Lecture 16: Basics of differential privacy

November 21, 2019

1 Differential Privacy

In this lecture we begin our final unit of the class on private machine learning. However, before we can get into the development of private algorithms for machine learning problems, we'll first need to cover a bit of background on the basics of differential privacy. Intuitively, privacy should ensure that the output of the algorithm should be similar on similar data points. Differential privacy has emerged as arguably the most commonplace way of rigorously arguing about privacy.

We first require some notation. Let $\mathcal{X}$ be a set, and let $X, X' \in \mathcal{X}^n$. We let $\|X - X'\|_0$ denote the number of indices where $X_i \neq X'_i$. Then, differential privacy can be defined as follows:

Definition 1.1 ((ε, δ)-differential privacy, [1])**.** A randomized algorithm $\mathcal{A} : \mathcal{X}^n \to Y$ is (ε, δ)-*differentially private* if for all $X, X' \in \mathcal{X}^n$ so that $\|X - X'\|_0 \leq 1$, and any $S \subseteq Y$, we have that

$$\Pr\left[\mathcal{A}(X) \in S\right] \leq e^{\varepsilon} \cdot \Pr\left[\mathcal{A}(X') \in S\right] + \delta \;,$$

where the probability is over the internal randomness of $\mathcal{A}$. When $\delta = 0$ we will refer to this as ε-differential privacy, or *pure differential privacy.* When $\delta > 0$, we refer to this as *approximate differential privacy.*

Notice by symmetry this also implies that

$$\Pr\left[\mathcal{A}(X') \in S\right] \leq e^{\varepsilon} \cdot \Pr\left[\mathcal{A}(X) \in S\right] + \delta \;.$$

When X, X' satisfy $\|X - X'\|_0 \leq 1$ we say that X, X' are *adjacent.* There are also other notions of adjacency that are sometimes considered in the differential privacy literature, especially in the private ML literature, but for now we will stick to the classic definition.

In the definition above, ε is a privacy loss parameter. It is not hard to see that if we wish for perfect privacy (i.e. $\varepsilon = \delta = 0$), then nothing non-trivial is possible, and so we quantify how much privacy we sacrifice with this parameter. It asks that any event of probability at least $O(\delta)$ is almost equally likely when the algorithm takes X or X'. In contrast, robust statistics asks for much less: only that the output of the algorithm is close to the truth with good probability. However, this comes at a cost: typically, in robust statistics, we think of corrupting a constant fraction of the data points, whereas in differential privacy, the guarantees tend to become meaningless after one has corrupted more than a constant number of data points.

1.1 Properties of differential privacy

One reason why differential privacy is such a popular notion is that it turns out to have very nice properties. Below we will list a number of these properties, and omit the proofs because they are standard and straightforward. The first property is that it is preserved under post-processing:

Fact 1.1 (Differential privacy is preserved under post-processing)**.** *Let $\mathcal{A} : \mathcal{X}^n \to Y$ be a (ε, δ)-differentially private algorithm, and let $f : Y \to Z$ be an arbitrary random function. Then $f \circ \mathcal{A}$ is (ε, δ)-differentially private.*

Another useful fact is *composition*:

Fact 1.2 (Basic composition of differential privacy)**.** *Let $\mathcal{A}_i : \mathcal{X}^n \to Y$ be $(\varepsilon_i, \delta_i)$-differentially private for $i \in [k]$. Then the algorithm $\mathcal{A} : \mathcal{X}^n \to Y^k$ defined by $\mathcal{A}(x) = (\mathcal{A}_1(x), \ldots, \mathcal{A}_k(x))$ is $(\sum \varepsilon_i, \sum \delta_i)$-differentially private.*

As we'll see later on, this is not the full picture for composition: for approximate DP we can actually get privacy loss which scale *sublinearly* with the number of things we are composing together. Finally, a similar-looking (but different) statement is that we also get privacy for larger distances:

