Journal of Applied Mathematics and Physics, 2017, 5, 900-911
http://www.scirp.org/journal/jamp
ISSN Online: 2327-4379
ISSN Print: 2327-4352

Error Analysis and Variable Selection for Differential Private Learning Algorithm

Weilin Nie[1], Cheng Wang[2]

Huizhou University, Huizhou, China
Email: niewl@hzu.edu.cn, wangch@hzu.edu.cn

How to cite this paper: Nie, W.L. and Wang, C. (2017) Error Analysis and Variable Selection for Differential Private Learning Algorithm. *Journal of Applied Mathematics and Physics*, **5**, 900-911.
https://doi.org/10.4236/jamp.2017.54079

Received: February 14, 2017
Accepted: April 27, 2017
Published: April 30, 2017

Abstract

In this paper, we construct a modified least squares regression algorithm which can provide privacy protection. A new concentration inequality is applied and the expected error bound is derived by error decomposition. Furthermore, via the error analysis, we find a method to choose an appropriate parameter ϵ to balance the error and privacy.

Keywords

Differential Privacy, Least Squares Regularization, Concentration Inequality, Error Decomposition

1. Introduction

Privacy protection attracts much attention in many branches of computer science. To deal with this, Dwork *et al.* proposed differential privacy in [1]. Soon [2] builds an exponential mechanism which is a useful approach to construct a differential private algorithm. The concept is introduced into learning theory in [3]. There, the authors consider output perturbation and object perturbation for ERM algorithms. Analysis of privacy and generalization for those algorithms also has been conducted. P. Jain and his collaborators have done a lot of work on differential private learning afterwards [4] [5] and etc. Recently, in [6], the authors find that the empirical average of the output from a differential private algorithm can converge to its expectation. And [7] provides another analysis of this convergence, which motivates our work.

In this paper, we consider the following statistical learning model (see [8] [9] for more details): The input space X is a compact metric space, and the output space is $Y \subset \mathbb{R}$ as a regression problem. Throughout the paper, we assume the output Y is uniformly bounded, *i.e.*, $|y| \le M$ for some $M > 0$ almost surely. On the sample space $Z := X \times Y$, we try to find a function $f : X \to Y$ via some

algorithms $\mathcal{A}$, reflecting the relationship between the input and output. Algorithm $\mathcal{A}$ relies on the random chosen sample $\mathbf{z}=\{z_i\}_{i=1}^m=\{(x_i,y_i)\}_{i=1}^m$, while the sample is drawn according to a distribution function ρ on Z. Furthermore, we assume there is a marginal distribution ρ_X on X and conditional distribution $\rho(y|x)$ on Y given some x.

Now we expect the algorithm can provide some privacy protection. We assume $\mathcal{A}$ satisfies the (ϵ,γ) differential private condition [1]. Denoting the Hamming distance between two sample sets $\{\mathbf{z}_1,\mathbf{z}_2\}$ is

$$d(\mathbf{z}_1,\mathbf{z}_2)=\#\{i=1,\cdots,m: z_{1,i}\neq z_{2,i}\},$$

i.e., there is only one element is different. Then (ϵ,γ)-differential private is defined as follows:

Definition 1 *A random algorithm $A: Z^m\to\mathcal{H}$ is (ϵ,γ)-differential private if for every two data sets $\mathbf{z}_1,\mathbf{z}_2$ satisfying $d(\mathbf{z}_1,\mathbf{z}_2)=1$, and every sets $\mathcal{O}\in\mathcal{H}$ we have*

$$\Pr\{A(\mathbf{z}_1)\in\mathcal{O}\}\leq \mathrm{e}^{\epsilon}\cdot\Pr\{A(\mathbf{z}_2)\in\mathcal{O}\}+\gamma.$$

Here $\mathcal{H}$ is a function space from X to Y, which is called the hypothesis space. In the sequel, we focus on the $(\epsilon,0)$-differential privacy with some $0<\epsilon<1$, which is always called ϵ-differential privacy for simplicity. How to choose an appropriate ϵ is a fundamental problem in differential private algorithms [10], and we will provide a method during our error estimation in the following sections.

2. Concentration Inequality

In this section, we study the error between average and expectation for an algorithm $\mathcal{A}$ providing ϵ-differential privacy. Our first result can be stated as follow:

Theorem 1 *If an algorithm $\mathcal{A}$ provides ϵ-differential privacy, and outputs a positive function $g_{z,\mathcal{A}}: X\times Y\to\mathbb{R}$ with bounded expectation $\mathbb{E}_{z,\mathcal{A}}g_{z,\mathcal{A}}\leq G$ for some $G>0$, where the expectation is taken over the sample via the algorithm output. Then*

$$\mathbb{E}_{z,\mathcal{A}}\left(\frac{1}{m}\sum_{i=1}^m g_{z,\mathcal{A}}(z_i)-\int_Z g_{z,\mathcal{A}}(z)\mathrm{d}\rho\right)\leq 2G\epsilon,$$

and

$$\mathbb{E}_{z,\mathcal{A}}\left(\int_Z g_{z,\mathcal{A}}(z)\mathrm{d}\rho-\frac{1}{m}\sum_{i=1}^m g_{z,\mathcal{A}}(z_i)\right)\leq 2G\epsilon.$$

