Private Empirical Risk Minimization with Analytic Gaussian Mechanism for Healthcare System

Jiahao Ding, *Student Member, IEEE*, Sai Mounika Errapotu, *Student Member, IEEE*, Yuanxiong Guo, *Member, IEEE*, Haixia Zhang, *Senior Member, IEEE*, Dongfeng Yuan, *Senior Member, IEEE* and Miao Pan, *Senior Member, IEEE*

***Abstract*—With the wide range application of machine learning in healthcare for helping humans drive crucial decisions, data privacy becomes an inevitable concern due to the utilization of sensitive data such as patients records and registers of a company. Thus, constructing a privacy preserving machine learning model while still maintaining high accuracy becomes a challenging problem. In this paper, we propose two differentially private algorithms, i.e., Output Perturbation with aGM (OPERA) and Gradient Perturbation with aGM (GRPUA) for empirical risk minimization, a useful method to obtain a globally optimal classifier, by leveraging the analytic Gaussian mechanism (aGM) to achieve privacy preservation of sensitive medical data in a healthcare system. We theoretically analyze and prove utility upper bounds of proposed algorithms and compare them with prior algorithms in the literature. The analyses show that in the high privacy regime, our proposed algorithms can achieve a tighter utility bound for both settings: strongly convex and non-strongly convex loss functions. Besides, we evaluate the proposed private algorithms on five benchmark datasets. The simulation results demonstrate that our approaches can achieve higher accuracy and lower objective values compared with existing ones in all three datasets while providing differential privacy guarantees.**

***Index Terms*—Differential Privacy, Analytic Gaussian Mechanism, Empirical Risk Minimization, Machine Learning, Healthcare.**

I. Introduction

In the big data era, data becomes incredibly easy to acquire and aggregate. Machine learning is being increasingly used to extract useful information from data aimed to benefit our lives in various aspects. Due to the high performance and great potential in different domains such as computer vision and speech recognition, machine learning brings lots of new opportunities to healthcare. For instance, Enlitic is using machine learning to spot nearly undetectable health problems from billions of medical images like X-rays, CT scans and MRIs [1]; Machine learning techniques are also used for calculating the mortality probability after a heart surgery [2] and the probability of patients suffering postpartum depression [3] and cardiovascular disease [4]. At the same time, however, machine learning algorithms used for healthcare may give rise to privacy concerns because the training data may contain sensitive information such as medical records, patients identification, and so on. Recently, Fredrikson et al. [5] used hill climbing on output probabilities of a computer vision classifier to recover the personal faces from the training set. Thus, the problem of data privacy in machine learning, especially for the privacy of training datasets, has attracted more attention.

Recently, differential privacy has been proposed as a de-facto privacy model that can offer a strong privacy guarantee when releasing sensitive results of statistical analysis. Differential privacy measures the difference in the output of an algorithm due to the presence of a single element in the original dataset, which ensures the adversary cannot infer any sensitive information. The mechanisms for achieving differential privacy mainly include the Laplace mechanism [6], the classical Gaussian mechanism [7], and the exponential mechanism [8]. Therefore, because of the powerful guarantee of differential privacy, research has been done on studying privacy preserving machine learning together with differential privacy, such as [9]–[11].

Empirical risk minimization (ERM) as a standard technique covers a wide set of learning tasks like classification and regression, etc., where the averaged loss of the model over a dataset is minimized. We can directly obtain private machine model by designing a private ERM algorithm, in other words, solving ERM problem in a differentially private way. A number of approaches have been proposed for designing a differentially private ERM algorithm that can be classified into three categories: output perturbation, objective perturbation and gradient perturbation. Output perturbation and objective perturbation first proposed by Chaudhuri et al. [9]. Output perturbation protects the privacy by using the Laplace mechanism or the classical Gaussian mechanism to perturb the output of the non-differentially private algorithm. Objective perturbation guarantees differential privacy by adding noise to the objective function of ERM and obtaining the precise solution to the noisy objective function. For output perturbation and objective perturbation proposed in [9], it is unnecessary for the practical problem to require precise solutions. Thus, an extended output perturbation approach for ERM problem has been proposed by [10], which obtained approximate solutions and achieved good utility and time complexity by running gradient descent algorithm for a fixed number of iterations. However, due to the limitations of the classical Gaussian mechanism, which is not suitable in the low privacy regime, the scope of applications of the above algorithms is limited. Gradient perturbation approach proposed by [12] achieves (ϵ, δ)-differential privacy by adding Gaussian noise to each iteration and derives good utility for both strongly convex and non-strongly convex loss functions. However, the classical Gaussian mechanism used in [12] is suboptimal in the high privacy regime. As a result,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2020.2997732, IEEE Transactions on Big Data

there is still much room for improvement concerning the utility bound.

To overcome above limitations of the classical Gaussian mechanism: loose in the high privacy regime ($\epsilon \to 0$) and not applicable in the low privacy regime ($\epsilon \to \infty$), the analytic Gaussian mechanism has been proposed by Balle et al. [13]. The improvement of the analytic Gaussian mechanism (aGM) is based on a noise calibration method that instead of relying on tail bounds of Gaussian distribution to calibrate the variance of Gaussian noise, focuses on numerical evaluations of the Gaussian cumulative density function (CDF).

In this paper, we utilize two private learning algorithms to build a privacy preserving classification model in the healthcare system for disease prognosis and diagnosis. The privacy level of the classifier is controlled by the healthcare server and the system will be distributed to each patient. Due to the strong privacy guarantee of the private classifier, the attacker with access to model parameters, training methods and model architectures still cannot infer any sensitive information. By utilizing the analytic Gaussian mechanism, the proposed efficient private ERM algorithms will provide tighter utility upper bounds. The main contributions of this paper are summarized as follows:

- We propose an efficient machine learning scheme for a healthcare system to train differentially private classification models, which helps doctors make diagnostic decisions and increase efficiency while protecting privacy of medical datasets.
- We design two differentially private machine learning algorithms, i.e., Output Perturbation with aGM (OPERA) and Gradient Perturbation with aGM (GRPUA). For both strongly convex and non-strongly convex problems, our theoretical analysis shows that the proposed algorithms have tighter utility bounds than previous works in the high privacy regime.
- Using real-world datasets, we apply our algorithms to two common machine learning objective functions: logistic regression, support vector machine and neural network to verify the effectiveness of OPERA and GRPUA algorithms. The results of performance evaluation show the proposed algorithms significantly outperform existing algorithms in the high privacy regime.

The rest of the paper is organized as follows. Section II presents the system architecture and threat model. We introduce preliminaries about the convex optimization and differential privacy in Section III. This is followed by two proposed private risk minimization algorithms in Section IV. We show performance evaluation and related work in Section V and Section VI. Section VII concludes the whole paper.

II. Problem Statement

In this section, we first describe the healthcare system architecture for classification and introduce the machine learning problem: empirical risk minimization, a learning principle that machine learning model is trained in regard to minimizing the average loss over the training data. We then describe the threat model of this healthcare system.

