1K2-2

差分プライベート最小二乗密度比推定

Differentially Private Least-squares Importance Fitting

中川 裕志 *1　高林 裕太 *2　荒井 ひろみ *3
Hiroshi Nakagawa　Yuta Takabayashi　Hiromi Arai

*1*3東京大学 情報基盤センター
Information Technology Center, The University of Tokyo

*2東京大学 大学院情報理工学系研究科
Dep. of Mathematical Informatics, The University of Tokyo

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 approaches 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 Differentially Private uLSIF and experimentally analyze it.

1. はじめに

プライバシー保護データマイニングのシナリオとして個人情報等を含んでおり非公開であるべきデータベース D に対する学習 $E_D[f(\boldsymbol{x})]$ を行いたいとき，直接このデータベース D を対象にすると、個人情報の漏洩リスクが高いと考えられる．

そこで，D と同じ属性を持つような公開済みデータベース E を用い，これに重み関数 w によって E への重み付けマイニング $E_E[f(\boldsymbol{x})w(\boldsymbol{x})]$ を行いデータベース D に対する学習 $E_D[f(\boldsymbol{x})]$ を近似する手法を考える．

$$\begin{aligned} E_D[f(\boldsymbol{x})] &= \int_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})p_D(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})w(\boldsymbol{x})p_E(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\ &= E_E[f(\boldsymbol{x})w(\boldsymbol{x})] \end{aligned} \tag{1}$$

式 (1) では，求めたい重み関数 w は $w(\boldsymbol{x}) = \frac{p_D(\boldsymbol{x})}{p_E(\boldsymbol{x})}$ という各データベースの確率密度の比の形をしており，w の推定を密度比推定と呼ぶ．

ただし，w と E の組み合わせから元のデータベース D のプライバシーが漏洩する可能性が残る．その対応策として，差分プライバシー [Dwork 06a] というプライバシー保護を適用した先行研究として Importance weighting mecanism[Ji and Elkan 13] と呼ばれる手法が存在する．以下の表 1 に密度比推定と差分プライバシーの組み合わせを示す．

表 1: 種々の密度比推定法

通常の手法	差分プライベート手法
確率的分類法　LogReg	DPLogReg
[Bickel et al. 07]	[Ji and Elkan 13]
最小二乗適合法　uLSIF	DPuLSIF（本論文）
[Kanamori et al. 09]	

表 1 で示すように，[Ji and Elkan 13] では密度比推定手法としてロジスティック回帰 (以下，LogReg と略記) を用いていたが，本研究ではより効率の良い手法として知られる最小二乗適合法 [Kanamori et al. 09](以下， uLSIF と略記) に差分プライバシーを適用した手法を提案する．

以下，本稿では 2 節で差分プライバシー，3 節で最小二乗密度比推定を説明した後に，4 節にて提案手法の差分プライベート最小二乗密度推定を説明し，5 節にて人工データと実データの場合の実験評価結果を示す．6 節はまとめである．

2. 差分プライバシー

1 レコード異なるデータベースの組 $D, D' \in \mathbb{D}$，($\mathbb{D}$ はデータベースの型) つまりハミング距離 $d_{\mathrm{H}}(D, D') = 1$ であるデータベースを隣接するいう．隣接するデータベースの組を用いると，差分プライバシーは定義 1 により定義される．

定義 1. あるメカニズム M が ε-差分プライバシーとは，任意の隣接するデータベースの組 D, D' について，M の任意の出力 $\boldsymbol{t} \subset \mathrm{Range}(M)$ に対して

$$\Pr[M(D) \in \boldsymbol{t}] \leq e^{\varepsilon} \times \Pr[M(D') \in \boldsymbol{t}] \tag{2}$$

が成り立つことである．

これは隣接するデータベースの組 D, D' に対しメカズム M を適用したとき，出力から元のデータベースを確率的に判別できないということを示す．また，この保証を満たす手法として，定理 1 の Laplacian Mechanism[Dwork et al. 06b] が知られている．

