IBISML研究会 2010年6月14日

統計的機械学習の新展開：確率密度比に基づくアプローチ

東京工業大学　計算工学専攻
杉山　将
sugi@cs.titech.ac.jp
http://sugiyama-www.cs.titech.ac.jp/~sugi

概要(1)

- 二つの確率密度関数の比を考える.

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

- 比によって，様々な学習問題が解決できる！
  - 非定常環境下での適応学習，ドメイン適応，マルチタスク学習
  - 二標本検定，異常値検出，時系列の変化点検知，
  - 相互情報量推定，独立性検定，特徴選択，十分次元削減，独立成分分析，因果推論，
  - 条件付き確率推定，確率的パターン認識
- “ひと粒で何度もおいしい”学習手法を紹介します

概要（2）

Vapnikの原理：ある問題を解くとき，その問題よりも難しい問題を途中段階で解いてはならない

$p_{\mathrm{nu}}(\boldsymbol{x}), p_{\mathrm{de}}(\boldsymbol{x})$

が分かる

$r(\boldsymbol{x}) = \dfrac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$

が分かる

- 密度を求めるよりも，密度比を求めるほうが易しい！
- 密度推定を経由せず，密度比を推定することにする．

概要(3)

- 2008年のIBISワークショップにて，同様の題目の発表をさせて頂いた.
- その時は「密度比を使えば面白いことができそう」という話題提供が中心でした.
- その後，IBIS業界の複数の方に興味を持って頂き，共同研究を進めてきました.
- その結果，「密度比を使えば本当に面白いことができる」ことがわかりました．本日はその報告です.
- 今日の発表をもとに，新たな共同研究パートナーが見つかれば，と期待しています.

発表の流れ

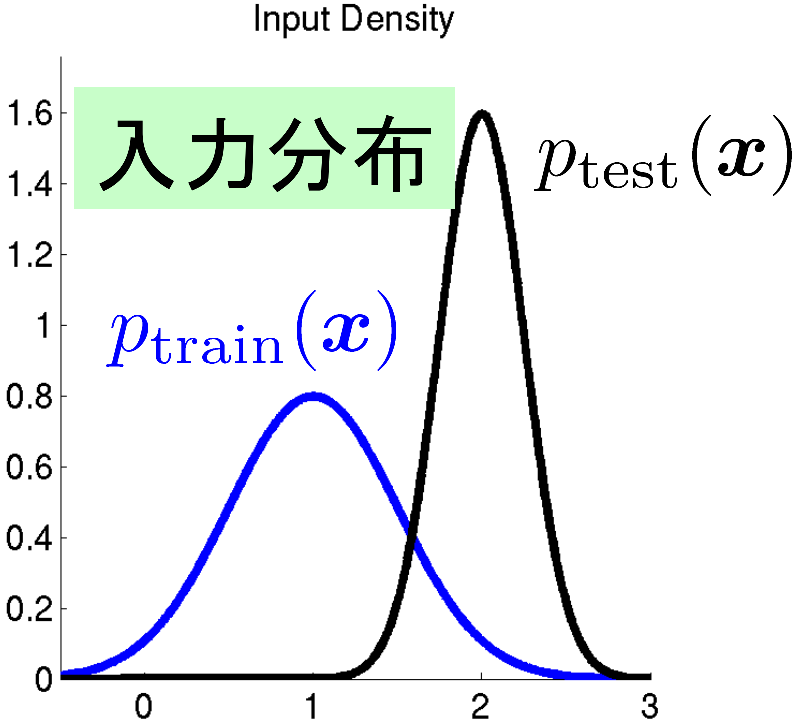
1. 密度比推定の応用例
   - A) 共変量シフト適応
   - B) 外れ値検出
   - C) 相互情報量推定
   - D) 条件付き確率推定
2. 密度比推定手法

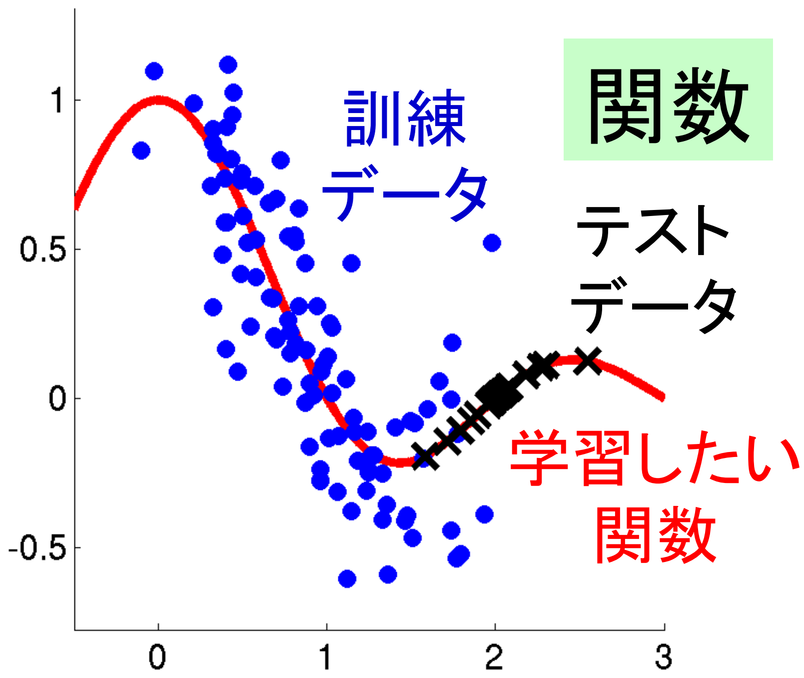
$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

共変量シフト適応

Shimodaira (JSPI2000)

- 共変量とは入力変数の別名.
- 共変量シフト：訓練時とテスト時で入力分布が変化するが，入出力関数は変わらない
- 外挿問題が典型的な例.