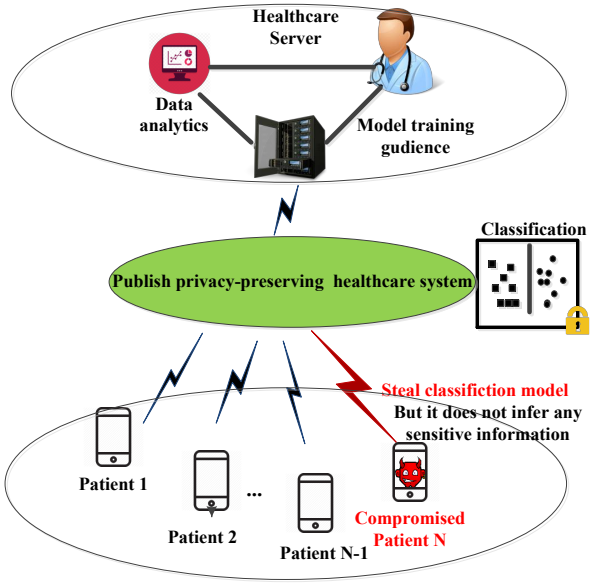


Fig. 1: Architecture of Healthcare System

A. System Architecture

In this work, we focus on the healthcare system where a healthcare server owns plenty of patients' data and utilizes them for model training. The healthcare system model is shown in Fig. 1. As shown in the figure, the sensitive health data that holds in the server includes glucose levels, heart rates and blood pressure, etc. The raw data obtained from different modalities like images and texts are preprocessed and represented with binary feature vectors. The dataset that stores in the server is denoted as $D = \{(x_1, y_1), ..., (x_n, y_n)\}$, where x is the feature vector, y is the corresponding label, and n is the size of D. The doctor will assist the healthcare server to perform machine learning over such feature vectors. After obtaining the trained model, the server will publish it to specific patients and doctors.

In particular, we assume that the doctor follows Empirical Risk Minimization (ERM) [14] rule, i.e., the machine learning model trained by minimizing the average loss error over the training data. Given a dataset $D = \{z_1, z_2, ..., z_n\}$, where each $z_i = (x_i, y_i)$ lies in a domain $\mathcal{X}$, the ERM problem is defined as follows

$$\min_{w \in \mathcal{C}} \; F(w, D) = \frac{1}{n} \sum_{i=1}^{n} f(w, z_i), \tag{1}$$

where $\mathcal{C} \subseteq \mathbb{R}^d$ is a convex set, $f(w, z_i)$ is a loss function for each z_i. The ERM problem (1) can be implemented for many important machine learning problems by choosing different loss functions, such as logistic regression, where the loss function is set to logistic loss; and support vector machine, in which the loss function is defined as hinge loss.

As we all know, medical resources are limited. The doctors with such a machine learning model can help them make diagnostic decisions and increase efficiency. As for patients, this machine learning model can help them do a health

assessment and boost chances of recovery. Specifically, in diabetes management, a research institute wants to develop a binary classifier that predicts whether the glucose level is normal or not. As a part of this procedure, the training dataset is collected from patients in a hospital including food intake, physical activity, and other biological and environmental factors. Moreover, the glucose levels gathered from the patients are labeled as '-1' or '1' (i.e., $y \in \{-1, 1\}$) based on a safety threshold. Thus, the research institute constructs a classifier over the training dataset that can predict the glucose level. As a result, without the frequent blood tests, the health of diabetics is better monitored. In addition, after publishing the learning model, it would also benefit the whole society.

B. Threat Model

Our goal is to let the healthcare server perform the machine learning model where it would not expose any other information about patients' data. During the training procedure, the training samples composed of feature vectors and labels are private, because they include the daily activity and health status of patients. As we mentioned before, the learning model constructed over the sensitive training dataset is shared by healthcare server and stored on mobile devices, which makes inference power-efficient and contribute to privacy due to no need communicating patient data to the healthcare system. As shown in Fig. 1, an adversary may attack a patient so that it can obtain the healthcare system model, i.e., the adversary access to the model's parameters. Additionally, we consider the same assumption of an adversary as the previous work [15] that is a strong adversary with full knowledge of the training methods and it aims to learn the private information in training samples from the model.

Based on the above adversary, the published machine learning model gives rise to privacy concern especially when the training dataset contains the medical records. Therefore, it is highly desirable to construct a privacy preserving healthcare system.

III. Preliminaries

In this section, we first introduce some background knowledge of convex analysis. Then we present the start-of-the-art privacy preserving technology: differential privacy, which helps us to build the private machine learning model.

A. Convex Analysis

Definition 1 (L-Lipschitz Function). *A function $f : \mathcal{C} \times \mathcal{X} \to \mathbb{R}$ over $w \in \mathcal{C}$ is said to be L-Lipschitz if for all $w_1, w_2 \in \mathcal{C}$ and $z \in \mathcal{X}$,*

$$|f(w_1, z) - f(w_2, z)| \leq L\|w_1 - w_2\|_2,$$

where $\|\cdot\|_2$ is l_2-norm.

Definition 2 (β-Smooth Function). *A function $f : \mathcal{C} \times \mathcal{X} \to \mathbb{R}$ is β-smooth if for every $w_1, w_2 \in \mathcal{C}$, and $z \in \mathcal{X}$ we have*

$$|f(w_1, z) - f(w_2, z) - \langle \nabla f(w_2), w_1 - w_2 \rangle| \leq \frac{L}{2}\|w_1 - w_2\|_2^2.$$

Definition 3 (μ-Strongly Convex). *A function $f : \mathcal{C} \times \mathcal{X} \to \mathbb{R}$ is μ-strongly convex if, for every $w_1, w_2 \in \mathcal{C}$, and for $z \in \mathcal{X}$ and every subgradient $\nabla f(w, z) \in \partial f(w, z)$, we have*

$$\partial f(w_2, z) \geq \langle \nabla f(w_1, z), w_2 - w_1 \rangle + \frac{\mu}{2}\|w_2 - w_1\|_2^2.$$

B. Differential Privacy

Differential privacy is introduced by Dwork [16], which provides strong privacy guarantees by measuring the privacy risk of each single element in the dataset. The concept of differential privacy is defined as:

Definition 4 ((ϵ,δ)-Differential Privacy). *A randomized Mechanism $\mathcal{A}$ is (ϵ,δ)-differentially private if for any two neighboring inputs $D, D' \in \mathbb{D}$ that differ in at most one single element, and for any possible output s in the output space of $\mathcal{A}$, it holds that*

$$Pr(\mathcal{A}(D) = s) \leq e^{\epsilon} Pr(\mathcal{A}(D') = s) + \delta.$$

To design a (ϵ,δ)-differentially private algorithm, the common approach is the classical Gaussian mechanism [7], which adds noise from Gaussian distribution calibrated to the sensitivity of the query function.

Definition 5 (Sensitivity). *The sensitivity of a query function $q(\cdot)$ that takes a dataset D as input is defined as follows*

$$\Delta_q = \sup_{D,D'} \|q(D) - q(D')\|_2$$

The sensitivity of the query function means the maximum difference in the output of q when one element of input data is changed.

Theorem 1 (Classical Gaussian Mechanism). *Given any function $q : \mathcal{X}^n \to \mathbb{R}^d$ and for any $\epsilon, \delta \in (0, 1)$, the Gaussian mechanism defined by $\mathcal{M}_G(D, q, \epsilon) = q(D) + \mathcal{N}(0, \sigma^2 I_d)$ is (ϵ, δ)-differential privacy, where $\mathcal{N}(0, \sigma^2 I_d)$ is the Gaussian distribution and $\sigma \geq \frac{\sqrt{2\log(1.25/\delta)}\Delta_q}{\epsilon}$.*

We can notice that the classical Gaussian mechanism is limited because of the range of privacy budget ϵ. It is shown that the classical Gaussian mechanism cannot achieve (ϵ, δ)-DP at the case $\epsilon \geq 1$. Thus, the analytic Gaussian mechanism (aGM) is presented in [13] to solve the limitations of the classical Gaussian mechanism.

Theorem 2 (Analytic Gaussian Mechanism). *For any $\epsilon > 0$ and $\delta \in [0, 1]$ and given any function $q : \mathcal{X}^n \to \mathbb{R}^d$, the Gaussian mechanism $\mathcal{M}_G(D, q, \epsilon) = q(D) + \mathcal{N}(0, \sigma^2 I_d)$ is (ϵ, δ)-differential privacy if and only if*

$$\Phi\left(\frac{\Delta_q}{2\sigma} - \frac{\epsilon\sigma}{\Delta_q}\right) - e^{\epsilon}\Phi\left(-\frac{\Delta_q}{2\sigma} - \frac{\epsilon\sigma}{\Delta_q}\right) \leq \delta, \qquad (2)$$

where Φ is the CDF of the standard Gaussian distribution: $\Phi(t) = Pr[\mathcal{N}(0, 1) \leq t] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-y^2/2} dy$.

