Homework 1 of CS291A Introduction to Differential Privacy (Fall 2021)

University of California, Santa Barbara

Assigned on Oct 4, 2021(Monday)

Due at 11:59 pm on Oct 20, 2021 (Thursday)

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 Differential Privacy basics [25 Pts]

(I have covered many of these in the lectures. The point of this question is to make sure you understand the lecture. If you don't follow the lecture, feel free to read the textbook, understand the proof and write them down.)

(a) (5 pts) Prove that differential privacy satisfies closure to post-processing.

(b) (5 pts) Prove that ϵ-DP mechanisms for add/remove one individual satisfies $(k\epsilon)$-DP for adding or removing a maximum of k individuals.

(c) (5 pts) Prove that Randomized Response with parameter $p > 0.5$ satisfies ϵ-DP. Parameterize ϵ as a function of p and come up with a formula to choose p according to a privacy budget ϵ.

(d) (5 pts) Prove that Laplace mechanism satisfies ϵ-DP.

(e) (5 pts) Prove that exponential mechanism satisfies ϵ-DP.

2 Adversary's point of view [20 Pts]

(a) (5 pts) Consider the randomized response mechanism that satisfy ϵ-DP. You output $y \sim \mathcal{M}(x)$ where you data $x \in \{1 : \text{Heart-desease}, 0 : \text{ no-heart-desease}\}$. Consider an adversary who wants to make an inference about your data i.e., this adversary is a binary classifier. Let's say that this classifier outputs $\hat{x} = 1$ if $y = 1$ and $\hat{(x)} = 0$ if $y = 0$. Show that this adversary's rule is the one that maximizes the likelihood.

(b) (5 pts) What is the classification error of the specific adversary above on your data?

(c) (5 pts) Show (using the definition of differential privacy) that no-adversary can have a smaller classification error on **all possible inputs** at the same time, i.e., show that the above strategy of the adversary is optimal.

(Hint: if an adversary predicts more accurately in the world when $x = 0$, what does it mean for the adversary when $x = 1$? Notice that the classifier needs to output the same thing, or sample $\hat{x}$ from the same probability distribution if y is the same.).

(d) (5 pts) Now let's consider a general ϵ-DP mechanism with $y \sim \mathcal{M}(x)$. $\mathcal{Y}$ can be space as long as we can define the probability distribution induced by $\mathcal{M}(x)$ and $\mathcal{M}(x')$. Consider an adversary who have two candidate datasets x, x' that are neighbors. Show that (using the definition of differential privacy) that the classification error of any such adversary must be larger than $\frac{1}{e^\epsilon+1}$. Can you explicitly construct the optimal attack?

(Hint: consider the event of the adversary is correct under $\mathcal{M}(x)$, that does the event imply under $\mathcal{M}(x')$?)

3 Privacy loss random variables [20 Pts + 5 Bonus point]

A central concept in DP analysis is the so-called privacy loss random variable $\boldsymbol{\varepsilon}$. Let randomized algorithm $\mathcal{M} : \mathcal{X}^* \to \mathcal{Y}$, where $\mathcal{X}^*$ is the space of the datasets and $\mathcal{Y}$ is the space of the outputs. The privacy loss random variable is defined for a fixed pair of neighboring datasets $x, x' \in \mathcal{X}^*$ as follows.

$$\boldsymbol{\varepsilon}_{\mathcal{M}}^{x,x'} = \log(\frac{p(\mathbf{y})}{p'(\mathbf{y})}) \text{ where random variable } \mathbf{y} \sim \mathcal{M}(x).$$

In the above, p, p' are the probability density function (or probability mass function if $\mathcal{Y}$ is discrete) of $\mathcal{M}(x)$ and $\mathcal{M}(x')$ respectively. For simplicity, you may ignore measure theoretic-considerations and just consider a finite discrete output space.

(a) (5 pts) Show that for all neighboring pair x, x' $\mathbb{P}(\boldsymbol{\varepsilon}_{\mathcal{M}}^{x,x'} > \epsilon) = 0$ *if and only if* $\mathcal{M}$ satisfies ϵ-DP.

(b) (5 pts) Prove the following useful lemma.

Lemma 1 (Tail bound to (ϵ, δ)-DP conversion)**.** *Let* $\boldsymbol{\varepsilon}_{\mathcal{M}}^{x,x'}$ *be the privacy loss RV defined above. If*

$$\mathbb{P}(\boldsymbol{\varepsilon}_{\mathcal{M}}^{x,x'} \geq \epsilon) \leq \delta$$

for all *pair of neighboring* x, x' *then* $\mathcal{M}$ *satisfies* (ϵ, δ)*-DP.*

(Hint: For any set of the output space $\mathcal{S}$, consider both $\mathcal{S}$ and its complement $\mathcal{S}^c$. Follow the definition of DP.)

