1H4-4

差分プライベート統合を用いた分散データからの線形回帰

Linear Regression for Distributed Data Set Using Differentially Private Aggregation

南 賢太郎 *1 荒井 ひろみ *2 佐藤 一誠 *2 中川 裕志 *2
Kentaro Minami Hiromi Arai Issei Sato Hiroshi Nakagawa

*1東京大学大学院 情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

*2東京大学情報基盤センター
Information Technology Center, The University of Tokyo

In many fields of statistical data analysis, inter-organizational information sharing has been important problem. For example, in the field of clinical data, only a few "case reports" tend to be available in a single medical institution (i.e. the entire data set is distributed over many organizations). Hence data sharing between institutions is quite important issue for accurate data analysis. However, since clinical data often consist of personal information, we have to take into account the privacy protection. In this paper, we consider the linear regression problem for distributed data sets. We propose a method to aggregate differentially private ridge regression estimators, which are made from the data of each organization.

1. はじめに

データ解析の高精度化にあたり，複数の組織間にまたがる情報共有の問題は重要である．特に，医療データ解析の分野では，特定の症例に関する解析を行いたい場合，単一の病院または研究機関内で取得可能なデータ数が少ないという状況が考えられ，複数組織で情報が共有できることの意義は大きい．しかし，そのようなデータは患者の年齢や性別などに加え，既往歴やゲノム情報といった非常に機微な個人情報を含んでいることが多く，組織間での直接的な情報共有は患者のプライバシ侵害に繋りうる．一般に，個人情報を含むデータを利用した解析には，このようなプライバシの問題が常に伴う．

本稿ではデータセットが複数組織にわたり分散している場合の線形回帰について論じる．線形回帰は統計データ解析としては古典的なものであるが，医療データにおいては薬効の解析や投薬量の見積もりに利用されるなど，その位置づけは依然として重要である．しかし，複数組織の情報を統合して学習を行う場合，データそのものを直接的に共有することはできない．そこで，情報交換に関するプライバシ保護基準として，差分プライバシ [Dwork 06] をみたす値の共有は許容されるものとする．この仮定のもとで，本稿では，差分プライバシをみたす回帰係数を交換した後に，それらを学習器統合の手法を用いて平均化することにより，プライバシ保護の制約下で線形回帰の問題を扱う手法を提案する．

差分プライバシの基本的なアイデアは，真の統計量に対してノイズを加えることにより，データセットに含まれる個々人のデータの摂動については出力分布がロバストになることを保証するというものである．一方，医療データ解析の文脈では，解析結果が将来の患者の投薬量の決定など非常に重要な判断に影響しうるため，データ解析プロセスには慎重な精度保証が要求される．そのため先行研究においては，ノイズ加算をもとにした差分プライバシの手法は，その有用性が疑問視されることも多かった．[Fredrikson 14] 本稿で提案する統合手法は，組織間の情報共有によるサンプル数増加の効果によって，差分プライバシ保証のもとでも学習精度を改善できることを期待しており，実問題に対して差分プライバシ手法を活用する場合の新たな方向性を示唆している．

連絡先: kentaro_minami@mist.i.u-tokyo.ac.jp

2. 問題設定

2.1 線形回帰

本稿ではランダムデザインの線形回帰問題を扱う．データセット $D_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ は $\mathcal{X} \times \mathbb{R}$ 上の独立なサンプルで，線形回帰モデル

$$y_i = \beta^\top x_i + w_i,\ i = 1, \ldots, n \tag{1}$$

の実現値であるとする．ここで，$w_1, \ldots, w_n$ は説明変数 $x_1, \ldots, x_n$ とは独立なノイズ項である．このとき，与えられたデータ D_n を用いて回帰係数 β を求めたい．

損失関数としては二乗誤差 $\ell(\hat{\beta}, (x, y)) = |y - \hat{\beta}^\top x|^2$ を用いるとし，平均二乗誤差（リスク）$\mathcal{R}(\hat{\beta}) = \mathbb{E}_{(X,Y)}[|y - \hat{\beta}^\top x|^2]$ をなるべく小さくする推定量を構成したい．通常用いられる推定量 $\hat{\beta}_\lambda$ は次のような形をしている．

$$\hat{\beta}_\lambda = \underset{\theta \in \Theta}{\operatorname{argmin}} \left\{ \hat{\mathbb{E}}_n |y_i - \theta^\top x_i|^2 + \lambda \parallel \theta \parallel_2^2 \right\} \tag{2}$$