Fact 1.3 (Group privacy)**.** *Let $\mathcal{A} : \mathcal{X}^n \to Y$ be ε-differentially private. Then, for any $X, X' \in \mathcal{X}^n$ so that $\|X - X'\|_0 \leq k$, and any event $S \subseteq Y$, we have*

$$\Pr\left[\mathcal{A}(X) \in S\right] \leq e^{k\varepsilon} \cdot \Pr\left[\mathcal{A}(X') \in S\right] \; .$$

2 Private mechanisms

Here we will go over some of the basic ways to convert a non-private algorithm into a private algorithm. We will focus primarily on numeric problems, namely, ones where the range is $Y = \mathbb{R}^d$. Let $f : \mathcal{X}^n \to \mathbb{R}^d$. Our goal is to convert f into a random function which is differentially private, while not distorting it too much. It is not hard to see that this is not easy to do for all functions; for instance, consider the function f which is very large at a specific $x \in \mathcal{X}^n$ and zero elsewhere; then since f is not very smooth, converting it into a private function will have to distort f a lot. To quantify this, for any $p \in [1, \infty)$, we define the parameter

$$\Delta_p(f) = \sup_{\|X - X'\|_0 = 1} \|f(X) - f(X')\|_p$$

to be the ℓ_p-sensitivity of f. Intuitively, if Δ_p is small, then f is smoother, and it should be easier to make f private. This is exactly what the Laplace and Gaussian mechanisms achieve.

2.1 The Laplace mechanism

We start with the following definition:

Definition 2.1 (The Laplace distribution)**.** Let $b > 0$. Then the (univariate) Laplace distribution (centered at 0) with scale b is the distribution over $\mathbb{R}$ with probability density function given by

$$p(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right) \; .$$

With this definition, we can now define the Laplace mechanism:

Definition 2.2. Given any function $f : \mathcal{X}^n \to \mathbb{R}^d$, the Laplace mechanism with parameter $\varepsilon > 0$ is the random function $\mathcal{M}_{f,\varepsilon} : \mathcal{X}^n \to \mathbb{R}^d$ defined by

$$\mathcal{M}_{f,\varepsilon}(X) = f(X) + Y \; ,$$

where $Y \in \mathbb{R}^d$ is a random vector whose coordinates are drawn independently from $\mathrm{Lap}(\Delta_1(f)/\varepsilon)$.

Often times we will just refer to this as the Laplace mechanism and denote it $\mathcal{M}_f$, when the parameter $\varepsilon > 0$ is understood. The Laplace mechanism satisfies the following privacy guarantee:

Theorem 2.1. *Let $f : \mathcal{X}^n \to \mathbb{R}^d$. Then $\mathcal{M}_{f,\varepsilon}$ is $\varepsilon > 0$ is $(\varepsilon, 0)$-differentially private.*

Proof. Let X, X' be adjacent. Let p_X denote the pdf of $\mathcal{M}_{f,\varepsilon}(X)$, and similarly let $p_{X'}$ denote the pdf of $\mathcal{M}_{f,\varepsilon}(X')$. Then at any point $y \in \mathbb{R}^d$, we have

$$\begin{aligned}\frac{p_X(z)}{p_{X'}(z)} &= \prod_{i=1}^{d} \left(\frac{\exp\left(-\frac{\varepsilon}{\Delta_1(f)} \left|f(X)_i - z_i\right|\right)}{\exp\left(-\frac{\varepsilon}{\Delta_1(f)} \left|f(X')_i - z_i\right|\right)} \right) \\ &= \exp\left(\frac{\varepsilon}{\Delta_1(f)} \left(\left\|f(X') - z\right\|_1 - \left\|f(X) - z\right\|_1 \right) \right) \\ &\leq \exp\left(\frac{\varepsilon}{\Delta_1(f)} \left\|f(X') - f(X)\right\|_1 \right) = e^{\varepsilon} \ ,\end{aligned}$$

as claimed. □

On the other hand, we also know that (as long as $\Delta_1(f)$) is not too large, that this is still pretty close to the true value of f, with good probability. This follows as a corollary of the following standard fact:

