最小二乗密度比推定における差分プライバシー

高林 裕太 †　　荒井 ひろみ ††　　中川 裕志 †††

† 東京大学　大学院情報理工学系研究科
〒 113–8656 東京都文京区本郷 7–3–1
yuta_takabayashi@mist.i.u-tokyo.ac.jp

†† 東京大学 情報基盤センター
〒 113–8658　東京都文京区弥生 2–11–16
arai@dl.itc.u-tokyo.ac.jp

††† 東京大学 情報基盤センター
〒 113–8658　東京都文京区弥生 2–11–16
nakagawa@dl.itc.u-tokyo.ac.jp

あらまし プライバシー保護データマイニングにおいて，非公開データベースに対する学習を既に公開されているデータベースで近似するというのは，現実的なシナリオである．このシナリオに対して差分プライバシーと呼ばれる保護概念からアプローチした Importance weighting mechanism は，確率的分類法と呼ばれる密度比推定手法に差分プライバシーを適用した手法である．一方，密度比推定自体の手法としては，より計算効率や精度の良い手法として，最小二乗密度比適合法；uLSIF が既に知られている．本稿では，この uLSIF に対して差分プライバシーを適用した手法を提案する．

Differentially Private Least-squares Importance Fitting

Yuta Takabayashi†　　Hiromi Arai††　　Hiroshi Nakagawa†††

†Dep. of Mathematical Informatics, The University of Tokyo
7-3-1 Hongou, Bunkyou-ku, Tokyo, Japan
yuta_takabayashi@mist.i.u-tokyo.ac.jp

††Information Technology Center, The University of Tokyo
2-11-16 Yayoi, Bunkyou-ku, Tokyo, Japan
arai@dl.itc.u-tokyo.ac.jp

†††Information Technology Center, The University of Tokyo
2-11-16 Yayoi, Bunkyou-ku, Tokyo, Japan
nakagawa@dl.itc.u-tokyo.ac.jp

Abstract In the privacy-preserving data mining, it is a possible scenario to approximate learning on a private database with a published database. The importance weighting mechanism, which approach the scenario with the privacy notion called Differential Privacy, is a differentially private extension of the importance estimation method called the Probabilistic Classification method. Meanwhile, another importance estimation method, the Least-squares Importance Fitting method called uLSIF, is already known as a more computationally efficient and accurate method. This paper proposes a differentially private uLSIF.

1 はじめに

プライバシー保護データマイニングにおいては，ある非公開なデータベースを，そこに含まれる個人情報のプライバシーは保護しつつも，なんらかの情報処理を行いたいという要請がある．特に，施される情報処理を特定せず，汎用的にデータベース自体を加工する形で出力する技術は，Privacy-Preserving Data Publishing（PPDP）と呼ばれ研究される分野である．

その中でも，特に差分プライバシー [3]（Differential Privacy，DP）と呼ばれる保護指標を満たす手法としては，単純に全てのデータに雑音を加える Laplace Mechanism [4] などがあるが，加える雑音量が非常に多くなってしまい，有用性の面で問題がある．これに対し，非公開データベースそのものではなく，既に公開されている別のデータベースを用いて，本来の非公開データベースへの処理出力を近似するという Importance weighting mechanism [7] が提案されている．これは本質的にはこの２つのデータベース間の密度比推定を DP 的に行う手法であり，PPDP 手法の中ではかなり良い有用性を示すことから，本論文ではこの枠組に着目する．

また，今紹介したように，差分プライバシーのシナリオで密度比推定を考えることは，PPDP 手法としてそのまま有用性のあるものであると同時に，そもそも密度比推定自体にも共変量シフト等様々な活用が考えられている [12].