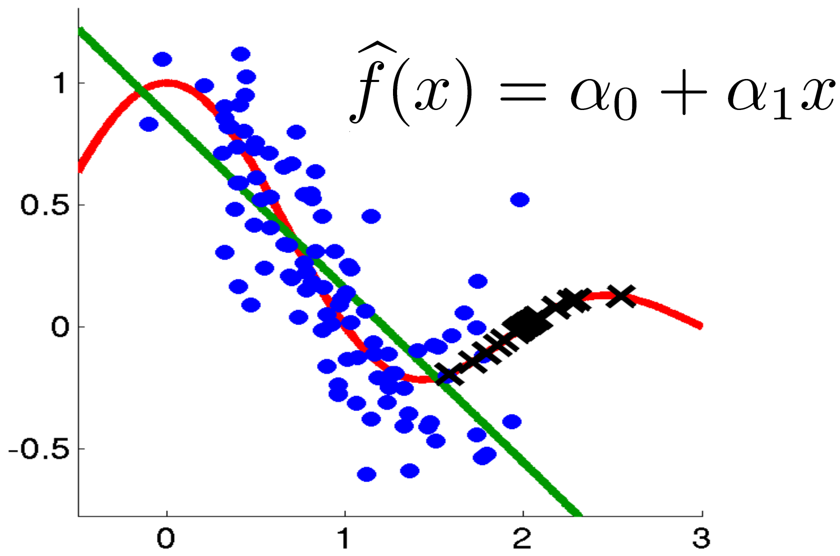
共変量シフト下での学習

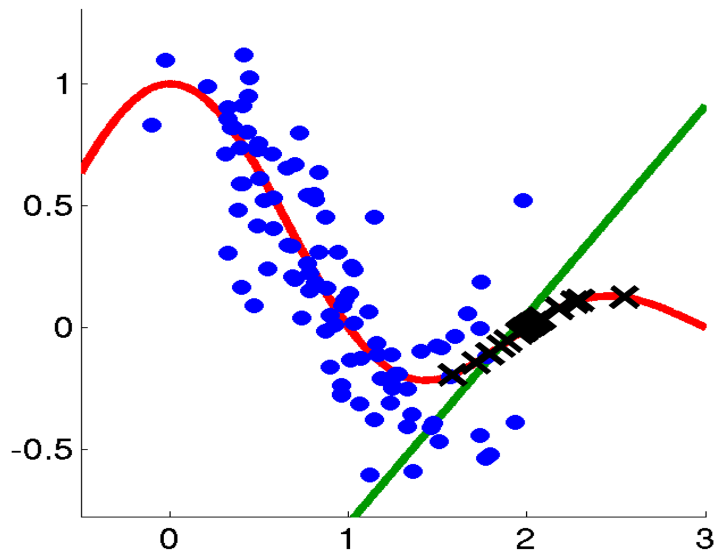
- 通常の最小二乗法は最適解に収束しない

$$\min_{\boldsymbol{\alpha}}\left[\sum_{i=1}^{n}\left(\hat{f}(\boldsymbol{x}_i)-y_i\right)^2\right]$$

- 密度比重み付き最小二乗法は最適解に収束

$$\min_{\boldsymbol{\alpha}}\left[\sum_{i=1}^{n}\frac{p_{\mathrm{test}}(\boldsymbol{x}_i)}{p_{\mathrm{train}}(\boldsymbol{x}_i)}\left(\widehat{f}(\boldsymbol{x}_i)-y_i\right)^2\right]$$