Denote sample sets $\mathbf{w}_j=\{z_1,z_2,\cdots,z_{j-1},z_j',z_{j+1},\cdots,z_m\}$ for $j\in\{1,2,\cdots,m\}$. We observe that

$$\begin{aligned}\mathbb{E}_{z,\mathcal{A}}\left(\frac{1}{m}\sum_{i=1}^m g_{z,\mathcal{A}}(z_i)\right)&=\frac{1}{m}\sum_{i=1}^m\mathbb{E}_z\mathbb{E}_{\mathcal{A}}\left(g_{z,\mathcal{A}}(z_i)\right)\\&=\frac{1}{m}\sum_{i=1}^m\mathbb{E}_z\mathbb{E}_{z_i'}\int_0^{+\infty}\Pr_{\mathcal{A}}\{g_{z,\mathcal{A}}(z_i)\geq t\}\mathrm{d}t\end{aligned}$$

$$\le \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_z\mathbb{E}_{z_i'}\int_0^{+\infty}\mathrm{e}^{\epsilon}\Pr_{\mathcal{A}}\left\{g_{w_i,\mathcal{A}}(z_i)\ge t\right\}\mathrm{d}t$$
$$= \mathrm{e}^{\epsilon}\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_{w_i}\mathbb{E}_{z_i}\mathbb{E}_{\mathcal{A}}\left(g_{w_i,\mathcal{A}}(z_i)\right) = \mathrm{e}^{\epsilon}\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_{w_i,\mathcal{A}}\mathbb{E}_{z_i}\left(g_{w_i,\mathcal{A}}(z_i)\right)$$
$$= \mathrm{e}^{\epsilon}\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_{w_i,\mathcal{A}}\int_Z g_{w_i,\mathcal{A}}(z)\mathrm{d}\rho = \mathrm{e}^{\epsilon}\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_{z,\mathcal{A}}\int_Z g_{z,\mathcal{A}}(z)\mathrm{d}\rho$$
$$= \mathrm{e}^{\epsilon}\mathbb{E}_{z,\mathcal{A}}\int_Z g_{z,\mathcal{A}}(z)\mathrm{d}\rho.$$

Then

$$\mathbb{E}_{z,\mathcal{A}}\left(\frac{1}{m}\sum_{i=1}^{m}g_{z,\mathcal{A}}(z_i)-\int_Z g_{z,\mathcal{A}}(z)\mathrm{d}\rho\right)$$
$$\le\left(\mathrm{e}^{\epsilon}-1\right)\mathbb{E}_{z,\mathcal{A}}\left(\int_Z g_{z,\mathcal{A}}(z)\mathrm{d}\rho\right)\le 2G\epsilon.$$

On the other hand,

$$\mathbb{E}_{z,\mathcal{A}}\int_Z g_{z,\mathcal{A}}(z)\mathrm{d}\rho = \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_z\mathbb{E}_{\mathcal{A}}\int_Z g_{z,\mathcal{A}}(z)\mathrm{d}\rho$$
$$= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_{w_i}\mathbb{E}_{\mathcal{A}}\int_Z g_{w_i,\mathcal{A}}(z)\mathrm{d}\rho = \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_{w_i}\mathbb{E}_{\mathcal{A}}\int_Z g_{w_i,\mathcal{A}}(z_i)\mathrm{d}\rho(z_i)$$
$$= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_{w_i}\mathbb{E}_{z_i}\mathbb{E}_{\mathcal{A}}\left(g_{w_i,\mathcal{A}}(z_i)\right) = \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_z\mathbb{E}_{z_i'}\int_0^{+\infty}\Pr_{\mathcal{A}}\left\{g_{w_i,\mathcal{A}}(z_i)\ge t\right\}\mathrm{d}t$$
$$\le \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_z\mathbb{E}_{z_i'}\mathrm{e}^{\epsilon}\int_0^{+\infty}\Pr_{\mathcal{A}}\left\{g_{z,\mathcal{A}}(z_i)\ge t\right\}\mathrm{d}t$$
$$= \mathrm{e}^{\epsilon}\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_z\mathbb{E}_{\mathcal{A}}\left(g_{z,\mathcal{A}}(z_i)\right) = \mathrm{e}^{\epsilon}\mathbb{E}_{z,\mathcal{A}}\frac{1}{m}g_{z,\mathcal{A}}(z_i).$$

This leads to

$$\mathbb{E}_{z,\mathcal{A}}\left(\int_Z g_{z,\mathcal{A}}(z)\mathrm{d}\rho-\frac{1}{m}\sum_{i=1}^{m}g_{z,\mathcal{A}}(z_i)\right)$$
$$=\left(\mathrm{e}^{\epsilon}-1\right)\mathbb{E}_{z,\mathcal{A}}\frac{1}{m}\sum_{i=1}^{m}g_{z,\mathcal{A}}(z_j)\le 2G\epsilon.$$

These verify our results.

Remark 1 *Similar results are proposed in* [6] *and* [7]. *However, there the authors limits the function to take value in* $[0,1]$ *or* $\{0,1\}$, *our result here extends theirs to the function taking value in* $\mathbb{R}^+$. *This makes our following error analysis implementable.*

3. Differential Private Learning Algorithm

In this section we consider the differential private least squares regularization algorithm. For a Mercer kernel K defined on $X\times X$, the function space $\mathcal{H}_K := \overline{\mathrm{span}\left\{K(x,\cdot), x\in X\right\}}$ is the induced reproducing kernel Hilbert space (RKHS). Denote $K_x(y)=K(x,y)$ for any $x,y\in X$, and $\kappa=\sup_{x,y\in X}\sqrt{K(x,y)}$. It is well known that $f(x)=\langle f, K_x\rangle_K$ as the reproducing property. In the sequel, we always assume $|y|\le M$ for some constant $M>0$. The least squares regularization algorithm, which has been extensively studied in such as [8] [11] [12] and etc. is:

$$f_{z,\lambda} = \arg\min_{f\in\mathcal{H}_K} \frac{1}{m}\sum_{i=1}^{m}\left(f(x_i)-y_i\right)^2 + \lambda\|f\|_K^2. \quad (1)$$

Denote π as a projection operator as we did in [13] [14]:

$$\pi\left(f(x)\right) = \begin{cases} M, & f(x) > M \\ f(x), & -M \le f(x) \le M \\ -M, & f(x) < -M \end{cases}.$$

Then we add a noise term b in the original algorithm (1) like the output perturbation algorithm in [3]:

$$f_{z,\mathcal{A}}(x) = \pi\left(f_{z,\lambda}(x)\right) + b \quad (2)$$

where the density of b is independent with z which will be clarified in the following analysis. Moreover, we take the following notation for simplicity:

$$\mathcal{E}(f) = \int_Z \left(f(x)-y\right)^2 \mathrm{d}\rho,\ \mathcal{E}_z(f) = \frac{1}{m}\sum_{i=1}^{m}\left(f(x_i)-y_i\right)^2.$$

