Homework 3 of CS291A Introduction to Differential Privacy (Fall 2021)

University of California, Santa Barbara

Assigned on Nov 21, 2021(Sunday)

Due at 11:59 pm on Dec 3, 2021 (Wednesday)

Notes:

- Be sure to read "Policy on Academic Integrity" on the course syllabus.
- There are ***[100 points]*** in this homework, and a bonus ***[5 points]***.
- You need to submit your homework via Gradescope.
- Contract the instructor if you spot typos. Any updates or correction will be posted on the course Announcements page and piazza, so check there occasionally.

---

0. Acknowledgment *[0 points]*

For each question in this HW, please list all your collaborators and reference materials (beyond those specified on the website) that were used for this homework.

1. **List of Collaborators** List the names of all people you have collaborated with and for which question(s).
2. **List of Acknowledgements.** If you find an assignment's answer or use a another source for help, acknowledge for which question and provide an appropriate citation (there is no penalty, provided you include the acknowledgement). If not, then write "none".

1 General facts about Differential Privacy [33 pts]

Please answer True or False and provide a short (one sentence explanation). (3 pts each)

(a) Differential privacy prevents all harms that could incur to an individual when a dataset including this individual is analyzed (differentially privately).

(b) Differentially Private machine learning algorithm cannot possibly predict my attribute accurately because otherwise it violates the DP guarantee.

(c) Differential Privacy prevents attackers from identifying Bob, even if the rest of the dataset (except Bob) is known.

(d) If one can implement homomorphic encryption computationally efficiently to train a machine learning model, then it provides stronger privacy guarantee than differential privacy.

(e) DP algorithms must add noise.

(f) Deterministic algorithms cannot be differentially private.

(g) (ϵ, δ)-Approximate differential privacy provides DP guarantee with probability $1 - \delta$.

(h) k-anonymity could fail completely upon composition, whereas differential privacy degrades more gracefully under composition.

(i) It is often worthwhile to exploit the sparsity in the problem of differentially private histogram release. Adding noise only to those elements of the histogram that are non-zero helps to improve the utility of the DP release.

(j) Gaussian mechanism is the only mechanism satisfying a linear upper bound of the Renyi differential privacy (i.e., Concentrated Differential Privacy).

(k) Objective perturbation mechanism dominates output perturbation mechanism in differentially private (convex) empirical risk minimization.

2 Differentially private Linear Regression [21 pts]

In this question, you will see the connection of various mechanisms for differentially private linear regression and how they are related to each otehr. The goal is to solve

$$\min_\theta \sum_{i=1}^n (x_i^T \theta - y_i)^2 + \lambda \|\theta\|^2 = \|X\theta - y\|^2 + \lambda\|\theta\|^2$$

This this question, we do not concern choosing the parameter of the randomization to achieve a particular (ϵ, δ)-DP. You should just write down the form of the output using the native parameters of the algorithms.

(a) (7 pts) Write down the output perturbation mechanism that adds noise to the solution in terms of X, y, λ, you may assume the noise you add is $\mathcal{N}(0, \sigma^2 I_d)$

(b) (7 pts) Express the objective perturbation algorithm as a data-dependent noise-adding procedure. You may assume the noise vector $b \sim \mathcal{N}(0, \sigma^2 I_d)$

(c) (7 pts) Express the posterior sampling algorithm as a data-dependent noise adding procedure. You may just express things as scale parameter γ such that we are outputting

$$\hat{\theta} \sim P(\theta|X, y) \propto \exp(-\gamma(\sum_{i=1}^n (x_i^T \theta - y_i)^2 + \lambda\|\theta\|^2)).$$

(Hint: the above samples from a multivariate Gaussian mechanism. What is the mean and what is the covariance matrix?)

3 Private selection and data-dependent DP [25 pts +5 bonus]

Consider the problem of private voting. Each voter can choose only one out of m candidates. The voting scores can be represented as a histogram. Assume that on our particular dataset, the most popular candidate attracts k more voters than the second most popular candidate.

We measure the utility by the probability of outputting the correct argmax.

(a) Write down how Dist2Instability mechanism works and replicate its privacy analysis. (The cleanest description of it is in the recap of Lecture 16)

(b) What is the utility of Dist2Instability mechanism with DP parameter (ϵ, δ)? (Hint: Discuss what happens when you vary $0 \leq k \leq n$).

(c) Consider Laplace mechanism (with DP parameter ϵ) for releasing the histogram, then output argmax as a post-processing. Lower bound the utility of this approach as a function of k? (Notice that this mechanism is the same as report-noisy-max).

(d) Consider exponential mechanism with DP parameter ϵ. Lower bound the utility of this approach as a function of k. (Hint: you could directly apply the theorem in Slide 12 of Lecture 5)

(e) Draw the utility as a function of k (by hand or by matplotlib). When shall we use which algorithm when we need (ϵ, δ)-DP and want to have the largest utility (assuming we have a good guess what k is)?

(f) (Bonus 5 pts) Derive the Gaussian mechanism version of Report-Noisy-Max and its utility.

4 Privacy amplification by sampling [21 pts]

Consider a dataset $x \in \{0, 1\}^n$ (represented as a binary bit vector), and the following algorithm:

1. Randomly select one coordinate of this dataset.
2. Run randomized response on the selected coordinate with parameter ϵ (i.e., output the correct value with probability $e^\epsilon/(1 + e^\epsilon)$)

In this question, we will work out the privacy parameter of this algorithm under the "Replace-One" neighboring relationship.

(a) (7 pts) We will describe a dataset x under the assumption that the ordering of the coordinates does not matter. Let $0 \leq m \leq n$ be the number of 1s . Without loss of generality, assume the last coordinate is what differs between x and its neighbor x'. What is the probability of outputting 1?

(b) (7 pts) Show that x and x' are ϵ' indistinguishable. Parameterize ϵ' by n, m, ϵ.

(c) (7 pts) Show that this algorithm is ϵ''-DP. Work out the parameter ϵ''. Notice that you cannot use the general theorem of privacy amplification by sampling, but you may use it to check if you solution is correct.