Mathematical Foundations of Machine Learning (CS 4783/5783)

Lecture 21: Differential Privacy

1 Differential Privacy

Differential Privacy is a strong notion of privacy for an algorithm that ensures that we cannot detect if one entry of a dataset is replaced. Specifically, let A be a randomized algorithm that takes as input a sample $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ and outputs $A(S)$ in some arbitrary outcome space.

Definition 1. *We say that A is (ϵ, δ) differentially private if for any sample S and sample S' that differ on at most one data point, and for any set C over the space of outcomes,*

$$P(A(S) \in C) \leq e^{\epsilon} P(A(S') \in C) + \delta$$

Note that since S and S' differ on at most one data point, the above definition tells us that both

$$P(A(S) \in C) \leq e^{\epsilon} P(A(S') \in C) + \delta$$

and that

$$P(A(S') \in C) \leq e^{\epsilon} P(A(S) \in C) + \delta$$

Specifically, as ϵ and δ are taken to be very small this says $P(A(S) \in C)$ and $P(A(S') \in C)$ are very close and so we cant distinguish if we have run our method on S or S'.

2 The Laplace Mechanism

Say we want a differentially private version of a real valued function f on a given sample S. One way to obtain such a version is to first evaluate f on a given sample S then add noise to it to guarantee differential privacy. Specifically, say we want a differentially private version of function f. In this case, let

$$M = \max_{S,S' \text{ s.t. } S',S \text{ vary on one point}} f(S) - f(S')$$

Now we could set

$$A(S) = f(S) + \frac{M}{\epsilon} X$$

where X is drawn from the Laplace distribution $\mathrm{Laplace}(0, 1)$. That is, distribution with density function

$$p(X) = \frac{1}{2} e^{-|X|}$$

Lemma 1. *Let*

$$A(S) = f(S) + \frac{M}{\epsilon} X$$

where $X \sim \mathrm{Laplace}(0, 1)$ and $M = \max_{S,S' \text{ s.t. } S',S \text{ vary on one point}} f(S) - f(S')$. The algorithm A is $(\epsilon, 0)$ differentially private.

Proof. Since $A(S) = f(S) + \frac{M}{\epsilon}\ X$, we have that $A(S) \sim \text{Laplace}(f(S), \frac{M}{\epsilon})$. Hence, we have that the probability density function of $A(S)$ is given by

$$p_{A(S)}(x) = \frac{\epsilon}{2M} e^{-\frac{\epsilon|x-f(S)|}{M}}$$

Similarly, the density function for $A(S')$ for any S' that differs from S on at most one point if given by

$$p_{A(S')}(x) = \frac{\epsilon}{2M} e^{-\frac{\epsilon|x-f(S')|}{M}}$$

Hence,

$$\frac{p_{A(S)}(x)}{p_{A(S')}(x)} = \frac{e^{-\frac{\epsilon|x-f(S)|}{2M}}}{e^{-\frac{\epsilon|x-f(S')|}{2M}}} = e^{\frac{\epsilon}{2M}(|x-f(S')|-|x-f(S)|)} \leq e^{\frac{\epsilon}{M}|f(S)-f(S')|} \leq e^{\epsilon}$$

Next note that for any set C, using the above,

$$P(A(S) \in C) = \int_C p_{A(S)}(x)dx \leq e^{\epsilon} \int_C p_{A(S')}(x)dx = e^{\epsilon} P(A(S') \in C)$$

Thus we have proved that the algorithm is $(\epsilon, 0)$ differentially private. $\square$

An example application is when $S = \{x_1, \ldots, x_n\}$ where each $x_t \in [-1, 1]$ and $f(S) = \frac{1}{n}\sum_{t=1}^{n} x_t$. In this case note that if $S' = \{x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_n\}$, then,

$$f(S) - f(S') = \frac{1}{n}(x_i - x'_i) \leq \frac{2}{n}$$

Hence $M \leq \frac{2}{n}$ and so in this case, to make mean $\epsilon, 0$ differentially private, we need to add Laplace noise of Laplace$(0, \frac{2}{\epsilon n})$

2.1 The Multidimensional Laplace Mechanism

Say function F now maps to a K-dimensional vector. The Laplace mechanism then easily extends to this multi-dimensional setting as well. In this case, define

$$B = \max_{S,S' \text{ s.t. } S',S \text{ vary on one point}} \|F(S) - F(S')\|_1$$