このように密度比推定を差分プライバシーを満たして上で行うことは様々な活用の場面が考えられ，またこの密度比推定自体にも既に確率的分類法，積率適合法，密度比適合法などの様々な手法が提案されている．例えば確率的分類法とは，密度比を推定したい二つの標本を，分類問題として確率的なモデルに当てはめるものである．特に有名なものとしてロジスティック回帰モデルを用いるもの [1] が知られており，[7] でもこれと同様な手法を用いてプライバシーを考慮した密度比推定を行っていた．

しかし，近年ではより統計的に良い性質を持ち，高速な計算が可能な手法として最小二乗密度比適合法（LSIF，uLSIF）[8] が提案されている．特に uLSIF は，パラメータの非負制約をなくすことで，解が解析的に求まると同時に，モデル選択としての交差確認法すらも解析的に計算できることから，非常に高速な推定を可能にしており，密度比推定手法のスタンダードとなっている．

表 1 に示すように，これまで差分プライバシーを考慮した密度比推定手法としては，確率的分類法に対する Importance weighting mechanism [7] 程度しか知られていない．これを踏まえ本論文では，特に先ほど述べた最小二乗適合法（uLSIF [8]）を差分プライバシー的な文脈で行う手法について提案し，有用性や雑音量を分析する．

表 1: 種々の密度比推定法

通常の手法	DP 的手法
確率的分類法（LogReg [1]）	[7]
最小二乗適合法（uLSIF [8]）	本論文

本論文の構成としては以下のようになっている．まず２節で説明に必要な概念を定義・紹介する．具体的には，2.1 節で保護基準としての差分プライバシー [3] を定義する．2.2 節では差分プライバシーを用いた重み付けマイニングの枠組み [7] を紹介し，2.3 節で今回主に用いる密度比推定手法である uLSIF [8] を説明する．3 節では，実際に差分プライベートな uLSIF を提案する．4 節では，プライバシー保護の際加える雑音量や有用性の分析を行う．5 節では，実際の数値実験により有用性の評価を行う．最後に 6 節で本論文をまとめ，また今後の課題について述べる．

2 準備

2.1 差分プライバシー

本節では，プライバシー保護基準・尺度としての差分プライバシー（Differential Privacy，DP）の概念を説明する．まずデータベース全体の集合を $\mathcal{D}$ と表し，それに対しある情報処理を行うアルゴリズム（クエリ）$f : \mathcal{D} \to \mathcal{A}$ が投げられるとする．また，$\varepsilon > 0$ をプライバシー保護の強度を表すパラメータとする．これ

は基本的に ε が小さいほど保護強度が高いことを表す．差分プライバシーは基本的に，データベース同士が一人，すなわち 1 レコードのみの差分を含む（隣接する）場合の保護を考えるものである．そのようなデータベースの例をここで $D, D' \in \mathcal{D}$ とする．これらに対し，差分プライバシーを以下のように定義する．

定義 1. あるアルゴリズム f が ε-差分プライバシーを満たすとは，任意の隣接するデータベース $D, D' \in \mathcal{D}$ について，アルゴリズム f による任意の出力の集合 $A \subset \mathcal{A}$ に対して以下の式が成り立つことである．

$$\Pr[f(D) \in A] \leq e^{\varepsilon} \times \Pr[f(D') \in A] \quad (1)$$

この式はつまり，出てきた回答がデータベース D によるものなのか，データベース D' によるものなのかが確率的に判別できないということを意味している．つまりその差分にいるような個人の情報に対しても保護されているということになる．

更にこの ε-差分プライバシーを，実際にある関数 f に対して満たす際の最も単純な手法として，出力摂動（Laplacian Mechanism とも）[4] を紹介する．まず，クエリ f に対して定義される sensitivity 値という量を導入する．

定義 2. 隣接するような任意のデータベース $D, D' \in \mathcal{D}$ に対し，クエリ f の（L1-）sensitivity $S(f)$ とは

$$S(f) = \sup_{D, D' \in \mathcal{D}} ||f(D) - f(D')||_1. \quad (2)$$

これは直感的には，隣接するようなデータベース同士に対するクエリの返答の最大差を意味する．このような sensitivity 値を定義した上で，出力摂動とは以下のような定理で表される．

定理 1. 元の関数 f の出力に，その f の sensitivity $S(f)$ を用いてパラメータ $\lambda = S(f)/\varepsilon$ としたラプラスノイズ

$$Lap(\lambda) = \frac{1}{2\lambda} \exp\left(-\frac{|t|}{\lambda}\right) \quad (3)$$

を加えたものは ε-差分プライバシーを満たす．ただし，多次元に対する（L1-）ラプラスノイズとは，確率密度が

$$p(\boldsymbol{t}) \propto \exp(-\frac{||\boldsymbol{t}||_1}{\lambda}) \quad (4)$$