Definition 2 *We denote Δf_z as the maximum infinite norm of difference when changing one sample point in z, i.e., if $d(\mathbf{z},\mathbf{z}')=1$,*

$$\Delta f_z = \sup_{z,z'}\|f_z - f_{z'}\|_\infty.$$

Then we have the following result:

Lemma 1 *Assume $\Delta\pi\left(f_{z,\lambda}(x)\right)$ is bounded, and b has density function proportion to $\exp\left\{-\frac{\epsilon|b|}{\Delta\pi\left(f_{z,\lambda}\right)}\right\}$, then algorithm (2) provides ϵ-differential privacy.*

The proof is just as Theorem 4 in [15]. For all possible function r, and $\mathbf{z},\mathbf{z}'$ differ in one element, then

$$\Pr\left\{f_{z,\mathcal{A}} = r\right\} = \Pr_b\left\{b = r - \pi\left(f_{z,\lambda}\right)\right\} \propto \exp\left(-\frac{\epsilon\left\|r-\pi\left(f_{z,\lambda}\right)\right\|_\infty}{\Delta\pi\left(f_{z,\lambda}\right)}\right),$$

and

$$\Pr\left\{f_{z',\mathcal{A}} = r\right\} = \Pr_b\left\{b = r - \pi\left(f_{z',\lambda}\right)\right\} \propto \exp\left(-\frac{\epsilon\left\|r-\pi\left(f_{z',\lambda}\right)\right\|_\infty}{\Delta\pi\left(f_{z',\lambda}\right)}\right).$$

So

$$\Pr\left\{f_{z,\mathcal{A}} = r\right\} \le \Pr\left\{\pi\left(f_{z',\mathcal{A}}\right) = r\right\} \times \mathrm{e}^{\frac{\epsilon\left\|\pi\left(f_{z,\lambda}\right)-\pi\left(f_{z',\lambda}\right)\right\|_\infty}{\Delta\pi\left(f_{z,\lambda}\right)}} \le \mathrm{e}^{\epsilon}\Pr\left\{f_{z',\mathcal{A}} = r\right\}.$$

Then the lemma is proved by a union bound.

Now we will bound the term $\Delta f_{z,\lambda}$.

Lemma 2 *For the function $f_{z,\lambda}$ obtained from algorithm (1), assume $\|f_{z,\lambda}\|_K \le R$ for any $\mathbf{z}\in Z^m$ for some $R \ge M$, and $0<\lambda\le 1$, we have*

$$\Delta f_{z,\lambda} \le \frac{2R\kappa^2(\kappa+1)}{\lambda m}.$$

Assume $f_{z,\lambda}$ and $f_{z',\lambda}$ are two results derived via algorithm (1) given any sample set $\boldsymbol{z}, \boldsymbol{z}'$ satisfying $d(\boldsymbol{z}, \boldsymbol{z}') = 1$. Without loss of generality, we set $\boldsymbol{z}' = (z_1, z_2, \cdots, z_{m-1}, z_{m'})$. Since the two functions are both the optimizer of algorithm (1), take derivative for f we have

$$\frac{2}{m}\sum_{i=1}^{m}\left(f_{z,\lambda}(x_i) - y_i\right)K_{x_i} + 2\lambda f_{z,\lambda} = 0$$

and

$$\frac{2}{m}\sum_{i=1}^{m-1}\left(f_{z',\lambda}(x_i) - y_i\right)K_{x_i} + \frac{2}{m}\left(f_{z',\lambda}(x'_m) - y'_m\right)K_{x_m} + 2\lambda f_{z',\lambda} = 0.$$

These lead to

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^{m}\left(f_{z,\lambda}(x_i) - f_{z',\lambda}(x_i)\right)K_{x_i} + \lambda\left(f_{z,\lambda} - f_{z',\lambda}\right)\\ &= \frac{1}{m}\left[\left(f_{z',\lambda}(x'_m) - y'_m\right)K_{x'_m} - \left(f_{z,\lambda}(x_m) - y_m\right)K_{x_m}\right].\end{aligned}$$

Take inner product with $f_{z,\lambda} - f_{z',\lambda}$ by both sides we have

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^{m}\left(f_{z,\lambda}(x_i) - f_{z',\lambda}(x_i)\right)^2 + \lambda\left\|f_{z,\lambda} - f_{z',\lambda}\right\|_K^2\\ &= \frac{1}{m}\left[\left(f_{z',\lambda}(x'_m) - y'_m\right)\left(f_{z,\lambda}(x'_m) - f_{z',\lambda}(x'_m)\right)\right.\\ &\qquad \left. - \left(f_{z,\lambda}(x_m) - y_m\right)\left(f_{z,\lambda}(x_m) - f_{z',\lambda}(x_m)\right)\right].\end{aligned}$$

This means

$$\begin{aligned}\lambda\left\|f_{z,\lambda} - f_{z',\lambda}\right\|_K^2 &\le \frac{1}{m}\left[\left|f_{z',\lambda}(x'_m) - y'_m\right| + \left|f_{z,\lambda}(x_m) - y_m\right|\right]\cdot\left\|f_{z,\lambda} - f_{z',\lambda}\right\|_\infty\\ &\le \frac{1}{m}\left(\left\|f_{z',\lambda}\right\|_\infty + \left\|f_{z,\lambda}\right\|_\infty + 2M\right)\kappa\left\|f_{z,\lambda} - f_{z',\lambda}\right\|_K.\end{aligned}$$

The last inequality is from the fact that

$$\|f\|_\infty = \sup_{x\in X} f(x) = \sup_{x\in X}\langle f, K_x\rangle_K \le \|K_x\|_K \cdot \|f\|_K \le \kappa\|f\|_K.$$

Since $\left\|f_{z,\lambda}\right\|_K \le R$, then $\left\|f_{z',\lambda}\right\|_K \le R$ as well. Therefore,

$$\left\|f_{z,\lambda} - f_{z',\lambda}\right\|_K \le \frac{1}{\lambda m}(2R\kappa + 2M)\kappa \le \frac{2R\kappa(\kappa+1)}{\lambda m}$$

for any $0 < \lambda \le 1$. So

$$\left\|f_{z,\lambda} - f_{z',\lambda}\right\|_\infty \le \frac{2R\kappa^2(\kappa+1)}{\lambda m}$$

for any $\boldsymbol{z}, \boldsymbol{z}'$, and our lemma holds.

