

Machine Learning Exercise Sheet 13

Advanced Topics

In-class Exercises

Differential Privacy

Problem 1: The goal is to prove that the Laplace mechanism is ϵ -Differentially Private.

From the definition we have that a randomized mechanism $\mathcal{M}_f : \mathcal{X} \rightarrow \mathcal{Y}$ is ϵ -differentially private if **for all** neighboring inputs $X \simeq X'$ and **for all** sets of outputs $Y \subset \mathcal{Y}$ we have:

$$\exp^{-\epsilon} \leq \frac{\mathbb{P}[\mathcal{M}_f(X) \in Y]}{\mathbb{P}[\mathcal{M}_f(X') \in Y]} \leq \exp^{\epsilon}$$

The Laplace mechanism is defined as follows $\mathcal{M}_f(X) = f(X) + Z$ where $Z \sim \text{Lap}(0, \frac{\Delta_1}{\epsilon})^d$ and the global l_1 sensitivity of a function $f : \mathcal{X} \rightarrow \mathbb{R}^d$ is $\Delta_1 = \sup_{X \simeq X'} \|f(X) - f(X')\|_1$.

We start by plugging in the definition $\text{Lap}(Z; 0, b) = \frac{1}{2b} \exp^{-\frac{|Z|}{b}}$ and using the fact that the noise is i.i.d. per dimension. We have:

$$\begin{aligned} \frac{\mathbb{P}[\mathcal{M}_f(X) \in Y]}{\mathbb{P}[\mathcal{M}_f(X') \in Y]} &= \prod_{i=1}^d \frac{\exp^{-\frac{\epsilon}{\Delta_1} |f(X)_i - Z_i|}}{\exp^{-\frac{\epsilon}{\Delta_1} |f(X')_i - Z_i|}} \\ &= \prod_{i=1}^d \exp^{\frac{\epsilon}{\Delta_1} [|f(X')_i - Z_i| - |f(X)_i - Z_i|]} \\ &\leq \prod_{i=1}^d \exp^{\frac{\epsilon}{\Delta_1} [|f(X')_i - f(X)_i|]} \\ &= \exp^{\frac{\epsilon}{\Delta_1} \sum_{i=1}^d [|f(X')_i - f(X)_i|]} \\ &= \exp^{\frac{\epsilon}{\Delta_1} \|f(X') - f(X)\|_1} \\ &\leq \exp^{\frac{\epsilon}{\Delta_1} \Delta_1} = \exp^{\epsilon} \end{aligned}$$

where the first inequality comes from the (reverse) triangle inequality and the second inequality is from the definition of global sensitivity.

Since the neighboring relation \simeq is symmetric we can repeat the above derivation to obtain

$$\frac{\mathbb{P}[\mathcal{M}_f(X') \in Y]}{\mathbb{P}[\mathcal{M}_f(X) \in Y]} \leq \exp^{\epsilon}$$

$$\Leftrightarrow \frac{\mathbb{P}[\mathcal{M}_f(X) \in Y]}{\mathbb{P}[\mathcal{M}_f(X') \in Y]} \leq e^{\epsilon}$$

where now $f(X')$ is in the numerator and $f(X)$ is in the denominator, which then gives us

$$\exp^{-\epsilon} \leq \frac{\mathbb{P}[\mathcal{M}_f(X) \in Y]}{\mathbb{P}[\mathcal{M}_f(X') \in Y]}$$

.