となるように生成してきた $\boldsymbol{t}$ のことを指すとする．結果的にはこれは各次元について i.i.d. に $Lap(\lambda)$ を加えることと等価である

2.2 Importance weighting mechanism

本節では，前節の差分プライバシーの概念を用いた Data publishing 手法の一つである，Importance weighting mechanism [7] の考え方を紹介する．

まずこの手法の前提として目指していることは，非公開データセット D についての，あるクエリ f の結果 $f(D)$ を得ることである．ここで，クエリ $f : \mathcal{X} \to \mathbb{R}^d$ はまずレコードに対する関数として定義され，これをデータセットに適用した場合，すなわち得たい結果はデータセット D についての平均

$$f(D) := E_D[f(\boldsymbol{x})] = \sum_{j=1}^{n} f(\boldsymbol{x}_j^D) \quad (5)$$

を返すものとする．

ここで，非公開データセット D についてプライバシーを保護するために，それと似たような別の公開済みデータセット E を用いて，先の D に対する結果 $f(D)$ を近似したいとする．更にそれを各レコード $\boldsymbol{x} \in \mathcal{X}$ に対して重み付け $w(\boldsymbol{x})$ をすることによって実現することを考える．つまり，

$$E_D[f(\boldsymbol{x})] = E_E[f(\boldsymbol{x})w(\boldsymbol{x})] \quad (6)$$

となる重み関数 $w(\boldsymbol{x})$ を求めればよい．

ここで各データセットに対し連続な確率密度 $p_D(\boldsymbol{x}), p_E(\boldsymbol{x})$ の存在を仮定すると，

$$\begin{aligned} E_D[f(\boldsymbol{x})] &= \int_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) p_D(\boldsymbol{x}) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) w(\boldsymbol{x}) p_E(\boldsymbol{x}) d\boldsymbol{x} \\ &= E_E[f(\boldsymbol{x})w(\boldsymbol{x})] \end{aligned} \quad (7)$$

となり，結局求めたかった重み関数 $w(\boldsymbol{x})$ というのは密度比

$$w(\boldsymbol{x}) = \frac{p_D(\boldsymbol{x})}{p_E(\boldsymbol{x})} \tag{8}$$

であることがわかる．

そこで，この重み関数 $w(\boldsymbol{x})$ をデータセット D から得られる情報として，データセット D に対する差分プライバシーを満たすようにリリースするという枠組みが，この Importance weighting mechanism である．

2.3 uLSIF

本節では密度比推定手法の中でも最小二乗密度比適合法の一つである uLSIF [8] について，プライバシーを考慮しない通常の場合の手順と導出を紹介する．まず，与えられてるものは非公開データベース $D = \{\boldsymbol{x}_j^D\}_{j=1}^n$ と公開データベース $E = \{\boldsymbol{x}_i^E\}_{i=1}^N$ であり，推定したいものは重み $w(\boldsymbol{x}) = \frac{p_D(\boldsymbol{x})}{p_E(\boldsymbol{x})}$（特に $\boldsymbol{x} \in E$ について）である．

始めに，推定する重み関数に線形モデル

$$\hat{w}(\boldsymbol{x}) = \sum_{l=1}^{b} \alpha_l \phi_l(\boldsymbol{x}) \quad (\boldsymbol{\alpha}, \boldsymbol{\phi} \geq \boldsymbol{0}) \tag{9}$$

を仮定する．もしくはカーネルモデルを用いて，

$$\hat{w}(\boldsymbol{x}) = \sum_{l=1}^{b} \alpha_l K(\boldsymbol{x}, \boldsymbol{c}_l). \tag{10}$$

$$(\boldsymbol{K}(\boldsymbol{x}) = \boldsymbol{K}(\boldsymbol{x}, \{\boldsymbol{c}_l\})$$
$$= (K(\boldsymbol{x}, \boldsymbol{c}_1), K(\boldsymbol{x}, \boldsymbol{c}_2), \cdots, K(\boldsymbol{x}, \boldsymbol{c}_b))^T)$$

ちなみに，カーネルとしてはガウスカーネル

$$K_\sigma(\boldsymbol{x}, \boldsymbol{c}_l) = \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{c}_l||^2}{2\sigma^2}\right) \tag{11}$$