Fact 2.2. *Let $Y \sim \mathrm{Lap}(b)$. Then for all $t \geq 0$, we have $\Pr\left[|Y| \geq tb\right] = \exp(-t)$.*

By combining this with a union bound, we immediately obtain:

Corollary 2.3. *Let $f : \mathcal{X}^n \to \mathbb{R}^d$. Then, for all $\eta \in (0, 1]$, we have:*

$$\Pr\left[\left\|f(x) - \mathcal{M}_{f.\varepsilon}(x)\right\|_\infty \geq \frac{\Delta_1(f)}{\varepsilon} \log \frac{d}{\eta} \right] \leq \eta \ .$$

Example: Histogram queries As an example which will be somewhat useful later on, consider the setting where we have subdivided $\mathcal{X}$ into disjoint subsets $A_1, \ldots, A_k$, and $f : \mathcal{X}^n \to \mathbb{R}^k$ simply reports the number of samples we see in each subset, i.e. $f(X)_j = |\{i : X_i \in A_j\}|$. It's not hard to see that $\Delta_1(f) \leq 2$, and hence the Laplace mechanism gives us an $(\varepsilon, 0)$-differentially private algorithm which reports each coordinate of f to accuracy $O\left(\frac{\log d}{\varepsilon}\right)$ with high probability. In particular, if $n \gtrsim \frac{\log d}{\varepsilon \alpha}$, then this gives us a $(1 + \alpha)$-multiplicative approximation to the true counts.

2.2 The Gaussian mechanism

A similar mechanism to the Laplace mechanism is the *Gaussian mechanism*:

Definition 2.3 (The Gaussian mechanism)**.** Given any function $f : \mathcal{X}^n \to \mathbb{R}^d$, the Gaussian mechanism with parameter $\sigma > 0$ is the random function $\mathcal{M}^{\text{gauss}}_{f,\sigma} : \mathcal{X}^n \to \mathbb{R}^d$ defined by

$$\mathcal{M}^{\text{gauss}}_{f,\sigma}(X) = f(X) + Y \ ,$$

where $Y \sim \mathcal{N}(0, \sigma^2 \cdot I)$.

The main advantage of this is that we only add coordinate-wise Gaussian noise that depends on the ℓ_2 sensitivity of f; often times this can be much smaller than the ℓ_1 sensitivity. However, the main downside of this is that it only provides approximate differential privacy. These two properties are captured by the following theorem:

Theorem 2.4. *Let $\varepsilon, \delta > 0$, and let $f : \mathcal{X}^n \to \mathbb{R}^d$. Let $\sigma = c\Delta_2(f)/\varepsilon$ for some c satisfying $c^2 \gtrsim \log(1/\delta)$. Then $\mathcal{M}^{\text{gauss}}_{f,\sigma}$ is (ε, δ)-differentially private.*

The proof of this theorem is slightly annoying (although very similar to the proof above), and we omit it for simplicity. We note that in fact the Gaussian mechanism satisfies a stronger notion of privacy, namely, *zero-concentrated differential privacy*:

Definition 2.4 (zero Concentrated Differential Privacy, [2])**.** Let $\rho > 0$. A randomized algorithm $\mathcal{A} : \mathcal{X}^n \to Y$ is ρ-zCDP if for all adjacent $X, X' \in \mathcal{X}^n$, we have that

$$D_\alpha \left(\mathcal{A}(X) || \mathcal{A}(X') \right) \leq \rho\alpha \qquad \forall \alpha \in (1, \infty) ,$$

where for any $\alpha > 1$ and two probability distributions P, Q,

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \mathop{\mathbb{E}}_{X \sim P} \left[\frac{dQ}{dP}(X)^{\alpha - 1} \right]$$

is the α-Renyi divergence between P, Q.