The result in Theorem 2 shows that it is enough to find a Gaussian noise with variance σ^2 satisfying the constraint (2) to achieve an (ϵ, δ)-differentially private Gaussian mechanism. Because the particularity of the Gaussian CDF Φ, a highly accurate calibration of the variance of Gaussian noise by using a numerical algorithm is proposed in [13].

IV. Private Risk Minimization

In this section, we provide two differentially private algorithms together with the analytic Gaussian mechanism to obtain differentially private machine learning models. And then we analyze their utilities for strongly convex and convex cases, respectively.

A. Private Convex Optimization via Output Perturbation

The output perturbation mechanism was first proposed in [9], which solves the ERM problem (1) to get the minimizer $\hat{w}$ and then adds random noise from a random variable. The mathematical description is as follows.

Definition 6 (Output Perturbation).

$$w_{priv}(D) = \hat{w}(D) + z, \tag{3}$$

where a random vector z is drawn from a Gaussian distribution $\mathcal{N}(0, \sigma^2 I)$ and $\hat{w}(D)$ is the minimizer $\hat{w}$ over a dataset D.

We first consider the case that the loss function $f(w, z_i)$ is μ-strongly convex, and then consider the non-strongly convex loss function.

1) Strongly Convex Case:

Lemma 1 ([10]). *Assume a loss function $f(w, z_i)$ is L-Lipschitz, β-smooth and μ-strongly convex for all z_i. Suppose we run gradient algorithm with constant step size $\eta \leq \frac{1}{\beta+\mu}$ for T steps. The sensitivity Δ then satisfies*

$$\Delta \leq \frac{5L(\mu+\beta)}{n\mu\beta}.$$

Theorem 3 (Privacy Guarantee). *In OPERA (Algorithm 1), for any $\epsilon > 0$ and $\delta \in [0, 1]$, it is (ϵ, δ)-differentially private.*

Proof. Theorem 3 can be directly derived by combining Theorem 2 and Lemma 1. □

Algorithm 1 Output Perturbation with aGM (OPERA)

1: **Input:** (ϵ, δ) is the privacy parameters and the sensitivity Δ. $f(w, z_i)$ is L-Lipschitz, β-smooth. η is the step size, T is the number of iteration and w_0 is the initial point.
2: Let $\delta_0 = \Phi(0) - e^{\epsilon}\Phi\left(-\sqrt{2\epsilon}\right)$.
3: **if** $\delta \geq \delta_0$ **then**
4: Define $B^+(v) = \Phi\left(\sqrt{\epsilon v}\right) - e^{\epsilon}\Phi\left(-\sqrt{\epsilon(v+2)}\right)$;
5: Compute $v^* = \sup\{v \in \mathbb{R}_{\geq 0} : B^+(v) \leq \delta\}$;
6: Let $\alpha = \sqrt{1 + v^*/2} - \sqrt{v^*/2}$.
7: **else**
8: Define $B^-(v) = \Phi\left(-\sqrt{\epsilon v}\right) - e^{\epsilon}\Phi\left(-\sqrt{\epsilon(v+2)}\right)$;
9: Compute $v^* = \inf\{v \in \mathbb{R}_{\geq 0} : B^-(v) \leq \delta\}$;
10: Let $\alpha = \sqrt{1 + v^*/2} + \sqrt{v^*/2}$.
11: **end if**
12: Set noise variance $\sigma^2 = \frac{\alpha^2\Delta^2}{2\epsilon}$.
13: **for** $t = 0, 1, ..., T-1$ **do**
14: $w_{t+1} = w_t - \frac{\eta}{n}\sum_{i=1}^{n}\nabla f(w_t, z_i)$.
15: **end for**
16: **Output:** $w_{priv} = w_T + \mathcal{N}(0, \sigma^2 I)$.

Theorem 4 (Utility Guarantee). *Suppose that for any z_i and $\|w\|_2 \leq R$, the loss function $f(w, z_i)$ is L-Lipschitz, β-smooth and μ-strongly convex. In OPERA (Algorithm 1), if one choose $\eta \leq \frac{1}{\beta+\mu}$ and $\Delta \leq \frac{5L(\mu+\beta)}{n\mu\beta}$, the following holds for $T = \Theta\left(\frac{\mu^2+\beta^2}{\mu\beta}\log(\frac{\mu^2 n^2 R^2 \epsilon}{L^2 d\alpha})\right)$,*

$$\mathbb{E}[F(w_{priv}, D)] - F(\hat{w}, D) \leq O\left(\frac{\beta L^2 d\alpha^2}{n^2\mu^2\epsilon}\right), \tag{4}$$

where α is parameter calculated in OPERA (Algorithm 1) for obtaining the noise variance σ^2.

Proof. By β-smoothness of f and $w_{priv} = w_T + z$, where $z \sim N(0, \sigma^2 I)$, we have

$$\begin{aligned}
&\mathbb{E}[F(w_{priv}, D)] - F(\hat{w}, D) \\
&\leq \mathbb{E}[F(w_T, D) + \langle \nabla F(w_T, D), z\rangle + \frac{\beta}{2}\|z\|_2^2] - F(\hat{w}, D) \\
&= F(w_T, D) - F(\hat{w}, D) + \frac{\beta}{2}\mathbb{E}\|z\|_2^2 \\
&\leq \frac{\beta}{2}exp(-\frac{2\mu\beta T}{(\mu+\beta)^2}) + \frac{25L^2(\mu+\beta)^2 d\alpha^2}{4n^2\mu^2\beta\epsilon},
\end{aligned}$$

where the second equality is due to Lemma 5 in [10].

Thus if we take $T = \Theta\left(\frac{\mu^2+\beta^2}{\mu\beta}\log(\frac{\mu^2 n^2 R^2 \epsilon}{L^2 d\alpha})\right)$, we can get

$$\mathbb{E}[F(w_{priv}, D)] - F(\hat{w}, D) \leq O\left(\frac{\beta L^2 d\alpha^2}{n^2\mu^2\epsilon}\right). \qquad \square$$

Remark 1. *For strongly convex loss function, our proposed algorithm OPERA achieves (ϵ, δ)-differential privacy and has the utility bound $\tilde{O}\left(\frac{d}{n^2\epsilon}\right)$ ($\tilde{O}$ means we ignore all log factors and α). Comparing with the best known utility upper bound $\tilde{O}\left(\frac{d}{n^2\epsilon^2}\right)$ for output perturbation approach in [10], OPERA yields a tighter bound by a factor of $\frac{1}{\epsilon}$ in the high privacy case ($\epsilon \to 0$). And also, our OPERA algorithm can extend to the low privacy regime ($\epsilon \to \infty$) with the help of the analytic Gaussian mechanism while providing (ϵ, δ)-differential privacy. The time complexity of our algorithm OPERA can be easily obtained as $time(\alpha) + \mathcal{O}(nd\log(\frac{n\epsilon}{\sqrt{d\log(\delta)}}))$, where $time(\alpha)$ is the time for computing α in Step 2-11, while the time complexity of [10] is $\mathcal{O}(nd\log(\frac{n\epsilon}{\sqrt{d\log(\delta)}})$. Hence, to provide a very efficient and robust way to find the solutions up to arbitrary accuracies for computing α, we adopt the the simple scheme in [13] based on binary search method.*

2) Non-strongly Convex Case:

Lemma 2 ([10]). *Assume a loss function $f(w, z_i)$ is L-Lipschitz and β-smooth for all z_i. Suppose we run gradient algorithm with constant step size $\eta \leq \frac{1}{\beta}$ for T steps. The sensitivity Δ then satisfies*

$$\Delta \leq \frac{3LT\eta}{n}.$$

Theorem 5 (Privacy Guarantee). *In OPERA (Algorithm 1), for any $\epsilon > 0$ and $\delta \in [0, 1]$, it is (ϵ, δ)-differentially private.*

Proof. While using Theorem 2 and Lemma 2, Theorem 5 can be directly derived. □

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2020.2997732, IEEE Transactions on Big Data

Theorem 6 (Utility Guarantee). *Suppose that for any z_i and $\|w\|_2 \leq R$, the loss function $f(w, z_i)$ is L-Lipschitz and β-smooth. In OPERA (Algorithm 1), if we choose $\eta \leq \frac{1}{\beta}$ and $\Delta \leq \frac{3LT\eta}{n}$, the following holds for $T = \Theta\left([\frac{\beta^2 n^2 R^2 \epsilon}{L^2 d\alpha^2}]^{\frac{1}{3}}\right)$,*

$$\mathbb{E}[F(w_{priv}, D)] - F(\hat{w}, D) \leq O\left(\left[\frac{LR^2\alpha}{n}\sqrt{\frac{\beta d}{\epsilon}}\right]^{\frac{2}{3}}\right), \quad (5)$$

where α is parameter calculated in OPERA (Algorithm 1) for obtaining the noise variance σ^2.