It can be easily verified by discussion that

$$\left\|\pi\left(f_{z,\lambda}\right) - \pi\left(f_{z',\lambda}\right)\right\|_\infty \le \left\|f_{z,\lambda} - f_{z',\lambda}\right\|_\infty$$

for any $\boldsymbol{z}, \boldsymbol{z}'$, so we have the choice of noise b and the result for algorithm (2).

Proposition 1 *Assume* $\left\|f_{z,\lambda}\right\|_K \le R$ *for any* $\boldsymbol{z} \in Z^m$ *for some* $R \ge M$, *and* b *takes value in* $(-\infty, +\infty)$, *we choose the density of* b *to be*

$\frac{1}{\alpha}\exp\left(-\frac{\lambda m\epsilon|b|}{2R\kappa^2(\kappa+1)}\right)$, *where* $\alpha=\frac{4R\kappa^2(\kappa+1)}{\lambda m\epsilon}$, *then the algorithm* (2) *provides* ϵ *-differential privacy.*

The proof is by combining the two lemmas and the inequality above. And by simply calculation we can get the expression of α.

4. Error Analysis for Differential Private Learning Algorithm

In this section, we will study the expectation of the error between $\mathcal{E}(f_{z,\mathcal{A}})-\mathcal{E}(f_\rho)$, where $f_\rho=\int_Y y\mathrm{d}\rho(y|x)$ is the regression function which minimizes $\mathcal{E}(f)$. Firstly we shall introduce the error decomposition:

$$\begin{aligned}
&\mathcal{E}(f_{z,\mathcal{A}})-\mathcal{E}(f_\rho)\le\mathcal{E}(f_{z,\mathcal{A}})-\mathcal{E}(f_\rho)+\lambda\|f_{z,\lambda}\|_K^2\\
&\le\mathcal{E}(f_{z,\mathcal{A}})-\mathcal{E}_z(f_{z,\mathcal{A}})+\mathcal{E}_z(f_{z,\mathcal{A}})-\mathcal{E}_z(\pi(f_{z,\lambda}))\\
&\quad+\mathcal{E}_z(\pi(f_{z,\lambda}))+\lambda\|f_{z,\lambda}\|_K^2-\mathcal{E}(f_\rho)\\
&\le\mathcal{E}(f_{z,\mathcal{A}})-\mathcal{E}_z(f_{z,\mathcal{A}})+\mathcal{E}_z(f_{z,\mathcal{A}})-\mathcal{E}_z(\pi(f_{z,\lambda}))\\
&\quad+\mathcal{E}_z(f_{z,\lambda})+\lambda\|f_{z,\lambda}\|_K^2-\mathcal{E}(f_\rho)\\
&\le\mathcal{E}(f_{z,\mathcal{A}})-\mathcal{E}_z(f_{z,\mathcal{A}})+\mathcal{E}_z(f_{z,\mathcal{A}})-\mathcal{E}_z(\pi(f_{z,\lambda}))\\
&\quad+\mathcal{E}_z(f_\lambda)+\lambda\|f_\lambda\|_K^2-\mathcal{E}(f_\rho)\\
&\le\mathcal{R}_1+\mathcal{R}_2+\mathcal{S}+D(\lambda),
\end{aligned}\tag{3}$$

where f_λ is a function in $\mathcal{H}_K$ to be determined and

$$\mathcal{R}_1=\mathcal{E}(f_{z,\mathcal{A}})-\mathcal{E}_z(f_{z,\mathcal{A}}),$$

$$\mathcal{R}_2=\mathcal{E}_z(f_{z,\mathcal{A}})-\mathcal{E}_z(\pi(f_{z,\lambda})),$$

$$\mathcal{S}=\mathcal{E}_z(f_\lambda)-\mathcal{E}(f_\lambda),$$

$$D(\lambda)=\mathcal{E}(f_\lambda)-\mathcal{E}(f_\rho)+\lambda\|f_\lambda\|_K^2.$$

Here $\mathcal{R}_1$ and $\mathcal{R}_2$ involve the function $f_{z,\mathcal{A}}$ from random algorithm (2) so we call them random errors. $\mathcal{S}$ and $D(\lambda)$ are similar as classical ones in the past literature in learning theory and we still call them sample error and approximation error. In the following, we will study these errors respectively.

4.1. Error Bounds for Random Errors

Proposition 2 *For function* $f_{z,\mathcal{A}}$ *obtained from algorithm* (2) *with density of* b *as described in Proposition* 1, *we have*

$$\mathbb{E}_{z,\mathcal{A}}\mathcal{R}_1\le 8\epsilon\left(\frac{2R^2\kappa^4(\kappa+1)^2}{\lambda^2m^2\epsilon^2}+M^2\right).$$

Note that

$$\mathcal{R}_1=\int_Z(f_{z,\mathcal{A}}(x)-y)^2\mathrm{d}\rho-\frac{1}{m}\sum_{i=1}^m(f_{z,\mathcal{A}}(x_i)-y_i)^2,$$

analogous analysis to the proof of Theorem 1 tells us that

$$\begin{aligned}
&\mathbb{E}_{z,\mathcal{A}}\left(\int_Z\left(f_{z,\mathcal{A}}(x)-y\right)^2\mathrm{d}\rho-\frac{1}{m}\sum_{i=1}^m\left(f_{z,\mathcal{A}}(x_i)-|y_i\right)^2\right)\\
&\le\left(\mathrm{e}^{\epsilon}-1\right)\mathbb{E}_z\mathbb{E}_{\mathcal{A}}\frac{1}{m}\sum_{i=1}^m\left(\pi\left(f_{z,\lambda}(x_i)\right)+b-y_i\right)^2\mathrm{d}\rho\\
&=2\epsilon\mathbb{E}_z\mathbb{E}_b\left(b^2+b\left(\pi\left(f_{z,\lambda}(x_i)\right)-y_i\right)+\left(\pi\left(f_{z,\lambda}(x_i)\right)-y_i\right)^2\right)\\
&\le 2\epsilon\left(\frac{8R^2\kappa^4(\kappa+1)^2}{\lambda^2m^2\epsilon^2}+4M^2\right),
\end{aligned}$$

which verifies the proposition.

