CS-5331 (4331) Assignment 2

Due on 03/26/2025, submit to Blackboard

1. **(20 points)** Suppose a data analyst uses logistic regression to construct a binary classifier on some sensitive dataset $\{\boldsymbol{x}_i | i \in [1, n]\}$. A binary label is represented as $y_i \in \{-1, +1\}$. Prove that the empirical loss function on a data sample takes the form of
$$\ell(\boldsymbol{\omega}; \boldsymbol{x}_i, y_i) = \log(1 + \exp(-y_i \boldsymbol{\omega}^T \boldsymbol{x}_i)).$$
[Hint: maximize the log likelihood of data samples.]

2. **(10 points)** Write the objective function of the logistic regression (consider the averaged empirical loss of the data samples and use l_2 norm to control over-fitting).

3. **(30 points)** Consider the KDDCup99 dataset (provided on blackboard) as an example and use output perturbation to generate the classifier under ϵ-differential privacy guarantee. Split the dataset into training and testing sets using a 2:1 ratio. Plot the testing accuracy when $\epsilon \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$.
[Hint: you can refer to the provided .py file for the python implementations of loss and gradient.]

4. **(40 points)** Consider the KDDCup99 (provided on blackboard) dataset as an example and use objective perturbation to generate the classifier under ϵ-differential privacy guarantee. Split the dataset into training and testing sets using a 2:1 ratio. (1) Show the perturbed objective function. (2) Plot the testing accuracy when $\epsilon \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$.

5. **(20 points)**[1] Read paper https://wanglun1996.github.io/publication/sp19.pdf. Discuss (1) why the above objective perturbation mechanism may fail to ensure differential privacy in practice; (2) why differentially-private machine learning algorithms can sometimes outperform the non-private baseline (i.e., the one without privacy protection) in terms of accuracy.

[1] This question is optional for undergraduate students enrolled in CS 4331 and can be attempted for extra credit. However, it is mandatory for graduate students enrolled in CS 5331. The total score for this assignment, including this question, is 120, but for graduate students, it will be normalized to 100.