ただし $\hat{\mathbb{E}}_n[f(x, y)] = n^{-1} \sum_{i=1}^n f(x_i, y_i)$ は経験分布による期待値を表す．この推定量は，$\lambda = 0$ のとき最小二乗 (ordinary least square, OLS) 推定量，$\lambda > 0$ のときリッジ回帰と呼ばれる．OLS は線形回帰における古典的な推定量であり，ノイズ項 ε_i が分散既知の正規分布である場合の最尤推定量に相当する．OLS は線形方程式の解として得られるが，$x_1, \ldots, x_n$ の与えられ方によっては不良設定となり得るため，正則化によってそれを回避したものがリッジ回帰推定量である．$\hat{\Sigma} = \hat{\mathbb{E}}_n[xx^\top]$ とすると，$\hat{\beta}_\lambda$ は

$$\hat{\beta}_\lambda = (\hat{\Sigma} + \lambda I)^{-1} \hat{\mathbb{E}}_n[yx] \tag{3}$$

と書かれる．ここで右辺の逆行列はリッジ回帰 ($\lambda > 0$) に対しては常に定義される．以下では，主にリッジ回帰について扱うものとする．

2.2 分散データとプライバシの問題

本稿ではさらに，全データセット D_n が $M+1$ 個の組織にわたって分散している場合を考える．$m = 0, 1, \ldots, M$ はそれぞれの組織を表す添字とする．各組織 m はデータセット D^m を持っており，全データセットは $D_n = \bigcup_{i=0}^{M} D^m$ と表される．(ただし $D^m \cap D^{m'} = \emptyset$, $m \neq m'$ とする.)

D^m に属するデータは比較的少なく，組織 m は十分な精度で推定を行うことができない．一方，組織間でデータを共有することができたと仮定すると，全てのデータを集めた D_n は学習に十分なサンプル数になるという状況を考える．例えば，前節で導入した問題は「ランダムデザイン」の回帰であり，データセットとして与えられていない未知の x に対しても y の予測値を返すことを想定している．これは実問題への応用を考えると自然な設定である．しかし，もし自組織のデータにおいて，x の定義域に対して「狭い」領域内の教師データしか含まれていないとすると，その外の領域に対する y をうまく予測できる推定量を作ることは難しいと予想される．このとき，他組織のデータを借用することによって，x の定義域内に教師データがまんべんなく存在するようにできれば，線形回帰の精度も向上すると考えられる．

さて，ここで実問題におけるデータの意味を考慮し，各 (x_i, y_i) が個人情報を含んでいるとする．例えば，薬効 y の推定タスクにおいて，説明変数 x が患者の年齢，性別，血圧などといった情報を加工・整形して作られた特徴量である場合がこれに相当する．

近年，そのような「特徴量」であっても，個人特定に利用しうるものは保護されるべき個人情報であるとみなされることが多く，ある組織から別の組織に本人の同意なしに提供することはできない．したがって，上記のようにデータが複数組織に分散しているような状況では，データセットそのものを直接的に共有することはできない．ある組織 m から第三者組織 m' に提供可能であるのはデータセット D^m そのものではなく，そこから抽出された何らかの統計量のようなものに限る．さらに，それらの情報が既往歴や身体的特徴といった，差別や偏見に繋りうる情報（要配慮個人情報）を含む場合には，より注意深いデータの取り扱いが要求されると考えるのが自然である．例えば，統計量を第三者と共有する際に，共有した統計量と外部情報とを突き合わせることによりデータセット D^m が逆推定されるというような，二次的なリスクも考慮した何らかのプライバシ保護制約を満たしていることが望ましい．

そこで本稿では，次節で導入する差分プライバシを満たしていれば組織間で情報を交換してもよいと仮定する．したがって，このプライバシ制約のもとで，学習に必要な情報のみをいかに効率的に共有するかが主な問題となる．