For the term $\mathcal{R}_2$, we have the same analysis.

Proposition 3 *For function* $f_{z,\mathcal{A}}$ *obtained from algorithm* (2) *with density of* b *as described in Proposition 1, we have*

$$\mathbb{E}_{z,\mathcal{A}}\mathcal{R}_2\le\frac{8R^2\kappa^4(\kappa+1)^2}{\lambda^2m^2\epsilon^2}.$$

Since

$$\begin{aligned}
\mathcal{R}_2&=\mathcal{E}_z\left(f_{z,\mathcal{A}}\right)-\mathcal{E}_z\left(\pi\left(f_{z,\lambda}\right)\right)\\
&=\frac{1}{m}\sum_{i=1}^m\left[\left(f_{z,\mathcal{A}}(x_i)-y_i\right)^2-\left(\pi\left(f_{z,\lambda}(x_i)\right)-y_i\right)^2\right]\\
&=\frac{1}{m}\sum_{i=1}^m b\left(b+2\pi\left(f_{z,\lambda}(x_i)\right)-2y_i\right)\\
&=b^2+2b\cdot\frac{1}{m}\sum_{i=1}^m\left(\pi\left(f_{z,\lambda}(x_i)\right)-y_i\right),
\end{aligned}$$

we have

$$\mathbb{E}_{z,\mathcal{A}}\mathcal{R}_2=\mathbb{E}_z\mathbb{E}_bb^2\le\frac{8R^2\kappa^4(\kappa+1)^2}{\lambda^2m^2\epsilon^2}.$$

And the proposition is proved.

4.2. Error Estimates for Sample Error and Approximation Error

Error estimates for sample error and approximation error have been extensively studied since [8]. Here we provide the proof for completeness. It is known that f_λ in the error decomposition (3) can be arbitrarily chosen in $\mathcal{H}_K$ in [12] [13] [14] and etc. Here we simply choose it to be the classical one

$$f_\lambda=\arg\min_{f\in\mathcal{H}_K}\mathcal{E}(f)+\lambda\|f\|_K^2.$$

From [16] [17] we have the expression of f_λ is

$$f_\lambda=\left(L_K+\lambda\right)^{-1}L_Kf_\rho,$$

where L_K is the operator defined on $L^2_{\rho_X}$ as

$$L_Kf(t)=\int_Xf(x)K(x,t)\mathrm{d}\rho_X.$$

[8] told us that L_K has a eigenvalue sequence $\{\mu_i\}_{i\ge1}$ satisfies $\mu_i>0$ $\mu_i\to0$ when $i\to\infty$, and $\|L_K\|\le\kappa^2$. Now we recall the Hoeffding inequality [18].

Lemma 3 *Let* ξ *be a random variable on a probability space* Z *satisfying* $\left|\xi(z)-\mathbb{E}\xi\right|\le B$ *for some* $B>0$ *for almost all* $z\in Z$, *then*

$$\Pr\left\{\left|\frac{1}{m}\sum_{i=1}^{m}\xi(z_i)-\mathbb{E}\xi\ge\varepsilon\right|\right\}\le 2\exp\left\{-\frac{m\varepsilon^2}{2B^2}\right\}.$$

Then we have the following analysis.

Proposition 4 *For* f_λ *and* f_ρ *defined as above, assume* $f_\rho\in L_K^r\left(L^2_{\rho_X}\right)$, *we have*

$$\mathbb{E}_{z,\mathcal{A}}\mathcal{S}+D(\lambda)\le\frac{8\sqrt{2\pi}M^2}{\sqrt{m}}+\lambda^{\min\{2r,1\}}\left(\kappa^{4r-2}+\kappa^{4r-4}+2\right)\left\|L_K^{-r}f_\rho\right\|_\rho^2.$$

Firstly we bound the sample error.

$$\begin{aligned}\mathcal{S}&=\mathcal{E}(f_\lambda)-\mathcal{E}_z(f_\lambda)\\&=\int_Z\left(f_\lambda(x)-y\right)^2\mathrm{d}\rho-\frac{1}{m}\sum_{i=1}^{m}\left(f_\lambda(x_i)-y_i\right)^2.\end{aligned}$$

Let $\xi(z)=-\left(f_\lambda(x)-y\right)^2$, since $\left|f_\rho(x)\right|=\left|\int_Y y\mathrm{d}\rho(y|x)\right|\le M$, and

$$\begin{aligned}\left\|f_\lambda\right\|_\infty&=\left\|\left(L_K+\lambda I\right)^{-1}L_Kf_\rho\right\|_\infty\\&\le\left\|\left(L_K+\lambda I\right)^{-1}L_K\right\|\cdot\left\|f_\rho\right\|_\infty\le M,\end{aligned}$$

we have $\left|\xi-\mathbb{E}\xi\right|\le 8M^2$. So from Hoeffding inequality there holds

$$\begin{aligned}&\Pr_z\left\{\left|\int_Z\left(f_\lambda(x)-y\right)^2\mathrm{d}\rho-\frac{1}{m}\sum_{i=1}^{m}\left(f_\lambda(x_i)-y_i\right)^2\right|\ge\varepsilon\right\}\\&\le 2\exp\left\{-\frac{m\varepsilon^2}{128M^4}\right\}.\end{aligned}$$