The main properties that we'll need from this definition of privacy are that (1) it is a notion of privacy formally between pure and approximate differential privacy, and (2) the Gaussian mechanism satisfies zCDP. These two facts are captured in the following two theorems, whose proofs we will omit for the sake of time:

Theorem 2.5. *For any $\varepsilon > 0$, and any randomized algorithm $\mathcal{M}$, we have:*

1. *If $\mathcal{M}$ is $(\varepsilon, 0)$-differentially private, then $\mathcal{M}$ is $(\varepsilon^2/2)$-zCDP.*
2. *If $\mathcal{M}$ is $(\varepsilon^2/2)$-zCDP, then $\mathcal{M}$ is $(\varepsilon^2/2 + \varepsilon\sqrt{2 \log 1/\delta}, \delta)$-differentially private for all $\delta > 0$.*

Theorem 2.6. *Let $f : \mathcal{X}^n \to \mathbb{R}^d$ be arbitrary, let $\rho > 0$, and let $\sigma = \Delta_2(f)/\sqrt{2\rho}$. Then $\mathcal{M}^{\text{gauss}}_{f,\sigma}$ satisfies ρ-zCDP.*

Additionally, note that as a simple consequence of Gaussian concentration, we have the following analog of Corollary 2.3:

Fact 2.7. *Let $f : \mathcal{X}^n \to \mathbb{R}^d$. Then, for all $\eta \in (0, 1]$, we have:*

$$\Pr \left[\| f(x) - \mathcal{M}_{f.\varepsilon,\delta}(x) \|_\infty \gtrsim \frac{\Delta_2(f)}{\varepsilon} \sqrt{\log \frac{d}{\eta}} \right] \leq \eta \ .$$

3 Advanced composition theorems

As mentioned before, approximate differential privacy (as well as concentrated differential privacy) satisfy stronger composition theorems than presented in Fact 1.2. Moreover, in general, all differentially private algorithms compose well under adaptive composition. Formally, we say that $\mathcal{M}$ is an adaptive composition of $\mathcal{M}_1, \ldots, \mathcal{M}_T$ if $\mathcal{M}_t$ may depend on the outcomes of $\mathcal{M}_1, \ldots, \mathcal{M}_{t-1}$. Then, we have:

Theorem 3.1 (Advanced composition theorems, [1, 3, 2])**.** *Let $\mathcal{M}$ be an adaptive composition of $\mathcal{M}_1, \ldots, \mathcal{M}_T$. Then:*

- *If $\mathcal{M}_i$ is $(\varepsilon_i, \delta_i)$-differentially private for all $i \in [k]$, then $\mathcal{M}$ is $(\sum \varepsilon_i, \sum \delta_i)$-differentially private.*
- *If $\mathcal{M}_i$ satisfies ρ_i-zCDP for all $i \in [k]$, then $\mathcal{M}$ is $(\sum \rho_i)$-zCDP.*
- *If $\mathcal{M}_i$ is (ε, δ_i)-differentially private for all $i \in [k]$, then for all $\delta > 0$, M is $(\varepsilon_0\sqrt{6T \log 1/\delta}, \delta + \sum \delta_i)$-differentially private.*

4 Differentially private learning

We now wish to formulate machine learning problems with differential privacy in mind. The slightly incongruous thing about differential privacy for ML is that differential privacy is a worst-case guarantee, whereas typically in ML we deal with average case data, i.e. data that came from some nice distribution $\mathcal{D}$. However, it still makes sense to insist upon *worst-case* privacy guarantees. This is because we don't know for sure if the data comes from the distribution we think it comes from. Thus, it is natural to want that the algorithm is always private, no matter what. Of course, the accuracy of the algorithm will hinge on the fact that the data satisfies the distributional assumption, but that is inevitable.

References

[1] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[2] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer, 2016.

[3] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 51–60. IEEE, 2010.