DS 102 Discussion 11

Monday, 27 April 2020

In this discussion, we'll take a deeper look at differential privacy. For two datasets S and S' which differ in only one entry (*e.g.*, differing in one individual), an ϵ-**differentially private algorithm** $\mathcal{A}$ satisfies:

$$\mathbb{P}(\mathcal{A}(S) = a) \leq e^{\epsilon}\mathbb{P}(\mathcal{A}(S') = a),$$

for all possible output values a of the algorithm $\mathcal{A}$. In words, the probability of seeing any given output of a differentially private algorithm doesn't change much by replacing any one entry in the dataset.

Datasets that differ in only one entry are called **neighboring** datasets.

1. **Laplace mechanism.** One of the most popular mechanisms for differential privacy is the **Laplace mechanism**. Suppose we want to report a statistic $f(\cdot)$, which takes as input a dataset. For example, S could be a dataset with the salaries of all Berkeley residents, and $f(S)$ could be the average salary in S. Denote by S and S' generic neighboring datasets. Define the **sensitivity** of f as:

$$\Delta_f = \max_{\text{neighboring } S,S'} |f(S) - f(S')|.$$

The Laplace mechanism reports $\mathcal{A}_{\text{Lap}}(S) = f(S) + \xi_\epsilon$, where ξ_ϵ is distributed according to the zero-mean Laplace distribution with parameter $\frac{\Delta_f}{\epsilon}$, denoted $\text{Lap}(0, \frac{\Delta_f}{\epsilon})$. The Laplace distribution $\text{Lap}(\mu, b)$ has the following density:

$$p(x) = \frac{1}{2b}e^{-\frac{|x-\mu|}{b}}$$

and is essentially a two-sided exponential distribution.

(a) Prove that the Laplace mechanism is ϵ-differentially private. More precisely, show that for every dataset S' that neighbors our dataset S, we have

$$\frac{\mathbb{P}(\mathcal{A}_{\text{Lap}}(S) = a)}{\mathbb{P}(\mathcal{A}_{\text{Lap}}(S') = a)} \leq e^{\epsilon}.$$

(b) In Part (a), we convinced ourselves that the Laplace mechanism indeed ensures privacy. However, privacy alone is easy to ensure: one can always report random noise. For the reported values to also be useful, we have to consider a trade-off between privacy and **accuracy**. Accuracy means that $\mathcal{A}_{\text{Lap}}(S)$ is actually close to $f(S)$ with high probability.

Using the fact that $X \sim \text{Lap}(0, b)$ satisfies

$$\mathbb{P}(|X| \geq t) \leq 2e^{-\frac{t}{b}},$$

prove that the Laplace mechanism also enjoys a good accuracy guarantee:

$$\mathbb{P}(|\mathcal{A}_{\text{Lap}}(S) - f(S)| \geq t) \leq 2e^{-\frac{t\epsilon}{\Delta f}}.$$

(c) What can you conclude about the relationship between sensitivity Δ_f and accuracy, for a fixed level of privacy ϵ? Does this make intuitive sense?

(d) Suppose you want to report the average salary, i.e. $f(S) = \frac{1}{n}\sum_{i=1}^{n} s_i$, where s_i is the salary of the i-th individual in the dataset. Moreover, suppose that all salaries are in the range $[0, M]$. What is an appropriate parameter of the Laplace mechanism, if we want to report the average salary in an ϵ-differentially private way? What is the accuracy guarantee of this mechanism?

2. **Post-processing of differential privacy.** An important property of differential privacy is that it is preserved under post-processing: if $\mathcal{A}(S)$ is an ϵ-differentially private statistic, then $g(\mathcal{A}(S))$ is still differentially private, for any function g. Prove this fact.