CPSC 471/571: Trustworthy Deep Learning **Fall 2025**

Homework 4

Due TBD

This problem set should be completed individually.

General Instructions

These questions require thought, but do not require long answers. Please be as concise as possible.

Submission instructions: You should submit your answers in a PDF file written using LaTeX.

Submitting answers: Prepare answers to your homework in a single PDF file. Make sure that the answer to each sub-question is on a *separate page.* The number of questions should be at the top of each page.

Honor Code: When submitting the assignment, you agree to adhere to the Yale Honor Code. Please read carefully to understand what it entails!

Notations. Unless explicitly stated otherwise, we will adhere to the following conventions.

- A lowercase letter (e.g., n) denotes a number.
- An uppercase letter (e.g., N, S, T) denotes a set and its lowercase (e.g., n, s, t) denotes the set's size.
- A lowercase bold letter (e.g., $\mathbf{x}$) denotes a vector.
- An uppercase bold letter (e.g., $\mathbf{A}$) denotes a matrix.
- A math calligraphic letter denote (e.g., $\mathcal{X}$) denotes a space.

1 Differential Privacy

Recall the definition of (ϵ, δ)-DP

Definition 1.1 ((ε, δ)-Differential Privacy)**.** *A randomized mechanism $\mathcal{A}$ is (ε, δ)-DP if for all neighboring $D \sim D'$ and measurable events C,*

$$\Pr[\mathcal{A}(D) \in C] \leq e^{\varepsilon} \Pr[\mathcal{A}(D') \in C] + \delta.$$

Note that two datasets D, D' are considered neighboring (denoted $D \sim D'$) if they differ on one individual. In this homework, we will try to show the amplification by subsampling result. Considering the following sampling procedures:

- *Poisson/Bernoulli subsampling:* Each record is included independently with probability $q \in [0, 1]$. Denote the random subset by $\mathcal{S}_q(D)$, and write the subsampled mechanism as $\mathcal{A}(\mathcal{S}_q(D))$.
- *Fixed-size subsampling (without replacement):* For $|D| = n$, sample a size-m subset uniformly: $\mathcal{S}_m(D)$.

Questions. (60pts)

a) Let $D \sim D'$ differ in the record $x^\star$. Show there exists a joint coupling of (S, S') with $S \sim \mathcal{S}_q(D)$ and $S' \sim \mathcal{S}_q(D')$ such that:

   - with probability $1 - q$: $S = S'$ (the differing record is not sampled), and
   - with probability q: S, S' are neighboring (they differ by at most the presence of $x^\star$).

   (Hint: Couple the coin that decides whether $x^\star$ is included, then share all other inclusion coins.)

b) Suppose $\mathcal{A}$ is ε-DP ($\delta = 0$). For any random switch $B \in \{0, 1\}$ with $\Pr[B = 1] = q$ and any neighboring $D \sim D'$, show that

   $$\Pr[\mathcal{A}(D) \in C \mid B = b] \leq e^{\varepsilon} \Pr[\mathcal{A}(D') \in C \mid B = b] \quad \text{for } b \in \{0, 1\}.$$

   Conclude a bound comparing the mixtures $\Pr[\,\cdot\,] = (1 - q) \Pr[\,\cdot \mid B = 0] + q \Pr[\,\cdot \mid B = 1]$.

c) Now let's use the above two results to show a simple amplification for pure DP via Poisson subsampling.

   Theorem 1.2 (Subsampling Amplification for Pure DP)**.** *If $\mathcal{A}$ is ε-DP and we define $\mathcal{M}(D) = \mathcal{A}(\mathcal{S}_q(D))$, then $\mathcal{M}$ is ε'-DP with*

   $$\varepsilon' = \log\big(1 + q\,(e^{\varepsilon} - 1)\big)$$

   (i.e., for all neighboring $D \sim D'$ and events C, $\Pr[\mathcal{M}(D) \in C] \leq e^{\varepsilon'} \Pr[\mathcal{M}(D') \in C]$).

   Using the coupling from part a) and the DP guarantee for $\mathcal{A}$, show

   $$\Pr[\mathcal{M}(D) \in C] \leq \big(1 - q + q e^{\varepsilon}\big) \Pr[\mathcal{M}(D') \in C],$$

   then identify $\varepsilon' = \log(1 - q + q e^{\varepsilon})$. Finally, show that $\varepsilon' \leq (e - 1) q \varepsilon$ when $\varepsilon \in [0, 1]$.

d) Let's extend the result to (ϵ, δ)-DP. Prove the following theorem.

Theorem 1.3 (Subsampling Amplification for (ϵ, δ)-DP). *If $\mathcal{A}$ is (ε, δ)-DP and $\mathcal{M}(D) = \mathcal{A}(\mathcal{S}_q(D))$, then $\mathcal{M}(D)$ is (ϵ', δ')-DP where*

$$\varepsilon' = \log\big(1 + q\,(e^{\varepsilon} - 1)\big), \quad \delta' = q\,\delta.$$

Verify that when $\delta = 0$ you recover part c).

References