Proof. By β-smoothness of f and $w_{priv} = w_T + z$, where $z \sim N(0, \sigma^2 I_d)$, we have

$$\begin{aligned}
&\mathbb{E}[F(w_{priv}, D)] - F(\hat{w}, D) \\
&\leq \mathbb{E}[F(w_T, D) + \langle \nabla F(w_T, D), z\rangle + \frac{\beta}{2}\|z\|_2^2] - F(\hat{w}, D) \\
&= F(w_T, D) - F(\hat{w}, D) + \frac{\beta}{2}\mathbb{E}\|z\|_2^2 \\
&\leq \frac{2\beta R^2}{T} + \frac{9L^2 d\alpha^2}{2\beta\epsilon n^2},
\end{aligned}$$

where the second inequality is due to Lemma 8 in [10].

Let $T = \Theta\left([\frac{\beta^2 n^2 D^2 \epsilon}{L^2 d\alpha^2}]^{\frac{1}{3}}\right)$. We have

$$\mathbb{E}[F(w_{priv}, D)] - F(\hat{w}, D) \leq O\left(\left[\frac{LR^2\alpha}{n}\sqrt{\frac{\beta d}{\epsilon}}\right]^{\frac{2}{3}}\right).$$

□

Remark 2. *For non-strongly convex loss function under (ϵ, δ)-differential privacy, our OPERA algorithm yields an upper bound $\tilde{O}\left(\sqrt[3]{\frac{d}{n^2\epsilon}}\right)$. In the high privacy case ($\epsilon \to 0$), We improve the utility bound by a factor of $\frac{1}{\sqrt[3]{\epsilon}}$ in terms of the best known bound $\tilde{O}\left(\sqrt[3]{\frac{d}{n^2\epsilon^2}}\right)$ in [10]. Due to the advantages of the analytic Gaussian mechanism, our OPERA algorithm can extend to the low privacy case ($\epsilon \to \infty$).*

B. Private Convex Optimization via Gradient Perturbation

The gradient perturbation mechanism first proposed in [12] is using noisy gradient to minimize the ERM problem (1), and then obtains the output the differential private minimizer. The definition of the gradient perturbation mechanism is as follows.

Definition 7 (Gradient Perturbation).

$$w_{t+1}(D) = w_t - \eta[\nabla F(w_t, D) + z], \quad (6)$$

where z is a zero mean random noise and the private output is $w_{priv}(D) = w_T(D)$.

Theorem 7. *GRPUA (Algorithm 2) is $(\sqrt{8\log(n/\delta)}\epsilon + 2n\epsilon(e^{\frac{2\epsilon}{n}} - 1), n\delta)$-differentially private.*

Proof. Consider the t-th query: $M_t = n\nabla f(w_t, z_i) + z$, where $z \sim \mathcal{N}(0, \sigma^2 I_d)$. As a result of the analytic Gaussian mechanism, M_t is (ϵ, δ)-differentially private. Applying the privacy amplification method in Lemma 3 with $\gamma = \frac{1}{n}$, the query M_t ensures $(\frac{2\epsilon}{n}, \frac{\delta}{n})$-differential privacy. We then apply strong composition theorem in Lemma 4 to guarantee T composition of M_t queries is $(\sqrt{8\log(n/\delta)}\epsilon + 2n\epsilon(e^{\frac{2\epsilon}{n}} - 1), n\delta)$-differential private with $T = n^2$.

□

Algorithm 2 Gradient Perturbation with aGM (GRPUA)

Input: (ϵ, δ) is the privacy parameters and loss function $f(w, z_i)$ is L-Lipschitz. η is the step size and w_1 is the initial point.
Let $\delta_0 = \Phi(0) - e^{\epsilon}\Phi\left(-\sqrt{2\epsilon}\right)$.
if $\delta \geq \delta_0$ **then**
    Define $B^+(v) = \Phi\left(\sqrt{\epsilon v}\right) - e^{\epsilon}\Phi\left(-\sqrt{\epsilon(v+2)}\right)$;
    Compute $v^* = \sup\{v \in \mathbb{R}_{\geq 0} : B^+(v) \leq \delta\}$;
    Let $\alpha = \sqrt{1 + v^*/2} - \sqrt{v^*/2}$.
else
    Define $B^-(v) = \Phi\left(-\sqrt{\epsilon v}\right) - e^{\epsilon}\Phi\left(-\sqrt{\epsilon(v+2)}\right)$;
    Compute $v^* = \inf\{v \in \mathbb{R}_{\geq 0} : B^-(v) \leq \delta\}$;
    Let $\alpha = \sqrt{1 + v^*/2} + \sqrt{v^*/2}$.
end if
Set noise variance $\sigma^2 = \frac{\alpha^2 L^2}{2\epsilon}$.
for $t = 1, ..., n^2 - 1$ **do**
    Pick $i \in [n]$.
    $w_{t+1} = w_t - \eta(t)[n\nabla f(w_t, z_i) + z]$, where $z \sim \mathcal{N}(0, \sigma^2 I_d)$.
end for
Output: $w_{priv} = w_{n^2}$.

Lemma 3 ((Privacy Amplification [17]). *For any n-dataset D, if an (ϵ, δ)-differential private algorithm runs on uniformly random γn entries, this algorithm preserves $(2\gamma\epsilon, \gamma\delta)$-differential privacy.*

Lemma 4 (Strong Composition [18]). *For any $\epsilon > 0$, $\delta \in [0, 1]$, and $\delta' \in [0, 1]$, the class of (ϵ, δ)-differentially private mechanisms satisfies $(\epsilon\sqrt{2T\log(1/\delta)} + T\epsilon(e^{\epsilon} - 1), \delta' + T\delta)$-differential privacy under T-fold adaptive composition.*

In Theorem 8 and Theorem 9, we provide the utility upper bounds for our GPRUA algorithm under two different cases when the loss function is μ-strongly convex and when the loss function is non-strongly convex.

1) Strongly Convex Case:

Theorem 8 (Utility Guarantee). *Suppose that for any z_i and $\|w\|_2 \leq R$, the loss function $f(w, z_i)$ is L-Lipschitz and μ-strongly convex. In GRPUA (Algorithm 2), if one choose $\eta(t) = \frac{1}{\mu n t}$, the following holds*

$$\mathbb{E}[F(w_{priv}, D)] - F(\hat{w}, D) \leq O\left(\frac{L^2 d\alpha^2 \log(n)}{\mu\epsilon n^3}\right), \quad (7)$$

where α is parameter calculated in GRPUA (Algorithm 2) for obtaining the noise variance σ^2.