Lemma 2. *Let*

$$A(S) = F(S) + \frac{B}{\epsilon}\ (X_1, \ldots, X_K)$$

where $X_1, \ldots, X_K \sim \text{Laplace}(0, 1)$ *are* K *Laplace distributed random variables. and* $B = \max_{S,S' \text{ s.t. } S',S \text{ vary on one p}}$ $F(S')\|_1$. *The algorithm* A *is* $(\epsilon, 0)$ *differentially private.*

Proof. Since $A(S) = F(S) + \frac{B}{\epsilon}\ (X_1, \ldots, X_K)$, we have that for every $i \in [K]$, $A(S)[i] \sim \text{Laplace}(F(S)[i], \frac{B}{\epsilon})$. Hence, we have that the probability density function of $A(S)[i]$ is given by

$$p_{A(S)}(x) = \left(\frac{\epsilon}{2B}\right)^K \prod_{i=1}^{K} e^{-\frac{\epsilon|x[i]-F(S)[i]|}{B}}$$

Similarly, the density function for $A(S')$ for any S' that differs from S on at most one point if given by

$$p_{A(S')}(x) = \left(\frac{\epsilon}{2B}\right)^K \prod_{i=1}^{K} e^{-\frac{\epsilon|x[i]-F(S')[i]|}{B}}$$

Hence,

$$\frac{p_{A(S)}(x)}{p_{A(S')}(x)} = \prod_{i=1}^{K} e^{\frac{\epsilon}{2B}(|x[i]-F(S')[i]|-|x[i]-F(S)[i]|)} \leq \prod_{i=1}^{K} e^{\frac{\epsilon}{B}|F(S)[i]-F(S')[i]|} = e^{\frac{\epsilon}{B}\|F(S)-F(S')\|_1} e^{\epsilon} \leq e^{\epsilon}$$

Next note that for any set C, using the above,

$$P(A(S) \in C) = \int_C p_{A(S)}(x)dx \leq e^{\epsilon} \int_C p_{A(S')}(x)dx = e^{\epsilon} P(A(S') \in C)$$

Thus we have proved that the algorithm is $(\epsilon, 0)$ differentially private. □

An immediate question that one might have is how bad does the Laplace mechanism distort our outcome. Specifically, recall that we want out procedure to output $F(S)$ in a differentially private fashion. So we would hope that our algorithm $A(S)$ returns a vector that is close to $F(S)$. The following lemma provides such a bound.

Lemma 3. *For any F, the differentially private algorithm A obtained using Laplace mechanism satisfies the following bound*

$$P\left(\|F(S) - A(S)\|_\infty \geq \log\left(\frac{K}{\delta}\right)\frac{B}{\epsilon}\right) \leq \delta$$

where $B = \max_{S,S' \text{ s.t. } S',S \text{ vary on one point}} \|F(S) - F(S')\|_1$

Proof. Let $(X_1, \ldots, X_K)$ be K random variables each drawn from Laplace distribution. In this case note that,

$$\begin{aligned} P\left(\|F(S) - A(S)\|_\infty \geq \log\left(\frac{K}{\delta}\right)\frac{B}{\epsilon}\right) &= P\left(\max_{i\in[K]} \|X[i]\| \geq \log\left(\frac{K}{\delta}\right)\frac{B}{\epsilon}\right) \\ &\leq KP\left(\|X[i]\| \geq \log\left(\frac{K}{\delta}\right)\frac{B}{\epsilon}\right) \\ &\leq K\frac{\delta}{K} = \delta \end{aligned}$$

□

3 Some Properties

The first important property of differential privacy is that post processing preserves privacy. Say algorithm A is (ϵ, δ) differentially private and say we apply a function g on outcome of algorithm A and output $g(A(S))$. Such post processing preserves privacy.

Lemma 4. *Let A be an (ϵ, δ) differentially private algorithm. Let g be any function on the space of outcomes of the algorithm A. Then, the algorithm B that computes $B(S) = g(A(S))$ is also (ϵ, δ) differentially private.*

Proof. Consider any set C on the space of outcomes of algorithm B. Define the set

$$D = \{d : g(d) \in C\}$$

that is D is the set of entries such that g applied to an element in D returns an outcome in set C. Note that,

$$P(B(S) \in C) = P(g(A(S)) \in C) = P(A(S) \in D)$$

Now using the differential privacy of A, we have

$$P(B(S) \in C) = P(A(S) \in D) \leq e^{\epsilon} P(A(S') \in D) + \delta$$

But if $A(S') \in D$, then $g(A(S')) \in C$ by definition of set D and so

$$P(B(S) \in C) \leq e^{\epsilon} P(A(S') \in D) + \delta = e^{\epsilon} P(g(A(S')) \in C) + \delta = e^{\epsilon} P(B(S') \in C) + \delta$$

Thus we can conclude that B is ϵ, δ differentially private.