を，カーネルの中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ としては D から b 個ランダムサンプリングしたものを用いると良いことがヒューリスティックに知られており [8]，本論文でもこれを最も一般的な形とみなす．

このとき，基本的な発想としては，訓練データセット E についての二乗損失，すなわち推定される重み $\hat{w}(\boldsymbol{x})$ と真の重み $w(\boldsymbol{x})$ との二乗誤差

$$J_0(\boldsymbol{\alpha}) = \frac{1}{2}\int (\hat{w}(\boldsymbol{x}) - w(\boldsymbol{x}))^2 p_E(\boldsymbol{x}) d\boldsymbol{x} \tag{12}$$

を最小化するようなパラメータ $\boldsymbol{\alpha}$ を選ぶというものが最小二乗密度比適合法である．

ここでこの二乗誤差を分解すると

$$J_0(\boldsymbol{\alpha}) = \frac{1}{2}\int (\hat{w}(\boldsymbol{x}) - w(\boldsymbol{x}))^2 p_E(\boldsymbol{x}) d\boldsymbol{x}$$
$$= \frac{1}{2}\int (\hat{w}(\boldsymbol{x}))^2 p_E(\boldsymbol{x}) d\boldsymbol{x} - \int \hat{w}(\boldsymbol{x}) p_D(\boldsymbol{x}) d\boldsymbol{x}$$
$$+ \frac{1}{2}\int (w(\boldsymbol{x}))^2 p_E(\boldsymbol{x}) d\boldsymbol{x} \tag{13}$$

となる．第 3 項は $\boldsymbol{\alpha}$ に関して定数であることに注意し，第 1,2 項のみに線形モデル $\hat{w}(\boldsymbol{x}) = \sum_{l=1}^b \alpha_l \phi_l(\boldsymbol{x})$ を入れ，標本平均で近似すると，$b \times b$ 行列 $\hat{H}$ と b 次元ベクトル $\hat{\boldsymbol{h}}$

$$\hat{H} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\phi}(\boldsymbol{x}_i^E)^T \boldsymbol{\phi}(\boldsymbol{x}_i^E), \tag{14}$$

$$\hat{\boldsymbol{h}} = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{\phi}(\boldsymbol{x}_j^D) \tag{15}$$

を用いて，

$$\hat{J}(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^T \hat{H} \boldsymbol{\alpha} - \hat{\boldsymbol{h}}^T \boldsymbol{\alpha} \tag{16}$$

と表される目的関数が得られる．

ここで，本来は確率の非負性より得られるパラメータ $\boldsymbol{\alpha}$ の非負制約 $\boldsymbol{\alpha} \geq 0$ を外し，L2-正則化項 $\frac{\lambda}{2}||\boldsymbol{\alpha}||_2^2$ を加えた最適化問題

$$\min_{\tilde{\boldsymbol{\alpha}}} \left[\frac{1}{2}\boldsymbol{\alpha}^T \hat{H} \boldsymbol{\alpha} - \hat{\boldsymbol{h}}^T \boldsymbol{\alpha} + \frac{\lambda}{2}||\boldsymbol{\alpha}||_2^2\right] \tag{17}$$

を考えると，これは解析的に求められ，

$$\tilde{\boldsymbol{\alpha}} = (\hat{H} + \lambda I)^{-1} \hat{\boldsymbol{h}} \tag{18}$$

となる．その後に事後補正

$$\hat{\boldsymbol{\alpha}} = \max(\boldsymbol{0}, \tilde{\boldsymbol{\alpha}}) \tag{19}$$

によって非負条件を満たし，最終的にそれらのパラメータを用いて重み

$$\hat{w}(\boldsymbol{x}) = \hat{\boldsymbol{\alpha}}^T \boldsymbol{\phi}(\boldsymbol{x}) \quad (\boldsymbol{x} \in E) \tag{20}$$

を出力するという一連の手法を Unconstrained Least-squares Importance Fitting（uLSIF）と呼ぶ．

uLSIF は解やモデル選択（λ, σ 等）の交差確認が解析的な計算で求まり非常に高速な上，パラメータ λ, σ 等を適切に定めれば，非負制約を外したことによる影響も小さくなり，十分な精度が出ることが知られている．