定理 1. メカニズム M を関数 f の出力に，その f の sensitivity $S(f)$ とプライバシパラメータ ε を用いて分散を $\lambda = S(f)/\varepsilon$ としたラプラスノイズ

$$\mathrm{Lap}(\lambda) = \frac{1}{2\lambda}\exp\left(-\frac{|\delta|}{\lambda}\right) \tag{3}$$

を加える操作と定義する．このとき M の出力は ε-差分プライバシーを満たす．多次元の場合は各次元ごとに $\mathrm{Lap}(\lambda)$ を加える．

ただし，分散パラメータに含まれる値である f の sensitivity $S(f)$ とは定義 2 により与えられる．

定義 2. クエリ f の（L1-）sensitivity $S(f)$ とは，任意の隣接するデータベース D, D' に対し

$$S(f) = \sup_{D,D'} ||f(D) - f(D')||_1 \tag{4}$$

で定義される量である．

差分プライバシーには合成定理と呼ばれる重要な定理が存在する．[Dwork and Roth 14]

定理 2. データベース $D \in \mathbb{D}$ に対し，ε-差分プライバシーを満たすメカニズム $F : \mathbb{D} \to \mathrm{Range}(F)$ とアルゴリズム $G : \mathrm{Range}(F) \to \mathrm{Range}(G)$ があったとき，合成したメカニズム $G \circ F$ は ε-差分プライバシーを満たす．

定理 3. (composition theorem) データベース $D \in \mathbb{D}$ に対し，ε_1-差分プライバシーを満たすメカニズム $F : \mathbb{D} \to \mathrm{Range}(F)$ と ε_2-差分プライバシーを満たすメカニズム $G : \mathbb{D} \times \mathrm{Range}(F) \to \mathrm{Range}(G)$ があったとき，合成したメカニズム $G \circ F$ は $\varepsilon_1 + \varepsilon_2$-差分プライバシーを満たす．

3. 最小二乗密度比推定（uLSIF）

線型モデル $\hat{w}(\boldsymbol{x}) = \boldsymbol{\alpha}^T \boldsymbol{\phi}(\boldsymbol{x}) = \sum_{l=1}^{b} \alpha_l \phi_l(\boldsymbol{x})$ を仮定し，二乗損失

$$J_0(\boldsymbol{\alpha}) = \frac{1}{2} \int (\hat{w}(\boldsymbol{x}) - w(\boldsymbol{x}))^2 p_E(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{5}$$

を最小化するパラメータ $\boldsymbol{\alpha}$ を選ぶ．この二乗損失をそのまま計算することは困難なので，変形すると，

$$\begin{aligned} J_0(\boldsymbol{\alpha}) =& \frac{1}{2} \int (\hat{w}(\boldsymbol{x}) - w(\boldsymbol{x}))^2 p_E(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ =& \frac{1}{2} \int (\hat{w}(\boldsymbol{x}))^2 p_E(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} - \int \hat{w}(\boldsymbol{x}) p_D(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &+ \frac{1}{2} \int (w(\boldsymbol{x}))^2 p_E(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \end{aligned} \tag{6}$$

となる．ここで第 3 項は $\boldsymbol{\alpha}$ に関して定数であり最小化では無視できるので，第 1,2 項のみに上記の $\hat{w}(\boldsymbol{x})$ を代入すると，

$$\begin{aligned} J(\boldsymbol{\alpha}) =& \frac{1}{2} \int (\hat{w}(\boldsymbol{x}))^2 p_E(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} - \int \hat{w}(\boldsymbol{x}) p_D(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ =& \frac{1}{2} \sum_{l,l'=1}^{b} \alpha_l \alpha_{l'} \left(\int \phi_l(\boldsymbol{x}) \phi_{l'}(\boldsymbol{x}) p_E(\boldsymbol{x}) d\boldsymbol{x} \right) \\ &- \sum_{l=1}^{b} \left(\int \phi_l(\boldsymbol{x}) p_D(\boldsymbol{x}) d\boldsymbol{x} \right) \\ =& \frac{1}{2} \boldsymbol{\alpha}^T H \boldsymbol{\alpha} - \boldsymbol{h}^T \boldsymbol{\alpha} \end{aligned} \tag{7}$$