2.3 差分プライバシ

個人情報 z_i の組からなるデータセット $D_n = \{z_1, \ldots, z_n\}$ より抽出した統計量や，学習器などを外部に公開するとき，D_n に含まれる任意の個人 z_i に関する情報の漏洩を小さくしたい．もし，ある個人 z_i を除いた全てのデータ $D_n \setminus \{z_i\}$ を知っている「十分強い攻撃者」を仮定したとしても，公開された値から残りの z_i の値を逆に推定できなければ，プライバシが保護されていると考えることにする．

差分プライバシの基本的なアイデアは，公開する統計量にノイズを加えて撹乱することにより，1 個人のデータのみで異なる 2 つのデータセットからの出力を確率分布の意味で区別できないようにすることである．データセットに D_n に付随して決まる確率変数 $\mathcal{M}(D_n)$ があるとする．$\mathcal{M}(D_n)$ は興味のある統計量にノイズを加えたものに相当する．2 つのデータセット D_n, D'_n が隣接しているとは，ある 1 要素のみ異なり，残りの要素は全て等しくなることを言い，$D_n \sim D'_n$ と書く．

定義 1 (差分プライバシ)．$\varepsilon > 0$, $\delta \geq 0$ とする．データセット D_n に確率変数 $\mathcal{M}(D_n) : \Omega \to A$ を対応させるアルゴリズム $\mathcal{M}$ が (ε, δ)-差分プライバシを満たすとは，任意の隣接するデータセット $D \sim D'$ に対し，値域における任意の可測集合 $S \subset A$ について

$$\Pr(\mathcal{M}(D) \in S) \leq e^{\varepsilon} \Pr(\mathcal{M}(D') \in S) + \delta \qquad (4)$$

が成り立つことをいう．特に，$\delta = 0$ のとき $\mathcal{M}$ は ε-差分プライバシを満たすという．

$\varepsilon > 0$, $\delta \geq 0$ はプライバシ強度を決定するパラメータであり，これらが小さければ隣接する任意のデータセット D, D' に対する出力の確率分布が (4) の意味でより近くなることが要請される．したがって，例えば $\delta = 0$ のとき，ε が小さいほど本来興味のある統計量に大きなノイズを加え，結果を大幅に撹乱しなければならない．

式 (4) はある意味で確率分布間の擬距離を定め，差分プライバシによる保護は，統計的な手法による D と D' の識別不可能性を保証することにおおよそ相当する．このことの正当化としては，差分プライバシ保証によって D と D' を判別する仮説検定の検出力の上界が与えられることなどが知られている．[Wasserman 10, Hall 13]

3. プライベート線形回帰推定量の統合

前節での問題意識をまとめると次のようになる．

- 各組織 m $(m = 0, 1, \ldots, M)$ はそれぞれ線形回帰モデル (1) の β を推定する．
- 推定量の精度が D^m のみでは不十分であるため，情報共有により他組織を含む全データ D_n を利用したい．
- 組織間での情報共有は，(ε, δ)-差分プライバシを満たす統計量の交換に限り許容される．

組織間でデータを直接共有することが許されていないため，上の設定のもとで効率よく学習を行うためには，(ε, δ)-差分プライバシ制約のもとで D^m に含まれる情報のうち学習に必要な成分をなるべく抽出した統計量を共有すること，ならびに，他組織から受け取った統計量の組から効率的に推定量を復元することが必要である．本稿では，これに対する自然なアプローチとして，(ε, δ)-差分プライバシを満たす線形回帰推定量そのものを共有し，学習器統合の手法を用いてそれらを統合する手法を提案する．

3.1 提案手法

図 1 は本稿で提案する統合手法における情報の流れを表した概念図である．各組織 $m = 0, \ldots, M$ はまず，自分のデータ D^m に関して (ε, δ)-差分プライバシを満たすように線形回帰推定量 $\hat{\beta}_m$ を構成し，互いに利用できるように交換する．次に，組織 $m_0 = 0$ は，受け取った M 個の推定量 $\{\hat{\beta}_m\}_{m=1}^{M}$ を弱学習器とみなし，自分のデータ D^0 を用いて統合する．差分プライベート線形回帰推定量の作り方は 3.1.1 節，それらを統合するアルゴリズムについては 3.1.2 節でそれぞれ説明する．