3　差分プライベートな uLSIF

本節では，本論文のメインである，差分プライバシーの要件を満たすような uLSIF の構成を行う．

目標は，出力 $\hat{w}(\boldsymbol{x}) = \hat{\boldsymbol{\alpha}}^T \boldsymbol{\phi}(\boldsymbol{x})$（定義から $\boldsymbol{x} \in E$ のみを考えれば良い）について，非公開データベース D に対する差分プライバシーを満たすことである．出力摂動 [4] を用いるのが最も単純な方法であるが，直接 $\hat{w}(\boldsymbol{x})$ の sensitivity を計算するのは困難であるため結合定理 [5]，DP を満たす値同士の計算結果は DP，を用いると，このときに必要なことはパラメータ $\hat{\boldsymbol{\alpha}}$ と基底関数 $\boldsymbol{\phi}(\boldsymbol{x})$ それぞれについて差分プライバシーを満たすことである．

基底関数 $\boldsymbol{\phi}(\boldsymbol{x})$ の DP 要件については通常のパラメトリックモデルの場合は考える必要が無いが，カーネルモデルを用いる場合，特に中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ として D からサンプルした点等を用いる場合は，カーネル関数 $K(\boldsymbol{x}, \boldsymbol{c}_l)$ 自体が D に依存するため，sensitivity を計算する等，注意深い対処が必要である．よって，以下では，まず単純な例としてパラメトリックモデルの場合の手法を構成し，その次にカーネルモデルを用いる場合に考えられる手法を紹介する．

3.1　パラメトリックモデルの場合

まず解析が単純になる場合として，推定される重みが線形パラメトリックモデル $\hat{w}(\boldsymbol{x}) = \sum_{l=1}^b \alpha_l \phi_l(\boldsymbol{x})$ で表される場合を考える．つまり，基底関数 $\boldsymbol{\phi}$ は D の変化に非依存である．

このとき，DP を満たすべきはパラメータ $\hat{\boldsymbol{\alpha}}$ であるが，その計算過程 $\tilde{\boldsymbol{\alpha}} = (\hat{H} + \lambda I)^{-1} \hat{\boldsymbol{h}}$ において D に依存するのは $\hat{\boldsymbol{h}}$ のみなことから，結合定理 [5]，DP な値について D に非依存な任意の計算を施しても出力は DP を満たす，を用いると $\hat{\boldsymbol{h}}$ について出力摂動に用いる sensitivity 値を求めればよい．

今非公開データベース D から導かれる $\hat{\boldsymbol{h}}$ と，D から 1 レコード $\boldsymbol{x}'$ を除いたデータベース D' から導かれる $\hat{\boldsymbol{h}}'$ の差を考えると，今 $\hat{\boldsymbol{h}}$ が定義として

$$\hat{\boldsymbol{h}} = \frac{1}{n} \sum_{\boldsymbol{x} \in D} \boldsymbol{\phi}(\boldsymbol{x}) \tag{21}$$

となっていることから，L1-sensitivity は

$$\begin{aligned} S(\hat{\boldsymbol{h}}) &= \sup_{D, D'} ||\hat{\boldsymbol{h}} - \hat{\boldsymbol{h}}'||_1 \\ &= \sup_{\boldsymbol{x}'} ||\frac{1}{n} \boldsymbol{\phi}(\boldsymbol{x}')||_1 \\ &= \frac{1}{n} \sup_{\boldsymbol{x}'} ||\boldsymbol{\phi}(\boldsymbol{x}')||_1 \\ &\leq \frac{bC}{n} \quad (C = \sup \phi_l(\boldsymbol{x})) \end{aligned} \tag{22}$$

となる．よって基底関数の上界 $C = \sup \phi_l(\boldsymbol{x})$ が定まる場合は sensitivity が上記のように定まり，これを用いて雑音を加えた $\tilde{\boldsymbol{h}} = \hat{\boldsymbol{h}} + Lap(S(\hat{\boldsymbol{h}})/\varepsilon)$ を用いて通常通りの uLSIF を行えば，結果として出力される $\hat{\boldsymbol{\alpha}}$，さらに $\hat{w}(\boldsymbol{x}) = \hat{\boldsymbol{\alpha}}^T \boldsymbol{\phi}(\boldsymbol{x})$ は ε-差分プライバシーを満たす．