となる．ただし，H は $b \times b$ 行列で，第 (l, l')-成分が

$$H_{l,l'} = \int \phi_l(\boldsymbol{x}) \phi_{l'}(\boldsymbol{x}) p_E(\boldsymbol{x}) d\boldsymbol{x} \tag{8}$$

であり，$\boldsymbol{h}$ は b 次元ベクトルで，第 l-成分が

$$h_l = \int \phi_l(\boldsymbol{x}) p_D(\boldsymbol{x}) d\boldsymbol{x} \tag{9}$$

である．これを標本平均で近似することにより，最終的に

$$\hat{J}(\boldsymbol{\alpha}) = \frac{1}{2} \boldsymbol{\alpha}^T \hat{H} \boldsymbol{\alpha} - \hat{\boldsymbol{h}}^T \boldsymbol{\alpha} \tag{10}$$

という計算可能な目的関数が得られる．

また，uLSIF では本来存在する非負制約 $\boldsymbol{\alpha} \geq \mathbf{0}$ を外して最適化問題を解き，事後補正を加えるという手順にすることにより高速な計算を可能にしている．実際の手順はアルゴリズム 1 のようになる．

アルゴリズム 1 uLSIF

入力： データベース $\{\boldsymbol{x}_j^D\}_{j=1}^n$，$\{\boldsymbol{x}_i^E\}_{i=1}^N$，パラメータ λ, σ，基底関数列 $\{\phi_l\}_{l=1}^b$
出力： $\hat{w}(\boldsymbol{x})$
$\hat{H}_{l,l'} = \frac{1}{N} \sum_{i=1}^{N} \phi_l(\boldsymbol{x}_i^E) \phi_{l'}(\boldsymbol{x}_i^E)$ for $l, l' = 1, 2, \ldots, b$
$\hat{h}_l = \frac{1}{n} \sum_{j=1}^{n} \phi_l(\boldsymbol{x}_j^D)$ for $l = 1, 2, \ldots, b$
$\hat{\boldsymbol{\alpha}} \leftarrow \max(\mathbf{0}, (\hat{H} + \lambda I)^{-1} \hat{\boldsymbol{h}})$
$\hat{w}(\boldsymbol{x}) \leftarrow \sum_{l=1}^{b} \hat{\alpha}_l \phi_l(\boldsymbol{x})$

4. 差分プライベート最小二乗密度比推定

定理 3 の差分プライバシーの合成定理を用いることにより，パラメータ $\boldsymbol{\alpha}$ と基底関数 $\boldsymbol{\phi}(\boldsymbol{x})$ にそれぞれ差分プライバシーを満たすことにより，最終出力 $\hat{w}(\boldsymbol{x})$ の差分プライバシーを満たすことができる．ここで考える問題を 2 つに場合分けする．

4.1 パラメトリックモデルの場合

まず基底関数 $\boldsymbol{\phi}(\boldsymbol{x})$ がデータベース D に非依存な場合を考えると，このときは定理 2 を用いることによって最終出力 $\hat{w}(\boldsymbol{x})$ の差分プライバシーは $\hat{\boldsymbol{h}}$ への雑音加算のみで満たせることがわかる．加算する雑音量は定理 1 と次の定理 4 により与えられる．