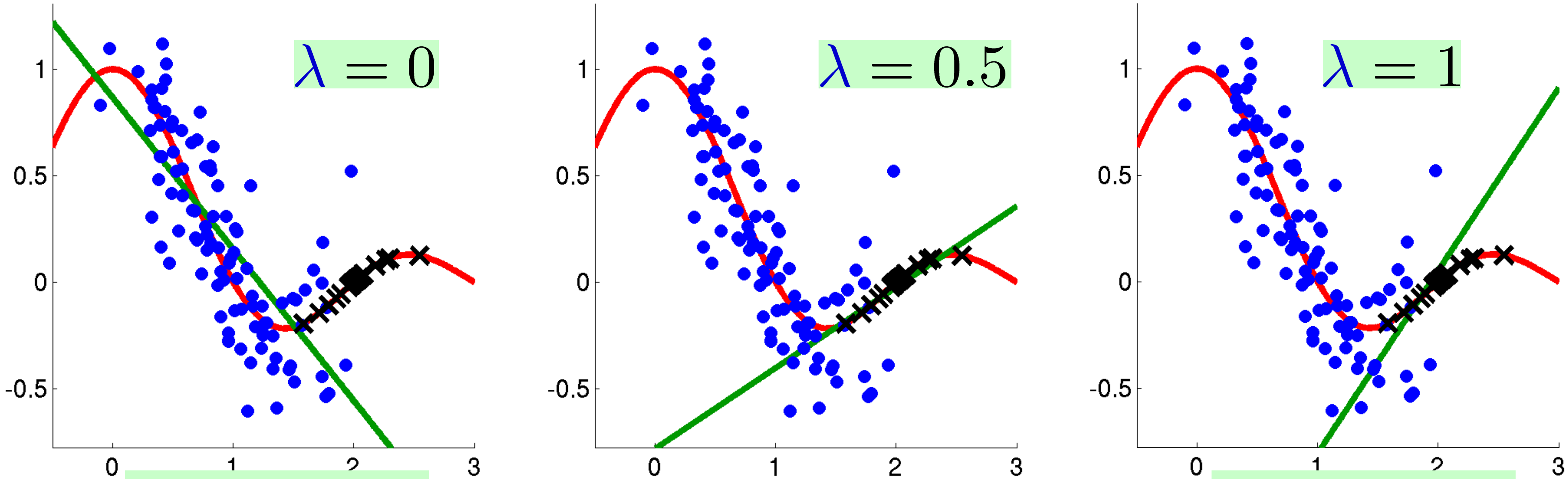
訓練誤差（尤度）に基づく学習法全てに適応可能！

“平坦化”密度比重み付け学習

Shimodaira (JSPI2000)

$$\min_{\boldsymbol{\alpha}} \left[\sum_{i=1}^{n} \left(\frac{p_{\mathrm{test}}(\boldsymbol{x}_i)}{p_{\mathrm{train}}(\boldsymbol{x}_i)} \right)^{\lambda} \left(\widehat{f}(\boldsymbol{x}_i) - y_i \right)^2 \right]$$

$0 \leq \lambda \leq 1$



バイアス：大
バリアンス：小

中間

バイアス：小
バリアンス：大

モデル選択が必要！

モデル選択

- 共変量シフト下では，通常のモデル選択規準は強いバイアスを持つ．
- 密度比重み付けは，モデル選択規準に不偏性を持たせるためにも用いられる：
  - Akaike information criterion（正則モデル）

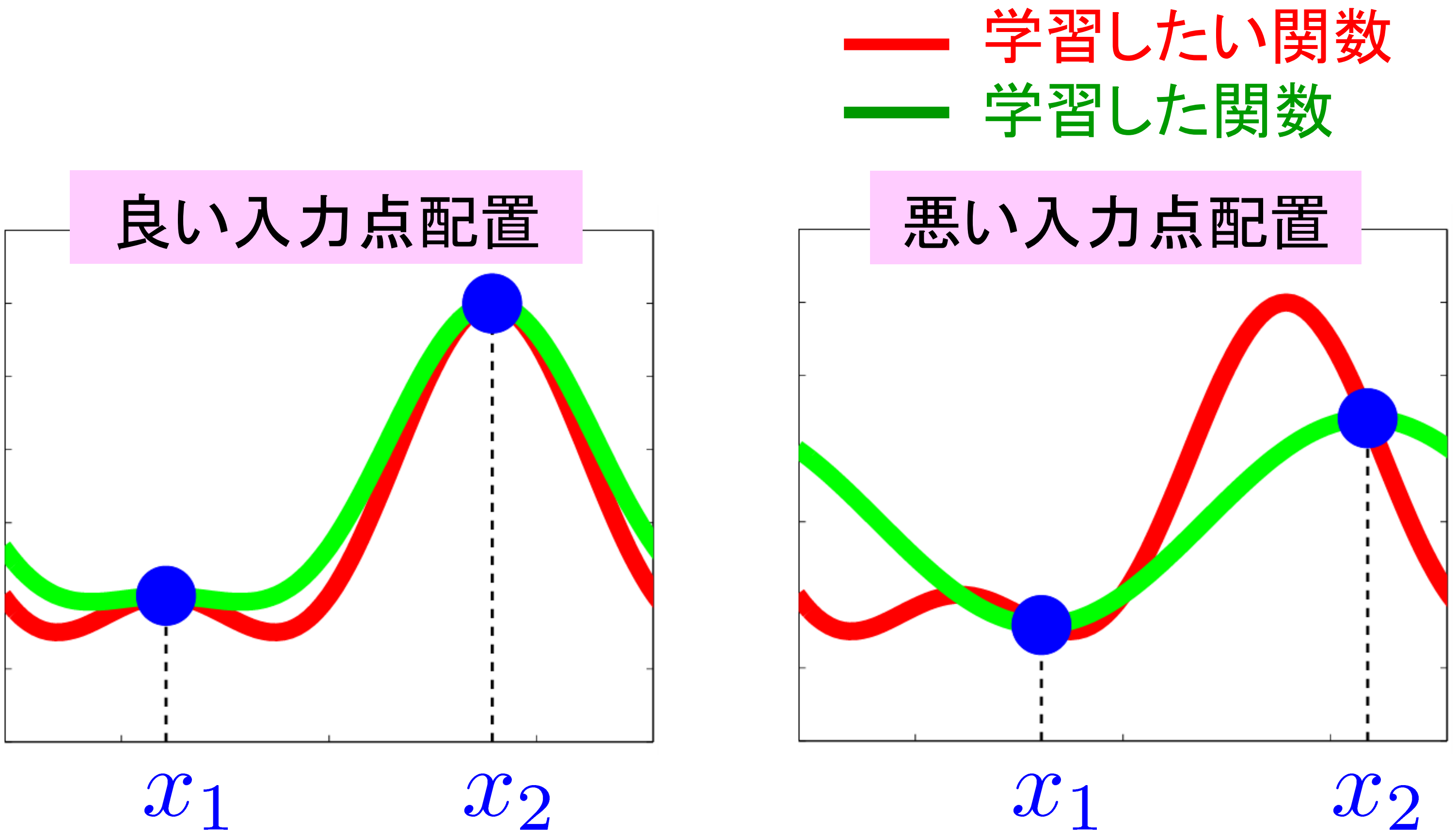
    Shimodaira (JSPI2000)
  - Subspace information criterion（線形モデル）