3.2　カーネルモデルを用いる場合

次に，重みのモデルにカーネルモデルを用いる場合 $\hat{w}(\boldsymbol{x}) = \hat{\boldsymbol{\alpha}}^T \boldsymbol{K}(\boldsymbol{x})$ について考察する．前節までと同様の議論により $\hat{\boldsymbol{\alpha}}$ についての DP は保証されるため，この節では基底関数となるカーネル関数 $K(\boldsymbol{x}, \boldsymbol{c}_l)$ についての DP を考える必要がある．更にこの際，できれば中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ として D の点を用いたい

以下では，基底関数としてカーネル関数 $K(\boldsymbol{x}, \boldsymbol{c}_l)$ を使うときの DP 的な対処法として，３つの構成例を紹介する．

手法 A　中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ に雑音を加える

まず最も単純に思いつきかつ最も実用的だと思われるのがこの手法は，中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ を D の分布にある程度沿うように選びながらも，それ自体に DP を満たすように雑音を加えてしまうというものである．その結果差分プライバシーの結合定理 [5] により，その中心を用いたカーネル関数 $K(\boldsymbol{x}, \boldsymbol{c}_l)$ も DP が満たされることになる．

ここでは最も単純に，D から b 点ランダムサンプルした後，DP を満たすよう出力摂動するという方法を考える．このとき，DP のサンプリング定理 [10] を用いることによって，元々ε-DP を満たしたい場合でも，D からのサンプル率を $p = b/n$ として，

$$\varepsilon' = \log\left(1 + \frac{1}{p}(e^{\varepsilon} - 1)\right) \tag{23}$$

となる ε'-DP を満たすように雑音を加えればよく，モデルの次元数 b を減らすほど，中心 $\{\boldsymbol{c}_l\}_{l=1}^{b}$ に対する雑音量も減らすことができるのが特徴である．

手法 B　中心 $\{\boldsymbol{c}_l\}_{l=1}^{b}$ を D に非依存にする

本来の最終目標は中心 $\{\boldsymbol{c}_l\}_{l=1}^{b}$ として D から b 個ランダムサンプリングしたものを用いることであるが，ここで中心にそもそも D に関わらない点を用いるという手法も考えられる．この場合は基底関数は形はカーネル関数ではあるものの，本質的にただのパラメトリックモデルの場合と同様であり $\boldsymbol{K}(\boldsymbol{x})$ については既に DP が満たされているため，3.1 節と同様に $\hat{\boldsymbol{\alpha}}$ に雑音を加えるのみで良い．

ちなみに D に非依存な中心 $\{\boldsymbol{c}_l\}_{l=1}^{b}$ の取り方としては，公開データベース E の点を使うのが良いと思われる．ただし，やはり D の点を用いるよりは精度は下がると予想される．

手法 C　(ε, δ)-差分プライバシーに緩和

最後に紹介するのは，今まで議論してきたプライバシー基準である ε-差分プライバシーを緩和してしまい，代わりに (ε, δ)-差分プライバシーという基準を用いるものである．ここでこの (ε, δ)-差分プライバシーの基準は以下のようになっている．

定義 3. あるアルゴリズム f が (ε, δ)-差分プライバシーを満たすとは，任意の隣接するデータベース $D, D' \in \mathcal{D}$，任意の出力の集合 $A \subset \mathcal{A}$ に対して以下の式が成り立つことである．

$$\Pr[f(D) \in A] \leq e^{\varepsilon} \times \Pr[f(D') \in A] + \delta \tag{24}$$

この定義の式の最後に δ が追加されている部分が緩和部分である．カーネルモデルを用い中心として D から b 点サンプルする場合を考えると，サンプル率 $p = b/n$ として確率 p 以外はパラメトリックモデルと同様，D の影響を受けていないと見なすことができるため，3.1 節と同様に $\hat{\boldsymbol{\alpha}}$ に雑音を加えるのみで (ε, p)-差分プライバシーを満たす事ができる．この手法は特に大きいデータセットを使っており，p が非常に小さい（$b \ll n$）場合に効果的といえる．