定理 4. $\hat{\boldsymbol{h}}$ の Sensitivity 値 $S(\hat{\boldsymbol{h}})$ は

$$\frac{bC}{n} \tag{11}$$

である．ただし，$C = \sup \phi_l(\boldsymbol{x})$ とする．

スペースの関係で証明は略す．以上より具体的な手順がアルゴリズム 2 のように与えられる．

アルゴリズム 2 DPuLSIF

入力： データベース $\{\boldsymbol{x}_j^D\}_{j=1}^n$，$\{\boldsymbol{x}_i^E\}_{i=1}^N$，パラメータ $\lambda, \sigma, \varepsilon$，基底関数列 $\{\phi_l\}_{l=1}^b$
出力： $\hat{w}(\boldsymbol{x})$
$C \leftarrow \sup \phi_l(\boldsymbol{x})$ { 定数として与えても良い }
$\hat{H}_{l,l'} = \frac{1}{N} \sum_{i=1}^{N} \phi_l(\boldsymbol{x}_i^E) \phi_{l'}(\boldsymbol{x}_i^E)$ for $l, l' = 1, 2, \ldots, b$
$\hat{h}_l = \frac{1}{n} \sum_{j=1}^{n} \phi_l(\boldsymbol{x}_j^D) + \mathrm{Lap}\left(\frac{bC}{\varepsilon n}\right)$ for $l = 1, 2, \ldots, b$
$\hat{\boldsymbol{\alpha}} \leftarrow \max(\mathbf{0}, (\hat{H} + \lambda I)^{-1} \hat{\boldsymbol{h}})$
$\hat{w}(\boldsymbol{x}) \leftarrow \sum_{l=1}^{b} \hat{\alpha}_l \phi_l(\boldsymbol{x})$

またこのとき，最終出力 $\hat{w}(\boldsymbol{x})$ への雑音量が次の定理 5 より与えられる．

定理 5. パラメトリックモデルを用いる場合の最終出力 $\hat{w}(\boldsymbol{x}) = \hat{\boldsymbol{\alpha}}^T \boldsymbol{\phi}(\boldsymbol{x})$ に影響する雑音量は

$$O\left(\frac{b}{\varepsilon \lambda n}\right) \tag{12}$$

である．

スペースの関係で証明は略す．

4.2 カーネルモデルを用いる場合

次に基底関数としてカーネル関数 $K(\boldsymbol{x}, \{\boldsymbol{c}_l\})$ を用いる場合を考えると，これは基本的にデータベース D に依存するため，前節同様パラメータ $\boldsymbol{\alpha}$ に差分プライバシーを満たすと同時にカーネル関数 $K(\boldsymbol{x}, \{\boldsymbol{c}_l\})$ にも差分プライバシーを満たす必要がある．以下の 3 通りの手法を提案する．

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

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

手法 C プライバシー条件を緩和する．

本稿では紙面の都合上，実験的に最も精度がよく ε-差分プライバシーを満たせる手法 A を説明する．

原理は単純であり，カーネル関数としてはガウシアンカーネルを用い、その中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ として D から b 個の点をランダムサンプリングした後，定理 1 により差分プライバシーを満たすように雑音を加えて使用するというものである．この際，定理 6 によって，D からのサンプル率 $\beta = b/n$ を下げるほど雑音量を減らせるのが特徴である．

定理 6 (DP のサンプリング定理 [Li et al. 12] の変形)．確率 1 のサンプリング，つまり全データベースを使った際に，アルゴリズム $f(D)$ が ε'-差分プライバシーを満たすならば，確率 β でサンプリングしたデータベースに対してアルゴリズム $f(D)$ は ε-差分プライバシーを満たす．ただし，

$$\varepsilon' = \log\left(1 + \frac{1}{\beta}(e^{p\varepsilon} - 1)\right) \tag{13}$$

である．

またこの際必要になる中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ の Sensitivity 値 $S(\{\boldsymbol{c}_l\}_{l=1}^b)$ は以下の定理 7 により与えられる．

定理 7. D からランダムサンプリングした中心 $\{\boldsymbol{c}_l\}_{l=1}^b$ の Sensitivity 値 $S(\{\boldsymbol{c}_l\}_{l=1}^b)$ は，

$$S(\{\boldsymbol{c}_l\}_{l=1}^b) = \max_{\boldsymbol{x},\boldsymbol{x}' \in D} ||\boldsymbol{x} - \boldsymbol{x}'||_1 \tag{14}$$

である．

これらを用いて，カーネルモデルを用いる手法 A をアルゴリズム 3 のように構成することができる．なお，中心のサンプル数 b は元の uLSIF の規定値である $\min(100, n)$ を用いているが，次節で述べる実験によりその妥当性を評価する．

アルゴリズム 3 DPuLSIF-A

入力： データベース $\{\boldsymbol{x}_j^D\}_{j=1}^n$，$\{\boldsymbol{x}_i^E\}_{i=1}^N$，パラメータ $\lambda, \sigma, \varepsilon, p$

出力： $\hat{w}(\boldsymbol{x})$

$b \leftarrow \min(100, n)$

b 個の中心 $\{\boldsymbol{k}_l\}_{l=1}^b$ を $\{\boldsymbol{x}_j^D\}_{j=1}^n$ から一様にサンプルする．