□

Lemma 5. *Let A_1 and A_2 be two ϵ_1 and ϵ_2 differentially private mechanisms. Then, $A(S) = (A_1(S), A_2(S))$ is an $\epsilon_1 + \epsilon_2$ differentially private mechanism.*

Proof. Below we do the proof assuming the outcome of the differentially private mechanism has a density function. (for the discrete setting the proof can also be easily extended and can also be extended more generally)

$$\begin{aligned}
\frac{p_{A(S)}(c_1, c_2)}{p_{A(S')}(c_1, c_2)} &= \frac{p_{A_1(S)}(c_1) \times p_{A_2(S)}(c_2)}{p_{A_1(S')}(c_1) \times p_{A_2}(c_2)} \\
&= \frac{p_{A_1(S)}(c_1)}{p_{A_1(S')}(c_1)} \times \frac{p_{A_2(S)}(c_2)}{p_{A_2}(c_2)} \\
&\leq e^{\epsilon_1} \times e^{\epsilon_2} = e^{\epsilon_1 + \epsilon_2}
\end{aligned}$$

□

4 Gaussian Mechanism

Lemma 6. *Let*

$$A(S) = F(S) + \frac{cB}{\epsilon} \, (X_1, \ldots, X_K)$$

where $X_1, \ldots, X_K \sim \mathrm{Normal}(0, 1)$ are K standard normal distributed random variables. and $B = \max_{S,S' \text{ s.t. } S',S \text{ vary on one point}} \|F(S) - F(S')\|_2$. and the constant $c = \sqrt{2\log(1.25/\delta)}$. Then the algorithm A is (ϵ, δ) differentially private.

Proof. For now say we use variance σ, we will later prove $\sigma = cB/\epsilon$ Note that,

$$\begin{aligned}
\frac{p_{A(S)}(x)}{p_{A(S')}(x)} &= \prod_{i=1}^{K} \frac{e^{-\frac{(x[i]-F(S)[i])^2}{2\sigma^2}}}{e^{-\frac{(x[i]-F(S')[i])^2}{2\sigma^2}}} = \prod_{i=1}^{K} e^{\frac{1}{2\sigma^2}\left((x[i]-F(S')[i])^2-(x[i]-F(S)[i])^2\right)} \\
&= e^{\frac{1}{2\sigma^2}\left(\|x-F(S')\|_2^2-\|x-F(S)\|_2^2\right)} \\
&= e^{\frac{1}{2\sigma^2}\left(\|x-F(S)\|_2^2+\|F(S)-F(S')\|_2^2+2(x-F(S))^\top(F(S)-F(S'))-\|x-F(S)\|_2^2\right)} \\
&= e^{\frac{1}{2\sigma^2}\left(\|F(S)-F(S')\|_2^2+2(x-F(S))^\top(F(S)-F(S'))\right)} \\
&\leq e^{\frac{1}{2\sigma^2}\left(B^2+2\|x-F(S)\|_2 B\right)}
\end{aligned}$$

Now note that $\frac{1}{2\sigma^2}\left(B^2 + 2\|x - F(S)\|_2 B\right) \leq \epsilon$ whenever,

$$\|x - F(S)\|_2 \leq \frac{\sigma^2\epsilon}{B} - \frac{B}{2}$$

On the other hand, note that $P_{A(S)}(\|x - F(S)\|_2 > \frac{\sigma^2\epsilon}{B} - \frac{B}{2})$ can be bounded using the fact that x is gaussian distributed with mean $F(S)$. Specifically for the setting of $\sigma = cB/\epsilon$ we can conclude that

$$P_{A(S)}(\|x - F(S)\|_2 > \frac{\sigma^2\epsilon}{B} - \frac{B}{2}) \leq \delta$$

Thus we conclude that the mechanism is (ϵ, δ) differentially private. $\square$