    Sugiyama & Müller (Stat&Dec.2005)
  - Cross-validation（任意のモデル）

    Sugiyama, Krauledat & Müller (JMLR2007)

能動学習

- 入力点の選び方は汎化能力に大きく影響する
- 目標:最適な入力点の場所を選ぶ



能動学習と共変量シフト

- 能動学習では共変量シフトが自然に起こる:
  - $p_{\mathrm{train}}(\boldsymbol{x})$ はユーザが設定.
  - $p_{\mathrm{test}}(\boldsymbol{x})$ は環境から自動的に決まる.
- 密度比重み付けは，能動学習のバイアスを軽減するのに役立つ:
  - 分布の基づく能動学習法:
    Wiens (JSPI2000)
    Kanamori & Shimodaira (JSPI2003)
    Sugiyama (JMLR2006)
  - ラベル無し標本に基づく能動学習法:
    Kanamori (NeuroComp2007)
    Sugiyama & Nakajima (MLJ2009)

実世界応用例(1)

- 顔画像からの年齢予測:
  - 照明や顔角度の変化(KRR&CV)

Ueki, Sugiyama & Ihara (ICPR2010)

- 話者識別:
  - 声質の変化(KLR&CV)

Yamada, Sugiyama & Matsui (SigPro2010)

- ブレイン・コンピュータインターフェース:
  - 心理状況の変化(LDA&CV)

Sugiyama, Krauledat & Müller (JMLR2007)
Li, Kambara, Koike & Sugiyama (IEEE-TBE2010)

実世界応用例（2）

- 日本語のテキスト分割:
  - 一般会話から医学用語へのドメイン適応(CRF&CV)

Tsuboi, Kashima, Hido, Bickel & Sugiyama (JIP2008)

- 半導体ウェハー位置合せ:
  - 最適なマーカーの選択(AL)

Sugiyama & Nakajima (MLJ2009)

- ロボット制御:
  - 標本の再利用(KLS&CV)

Hachiya, Akiyama, Sugiyama & Peters (NN2009)
Hachiya, Peters & Sugiyama (ECML2009)

  - 最適な探索政策の選択(AL)

Akiyama, Hachiya, & Sugiyama (NN2010)

発表の流れ

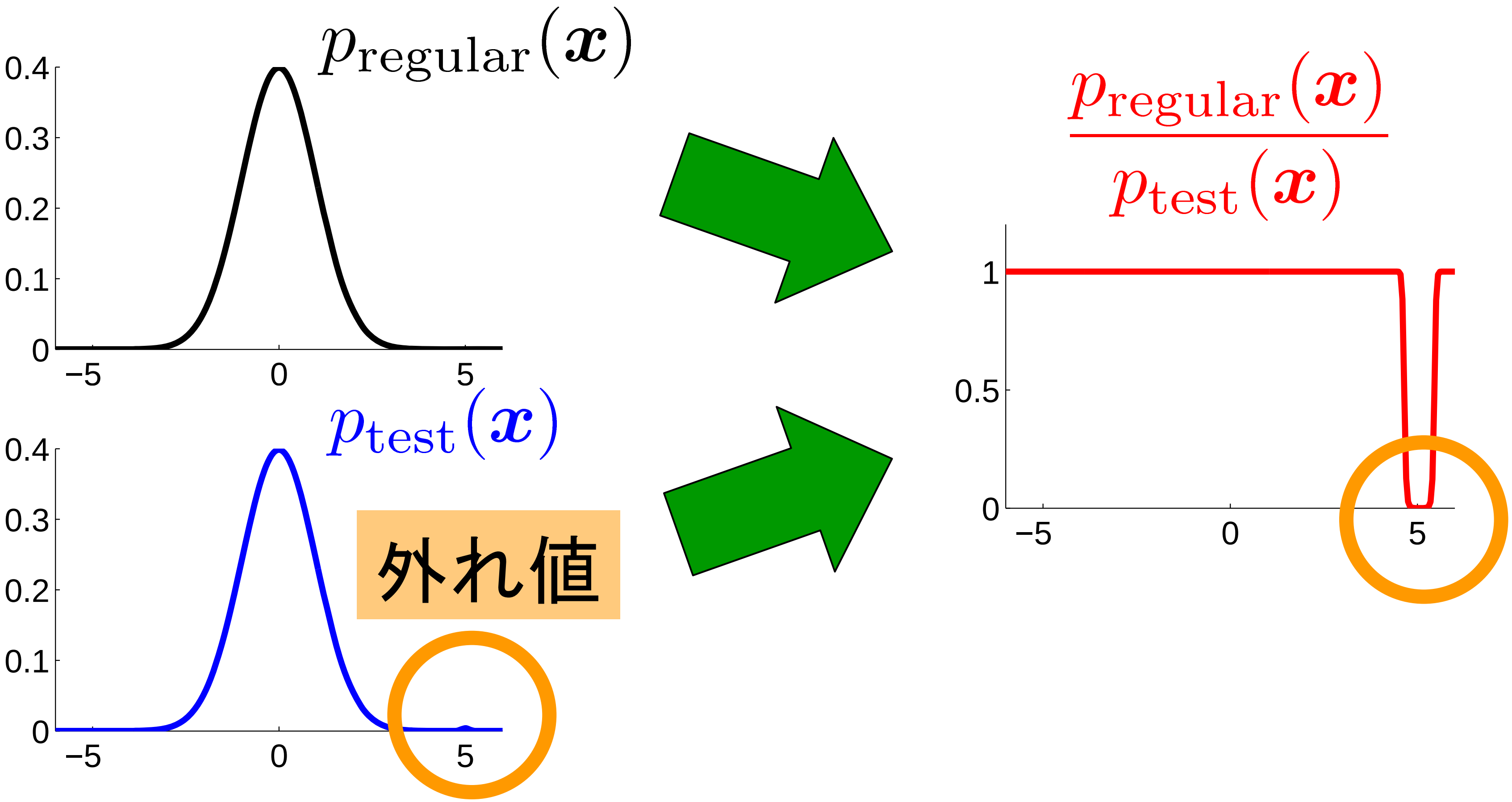
1. 密度比推定の応用例
   - A) 共変量シフト適応
   - B) 外れ値検出
   - C) 相互情報量推定
   - D) 条件付き確率推定
2. 密度比推定手法