$\boldsymbol{c}_l \leftarrow \boldsymbol{k}_l + \mathrm{Lap}\left(\frac{\max_{\boldsymbol{x},\boldsymbol{x}' \in D} ||\boldsymbol{x}-\boldsymbol{x}'||_1}{\varepsilon'}\right)$ for $l, l' = 1, 2, \ldots, b$

$\hat{H}_{l,l'} \leftarrow \frac{1}{N}\sum_{i=1}^N \exp\left(-\frac{||\boldsymbol{x}_i^E - \boldsymbol{c}_l||^2 + ||\boldsymbol{x}_i^E - \boldsymbol{c}_{l'}||^2}{2\sigma^2}\right)$ for $l, l' = 1, 2, \ldots, b$

$\hat{h}_l \leftarrow \frac{1}{n}\sum_{j=1}^n \exp\left(-\frac{||\boldsymbol{x}_j^D - \boldsymbol{c}_l||^2}{2\sigma^2}\right) + \mathrm{Lap}\left(\frac{bp}{\varepsilon n}\right)$ for $l = 1, 2, \ldots, b$

$\hat{\boldsymbol{\alpha}} \leftarrow \max(\boldsymbol{0}, (\hat{H} + \lambda I)^{-1}\hat{\boldsymbol{h}})$

$\hat{w}(\boldsymbol{x}) \leftarrow \sum_{l=1}^b \hat{\alpha}_l \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{c}_l||^2}{2\sigma^2}\right)$

5. 実験

5.1 人工データ

まず人工データによる実験を行う．非公開データベース D，公開データベース E を

$$D \sim \mathcal{N}(2, 1/4)$$
$$E \sim \mathcal{N}(1, 1) \tag{15}$$

という確率分布に従いデータ数の比率が $n : N = 10 : 1$ となるように生成し，提案手法を用いて密度比推定を行った際の概形を見る．

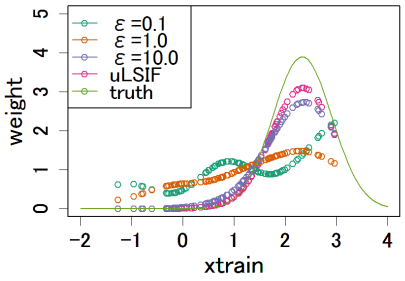

図 1: 手法 A による密度比推定の例

図 1 に示すのは手法 A を用いて密度比推定を行った際の典型的な結果である．黄緑の実線が理論的な密度比関数 $w(\boldsymbol{x})$ であり，これに近いほど精度が良い．基本的に提案手法においてはプライバシーの保護を強く，つまり ε を下げるほど精度が悪くなることが分かる．

次に平均二乗誤差

$$\mathrm{MSE} := \frac{1}{N}\sum_{\boldsymbol{x} \in E}(\hat{w}(\boldsymbol{x}) - w(\boldsymbol{x}))^2$$

を用いて，提案するカーネルモデルの 3 手法について相対的な有用性評価を行う．

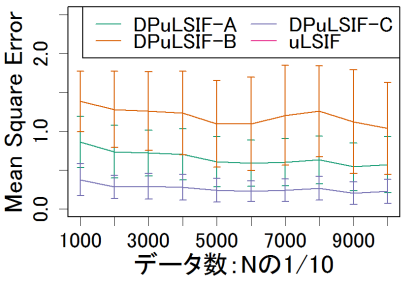

図 2: 手法 A による密度比推定の例

図 2 に示すように，各手法ともデータ数を増やすと誤差量が減少し，手法 C，A，B の順に精度が良い．

最後に，手法 A における中心のサンプル数 b を変化させたときの MSE の変化を見る．

図 3 に示すように，元手法 uLSIF については確かに規定値である $b = 100$ あたりが最も誤差が少ない結果となる．また，提案手法 DPuLSIF-A については b に比例した誤差が加わる

