Homework 1
CSCI 699: Privacy-Preserving Machine Learning

Instructor: Sai Praneeth Karimireddy
Due: Sep 15, 2025

Instructions: Answer the following questions clearly and concisely. Justify your answers with precise reasoning where necessary. Points for each question are indicated. For the theory part, type your answers in **Latex** (you can use overleaf for this) and submit the pdf on brightspace. For the practical part, please **upload a jupyter notebook** when submitting your solution on Brightspace. Do not just submit a colab link.

Questions

Question 1: K-Anonymity Interpretation (3 points)

Consider an anonymized dataset that has been released under the notion of k-anonymity. Explain if k-anonymity protects against each of the following privacy attacks.

(a) **Membership inference:** Can an attacker determine whether a specific individual is part of the dataset?

(b) **Sensitive attribute disclosure:** Can an attacker deduce whether a specific individual has a particular sensitive attribute (e.g., COVID positive/negative)?

(c) **Identity disclosure:** Can an attacker identify which specific data record corresponds to a particular individual?

Question 2: Differential Privacy for Datasets with Multiple Differences (2 points)

Let $A(D)$ be an algorithm that satisfies ε-differential privacy (DP) when the notion of "similar datasets" refers to datasets that differ in exactly one datapoint. Prove that the same algorithm $A(D)$ satisfies $k\varepsilon$-differential privacy when we redefine neighboring datasets to be those that differ in up to k datapoints.

Question 3: Hypothesis testing and Differential Privacy (2 points)

Let A be an algorithm that satisfies ε-differential privacy. Prove the following lower bound relationship between the type I and type II errors of any hypothesis test based on the output of A:

$$e^{\varepsilon} \cdot \text{Type I Error} + \text{Type II Error} \geq 1.$$

Question 4: Extracting Training Data from trained models (3 points)

See this notebook here and fill in the missing code (3 points):

`https://colab.research.google.com/drive/1oGrJ-Knu8KhW1SlbfUTIq7KtOoGdkrCI?usp=sharing`.

Please attach a jupyter notebook when submitting your solution on Brightspace. Do not just submit a colab link.