Proof. For t-th query $M_t = n\nabla f(w_t, z_i) + z$, we have $\mathbb{E}[M_t] = \nabla F(w_t, D)$ (the expectation is taken $w.r.t$ z_i and z). Also, $\mathbb{E}[\|M_t\|_2^2] = n^2\mathbb{E}[\|\nabla f(w_t, z_i)\|_2^2] +$

$2n\mathbb{E}[\langle \nabla f(w_t, z_i), z\rangle] + \mathbb{E}[\|z\|_2^2] \leq n^2L^2 + d\sigma^2$. Thus the theorem holds by directly using Lemma 5, where $G = \sqrt{n^2L^2 + d\sigma^2}$, $T = n^2$, and $\lambda = n\mu$ and $\eta(t) = \frac{1}{\mu nt}$. □

Lemma 5 ([19]). *Suppose $F(w)$ be a λ-strong convex function and $\hat{w} = \arg\min_{w\in\mathcal{C}} F(w)$. In a stochastic gradient algorithm, we have $w_{t+1} = w_t - \eta(t)M_t(w_t)$, where $\mathbb{E}[M_t(w_t)] = \nabla F(w_t)$, $\mathbb{E}[\|M_t(w_t)\|_2^2] \leq G^2$ and the learning rate $\eta(n) = \frac{1}{\lambda t}$. Then for any $T > 1$, the following holds*

$$\mathbb{E}[F(w_T) - F(\hat{w})] = O\left(\frac{G^2\log(T)}{\lambda T}\right).$$

Remark 3. *For $(\sqrt{8\log(n/\delta)}\epsilon + 2n\epsilon(e^{\frac{2\epsilon}{n}} - 1), n\delta)$-differential privacy, our GRPUA algorithm has a utility bound of $\tilde{O}\left(\frac{d}{n^3\epsilon}\right)$, which has much improvement of previous works $\tilde{O}\left(\frac{d}{n^2\epsilon^2}\right)$ in [12] and [11]. The time complexity of our GRPUA algorithm can be easily obtained as $time(\alpha) + \mathcal{O}(nd^2)$, where $time(\alpha)$ is the time for computing α in Step 2-11, while the time complexity of [12] is $\mathcal{O}(nd^2)$. We also use the simple approach proposed in [13] to improve the efficiency of computing $time(\alpha)$.*

2) Non-strongly Convex Case:

Theorem 9 (Utility Guarantee). *Suppose that for any z_i and $\|w\|_2 \leq R$, the loss function $f(w, z_i)$ is L-Lipschitz. In GRPUA (Algorithm 2), if one choose $\eta(t) = \frac{R}{\sqrt{t(n^2L^2+d\sigma^2)}}$, the following holds*

$$\mathbb{E}[F(w_{priv}, D)] - F(\hat{w}, D) \leq O\left(\frac{LR\alpha\log(n)}{n}\sqrt{\frac{d}{\epsilon}}\right), \quad (8)$$

where α is parameter calculated in GRPUA (Algorithm 2) for obtaining the noise variance σ^2.

Proof. The bound is obtained by directly using Lemma 6, where $G = \sqrt{n^2L^2 + d\sigma^2}$, and $T = n^2$, and $\eta(t) = R/\sqrt{t(n^2L^2 + d\sigma^2)}$. □

Lemma 6 ([19]). *Suppose $F(w)$ be a convex function with $\|w\|_2 \leq R$ and $\hat{w} = \arg\min_{w\in\mathcal{C}} F(w)$. In a stochastic gradient algorithm, we have $w_{t+1} = w_t - \eta(t)M_t(w_t)$, where $\mathbb{E}[M_t(w_t)] = \nabla F(w_t)$, $\mathbb{E}[\|M_t(w_t)\|_2^2] \leq G^2$ and the learning rate $\eta(n) = \frac{R}{G\sqrt{t}}$. Then for any $T > 1$, the following holds*

$$\mathbb{E}[F(w_T) - F(\hat{w})] = O\left(\frac{RG\log(T)}{\sqrt{T}}\right).$$

Remark 4. *For $(\sqrt{8\log(n/\delta)}\epsilon + 2n\epsilon(e^{\frac{2\epsilon}{n}} - 1), n\delta)$-differential privacy and the non-strongly convex loss function, our GRPUA algorithm yields an upper bound of $\tilde{O}\left(\sqrt{\frac{d}{n^2\epsilon}}\right)$. In the high privacy case, our GRPUA algorithm improve the previous results $\tilde{O}\left(\frac{\sqrt{d}}{n\epsilon}\right)$ by a factor of $\frac{1}{\sqrt{\epsilon}}$.*

V. Performance Evaluation

In this section, we will conduct experiments on real benchmarks and consider two machine learning models, logistic regression and support vector machine, to evaluate the performance of our proposed algorithms.

A. Experiment Settings

1) Dataset Description : The experiments involve five datasets:

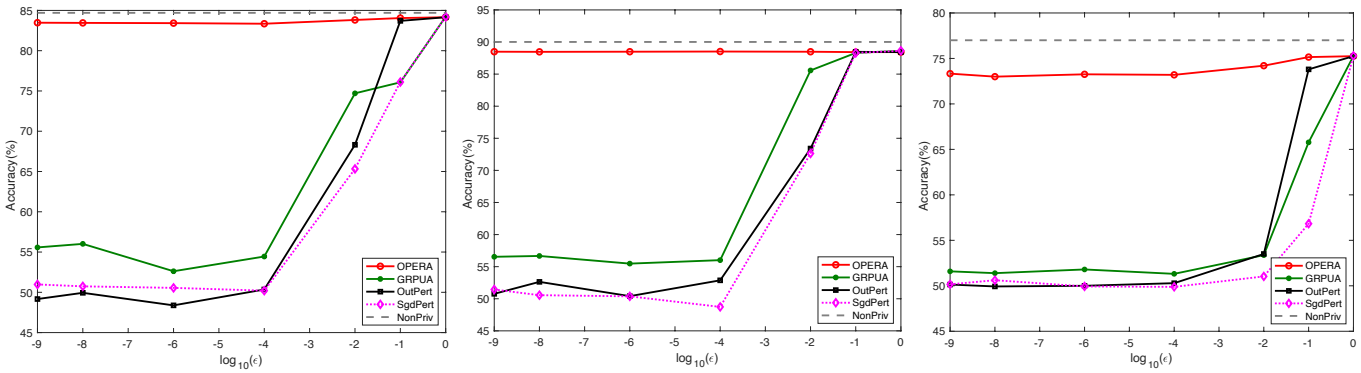
- **Adult**: The Adult dataset is from UCI Machine Learning Repository [20] and consists of 48,842 personal records, including age, education, occupation, work-class, sex, race, income, etc. The label is to predict whether the annual income is more than \$50k or not.
- **BANK**: The BANK dataset was collected from the marketing campaigns of a Portuguese banking institution, which includes 45,211 examples. Each example contains age, job, housing, loan, month, campaign, etc. The goal is to predict whether the product is subscribed.
- **IPUMS-BR**: The IPUMS-BR dataset is from IPUMS-International [21] and it contains 38,000 records of census microdata, which includes year, phone, sewage, cell, autos, etc. The label is to predict whether the monthly income of an individual is above \$300.
- **Cardio**: The Cardio dataset is from [22] and it contains 70,000 records of patients, which includes 11 features like Age, weight, Gender, Glucose, smoking, etc. The target is to predict whether an individual is presence or absence of cardiovascular disease.
- **Heart**: The Heart dataset is from [20] and it contains 303 records of patients, which includes 14 features like Age, fasting blood sugar, resting electrocardiographic results, exercise induced angina, etc. The goal is to decide the presence of heart disease in the patient.

2) Data Prepossessing: The following steps are used to prepossess the datasets. We first remove all individuals with missing values and then use the one-hot encoding method to convert every categorical attribute into a set of binary vectors. After that, we normalize all numerical attributes such that the range of each value is $[0, 1]$. Finally, we transform labels of Adult $\{>50\text{k}, \leq 50\text{k}\}$, BANK {subscribed, not-subscribed}, IPUMS-BR $\{>300, \leq 300\}$, to $\{+1, -1\}$, Cardio {presence, absence} to $\{+1, -1\}$, and Heart {presence, absence} to $\{+1, -1\}$.