Then we have

$$\begin{aligned}\mathbb{E}_{z,\mathcal{A}}\mathcal{S}&\le\mathbb{E}_z\left|\mathcal{S}\right|=\int_0^{+\infty}\Pr_z\left\{\left|\mathcal{S}\right|\ge t\right\}\mathrm{d}t\\&=\int_0^{+\infty}2\exp\left\{-\frac{mt^2}{128M^4}\right\}\mathrm{d}t\le\frac{8\sqrt{2\pi}M^2}{\sqrt{m}}.\end{aligned}$$

For the approximation error, note that $\mathcal{E}(f_\lambda)-\mathcal{E}(f_\rho)=\left\|f_\lambda-f_\rho\right\|_\rho^2$ [9] which is independent with z and b, we have

$$\begin{aligned}\mathbb{E}_{z,\mathcal{A}}\mathcal{E}(f_\lambda)-\mathcal{E}(f_\rho)&=\left\|f_\lambda-f_\rho\right\|_\rho^2\\&=\left\|\left(L_K+\lambda I\right)^{-1}\left(L_K-\left(L_K+\lambda I\right)\right)f_\rho\right\|_\rho^2\\&=\lambda^2\left\|\left(L_K+\lambda I\right)^{-1}L_K^rL_K^{-r}f_\rho\right\|_\rho^2\\&\le\lambda^2\left\|\left(L_K+\lambda I\right)^{-1}L_K^r\right\|^2\left\|L_K^{-r}f_\rho\right\|_\rho^2\\&\le\begin{cases}\lambda^{2r}\left\|L_K^{-r}f_\rho\right\|_\rho^2, & r\le 1\\ \lambda^2\kappa^{4(r-1)}\left\|L_K^{-r}f_\rho\right\|_\rho^2, & r>1\end{cases}\\&\le\lambda^{\min\{2r,2\}}\left(\kappa^{4(r-1)}+1\right)\left\|L_K^{-r}f_\rho\right\|_\rho^2.\end{aligned}$$

On the other hand, in [8], the authors pointed out that $\|f\|_K = \left\|L_K^{-\frac{1}{2}} f\right\|_\rho$ for any $f \in \mathcal{H}_K$. So

$$\begin{aligned}
\mathbb{E}_{z,\mathcal{A}} \lambda \|f_\lambda\|_K^2 &= \lambda \left\|\left(L_K + \lambda I\right)^{-1} L_K f_\rho\right\|_K^2 \\
&= \lambda \left\|\left(L_K + \lambda I\right)^{-1} L_K^{\frac{1}{2}} f_\rho\right\|_\rho^2 \\
&\le \lambda \left\|\left(L_K + \lambda I\right)^{-1} L_K^{\frac{1}{2}+r}\right\|^2 \cdot \left\|L_K^{-r} f_\rho\right\|_\rho^2 \\
&\le \begin{cases} \lambda^{2r} \left\|L_K^{-r} f_\rho\right\|_\rho^2, & r \le \frac{1}{2} \\ \lambda \cdot \kappa^{4r-2} \left\|L_K^{-r} f_\rho\right\|_\rho^2, & r > \frac{1}{2} \end{cases} \\
&\le \lambda^{\min\{2r,1\}} \left(\kappa^{4r-2} + 1\right) \left\|L_K^{-r} f_\rho\right\|_\rho^2.
\end{aligned}$$

Combining the 3 bounds above, we can verify the proposition.

4.3. Convergence Result with Fixed ϵ

In our analysis for $\mathbb{E}_{z,\mathcal{A}} \mathcal{R}_1$ above, we indeed have the following result

$$\mathbb{E}_{z,\mathcal{A}} \mathcal{R}_1 \le \frac{16R^2\kappa^4 (\kappa+1)^2}{\lambda^2 m^2 \epsilon} + 2\epsilon \mathbb{E}_z \mathcal{E}_z \left(\pi\left(f_{z,\lambda}\right)\right).$$

Therefore, the error decomposition can be

$$\begin{aligned}
&\mathbb{E}_{z,\mathcal{A}} \left(\mathcal{E}\left(f_{z,\mathcal{A}}\right) - (1+2\epsilon)\mathcal{E}\left(f_\rho\right)\right) \\
&= \mathbb{E}_{z,\mathcal{A}} \left(\mathcal{R}_1 + \mathcal{R}_2 + \mathcal{S} + D(\lambda) - 2\epsilon\mathcal{E}\left(f_\rho\right)\right) \\
&\le \frac{16R^2\kappa^4 (\kappa+1)^2}{\lambda^2 m^2 \epsilon} + \frac{8R^2\kappa^4 (\kappa+1)^2}{\lambda^2 m^2 \epsilon^2} + \mathbb{E}_z 2\epsilon \left(\mathcal{E}_z\left(\pi\left(f_{z,\lambda}\right)\right) - \mathcal{E}\left(f_\rho\right)\right) + \mathbb{E}_z \left(\mathcal{S} + D(\lambda)\right) \\
&\le \frac{24R^2\kappa^4 (\kappa+1)^2}{\lambda^2 m^2 \epsilon^2} + 2\epsilon \mathbb{E}_z \left(\mathcal{E}_z\left(f_{z,\lambda}\right) + \lambda \left\|f_{z,\lambda}\right\|_K^2 - \mathcal{E}\left(f_\rho\right)\right) + \mathbb{E}_z \left(\mathcal{S} + D(\lambda)\right) \\
&\le \frac{24R^2\kappa^4 (\kappa+1)^2}{\lambda^2 m^2 \epsilon^2} + 2\epsilon \mathbb{E}_z \left(\mathcal{E}_z\left(f_\lambda\right) + \lambda \left\|f_\lambda\right\|_K^2 - \mathcal{E}\left(f_\rho\right)\right) + \mathbb{E}_z \left(\mathcal{S} + D(\lambda)\right) \\
&\le \frac{24R^2\kappa^4 (\kappa+1)^2}{\lambda^2 m^2 \epsilon^2} + (1+2\epsilon) \mathbb{E}_z \left(\mathcal{S} + D(\lambda)\right) \\
&\le \frac{24M^2\kappa^4 (\kappa+1)^2}{\lambda^3 m^2 \epsilon^2} + \frac{3\sqrt{2\pi} M^2 (1+2\epsilon)}{\sqrt{m}} + \lambda^{\min\{1,2r\}} \left(\kappa^{4r-2} + \kappa^{4r-4} + 2\right) \left\|L_K^{-r} f_\rho\right\|_\rho^2.
\end{aligned}$$

Then by choosing $\lambda = \left(\frac{1}{m}\right)^{2/\min\{4,3+2r\}}$ for balance we have the following result.