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

正常値に基づく異常値検出

Hido, Tsuboi, Kashima, Sugiyama & Kanamori (ICDM2008, KAIS2010)
Smola, Song & Teo (AISTATS2009)

- 正常データと傾向が異なるテストデータを異常値とみなす.



実世界応用例

- 製鉄プロセスの異常診断
- 光学部品の品質検査
  Takimoto, Matsugu & Sugiyama (DMSS2009)
- ローン顧客の審査
  Hido, Tsuboi, Kashima, Sugiyama & Kanamori (KAIS2010)
- 生体データからの睡眠診断
  Kawahara & Sugiyama (SDM2009)

発表の流れ

1. 密度比推定の応用例
   - A) 共変量シフト適応
   - B) 外れ値検出
   - C) 相互情報量推定
   - D) 条件付き確率推定
2. 密度比推定手法

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

相互情報量推定

- 相互情報量: $\mathrm{MI} = \int p(\boldsymbol{x},\boldsymbol{y}) \log \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})} d\boldsymbol{x}d\boldsymbol{y}$

$\mathrm{MI} = 0$ $\Longleftrightarrow$ x と y は独立

- 相互情報量は密度比を用いて計算できる.

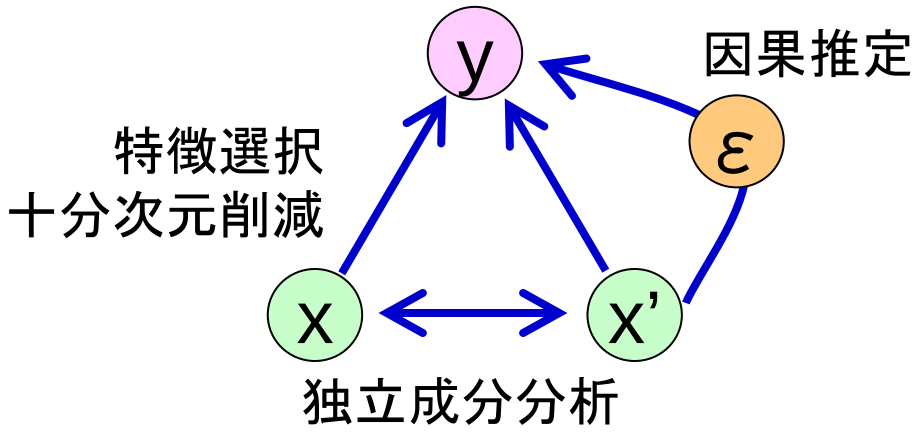
$$r(\boldsymbol{x},\boldsymbol{y}) = \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})} \qquad \widehat{\mathrm{MI}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{r}(\boldsymbol{x}_i,\boldsymbol{y}_i)$$

Suzuki, Sugiyama, Sese & Kanamori (FSDM2008)
Suzuki, Sugiyama & Tanaka (ISIT2009)

相互情報量推定量の応用例

- 入出力間の独立性判定:
  - 特徴選択 — Suzuki, Sugiyama, Sese & Kanamori (BMC Bioinformatics 2009)
  - 十分次元削減 — Suzuki & Sugiyama (AISTATS2010)
- 入力間の独立性判定:
  - 独立成分分析 — Suzuki & Sugiyama (NeuralComp, to appear)
- 入力と残差との独立性判定:
  - 因果推定 — Yamada & Sugiyama (AAAI2010)