3) Compared Methods: In this paper, we compare our OPERA and GRPUA algorithms with existing algorithms, namely, OutPert, SgdPert, NonPriv. OutPert is the name of output perturbation algorithm used in [10] that runs gradient descent algorithm for a fixed number of iterations and adding Gaussian noise from the classical Gaussian mechanism to the output. SgdPert [12] perturbs gradient of stochastic gradient algorithms and it computes the privacy budget using privacy amplification and strong composition methods to ensure the differential privacy. We set NonPriv as the optimal accuracy values that are obtained by running different optimization algorithms to do the classification tasks on these three datasets.

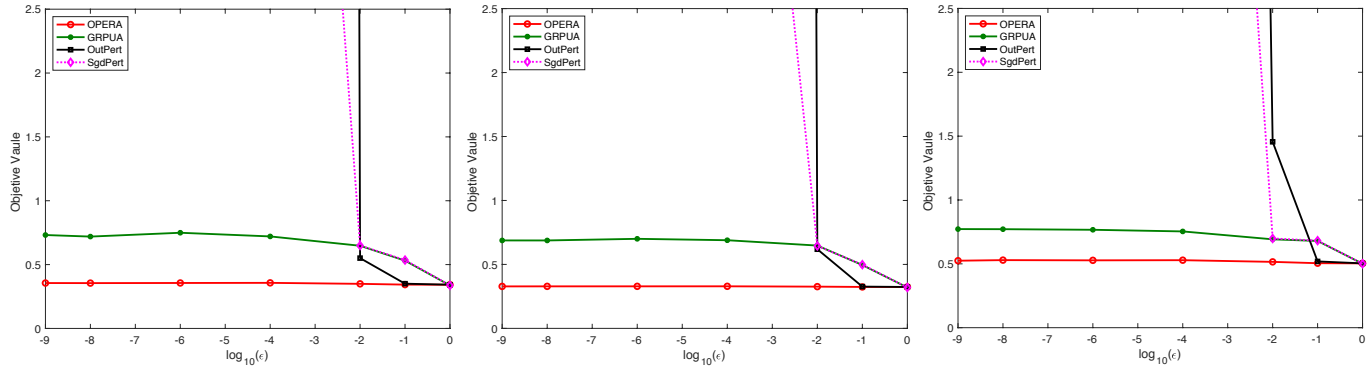
4) Parameter Setting: We perform 30 independent runs of algorithms, and record the mean values of objective value and accuracy. In all experiments, we set $\delta = 0.001$, $h = 0.5$ and the regularization term $\lambda = 0.001$. To enforce the Lipschitz constant L, we normalize each data sample to a unit norm (i.e., $\|x_i\|_2 = 1$, for $i = 1, \cdots, n$), which makes the Lipschitz constant $L = 1$. For logistic regression, we have smoothness

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2020.2997732, IEEE Transactions on Big Data



(a) Adult (b) BANK (c) IPUMS-BR

Fig. 2: Comparison of classification accuracies for Logistic regression on three datasets with different ϵ.



(a) Adult (b) BANK (c) IPUMS-BR

Fig. 3: Comparison of Objective values for Logistic regression on three datasets with different ϵ. (Note that the figure does not represent the objective values at much higher scale and only represents the objective values in the range of $[0, 2.5]$.)

constant $\beta_l \leq \frac{1}{4} + \lambda$. For SVM, we have smoothness constant $\beta_s \leq \frac{1}{2h} + \lambda$.

B. Logistic Regression

We apply our two algorithms OPERA and GRPUA to a logistic regression model for classification.

Given a dataset $\{(x_i, y_i)\}_{i=1}^n (x_i \in \mathbb{R}^d, y_i \in \{+1, -1\})$, the regularized logistic regression model is defined as

$$f(w, z_i) = \log(1 + \exp(-y_i w^T x_i)) + \frac{\lambda}{2}\|w\|_2^2,$$

where λ is the regularizer. The goal of classification is to minimize the logistic regression to obtain a minimizer, which is the weight vector.

Figure 2 shows classification accuracies of logistic regression on three datasets, i.e., Adult, BANK and IPUMS-BR. The proposed OPERA algorithm consistently outperforms OutPert in the high privacy regime ($\epsilon \to 0$). This is because OutPert uses the classical Gaussian mechanism to perturb gradients or outputs to achieve differential privacy. Compared with the analytical Gaussian mechanism, the classical Gaussian mechanism yields much more noise when ϵ is small. In other words, the noise calibration strategy of the analytical Gaussian mechanism provides a minimal amount of noise required to obtain (ϵ, δ)-differential privacy. As for GRPUA algorithm, it always outperforms the gradient perturbation algorithm: SgdPert, in the high privacy case. We also compare the computation time results of all methods on Adult dataset, as illustrated in Table I.

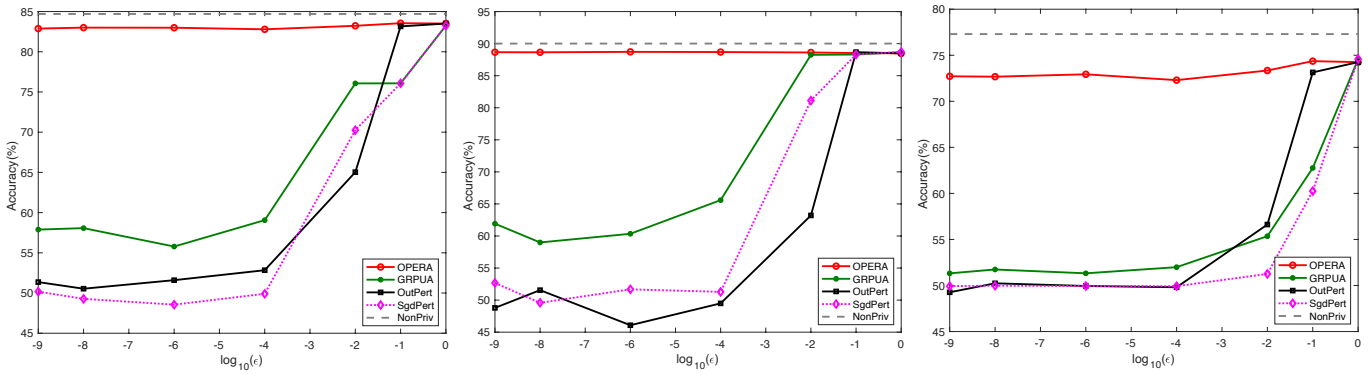
TABLE I: Computation Time.

	OPEERA	OutPert	GRPUA	SgdPert
$\epsilon = 0.01$	21.325s	21.282s	244.76635s	244.00784s
$\epsilon = 0.001$	67.976s	67.798s	245.04303s	244.30279s

We further explore how the final objective value of each algorithm changes as the value of ϵ increases in Fig. 3. As classification accuracy observed in experiments, OPERA nearly obtains the best achievable objective values in the high privacy regime due to the advantages of the analytic Gaussian mechanism. GRPUA also obtains low objective values but the values are much larger compared with OPERA. Due to adding noise to the gradient, it is difficult for the algorithm to find the right descent direction. Thus the objective values are higher compared with OPERA. However, GRPUA still performs better than other algorithms.

