr[F(S') \in K]} &\leq e^{\frac{\epsilon}{2}} \cdot e^{\frac{\epsilon}{2}} \\
&= e^\epsilon
\end{aligned} \tag{7}$$

So Alg 1 is a $\epsilon - DP$ algorith.

Prove the utility guarantee

2.2.1 Proof of a

According to the definition of PDF, for any $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(|Y| \geq \lambda t) &= \int_{\lambda t}^{\infty} \frac{1}{2\lambda} \exp(-\frac{|x|}{\lambda}) dx + \int_{-\infty}^{-\lambda t} \frac{1}{2\lambda} \exp(-\frac{|x|}{\lambda}) dx \\
&= 2 \int_{\lambda t}^{\infty} \frac{1}{2\lambda} \exp(-\frac{|x|}{\lambda}) dx \\
&= - \int_{\lambda t}^{\infty} \exp(-\frac{x}{\lambda}) d(-\frac{x}{\lambda}) \\
&= \int_{-t}^{-\infty} e^k dk \quad (k = -\frac{x}{\lambda}) \\
&= -(e^k \Big|_{-t}^{-\infty}) \\
&= e^t \\
&\leq e^t
\end{aligned} \tag{8}$$

2.2.2 Proof of b

$$\begin{aligned}
\mathbb{P}(|Y_{max}| > \lambda t') = 1 - \prod_{i=1}^{k} \mathbb{P}(|Y_i| \leq \lambda t') \\
\mathbb{P}(|Y_{max}| > \lambda(\log(k) + t) \leq 1 - (1 - \frac{e^{-t}}{k})^k \qquad (\text{let } t' = \log(k) + t)
\end{aligned} \tag{9}$$

Let $f(t) = 1 - (1 - e^{-t})^k - e^{-t}$, then $f'(t) = e^{-t}[1 - (1 - \frac{e^t}{k})^{k-1}] > 0$. So $f(t)$ is a Monotonic function. Given $\lim_{t \to \infty} f(t) = 1 - 0 - (1 - 0)^k = 0$, $f(t) \leq 0$. Hence, $1 - (1 - \frac{e^{-t}}{k})^k \leq e^{-t}$. Then $\mathbb{P}(|Y_{max}|) \leq e^{-t}$.

According to (b), we know:

$$\mathbb{P}[|Z_{max}| > \frac{2}{\epsilon}(\log n + t))] \leq \exp(-t) \tag{10}$$

Suppose $|Z_{max}| \leq \frac{2}{\epsilon}(\log n + t))$, then for all Z_i, $-\frac{2}{\epsilon}(\log n + t)) \leq Z_i \leq \frac{2}{\epsilon}(\log n + t))$. Then $c_i - \frac{2}{\epsilon}(\log n + t)) \leq \tilde{c}_i \leq c_i \frac{2}{\epsilon}(\log n + t))$.

Given that $\tilde{c}_{i^*} \geq \tilde{c}_{j^*}$,

$$c_{i^*} + \frac{2}{\epsilon}(\log n + t)) \geq c_{j^*} - \frac{2}{\epsilon}(\log n + t)) c_{i^*} \geq c_{j^*} - \frac{4}{\epsilon}(\log n + t))$$

So the even E_1 '$c_{i^*} > c_{j^*} - \frac{4}{\epsilon}(\log n + t))$' $\subseteq$ the event '$|Z_{max}| > \frac{2}{\epsilon}(\log n + t))$'. Hence $\mathbb{P}(\mathbb{E_1} \leq \exp(-t)$

Question 3

Compare

Algorithm in Q2 has the process of 'argmax' and outputs $i \in R$ while algorithm in Q1 outputs a list. Algorithm in Q1 should ensure that all outputs keep the privacy while algorithm in Q2 only ensure the final i.

Implementation of Exponetial Mechanism

Denotes the algorithm is G

$$\mathbb{P}[G(S) = i] = \frac{e^{\frac{\epsilon}{2\Delta\sigma} c_i}}{\sum_{k=0}^{n} e^{\frac{\epsilon}{2\Delta\sigma} c_k}} \tag{11}$$

And for the utility,

$$P\left[\sigma(G(S, C), S) \leq OPT_\sigma(S) - \frac{2\Delta\sigma}{\epsilon}\left(\log\left(\frac{n}{c_{i^*}}\right) + t\right)\right] \leq \exp(-t) \tag{12}$$

where $\sigma(iS) = c_i$, $OPT_\sigma(S) = i^*$, $\Delta\sigma = 1$

Preference

I will prefer algorithm in Q2, it provides a better utility guarantee.