- 明日のIBISML研究会で発表:
  - 強化学習における特徴選択(八谷大岳)
  - 因果推定(山田誠)



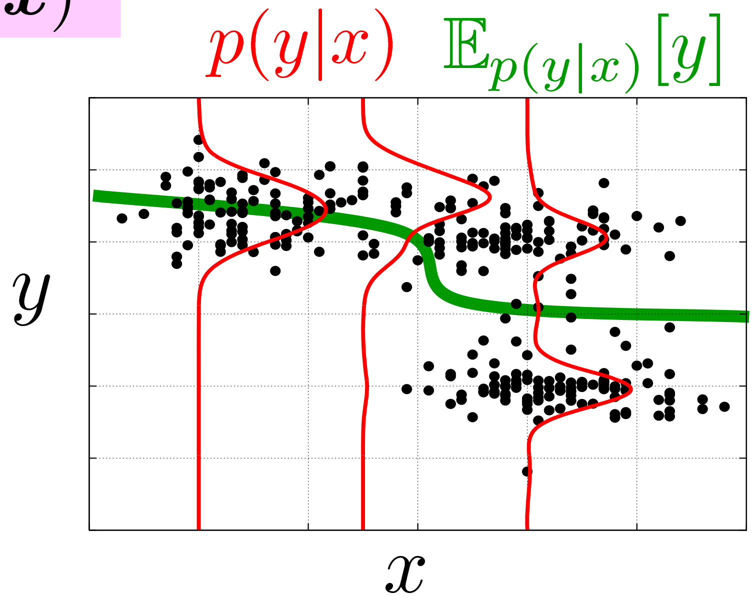
発表の流れ

1. 密度比推定の応用例
   - A) 共変量シフト適応
   - B) 外れ値検出
   - C) 相互情報量推定
   - D) 条件付き確率推定
2. 密度比推定手法

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

条件付き確率推定

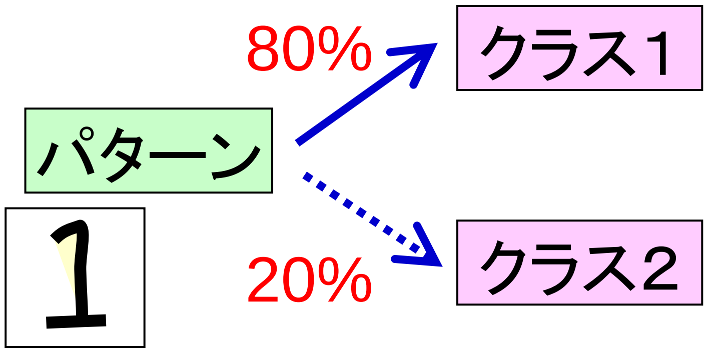
$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})}$$



■ 出力 y が連続のとき:
- 条件付き密度推定
- 移動ロボットの状態遷移推定

Sugiyama, Takeuchi, Suzuki, Kanamori, Hachiya & Okanohara (IEICE-ED2010)

■ 出力 y がカテゴリ値のとき:
- 確率的分類



■ 9月のIBISML研究会で発表:
- 汎化性能の更なる向上（山田誠）
- マルチタスク学習（Jaak Simm）

発表の流れ

1. 密度比推定の応用例
2. 密度比推定手法
   - A) 確率的分類法
   - B) 積率適合法
   - C) 密度比適合法
   - D) 比較
   - E) 次元削減

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

密度比推定

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

- 密度比をデータから推定する.

$$\{\boldsymbol{x}_i^{\mathrm{nu}}\}_{i=1}^{n_{\mathrm{nu}}} \overset{i.i.d.}{\sim} p_{\mathrm{nu}}(\boldsymbol{x})$$

$$\{\boldsymbol{x}_j^{\mathrm{de}}\}_{j=1}^{n_{\mathrm{de}}} \overset{i.i.d.}{\sim} p_{\mathrm{de}}(\boldsymbol{x})$$

- ナイーブな方法:分母と分子の密度をそれぞれ推定し, その比を取る

$$\widehat{r}(\boldsymbol{x}) = \frac{\widehat{p}_{\mathrm{nu}}(\boldsymbol{x})}{\widehat{p}_{\mathrm{de}}(\boldsymbol{x})}$$

Vapnikの原理

ある問題を解くとき，その問題よりも
難しい問題を途中段階で解いてはならない

$p_{\mathrm{nu}}(\boldsymbol{x}), p_{\mathrm{de}}(\boldsymbol{x})$

が分かる

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

が分かる

- 密度を求めるよりも，密度比を求めるほうが易しい！
- 密度推定を経由せず，密度比を推定することにする．

発表の流れ

1. 密度比推定の応用例
2. 密度比推定手法
    A) 確率的分類法
    B) 積率適合法
    C) 密度比適合法
    D) 比較
    E) 次元削減