4　雑音量の分析

前節で具体的に構成した DP 的 uLSIF について，加えた雑音量についての分析を行う．

パラメトリックモデル，もしくは手法 C を用いるとき，出力 $\hat{w}(\boldsymbol{x}) = \hat{\boldsymbol{\alpha}}^T \boldsymbol{\phi}(\boldsymbol{x})$ に結果的に影響している雑音量，つまりパラメータ $\hat{\boldsymbol{\alpha}}$ に影響している雑音量は $O(b/\varepsilon\lambda n)$ である．正則化パラメータ λ を大きくすると雑音量は減るが，最適解における $\boldsymbol{\alpha}$ の範囲が最大値側から狭まり，バイアスがかかる．これは $\boldsymbol{\alpha}$ が平均化すること，つまり基底関数列に対して一様になることにつながる．また，D のデータ数 n を増やせば雑音量は減る．プライバシーパラメータ ε を大きくすれば，すなわち必要な保護を弱めれば，当然ながら雑音量は減る．重みの線形モデルの次元 b を減らせば雑音量は減るが，モデルの表現力は落ちる．

5　実験

この節では実際に単純な人工データについて，数値計算を行った例を紹介する．

今回は，レコードは 1 次元の実数であり（$\boldsymbol{x} \in \mathbb{R}$），公開データベース E は平均 1，分散 1 の正規分布，非公開データベース D は平均 2，分散 1/4 の正規分布に従うとする．ここでそれぞれ E は $N = 50$ 点，D は $n = 2000$ 点ランダムに生成したインスタンスに対し，今まで紹介してきた，元の uLSIF，カーネルモデルを用いる場合の手法 3 つを用いて実際に密度比の推定を行っていく．なお，実行の際必要になるパラ

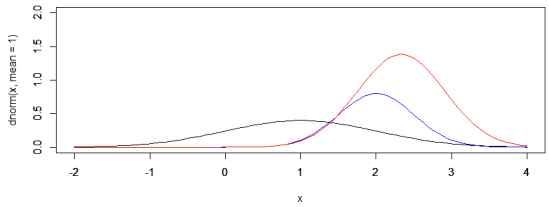

図 1: D, E の確率密度 $p_D(x)$, $p_E(x)$ と，理論的な密度比 $w(x) = p_D(x)/p_E(x)$

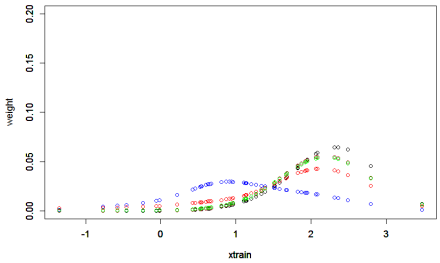

図 3: 中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ に E の点を用いる手法 B

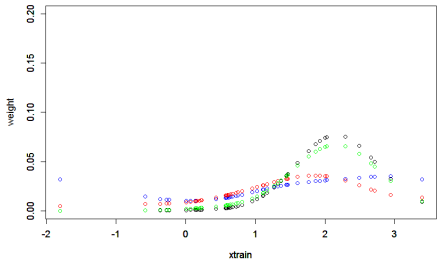

図 2: 中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ に雑音を加える手法 A

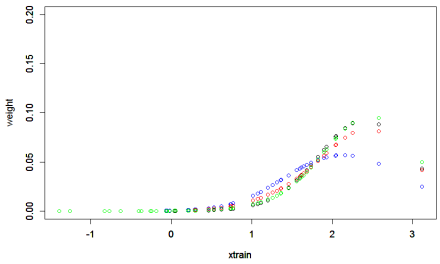

図 4: (ε, δ)-差分プライバシーに緩和する手法 C

メータ λ,σ は元の uLSIF で 1 つ抜き交差確認法を行い，最良となったものを選択した．なお，今回この手順については差分プライバシーを考慮していない．

図 1 に D, E の確率密度 $p_D(x)$, $p_E(x)$ と，理論的な密度比 $w(x) = p_D(x)/p_E(x)$ のイメージを示す．今回の計算例では，赤の線で表される密度比 $w(x) = p_D(x)/p_E(x)$ を精度良く求めることが目標となる．

図 2 に示すのは，中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ に雑音を加える手法 A での実行例である．ここで，黒は元の uLSIF でカーネルの中心 D を用いた場合（理想的な結果），青，赤，緑はそれぞれこの手法において保護の強さ ε を 0.1, 1, 10 にした場合である．(以下どの手法も同様) この手法に特徴的なのは，プライバシーの保護を強くすると，推定される重みが強く一様に近づくことである．これは，カーネルの中心が雑音により広く散ってしまうことに起因すると思われる．