3.1.1 差分プライベート弱学習器の構成

本節では差分プライバシを満たす線形回帰推定量の構成について説明する．リッジ回帰推定量 (2) は目的関数

$$\mathcal{J}(\hat{\beta}; D_n) = \hat{\mathbb{E}}_n |y_i - \hat{\beta}^{\top} x_i|^2 + \lambda \parallel \theta \parallel^2 \qquad (5)$$

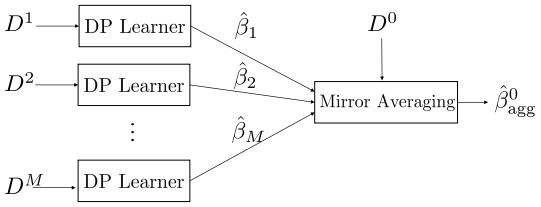

図 1　統合学習器の概略．各組織 $m = 0, 1, \ldots, M$ は自組織のデータ D^m から差分プライベート線形回帰推定量 $\hat{\beta}_m$ を構成し，互いに交換する．組織 0 はデータ D^0 を用いて $\{\hat{\beta}_m\}_{m=1}^M$ を統合する．

の最小化によって定義されている．したがって，差分プライベート線形回帰推定量は，差分プライバシ制約を満たしつつ目的関数 $\mathcal{J}$ の真の最小値になるべく近い値を達成するもの，として特徴づけられる．言い換えれば，データセット D_n が与えられたとき，(ε, δ)-差分プライバシを満たす確率変数 $\hat{\beta}_{\mathrm{priv}}$ で，プライバシリスク

$$\mathbb{E}_{\hat{\beta}_{\mathrm{priv}}}[\mathcal{J}(\hat{\beta}_{\mathrm{priv}}; D_n)] - \inf_{\theta;\Theta} \mathcal{J}(\theta; D_n) \tag{6}$$

を小さくするものを考えたい．この形式の問題は**差分プライベート経験リスク最小化**と呼ばれ，近年盛んに研究されている．[Chaudhuri 11, Kifer 12, Bassily 14, Talwar 14] また，ノイズに正規分布を仮定したパラメトリックな設定における線形回帰では，パラメータの事後分布からのサンプリングが自動的に差分プライバシを満たすという事実を利用した効率的な手法も提案されている．[Wang 15]

本稿では目的関数摂動法 [Chaudhuri 11, Kifer 12] を用いる．以下では $\delta = 0$ とし，ε-差分プライバシを満たす推定量のサンプリング手法について説明する．一般に，凸損失関数 $\ell(\cdot; (x, y))$ に関する正則化つき経験リスク最小化を考える．任意の (x, y) について $\zeta > 0$ が存在して $\| \nabla \ell(\hat{\beta}; (x, y)) \|_2 \leq \zeta$ を満たすとする．また，ヘッシアン $\nabla^2 \ell(\hat{\beta}, (x, y))$ の固有値は上界 Λ を持つとする．リッジ回帰 (2) の場合には，$x, y, \hat{\beta}$ の定義域が有界であることを仮定すればどちらの条件も満たす．このとき，$b \in \mathbb{R}^p$ を密度関数 $\nu(b; \varepsilon, \zeta) \propto \exp(-\varepsilon \| b \|_2 / 2\zeta)$ をもつ確率分布からサンプルし，新たな目的関数

$$\mathcal{J}(\hat{\beta}; D_n) + \frac{\Lambda}{\varepsilon n} \| \hat{\beta} \|_2^2 + \frac{1}{n} b^\top \hat{\beta} \tag{7}$$

を最小化する θ を推定量として公開する．よってとくに，差分プライベートリッジ回帰推定量は

$$\hat{\beta}_{\mathrm{priv}} = \left(\hat{\Sigma} + (\lambda + \Lambda/\varepsilon n) I \right)^{-1} (\hat{\mathbb{E}}_n[yx] + b/2n) \tag{8}$$