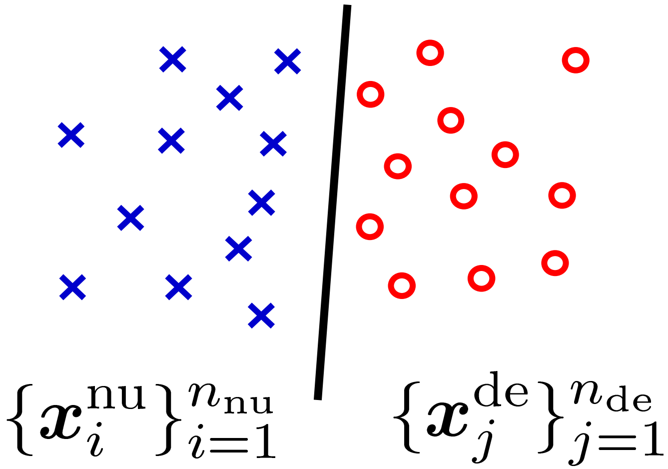
$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

確率的分類

- 分母と分子の標本を確率的に分類する:

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})} = \frac{p(\boldsymbol{x}|\text{`nu'})}{p(\boldsymbol{x}|\text{`de'})}$$

$$= \frac{p(\text{`nu'})}{p(\text{`de'})} \frac{p(\text{`de'}|\boldsymbol{x})}{p(\text{`nu'}|\boldsymbol{x})}.$$



- ロジスティック回帰:
  - モデルが正しいとき，漸近分散を最小にする.

    Qin (Biometrika1998)

  - モデルが正しくないときは，密度比適合法の方が良い.

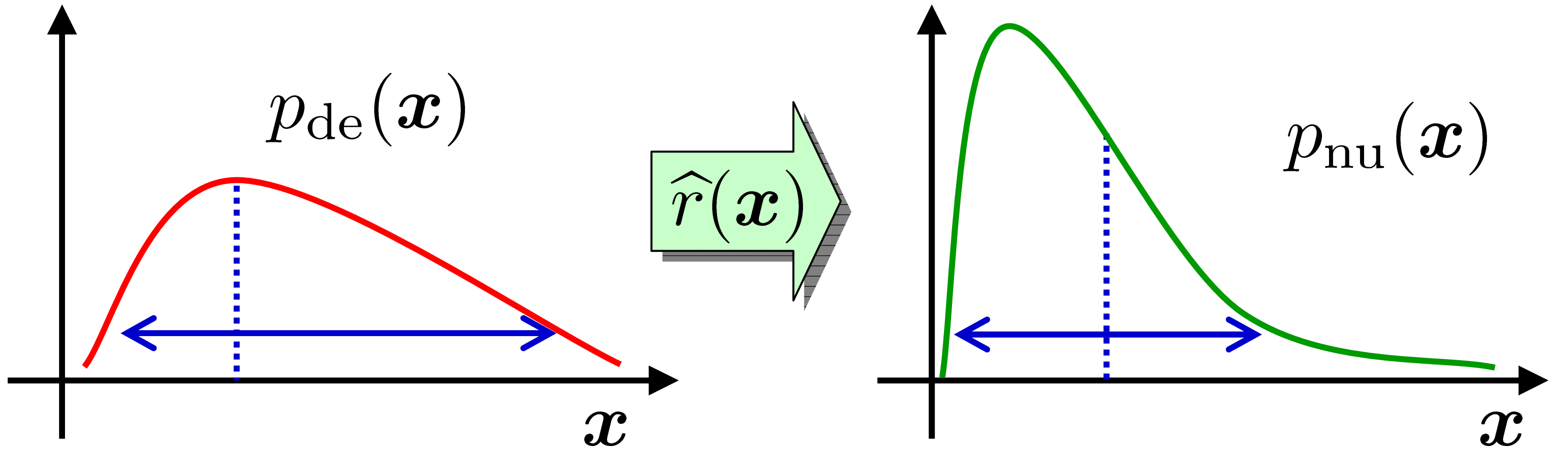
    Kanamori, Suzuki & Sugiyama (IEICE-EA2010)

発表の流れ

1. 密度比推定の応用例
2. 密度比推定手法
   - A) 確率的分類法
   - B) 積率適合法
   - C) 密度比適合法
   - D) 比較
   - E) 次元削減

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

積率適合法



- $p_{\mathrm{nu}}(\boldsymbol{x})$と $\widehat{p}_{\mathrm{nu}}(\boldsymbol{x}) = \widehat{r}(\boldsymbol{x})p_{\mathrm{de}}(\boldsymbol{x})$ の積率を適合する.
  - 例：平均（1次の積率）の適合

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

$$\min_{\widehat{r}} \left\| \int \boldsymbol{x}\widehat{r}(\boldsymbol{x})p_{\mathrm{de}}(\boldsymbol{x})d\boldsymbol{x} - \int \boldsymbol{x}p_{\mathrm{nu}}(\boldsymbol{x})d\boldsymbol{x} \right\|^2$$

カーネル平均適合

Huang, Smola, Gretton, Borgwardt & Schölkopf (NIPS2006)

- 有限個の積率を適合しても真の密度比が得られるとは限らない.
- カーネル平均適合:
  - ガウス再生核ヒルベルト空間内で平均を適合することにより，全ての次数の積率を効率よく適合できる.

$$\min_{\widehat{r}\in\mathcal{H}} \left\| \int K(\boldsymbol{x},\cdot)p_{\mathrm{nu}}(\boldsymbol{x})d\boldsymbol{x} - \int K(\boldsymbol{x},\cdot)\widehat{r}(\boldsymbol{x})p_{\mathrm{de}}(\boldsymbol{x})d\boldsymbol{x} \right\|_{\mathcal{H}}^2$$

  - 二次計画問題の解として，$\{r(\boldsymbol{x}_j^{\mathrm{de}})\}_{j=1}^{n_{\mathrm{de}}}$ の推定値が得られる.
  - 密度比関数全体の推定や，様々な損失関数にも拡張できる.