図 3 に示すのは，中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ に E の点を用いる手法 B での実行例である．この図にのみ登場する茶色の線は元の uLSIF でカーネルの中心に E を用いた場合であり，比較のために入れた．この手法に特徴的なことは，プライバシーの保護を強くすると，推定される重みの形状が E 寄りになって行くことである．これは中心として D ではなく E の点を用いていることに直接起因する．また，そもそもプライバシーを考慮しない場合でも中心に E の点を用いると，理想的な結果とは形が変わっていることが分かる．これは，この実験設定における E の点の少なさも影響していると思われる．

図 4 に示すのは，プライバシーの条件を緩和してしまう手法 C である．この手法は条件を緩和してるだけありこの中で最も精度が良いが，やはりプライバシーの保護を強くすると推定される重みが一様に近づくことが分かる．

6　まとめと今後の課題

Importance weighting mechanism [7] で提唱されているように，密度比推定を差分プライバ

シー [3] を満たすように行うことは実用的に意味がある．そんな中でも本研究は，密度比推定として現在最も効率の良い手法の 1 つである uLSIF [8] について，差分プライバシー的シナリオで行う手法を構成した．

今後の課題として，主に実用性の面での，既存手法との比較や，提案手法同士の比較，最適なパラメータの分析などが挙げられる．また，さらに根本的な手法の改善も考えられ，特に凸最適化問題に対する Objective Perturbation [2] 等の先行手法をうまく組み合わせていくことで，より効率よく差分プライバシーを満たすことができると考えられる．

謝辞

本研究は，科学研究費基盤研究 (B)「情報検索システムにおけるプライバシー保護に関する研究」の助成を受けました．

参考文献

[1] Bickel, Steffen, Michael Brückner, and Tobias Scheffer. “Discriminative learning for differing training and test distributions.” Proceedings of the 24th international conference on Machine learning. pp. 81-88. ACM, 2007.

[2] Chaudhuri, Kamalika, Claire Monteleoni, and Anand D. Sarwate. “Differentially private empirical risk minimization.” The Journal of Machine Learning Research 12 (2011): pp. 1069-1109.

[3] C. Dwork. “Differential privacy.” Automata, languages and programming, Springer Berlin Heidelberg. pp. 1-12. 2006.

[4] C. Dwork, F. McSherry, K. Nissim, and A. Smith. “Calibrating noise to sensitivity in private data analysis.” in Theory of Cryptography, vol. 3876. pp. 265-284. 2006.

[5] C. Dwork and Aaron Roth. “The algorithmic foundations of differential privacy.” Theoretical Computer Science 9.3-4 (2013): pp. 211-407.

[6] Huang, Jiayuan, et al. “Correcting sample selection bias by unlabeled data.” Advances in neural information processing systems. pp. 601-608. 2006.

[7] Ji, Zhanglong, and Charles Elkan. “Differential privacy based on importance weighting.” Machine learning 93.1 (2013): pp. 163-183.

[8] Kanamori, Takafumi, Shohei Hido, and Masashi Sugiyama. “A least-squares approach to direct importance estimation.” The Journal of Machine Learning Research 10 (2009): pp. 1391-1445.

[9] Kanamori, Takafumi, Taiji Suzuki, and Masashi Sugiyama. “Theoretical analysis of density ratio estimation.” IEICE transactions on fundamentals of electronics, communications and computer sciences 93.4 (2010): pp. 787-798.

[10] Li, Ninghui, Wahbeh Qardaji, and Dong Su. “On sampling, anonymization, and differential privacy or, k-anonymization meets differential privacy.” Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security. pp. 32-33. ACM, 2012.

[11] McSherry, Frank, and Kunal Talwar. “Mechanism design via differential privacy.” Foundations of Computer Science, 2007. FOCS’07. 48th Annual IEEE Symposium on. pp. 94-103. IEEE, 2007.

[12] Sugiyama, Masashi, et al. “Direct importance estimation with model selection and its application to covariate shift adaptation.” Advances in neural information processing systems. pp. 1433-1440. 2008.