CSC2412 F18: Assignment 1

Due: November 29, by midnight

Guidelines: (read fully!!)

- Your assignment solution must be submitted as a *typed* PDF document. Scanned handwritten solutions, solutions in any other format, or unreadable solutions will **not** be accepted or marked. You are encouraged to learn the LaTeX typesetting system and use it to type your solution. See the course website for LaTeX resources.
- To submit this assignment, use the MarkUs system, at URL `https://markus.teach.cs.toronto.edu/csc2412-2019-09`
- This is an *individual assignment*. You may consult any of the reading material posted on the course website. However, your solutions should show your individual work.
- You may use any result discussed in class or covered in the assigned reading by just referring to it. You do not need to reproduce proofs that we have covered in the lectures.
- Unless stated otherwise, you should justify all your answers using rigorous arguments. Your solution will be marked based both on its completeness and correctness, and also on the clarity and precision of your explanation.

Question 1. (6 marks)

Consider the least squares regression problem, in which the dataset is $X = ((x_1, y_1), \dots, (x_n, y_n))$, and each $x_i \in \mathcal{X} = [0,1]^d$, and $y_i \in [0,1]$, we consider functions of the type $f_\theta(x) = \theta^\top x$ for $\theta \in B_2(0,R)^d$, and the loss is given by

$$l(f_\theta(x), y) = (y - f_\theta(x))^2.$$

The goal is to minimize the empirical loss $L(\theta, X) = \frac{1}{n}\sum_{i=1}^n l(f_\theta(x_i), y_i)$ over all $\theta \in B_2(0,R)$.

Suppose we use the private stochastic gradient descent algorithm from the lecture notes to approximately minimize this loss. I.e. suppose we use the algorithm from the notes to give an (ε, δ)-differentially private algorithm which outputs $\theta^{\text{priv}} = \frac{1}{T}\sum_{t=0}^{T-1} \theta^t$ such that, for all large enough datasets X,

$$\mathbb{E}L(\theta^{\text{priv}}, X) - \min_{\theta \in B_2(0,R)^d} L(\theta, X) \le \alpha.$$

What is the value of n_0 for which this inequality holds for all datasets X of size $n \ge n_0$? Justify your answer.

You can use the Cauchy-Schwarz inequality: for any two vectors $u, v \in \mathbb{R}^d$, $|u^\top v| \le \|u\|_2 \|v\|_2$.

Question 2. (14 marks)

Your goal in this question is to give an $(\varepsilon, 0)$-differentially private algorithm for logistic regression, by implementing the private gradient descent algorithm using Laplace noise rather than Gaussian noise, and applying the standard (rather than advanced) composition theorem.

In particular, consider linear classifiers $f_\theta : [0,1]^d \to [-1,1]$ of the form $f_\theta(x) = \sum_{i=1}^d \theta_i x_i + \theta_0$ for $x \in [0,1]^d$, and let the loss of predicting label $f_\theta(x)$ instead of the true label y be given by

$$l(f_\theta(x), y) = \log(1 + e^{-y f_\theta(x)}).$$

For a dataset X consisting of labelled samples $(x_1, y_1), \dots, (x_n, y_n) \in [0,1]^d \times \{-1, +1\}$, this gives the empirical loss function

$$L(\theta, X) = \frac{1}{n}\sum_{i=1}^n l(f_\theta(x_i), y_i)$$

Part a. (3 marks)

Let $\nabla L(\theta, X)$ be the gradient of L with respect to θ. Given an upper bound on the ℓ_1 sensitivity of $\nabla L(\theta, X)$, i.e., on

$$\max_{X \sim X'} \|\nabla L(\theta, X) - \nabla L(\theta, X)\|_1,$$

with the maximum taken over neighbouring datasets X, X' differing in a single data point.

Part b. (4 marks)

Suppose we run the Noisy Gradient Descent algorithm from Section 4. of the notes on the course website (`http://www.cs.toronto.edu/~anikolov/CSC2412F19/notes/ERM.pdf`), but we sample each noise vector w^t by sampling each of its $d+1$ coordinates from the Laplace distribution. What is the scale of the Laplace noise you need to add so that the the final output $\frac{1}{T}\sum_{t=0}^{T-1} \theta^t$ is ε-differentially private? Assume we are using the logistic loss $L(\theta, X)$ above, and give your answer as a function of T, d, and ϵ.

Part c. (7 marks)

Show that, for an appropriate value of T, the ε-differentially private algorithm from the previous subquestion outputs $\theta^{\text{priv}} = \frac{1}{T}\sum_{t=0}^{T-1} \theta^t$ such that

$$\mathbb{E}L(\theta^{\text{priv}}, X) - \min_{\theta \in B_2^{d+1}(R)} L(\theta, X) \le \alpha.$$

as long as

$$n \ge \frac{KR^2(d+1)^2}{\varepsilon \alpha^2}.$$

for a sufficiently large constant K. You can use the analysis from the lecture notes. Be specific about how you choose the value of T.