(c) (5 pts) (non-adaptive composition) Use the above to show that for two *fixed* mechanisms $\mathcal{M}_1, \mathcal{M}_2$. Let's say $\mathcal{M}_1, \mathcal{M}_2$ satisfies that for any pair of neighboring input x, x',

$$\mathbb{P}(\boldsymbol{\varepsilon}_{\mathcal{M}_1}^{x,x'} \geq \epsilon) \leq \delta$$

and

$$\mathbb{P}(\boldsymbol{\varepsilon}_{\mathcal{M}_2}^{x,x'} \geq \epsilon) \leq \delta$$

show that the composition $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$ satisfies that $(2\epsilon, 2\delta)$-DP.

(d) (5 pts) (Adaptive composition) Let's say that $\mathcal{M}_2$ depends on the realized output of $\mathcal{M}_1$, but for each potential output $y_1 \in \mathcal{Y}_1$, $\mathcal{M}_2$ satisfies $\mathbb{P}(\boldsymbol{\varepsilon}_{\mathcal{M}_2}^{x,x'} \geq \epsilon) \leq \delta$ for any pair of neighboring input x, x', . Show that the composition $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$ satisfies $(2\epsilon, 2\delta)$-DP.

(e) (5 pts) (Bonus Question) Let $\mathcal{M}_1, ..., \mathcal{M}_k$ be a sequence of randomized algorithms, $\mathcal{M}_i$ may depend on the realized output of $\mathcal{M}_1, ..., \mathcal{M}_{i-1}$. Assume that for all i, $\mathcal{M}_i$ satisfies that $\mathbb{P}(\boldsymbol{\varepsilon}_{\mathcal{M}_1}^{x,x'} \geq \epsilon) \leq 0$ show (using Azuma-Hoeffding's inequality) that the composition $(\mathcal{M}_1, ..., \mathcal{M}_k)$ satisfies (ϵ', δ)-DP with

$$\epsilon' = k\epsilon(e^\epsilon - 1) + \sqrt{2k \log(1/\delta)}\epsilon.$$

(Hint: The challenge is in bounding the expected value of the privacy loss RV with $\epsilon(e^\epsilon - 1)$. Notice that the expected value of the privacy loss random variable is the KL-divergence, which is always larger than 0.)

4 Implement differential private data release for COVID'19 data / an introduction to *autodp* [35 Pts]

Set up your favorite Python programming environment. Install "numpy, scipy, pandas". Also install "autodp" by following instructions here `https://github.com/yuxiangw/autodp`.

Read the notebook here (thanks EricThomson for making it available) for `https://github.com/EricThomson/covid-mapping/blob/master/covid_map.ipynb`, modify it so that you are plotting the state-level daily COVD cases (normalized by the popular of each state) for 09/30/2021.

In this question we will implement how to differential-privately release and visualize this dataset. Answer the following questions using Jupyter notebook and save the notebook as a pdf file (combine with the rest of your written answers), then submit through Gradescope. Please also submit your code separately.

(a) (5 pts) Formulate the problem as a numerical query where f takes the dataset x of individuals and output the daily COVID case count by states on 09/30/2021. Write down the

mathematical formula for this query f. You may use your favorite definition of the dataset, but please specify the data-domain and how the output can be computed from x using formal mathematical notation. Explain any choices that are non-standard.

(b) (5 pts) Under the add/remove neighboring relationship, work out the L1 sensitivity of f. You may specify any additional assumption that you are willing to make.

(c) (5 pts) Figure out how to generate a Laplace random variable using `numpy.random`, and implement the Laplace mechanism to privately release $f(x)$ using the formula we learned from the class on how to calibrate the noise to a pre-defined privacy budget $\epsilon =$.

(d) (5 pts) Quantitatively measure the accuracy of the output. Plot the ground truth and the differentially private release of the output in terms of the mean normalized L1-distance

$$\frac{1}{\text{Number of states}} \sum_i \frac{|y_i - f(x)_i|}{|\text{Population of State } i|}.$$

Plot the above metric against $\epsilon = [1e-3, 1e-2, 1e-1, 1, 10, 100]$ with the horizontal and vertical axis on log-log-scale.

(e) (5 pts) Qualitatively visualizing the accuracy of the output. Plot the ground truth and the differentially private release of the output on the map using the bubble plot side-by-side for the case when $\epsilon = 0.5$.

(f) (5 pts) ((ϵ, δ)-DP of Laplace Mechanism) Review the first example in this autodp tutorial `https://github.com/yuxiangw/autodp/blob/master/tutorials/tutorial_new_api.ipynb`, use the b parameter you calculated from Part (c) for $\epsilon = 0.5$ to instantiate a `Mechanism` object using `autodp.mechanism_zoo.LaplaceMechanism`. Plot the ϵ parameter of the Laplace mechanism for $\delta = [0.5, 0.4, 0.3, 0.2, 0.1, 1e-8]$.

(g) (5 pts) Note that this is to release the data for 1 day. How much noise do you need to add if you need to release the data continuously every day for one year (365 days)? Assume that any individual could only get COVID once (because, e.g., then you developed a immune system that covers you for a lifetime[1]), then what is the L1 sensitivity of the query that output the case numbers for a indefinite period of time?

(There is no need to implement anything. A theoretical discussion is enough.)

[1]Note that this is just a hypothetical discussion as the assumption is not necessarily true!