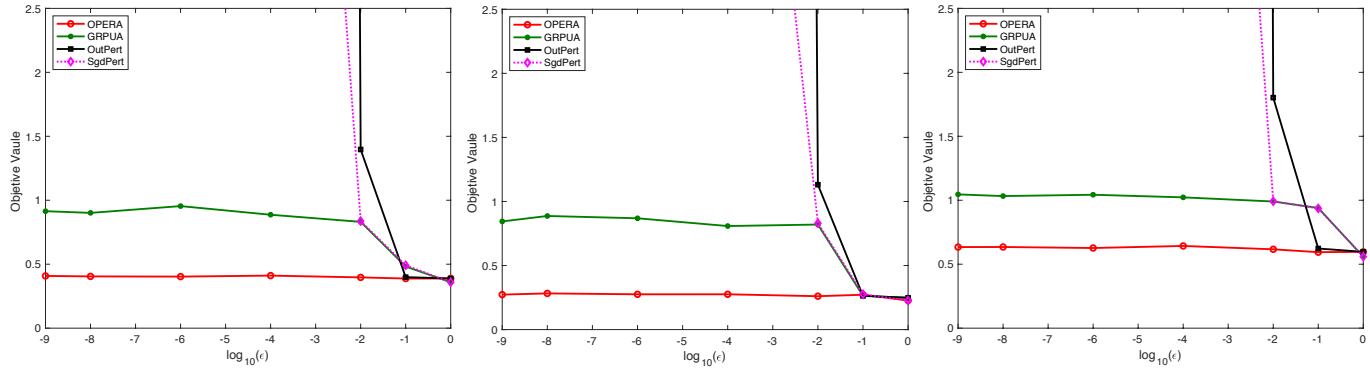
It should be emphasized that the analytic Gaussian mechanism has shown that the standard deviation of noise must scale like $\Omega(1/\sqrt{\epsilon})$ in order to extend the classical Gaussian mechanism ($\sigma = \Theta(1/\epsilon)$), from the range $\epsilon \in (0, 1)$ to $\epsilon \geq 1$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2020.2997732, IEEE Transactions on Big Data



(a) Adult (b) BANK (c) IPUMS-BR

Fig. 4: Comparison of classification accuracies for SVM on three datasets with different ϵ.



(a) Adult (b) BANK (c) IPUMS-BR

Fig. 5: Comparison of Objective values for SVM on three datasets with different ϵ. (Note that the figure does not represent the objective values at much higher scale and only represents the objective values in the range of $[0, 2.5]$.)

Moreover, as the value of privacy budget ϵ gets close to 1, the noise provided by analytic Gaussian mechanism is almost the same as the classical Gaussian mechanism, shown in [13]. Therefore, in both Fig. 2 and Fig. 3, the performance of GRPUA and OPERA algorithms is satisfactory as the value of privacy budget ϵ approaches 1.

C. Support Vector Machine

In this paper, we also perform our algorithms to a support vector machine (SVM) model for classification.

The regularized SVM model is given by

$$f(w, z_i) = H(y_i w^T x_i) + \frac{\lambda}{2}\|w\|_2^2$$

with

$$H(u) = \begin{cases} 1-u, & 1-u > h \\ 0, & 1-u < -h \\ \frac{(1-u)^2}{4h} + \frac{1-u}{2} + \frac{h}{4}, & otherwise \end{cases}$$

where $x_i \in \mathbb{R}^d, y_i \in \{+1, -1\}$, and λ is the regularizer.

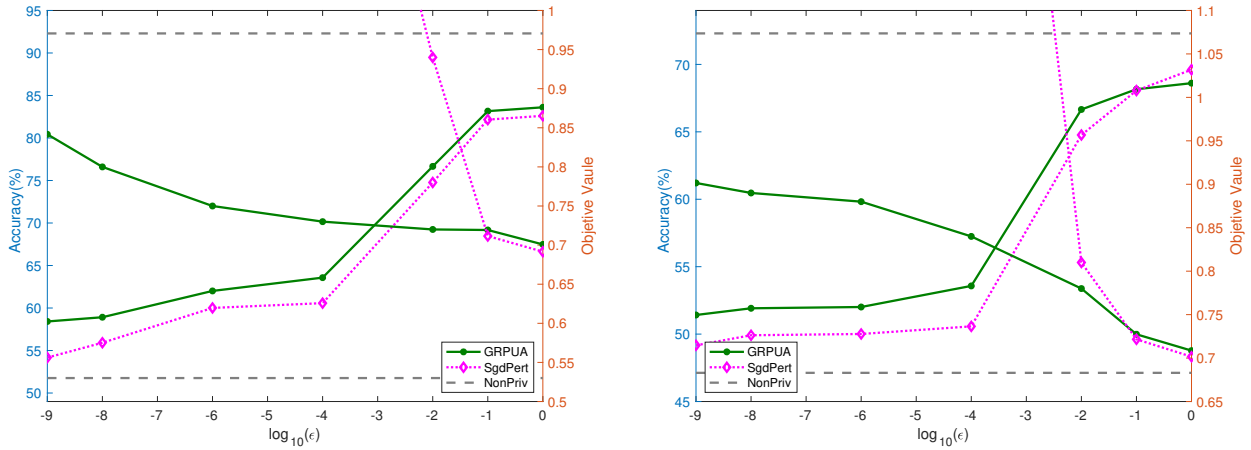
Figure 4 studies the performance of OPERA, GRPUA, Outpert, and SgdPert for SVM model on three datasets, i.e., Adult, BANK and IPUMS-BR. As it is clear from plots, OPERA also achieves the best accuracy in the whole privacy regime, i.e., the values of privacy budget ϵ from 10^{-9} to 1. Furthermore, as we decrease the value of privacy budget ϵ, GRPUA algorithm becomes better than OutPert and SgdPert algorithms on these datasets.

As shown in Fig. 5, our algorithms OPERA and GRPUA are quite robust for the very low privacy budget ϵ. For OPERA algorithm, it has obtained an absolutely low objective value when other algorithms have to sacrifice privacy to achieve this value. It should be mentioned that the objective values Outpert are lower than GRPUA when ϵ approximates to 1. At the same time, GRPUA has the same objective value as SgdPert. This is because the gain provided by analytic Gaussian mechanism is not more pronounced over the classical Gaussian mechanism in the low privacy regime, in other words, the amount of Gaussian noise used in OutpPert and SgdPert algorithms are nearly as much as the noise from the analytic Gaussian mechanism.

D. Neural Network

Here we consider a neural network architecture which includes two hidden layers with 100 hidden units each. We also use Rectified linear unit (ReLU) as the activation function, h_{relu}, and choose the deterministic cross-entropy loss as loss

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2020.2997732, IEEE Transactions on Big Data



(a) Heart (b) Cardio

Fig. 6: Comparison of accuracy and objective values for neural network on two medical datasets with different ϵ. (Note that the figure does not represent the objective values at much higher scale.)

function, which is given by

$$f(w, z_i) = -y_i \log(h(z_i)) - (1 - y_i) \log(1 - h(z_i))$$

where $h(z_i) = \frac{1}{1+exp(-z_i)}$, $z_i = w_3 h_{relu}(w_2 h_{relu}(w_1 x_i))$, and $w = \{w_1, w_2, w_3\}$ denotes the ensemble of weight matrices.

Since the neural network is non-convex problem, the sensitivity computations (i.e., Lemma 1, Lemm 2) for OPERA are not hold. Hence, we just adopt GRPUA to train the neural network and compare the results with SgdPert, where we use the norm gradient clipping C to bound the sensitivity of gradient $\Delta = C$. Here, we set clipping threshold $C = 1$. The experimental results on two medical datasets, Heart and Cardio, are shown in Fig. 6(a) and Fig. 6(b). We observe that our algorithm still outperforms baseline algorithm for neural network on medical datasets.

VI. Related Work

A. Privacy threats in machine learning

Many attacks have been proposed in the works which try to obtain sensitive information from the target model. Shokri et al. in [23] proposed membership inference attack to infer whether or not a given data sample is present in the training dataset. Fredrikson et al. in [5] presented an attribute inference attack (or model inversion attack) to infer the value of sensitive attribute of a test input. Wang et al. in [24] proposed a hyperparameter stealing attack to recover the underlying hyperparameters (such as model architectures) used during the model training. Moreover, an adversary may perform model stealing attack [25] to obtain a new model, whose performance is equivalent to that of a target model, via black-box access to the target model. As pointed out by [26], differential privacy has also been studied as a powerful method to defense against above tasks.

