Machine Learning Exercise Sheet 13

Privacy

In-class Exercises

Differential Privacy

Problem 1: Prove that the Laplace mechanism is ϵ-Differentially Private.
Note: 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 $f : \mathcal{X} \to \mathbb{R}^d$ is $\Delta_1 = \sup_{X \simeq X'} ||f(X) - f(X')||_1$.

Homework

Differential privacy

Problem 2: Assume that you have trained an univariate linear regression model $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x}$ with $\boldsymbol{w} \in \mathbb{R}^D$ for D-dimensional binary data from input space $\mathcal{X} = \{0,1\}^D$. You want to make its prediction ϵ-differentially private with respect to changes in a single input dimension, i.e. $\boldsymbol{x} \simeq \boldsymbol{x}' \iff ||\boldsymbol{x}-\boldsymbol{x}'||_0 = 1$ for all points from input space $\mathcal{X}$.

a) Compute the global Δ_1 sensitivity of f w.r.t. "$\simeq$".

b) To ensure differential privacy, you want to use the Laplace mechanism $\mathcal{M}_{f,\text{Lap}}(\boldsymbol{x}) = f(\boldsymbol{x}) + z$ with $z \sim \text{Lap}(\mu, b)$. Based on your result from a), which values do you have to use for μ and b to ensure ϵ-differential privacy w.r.t. to neighboring relation "$\simeq$?

c) Instead of randomizing the output of our model, we can also guarantee differential privacy by randomizing its inputs. Prove that the randomized mechanism $\mathcal{M}' = f(\boldsymbol{x}+\boldsymbol{z})$ with $\boldsymbol{z} \sim \text{Lap}\left(0, \frac{1}{\epsilon}\right)^D$ is ϵ-differentially private w.r.t to neighboring relation "$\simeq$".

Problem 3: You are given a dataset with n instances $\{x_1, \ldots, x_n\}$, with $x_i \in \mathcal{X}$. The instances are randomly split into disjoint groups $G_1, G_2, \ldots G_m$, each of size $\frac{n}{m}$ (assume that m divides n, i.e. $\frac{n}{m}$ is an integer).
First you apply an *arbitrary* function $f : \mathcal{X}^{\frac{n}{m}} \to [a, b]$ (where a and b are given constants) to each of the groups, i.e. you compute $g_1 = f(G_1), g_2 = f(G_2), \ldots, g_m = f(G_m)$. Then you compute the final output by aggregating the per-group outputs by computing either their mean or their median.

a) Derive the global Δ_1 sensitivity of the function $f' := \text{mean}(f(G_1), \ldots, f(G_m))$.

b) Derive the global Δ_1 sensitivity of the function $f'' := \text{median}(f(G_1), \ldots, f(G_m))$.

c) Can you make the function f' and/or f'' differentially private for any function $f : \mathcal{X}^{\frac{n}{m}} \to [a, b]$? If yes, specify the noise distribution from which we have to sample to obtain an ϵ-DP private mechanism. If no, why not?

Upload a single PDF file with your homework solution to Moodle by 02.02.2022, 11:59pm CET. We recommend to typeset your solution (using LaTeX or Word), but handwritten solutions are also accepted. If your handwritten solution is illegible, it won't be graded and you waive your right to dispute that.

Problem 4: One of the fundamental properties of differential privacy is "group privacy" (see p.22): If mechanism $\mathcal{M}$ is ϵ-DP w.r.t $X \simeq X'$, then $\mathcal{M}$ is $(t\epsilon)$-DP w.r.t. changes of t instances/individuals.

Prove that group privacy holds when using the l_0 norm as the neighboring relation for vector data. That is: If mechanism $\mathcal{M}$ is ϵ-DP w.r.t. "$\simeq$", where $\boldsymbol{x} \simeq \boldsymbol{x}' \iff ||\boldsymbol{x} - \boldsymbol{x}'||_0 = 1$, then $\mathcal{M}$ is $(t\epsilon)$-DP w.r.t. "$\simeq_t$", where $\boldsymbol{x} \simeq_t \boldsymbol{x}' \iff ||\boldsymbol{x} - \boldsymbol{x}'||_0 = t$.