で与えられる．

3.1.2　弱学習器の統合

データ D^0 と M 個の推定量 $\{\hat{\beta}_m\}_{m=1}^M$ が与えられたとき，凸結合 $\hat{\beta}_{\mathrm{agg}} = \sum_{m=1}^M a_m \hat{\beta}_m$ によって新たな推定量を作る問題を学習器統合という．推定量の混合比 $a = (a_1, \ldots, a_M) \in \mathbb{R}_+^M$, $\sum_m a_m = 1$ を決定する手法について考えよう．

D^0 に含まれるサンプル数を $n_0 = |D^0|$ とおく．与えられた推定量の中でリスクの意味で最良のものと比較した不等式

$$\mathbb{E}^{n_0}_{(X,Y)} \mathcal{R}(\hat{\beta}_{\mathrm{agg}}) \leq \min_{1 \leq m \leq M} \mathcal{R}(\hat{\beta}_m) + \Delta(n_0, M) \tag{9}$$

において，残差 $\Delta(n_0, M)$ が小さいものが良い統合手法であると言える．回帰モデル (1) では，残差の下界は $\Delta(n_0, M) \gtrsim \log M / n_0$ であることが知られている．[Tsybakov 03] この最適なレートを達成させる統合手法はいくつか知られており，ミラーアベレージング [Juditsky 08] はその中でも利用しやすいもののひとつである．本節ではこれについて説明する．

ミラーアベレージング推定量は次のように構成される．$\tau > 0$ を温度パラメータとする．このとき，推定量の混合比 a を次のように定める．

$$a_m = \frac{1}{n_0} \sum_{t=1}^{n_0} \frac{\exp(-\tau^{-1} \sum_{i=1}^t |y_i - \hat{\beta}_m^\top x_i|^2)}{\sum_{l=1}^M \exp(-\tau^{-1} \sum_{i=1}^t |y_i - \hat{\beta}_l^\top x_i|^2)}. \tag{10}$$

ここで，$(x_1, y_1), \ldots, (x_{n_0}, y_{n_0})$ は D^0 に属するデータである．(10) はエネルギーが経験リスク $\sum_{i=1}^{n_0} |y_i - \hat{\beta}_l^\top x_i|^2$ であるような Gibbs 分布によって平均を考えたものに近い．このような推定量は Gibbs 推定量，あるいは指数型重み付け統合 (exponential weighting aggregation) などと呼ばれる．

3.2　汎化誤差の評価

統合によって作られた推定量 $\hat{\beta}^0_{\mathrm{agg}}$ の汎化誤差を解析する．直感的には，統合学習器 $\hat{\beta}^0_{\mathrm{agg}}$ の性能は要素ごとの性能，すなわち

1. 各組織 $m \in \{1, \ldots, M\}$ ごとに作られた，個々の差分プライベート推定量 $\hat{\beta}_m$ の汎化誤差をいかに小さくできるか
2. 所与の弱学習器の組 $\{\hat{\beta}_m\}_{m=1}^M$ に対して，データセット D^0 を用いてそれらをいかに上手く統合できるか

に依存すると考えられ，汎化誤差の評価はそれらの組み合わせによって得られる．

β に対する任意の推定量 $\hat{\beta}$ に対して，汎化誤差は

$$\begin{aligned} \mathcal{R}(\hat{\beta}) - \mathcal{R}(\beta) &= \int_{\mathcal{X}} |\hat{\beta}^\top x - \beta^\top x|^2 dP_X \\ &= \| (\hat{\beta} - \beta)^\top x \|^2_{L^2(P_X)} \end{aligned} \tag{11}$$

で与えられることに注意する．

本節では簡単のため，次の仮定をおく．

仮定 2.　1. 説明変数の空間 $\mathcal{X}$ は $\mathbb{R}^p$ の単位球に含まれる閉凸集合とする．また，$\hat{\beta}$ の候補の空間 $\Theta \subset \mathbb{R}^p$ は半径 $B > 0$ の球に含まれる閉凸集合とする．

2. 真の回帰係数 β は Θ に含まれる．したがって特に，任意の $\hat{\beta} \in \Theta$ に対して $\| \hat{\beta} - \beta \| \leq 2B$.