Theorem 2 *Let* $f_{z,\mathcal{A}}$ *derived from algorithm* (2), $f_{z,\lambda}$, f_λ *defined in the above subsections, and assume* $f_\rho \in L_K^r \left(L_{\rho_X}\right)$, *take* $\lambda = \left(\frac{1}{m}\right)^{2/\min\{4,3+2r\}}$, *there holds*

$$\mathbb{E}_{z,\mathcal{A}}\left(\mathcal{E}\left(f_{z,\mathcal{A}}\right)-(1+2\epsilon)\mathcal{E}\left(f_\rho\right)\right)\le C_\epsilon\left(\frac{1}{m}\right)^{\min\left\{\frac{1}{2},\frac{4r}{3+2r}\right\}},$$

where constant

$$C_\epsilon=\frac{24M^2\kappa^4(\kappa+1)^2}{\epsilon^2}+8\sqrt{2\pi}M^2(1+2\epsilon)+\left(\kappa^{4r-2}+\kappa^{4r-4}+2\right)\left\|L_K^{-r}f_\rho\right\|_\rho^2.$$

4.4. Selection of ϵ and Total Error Bound

From the analysis for random error, sample error and approximation error above, we can obtain the whole error bound as follow.

Theorem 3 *Let* $f_{z,\mathcal{A}}$ *derived from algorithm* (2), $f_{z,\lambda}$, f_λ *defined in the above subsections, and assume* $f_\rho\in L_K^r\left(L_{\rho_X{}^2}\right)$, *take*

$$\lambda=\left(\frac{1}{m\epsilon}\right)^{2/\min\{4,3+2r\}},$$

and

$$\epsilon=\left(\frac{1}{\sqrt{m}}\right)^{\min\{1/3,4r/(3+6r)\}}$$

we have

$$\mathbb{E}_{z,\mathcal{A}}\left(\mathcal{E}\left(f_{z,\mathcal{A}}\right)-\mathcal{E}\left(f_\rho\right)\right)\le\tilde{C}\left(\frac{1}{m}\right)^{\min\left\{\frac{1}{3},\frac{4r}{3+6r}\right\}},$$

where constant

$$\tilde{C}=8\left(1+\sqrt{2\pi}\right)M^2+24M^2\kappa^4(\kappa+1)^2+\left(\kappa^{4r-2}+\kappa^{4r-4}+2\right)\left\|L_K^{-r}f_\rho\right\|_\rho^2.$$

It can be seen from error decomposition (3) that

$$\begin{aligned}\mathbb{E}_{z,\mathcal{A}}\left(\mathcal{E}\left(f_{z,\mathcal{A}}\right)-\mathcal{E}\left(f_\rho\right)\right)&\le\mathbb{E}_{z,\mathcal{A}}\left(\mathcal{E}\left(f_{z,\mathcal{A}}\right)-\mathcal{E}\left(f_\rho\right)+\lambda\left\|f_{z,\lambda}\right\|_K^2\right)\\&\le\mathbb{E}_{z,\mathcal{A}}\left(\mathcal{R}_1+\mathcal{R}_2+\mathcal{S}+D(\lambda)\right)\\&\le 8\epsilon\left(\frac{2R^2\kappa^4(\kappa+1)^2}{\lambda^2m^2\epsilon^2}+M^2\right)+\frac{8R^2\kappa^4(\kappa+1)^2}{\lambda^2m^2\epsilon^2}\\&\quad+\frac{8\sqrt{2\pi}M^2}{\sqrt{m}}+\lambda^{\min\{2r,1\}}\left(\kappa^{4r-2}+\kappa^{4r-4}+2\right)\left\|L_K^{-r}f_\rho\right\|_\rho^2\\&\le 8M^2\epsilon+\frac{24R^2\kappa^4(\kappa+1)^2}{\lambda^2m^2\epsilon^2}+\frac{8\sqrt{2\pi}M^2}{\sqrt{m}}\\&\quad+\lambda^{\min\{2r,1\}}\left(\kappa^{4r-2}+\kappa^{4r-4}+2\right)\left\|L_K^{-r}f_\rho\right\|_\rho^2.\end{aligned}$$

Since $\lambda\left\|f_{z,\lambda}\right\|_K^2\le\mathcal{E}_z\left(f_{z,\lambda}\right)+\lambda\left\|f_{z,\lambda}\right\|_K^2\le\mathcal{E}_z(0)\le M^2$, we have $\left\|f_{z,\lambda}\right\|_K\le\frac{M}{\sqrt{\lambda}}$, *i.e.*, we can choose $R=\frac{M}{\sqrt{\lambda}}$. Now take $\lambda=\left(\frac{1}{m\epsilon}\right)^{2/\min\{4,3+2r\}}$ and

$\epsilon = \left(\frac{1}{\sqrt{m}}\right)^{\min\{1/3, 4r/(3+6r)\}}$ for balance, and the result is proved.

5. Conclusions

Theorem 2, where ϵ is taken as a constant, reveals that the generalization error $\mathcal{E}\left(\pi\left(f_{z,\mathcal{A}}\right)\right)$ converges not to the one of regression function $\mathcal{E}\left(f_\rho\right)$, but a little different one $\left(1+2\epsilon\right)\mathcal{E}\left(f_\rho\right)$ in expectation.