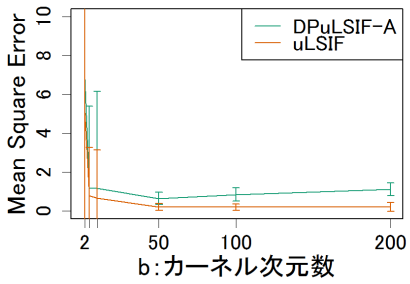

図 3: 手法 A による中心のサンプル数 b と MSE の推移

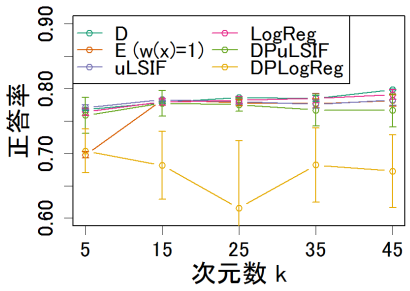

図 5: adult について次元数 k と各正答率の推移

影響により，$b = 50$ あたりから徐々に誤差が増大していくことがわかる．この結果を踏まえた上で，以降の実験も元手法と同じ $b = 100$ を用いて行う．

5.2 実データ

次に実データを用いた実験を行う．UCI Machine Learning Repository の adult データセットを用い，男性の 9 割と女性の 1 割を非公開データベース D に，残りを公開データベース E に振り分けることによって確率分布の違いをシミュレートする．これらに対し各密度比推定手法を適用して得られる重み関数を用いた上で，各種のタスクを実行し有用性を評価する．ここでは，主成分分析を適用し，第 k 主成分までの k 次元のデータを用いたときの k と精度の関係を調べる．

まず，簡単なタスクの例としてカウントクエリを考える．

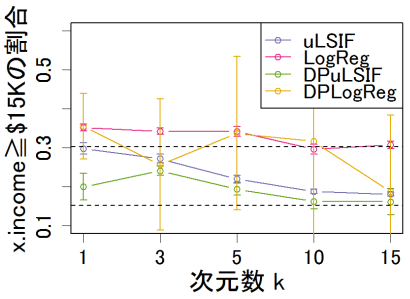

図 4: adult について次元数 k とカウントクエリの結果の推移

図 4 に示すのは，各手法で推定された密度比を用いて非公開データベース D のうち収入属性値 : income $\geq \$15K$ が真であるものの割合を推定したものである．また，上の点線が D における真の値，下の点線が E における値を表しており，密度比推定により上の線に近づいているほど有用性が高いと言える．提案手法 DPuLSIF や元の uLSIF はデータベースの次元が低いほど良い精度が出せることが分かる．

次に，最もよく使われる機械学習のタスクである分類のうち，サポートベクターマシンを用いたものを考える．

図 5 に示すのは，各手法を用いてデータベース D の収入属性値 income $\geq \$15K$ 属性についての分類器を作り，それを実際にデータベース D に適用した正答率である．提案手法 DPuLSIF は次元数 $k = 15$ で差分プライバシーを用いない uLSIF と同程度の正答率を示した．

6. まとめと今後の課題

本研究では現在最も効率の良い密度比推定手法として知られる uLSIF について，差分プライベートな手法を構成した．人工データや実データを用いた実験により，特に低次元なデータについて良い性能が出ることを示した．

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

[Bickel et al. 07] Steffen Bickel, Michael Brückner, and Tobias Scheffer. Discriminative learning for differing training and test distributions. *Proceedings of the 24th international conference on Machine learning. ACM*, 2007.

[Dwork 06a] Cynthia Dwork. Differential privacy. *Automata, languages and programming, Springer Berlin Heidelberg*. pp. 1-12. 2006.

[Dwork et al. 06b] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. *in Theory of Cryptography, vol. 3876*. pp. 265-284. 2006.

[Dwork and Roth 14] Cynthia Dwork, and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science 9.3-4*. pp. 211-407. 2014.

[Ji and Elkan 13] Zhanglong Ji, and Charles Elkan. Differential privacy based on importance weighting. *Machine learning 93.1*. pp. 163-183. 2013.

[Kanamori et al. 09] Takafumi Kanamori, Shohei Hido, and Masashi Sugiyama. A least-squares approach to direct importance estimation. *The Journal of Machine Learning Research 10*. pp. 1391-1445. 2009.

[Li et al. 12] Ninghui Li, 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. ACM*, 2012.