3. 回帰モデル (1) において，ノイズ変数 w_i は sub-Gaussian であるとする．すなわち，$\sigma^2 > 0$ が存在して

$$\mathbb{E}[\exp(t w_i)] \leq \exp(\sigma^2 t^2 / 2), \ \forall t \in \mathbb{R}$$

を満たす．σ^2 は既知とする．

統合推定量の汎化誤差の確率的上界を評価しよう．組織 $m = 1, \ldots, M$ は目的関数摂動 () によって ε-差分プライベート推定量 $\hat{\beta}_m$ を構成する．組織 $m = m_0 = 0$ は与えられた弱学習器 $\{\hat{\beta}_m\}_{m=1}^M$ と自分のデータ D^0 を用いて，ミラーアベレージン

グ推定量 $\hat{\beta}^0_{\mathrm{agg}}$ を構成する．また，L_2-正則化のパラメータ λ に対して，近似誤差関数 $A_2(\lambda)$ を

$$A_2(\lambda) = \inf_{\hat{\beta}\in\Theta} \left\{ \lambda \parallel \hat{\beta} \parallel_2^2 + \parallel (\hat{\beta} - \beta)^\top x \parallel^2_{L^2(P_X)} \right\} \quad (12)$$

で定義する．以上の準備のもとで，次の定理が成立する．

定理 3. $\varepsilon > 0$ をプライバシ強度パラメータとする $0 < \alpha < 1$ とする．ミラーアベレージングの温度パラメータ τ を $\tau \geq 2\sigma^2 + 8B^2$ となるように選び，上記のようにして統合推定量 $\hat{\beta}^0_{\mathrm{agg}}$ を構成する．このとき，n, n_0, ε, M のいずれにも依らない正の定数 C_1, C_2 が存在して，m_0 以外のデータ $D_n \setminus D^0$ の分布に関して，確率 $1-\alpha$ 以上で次の不等式が成立する．

$$\begin{aligned} &\mathbb{E}^{n_0}_{(X,Y)}\mathbb{E}^M_b \parallel (\hat{\beta}^0_{\mathrm{agg}} - \beta)^\top x \parallel^2_{L^2(P_X)} \leq A_2(\lambda) \\ &+\frac{1}{\lambda}\left(\sqrt{\frac{M\log(\frac{2M}{\alpha})}{n-n_0}} + \sqrt{\frac{4M}{n-n_0}} + \frac{C_1 M \log(\frac{2M}{\alpha})}{n-n_0} \right) \\ &+\frac{C_2 M}{\varepsilon(n-n_0)} + \frac{\tau \log M}{n_0}. \end{aligned} \quad (13)$$

ここで，期待値 $\mathbb{E}^{n_0}_{(X,Y)}$ はデータセット D^0 に関してとり，$\mathbb{E}^M_b$ は目的関数摂動法のノイズに関してとる．

証明. 長いので概略のみ述べる．個々の弱学習器 $\hat{\beta}_m$ の汎化誤差のバウンドと，統合手法が満たすオラクル不等式の組み合わせによって示す．まず，リッジ回帰推定量 $\hat{\beta}_\lambda$ を線形カーネル $k(x,x') = x^\top x'$ に対するカーネル推定量だとみなし，カーネル法の一般論から汎化誤差のバウンドを導出する．これには仮定 2-1,2 および [Steinwart 08], Theorem 6.24 を用いれば良い．次に，本来のリッジ回帰推定量 $\hat{\beta}_\lambda$ と差分プライベート推定量 $\hat{\beta}_m$ のリスクの差を，[Kifer 12], Theorem 4 および [Shalev-Shwartz 09], Theorem 2 より計算する．これらの和が個々の $\hat{\beta}_m$ の汎化誤差のバウンドを与える．最後に，仮定 2-3 および [Juditsky 08], Corollary 5.5 によって不等式 (13) を得る． □