It can be seen from the definition of differential privacy that algorithms will provide more privacy when ϵ tends to 0. However, Theorem 3 shows that ϵ cannot be too small, since the expected error will be very large accordingly. Hence our choice can be regarded as a balance between privacy protection and the expected error. In [19], the authors announce that ϵ also needs tend to 0 in some rates to keep generalization which matches our result.

Compared with previous learning theory results [12] [20] [21] [22] and etc., our learning rate is not so good since a perturbation term is introduced. However, in our result Theorem 1, we did not need a capacity condition as what we did in classical error analysis, *i.e.*, conditions on covering numbers, VC or Vγ dimensions. Instead the ϵ-differential private condition is adopted. So it may be capable and interesting for us to apply such condition to other learning algorithms.

Acknowledgements

This work is supported by NSFC (Nos. 11326096, 11401247), NSF of Guangdong Province in China (No. 2015A030313674), National Social Science Fund in China (No. 15BTJ024), Planning Fund Project of Humanities and Social Science Research in Chinese Ministry of Education (No. 14YJAZH040), Foundation for Distinguished Young Talents in Higher Education of Guangdong, China (No. 2016KQNCX162) and the Major Incubation Research Project of Huizhou University (No. hzux1201619).

References

[1] Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006) Calibrating Noise to Sensitivity in Private Data Analysis. In: Halevi, S. and Rabin, T., Eds., *Theory of Cryptography*, Springer, Berlin, 265-284.

[2] McSherry, F. and Talwar, K. (2007) Mechanism Design via Differential Privacy. *Proceedings of the* 48*th Annual Symposium on Foundations of Computer Science*, Providence, 21-23 October 2007, 94-103. https://doi.org/10.1109/focs.2007.66

[3] Chaudhuri, K., Monteleoni, C. and Sarwate, A.D. (2011) Differentially Private Empirical Risk Minimization. *Journal of Machine Learning Research*, **12**, 1069-1109.

[4] Jain, P. and Thakurta, A.G. (2013) Differentially Private Learning with Kernels. *JMLR*: *Workshop and Conference Proceedings*, **28**, 118-126.

[5] Jain, P. and Thakurta, A.G. (2014) Dimension Independent Risk Bounds for Differentially Private Learning. *Proceedings of the* 31*st International Conference on Machine Learning*, Beijing, 21-26 June 2014, 476-484.

[6] Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O. and Roth, A. (2015)

Preserving Statistical Validity in Adaptive Data Analysis. *ACM Symposium on the Theory of Computing*, Portland, 14-17 June 2015, 117-126. https://doi.org/10.1145/2746539.2746580

[7] Bassily, R., Nissim, K., Smith, A., Steinke, T., Stemmer, U. and Ullman, J. (2015) Algorithmic Stability for Adaptive Data Analysis.

[8] Cucker, F. and Smale, S. (2002) On the Mathematical Foundations of Learning. *Bulletin of the AMS*, **39**, 1-49. https://doi.org/10.1090/S0273-0979-01-00923-5

[9] Cucker, F. and Zhou, D.X. (2007) Learning Theory: An Approximation Theory Viewpoint. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511618796

[10] Dwork, C. (2008) Differential Privacy: A Survey of Results. *International Conference on Theory and Applications of Models of Computation*, Xi'an, 25-29 April 2008, 1-19.

[11] Steinwart, I., Hush, D. and Scovel, C. (2009) Optimal Rates for Regularized Least Squares Regression. In: Dasgupta, S. and Klivans, A., Eds., *Proceedings of the* 22*nd Annual Conference on Learning Theory*, Montreal, 18-21 June 2009, 79-93.

[12] Wu, Q., Ying, Y. and Zhou, D.X. (2006) Learning Rates of Least-Square Regularized Regression. *Foundations of Computational Mathematics*, **6**, 171-192. https://doi.org/10.1007/s10208-004-0155-9

[13] Nie, W.L. and Wang, C. (2015) Constructive Analysis for Coefficient Regularization Regression Algorithms. *Journal of Mathematical Analysis and Applications*, **431**, 1153-1171.

[14] Wang, C. and Nie, W.L. (2014) Constructive Analysis for Least Squares Regression with Generalized K-Norm Regularization. *Abstract and Applied Analysis*, **2014**, Article ID: 458459. https://doi.org/10.1155/2014/458459

[15] Dwork, C. (2006) Differential Privacy. Springer, Berlin, 1-12.

[16] Smale, S. and Zhou, D.X. (2003) Estimating the Approximation Error in Learning Theory. *Analysis and Applications*, **1**, 17-41. https://doi.org/10.1142/S0219530503000089

[17] Smale, S. and Zhou, D.X. (2007) Learning Theory Estimates via Integral Operators and Their Applications. *Constructive Approximation*, **26**, 153-172. https://doi.org/10.1007/s00365-006-0659-y

[18] Hoeffding, W. (1963) Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, **58**, 13-30. https://doi.org/10.1080/01621459.1963.10500830

[19] Wang, Y.-X., Lei, J. and Fienberg, S.E. (2015) Learning with Differential Privacy: Stability, Learn Ability and the Sufficiency and Necessity of ERM Principle.

[20] Wang, C. and Zhou, D.X. (2011) Optimal Learning Rates for Least Squares Regularized Regression with Unbounded Sampling. *Journal of Complexity*, **27**, 55-67.

[21] Hu, T., Fan, J., Wu, Q. and Zhou, D.X. (2015) Regularization Schemes for Minimum Error Entropy Principle. *Analysis and Applications*, **13**, 437-455. https://doi.org/10.1142/S0219530514500110

[22] Christmann, A. and Zhou, D.X. (2016) Learning Rates for the Risk of Kernel-Based Quantile Regression Estimators in Additive Models. *Analysis and Applications*, **14**, 449-477. https://doi.org/10.1142/S0219530515500050