B. Privacy preserving ERM

A more recent line of research focuses on developing privacy preserving algorithms for convex ERM problems which tackle the problem from different aspects. In online settings, [27] investigated generic regret bound of online linear optimization problems. [28] focused on the learnability and stability under differential privacy. In incremental settings, [29] studied private incremental regression combining continual release to analyze the utility bound of several algorithms. For the special case of "generalized linear models", [30] gave dimension-independent expected excess risk bounds by using a sampling technique for an ERM with entropy regularizer. Some papers also consider the private ERM learning in a high dimension dataset. For a private high-dimensional sparse regression algorithm, [31] consider the convergence of parameter. The following work [32] used an algorithm based on a sample efficiency test of stability to extend and improve the results. [33] introduced Gaussian width of the parameter space in the random projection to derive a risk bound by using a private compress learning method in ERM algorithms.

C. Privacy preserving deep learning

There have been several techniques to develop privacy preserving deep learning mechanisms using differential privacy. Abadi et al. [15] trained deep neural networks with non-convex objective functions using differentially private stochastic gradient descent algorithm and they also implemented moments accountant to calculate the privacy budget. In the work of [34], the authors trained an ensemble teacher model by combining a set of teacher models, which are trained over disjoint training datasets and the author also trained the differentially private student model by querying the ensemble teacher to label public data. Moreover, Xie et al. [35] and Zhang et al. [36] focused on achieving differential privacy on Generative Adversarial Nets (GAN). However, none of these works provide utility guarantees for their algorithms.

VII. Conclusion

In this paper, we present a machine learning approach with differential privacy to healthcare scenario for increasing diagnostic efficiency while preserving training data privacy. We first develop two algorithms, OPERA and GRPUA, to obtain differentially private machine learning models, which provide privacy guarantee by applying the analytic Gaussian mechanism to the output and the gradient, respectively. In addition, we theoretically analyze the utility guarantee and privacy guarantee of proposed algorithms. For strongly convex and non-strongly convex loss functions, OPERA provides tighter utility bounds than existing output perturbation methods and GRPUA also achieves tighter utility bounds compared with previous gradient perturbation methods, especially in the high privacy regime. At last, we evaluate our algorithms on five datasets and the experiment results show that the proposed algorithms outperform the existing methods and guarantee the privacy of training datasets at the same time.

References

[1] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in bioinformatics*, vol. 1, no. 11, May 2017.

[2] R. Blankstein, R. P. Ward, M. Arnsdorf, B. Jones, Y.-B. Lou, and M. Pine, "Female gender is an independent predictor of operative mortality after coronary artery bypass graft surgery: contemporary analysis of 31 midwestern hospitals," *Circulation*, vol. 112, no. 9, pp. 323–327, 2005.

[3] S. Jiménez-Serrano, S. Tortajada, and J. M. García-Gómez, "A mobile health application to predict postpartum depression based on machine learning," *Telemedicine and e-Health*, vol. 21, no. 7, pp. 567–574, 2015.

[4] R. B. D'Agostino Sr, M. J. Pencina, J. M. Massaro, and S. Coady, "Cardiovascular disease risk assessment: insights from framingham," *Global heart*, vol. 8, no. 1, pp. 11–23, 2013.

[5] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, Denver, CO, October 2015.

[6] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of cryptography conference*, New York, NY, March 2006.

[7] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, August 2014.

[8] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Foundations of Computer Science*, Providence, RI, October 2007.

[9] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate, "Differentially private empirical risk minimization," *Journal of Machine Learning Research*, vol. 12, pp. 1069–1109, Mar 2011.

[10] J. Zhang, K. Zheng, W. Mou, and L. Wang, "Efficient private erm for smooth objectives," *arXiv preprint arXiv:1703.09947*, 2017.

[11] D. Wang, M. Ye, and J. Xu, "Differentially private empirical risk minimization revisited: Faster and more general," in *Advances in Neural Information Processing Systems*, Long Beach, CA, December 2017.

[12] R. Bassily, A. Smith, and A. Thakurta, "Private empirical risk minimization: Efficient algorithms and tight error bounds," in *Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, Philadelphia, PA, October 2014.

[13] B. Balle and Y.-X. Wang, "Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising," in *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, Sweden, July 2018.

[14] H. Trevor, T. Robert, and F. JH, *The elements of statistical learning: data mining, inference, and prediction*. Springer-Verlag New York, 2009.

[15] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, Vienna, Austria, October 2016.

[16] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Advances in Cryptology - EUROCRYPT 2006*. Berlin, Heidelberg: Springer Berlin Heidelberg, May 2006, pp. 486–503.

[17] A. Beimel, H. Brenner, S. P. Kasiviswanathan, and K. Nissim, "Bounds on the sample complexity for private learning and private data release," *Machine learning*, vol. 94, no. 3, pp. 401–437, March 2014.

[18] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and differential privacy," in *IEEE 51st Annual Symposium on Foundations of Computer Science*, Las Vegas, NV, October 2010.

[19] O. Shamir and T. Zhang, "Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes," in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, GA, Jun 2013.

[20] D. Dheeru and E. Karra Taniskidou, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[21] M. P. Center, "Integrated public use microdata series, international: Version 7.0," 2018. [Online]. Available: http://international.ipums.org

[22] Kaggle, "UCI machine learning repository." [Online]. Available: https://www.kaggle.com/sulianova/cardiovascular-disease-dataset

[23] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *2017 IEEE Symposium on Security and Privacy (SP)*, San Jose, CA, May 2017.

[24] B. Wang and N. Z. Gong, "Stealing hyperparameters in machine learning," in *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018, pp. 36–52.

[25] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Stealing machine learning models via prediction apis," in *25th {USENIX} Security Symposium ({USENIX} Security 16)*, 2016, pp. 601–618.

[26] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, 2019.

[27] N. Agarwal and K. Singh, "The price of differential privacy for online learning," in *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, August 2017.

[28] Y.-X. Wang, J. Sharpnack, A. J. Smola, and R. J. Tibshirani, "Learning with differential privacy: Stability, learnability and the sufficiency and necessity of erm principle," *Journal of Machine Learning Research*, vol. 17, no. 183, pp. 1–40, 2016.

[29] S. P. Kasiviswanathan, K. Nissim, and H. Jin, "Private incremental regression," in *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, Chicago, IL, May 2017.

[30] P. Jain and A. G. Thakurta, "(near) dimension independent risk bounds for differentially private learning," in *Proceedings of the 31st International Conference on Machine Learning*, Beijing, China, June 2014.

[31] D. Kifer, A. Smith, and A. Thakurta, "Private convex empirical risk minimization and high-dimensional regression," in *Proceedings of the 25th Annual Conference on Learning Theory*, Edinburgh, Scotland, June 2012.

[32] A. G. Thakurta and A. Smith, "Differentially private feature selection via stability arguments, and the robustness of the lasso," in *Proceedings of the 26th Annual Conference on Learning Theory*, Princeton, NJ, June 2013.

[33] S. P. Kasiviswanathan and H. Jin, "Efficient private empirical risk minimization for high-dimensional learning," in *Proceedings of The 33rd International Conference on Machine Learning*, New York, NY, June 2016.

[34] N. Papernot, M. Abadi, Ú. Erlingsson, I. Goodfellow, and K. Talwar, "Semi-supervised knowledge transfer for deep learning from private training data," in *The 5th International Conference on Learning Representations*, Toulon, France, April 2017.

[35] L. Xie, K. Lin, S. Wang, F. Wang, and J. Zhou, "Differentially private generative adversarial network," *arXiv preprint arXiv:1802.06739*, 2018.

[36] X. Zhang, J. Ding, S. M. Errapotu, X. Huang, P. Li, and M. Pan, "Differentially private functional mechanism for generative adversarial networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, 2019, pp. 1–6.