不等式 (13) の右辺は大きくわけて 3 つの部分に分かれる．$O(M/\varepsilon(n-n_0))$ の項は差分プライバシを達成するためにノイズを付与したことによる「対価」を表す．$\tau \log M/n_0$ は，手元の n_0 個のデータを用いて，他組織から与えられた推定量の統合をどれだけ上手く行えるかを表している．残りの項は本来のリッジ回帰推定量の収束レートに相当し，目的関数摂動法やミラーアベレージングに関係なく決まる部分である．なお，本稿では簡単のため，この部分の評価には [Steinwart 08], Theorem 6.24 を利用しているが，これはサンプル数に対するオーダーの意味で最適とは限らない．詳しくは [Audibert 11] などを参照されたい．

4. 結論

差分プライバシをみたす線形回帰推定量を互いに交換し，それらをミラーアベレージングによって統合することで分散データからプライバシを保護しつつ効率的に学習を行う手法の提案を行った．汎化誤差の理論バウンドは，個々の差分プライベート推定量および統合手法の性能の組み合わせによって求まる．よって特に，組織間で交換する差分プライベート推定量の性能を上げることによって，分散学習の効果も高まることが期待できる．本稿では主に目的関数摂動について考察を行ったが，種々の差分プライベート推定量に関して理論的・実験的な性能を調べることは今後の課題である．

参考文献

[Audibert 11] Audibert, J.-Y. and Catoni, O.: Robust Linear Least Square Regression, *The Annals of Statistics*, Vol. 39, No. 5, pp. 2766–2794 (2011)

[Bassily 14] Bassily, R., Smith, A., and Thakurta, A.: Differentially Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds, in *IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS)* (2014)

[Chaudhuri 11] Chaudhuri, K., Monteleoni, C., and Sarwate, A.: Differentially private empirical risk minimization, *Journal of Machine Learning Research*, Vol. 12, pp. 1069–1109 (2011)

[Dwork 06] Dwork, C.: Differential privacy, in *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP)*, pp. 1–12 (2006)

[Fredrikson 14] Fredrikson, M., Lantz, E., Jha, S., Lin, S., Page, D., and Rintenpart, T.: Privacy in Pharmocogenetics: An End-to-End Case Study of Personalized Warfalin Dosing, in *23rd USENIX Security Symposium* (2014)

[Hall 13] Hall, R., Rinaldo, A., and Wasserman, L.: Differential privacy for functions and functional data, *Journal of Machine Learning Research*, Vol. 14, pp. 703–727 (2013)

[Juditsky 08] Juditsky, A., Rigollet, P., and Tsybakov, A. B.: Learning by mirror averaging, *The Annals of Statistics*, Vol. 36, No. 5, pp. 2183–2206 (2008)

[Kifer 12] Kifer, D., Smith, A., and Thakurta, A.: Private convex empirical risk minimization and high-dimensional regression, in *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, pp. 25.1–25.40 (2012)

[Shalev-Shwartz 09] Shalev-Shwartz, S., Shamir, O., Srebro, N., and Sridharan, K.: Stochastic Convex Optimization, in *Proceedings of the 22nd Annual Conference on Learning Theory (COLT)* (2009)

[Steinwart 08] Steinwart, I. and Christmann, A.: *Support Vector Machines*, Springer (2008)

[Talwar 14] Talwar, K., Thakurta, A., and Zhang, L.: Private Empirical Risk Minimization Beyond the Worst Case: The Effect of the Constraint Set Geometry (2014), Available at `http://arxiv.org/abs/1411.5417`

[Tsybakov 03] Tsybakov, A. B.: Optimal rates of aggregation, in Schoölkapf, B. and Warmuth, M. eds., *Lecture notes in artificial intelligence: Vol.2777. Computational learning theory and kernel machines*, pp. 303–313 (2003)

[Wang 15] Wang, Y., Fienberg, S., and Smola, A.: Privacy for Free: Posterior Sampling and Stochastic Gradient Monte Carlo (2015), Available at `http://arxiv.org/abs/1502.07645`

[Wasserman 10] Wasserman, L. and Zhou, S.: A statistical framework for differential privacy, *The Journal of The Americal Statistical Association*, Vol. 105, pp. 375–289 (2010)