Kanamori, Suzuki & Sugiyama (arXiv2009)

発表の流れ

1. 密度比推定の応用例
2. 密度比推定手法
   - A) 確率的分類法
   - B) 積率適合法
   - C) 密度比適合法
     - i. カルバック･ライブラー距離
     - ii. 二乗距離
   - D) 比較
   - E) 次元削減

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

密度比適合法

Sugiyama, Suzuki & Kanamori (RIMS2010)

- ブレグマン距離(BR)のもとで，密度比のモデル $r(\boldsymbol{x})$ を真の密度比 $\widehat{r}(\boldsymbol{x})$ に適合させる：

$$\mathrm{BR}_f\,(\widehat{r}) = \int p_{\mathrm{de}}(\boldsymbol{x})\nabla f(\widehat{r}(\boldsymbol{x}))\widehat{r}(\boldsymbol{x})d\boldsymbol{x} - \int p_{\mathrm{de}}(\boldsymbol{x})f(\widehat{r}(\boldsymbol{x}))d\boldsymbol{x}$$

$$-\int p_{\mathrm{nu}}(\boldsymbol{x})\nabla f(\widehat{r}(\boldsymbol{x}))d\boldsymbol{x}$$ （定数項は無視）

- BRの標本近似は次式で与えられる：

$$\widehat{\mathrm{BR}}_f\,(\widehat{r}) = \frac{1}{n_{\mathrm{de}}}\sum_{j=1}^{n_{\mathrm{de}}}\nabla f(\widehat{r}(\boldsymbol{x}_j^{\mathrm{de}}))\widehat{r}(\boldsymbol{x}_j^{\mathrm{de}}) - \frac{1}{n_{\mathrm{de}}}\sum_{j=1}^{n_{\mathrm{de}}} f(\widehat{r}(\boldsymbol{x}_j^{\mathrm{de}}))$$

$$-\frac{1}{n_{\mathrm{nu}}}\sum_{i=1}^{n_{\mathrm{nu}}}\nabla f(\widehat{r}(\boldsymbol{x}_i^{\mathrm{nu}}))$$

発表の流れ

1. 密度比推定の応用例
2. 密度比推定手法
   A) 確率的分類法
   B) 積率適合法
   C) 密度比適合法
      i. カルバック・ライブラー距離
      ii. 二乗距離
   D) 比較
   E) 次元削減

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

カルバック・ライブラー距離(KL)

Sugiyama, Nakajima, Kashima, von Bünau & Kawanabe (NIPS2007)
Nguyen, Wainwright & Jordan (NIPS2007)

- $f(t) = t \log t - t$ とおけば, BRは

$$\widehat{\mathrm{KL}}(\widehat{r}) = \frac{1}{n_{\mathrm{de}}} \sum_{j=1}^{n_{\mathrm{de}}} \widehat{r}(\boldsymbol{x}_j^{\mathrm{de}}) - \frac{1}{n_{\mathrm{nu}}} \sum_{i=1}^{n_{\mathrm{nu}}} \log \widehat{r}(\boldsymbol{x}_i^{\mathrm{nu}})$$

- 線形モデル $\widehat{r}(\boldsymbol{x}) = \sum_{\ell=1}^{b} \alpha_\ell \phi_\ell(\boldsymbol{x})$ に対しては,

$\phi_\ell(\boldsymbol{x}) \geq 0$

KL最小化は**凸最適化**になる. （例：ガウスカーネル）

KL法の性質

Nguyen, Wainwright & Jordan (NIPS2007)
Sugiyama, Suzuki, Nakajima, Kashima, von Bünau & Kawanabe (AISM2008)

■ パラメトリックモデルの場合： $\widehat{r}(\boldsymbol{x}) = \sum_{\ell=1}^{b} \alpha_\ell \phi_\ell(\boldsymbol{x})$

- 学習したパラメータは $1/\sqrt{n}$ の速さで最適値に収束.
- 最適な収束率を達成している.

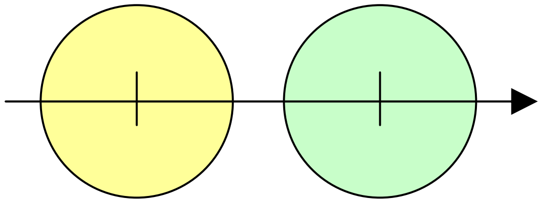
$n = \min(n_{\mathrm{nu}}, n_{\mathrm{de}})$

■ ノンパラメトリックモデルの場合： $\widehat{r}(\boldsymbol{x}) = \sum_{\ell=1}^{n_{\mathrm{nu}}} \alpha_\ell K(\boldsymbol{x}, \boldsymbol{x}_\ell^{\mathrm{nu}})$

- 学習した関数は $1/\sqrt{n}$ より少し遅い速さで真の関数に収束（被覆数やブラケットエントロピーに依存）.
- 最適な収束率を達成している.

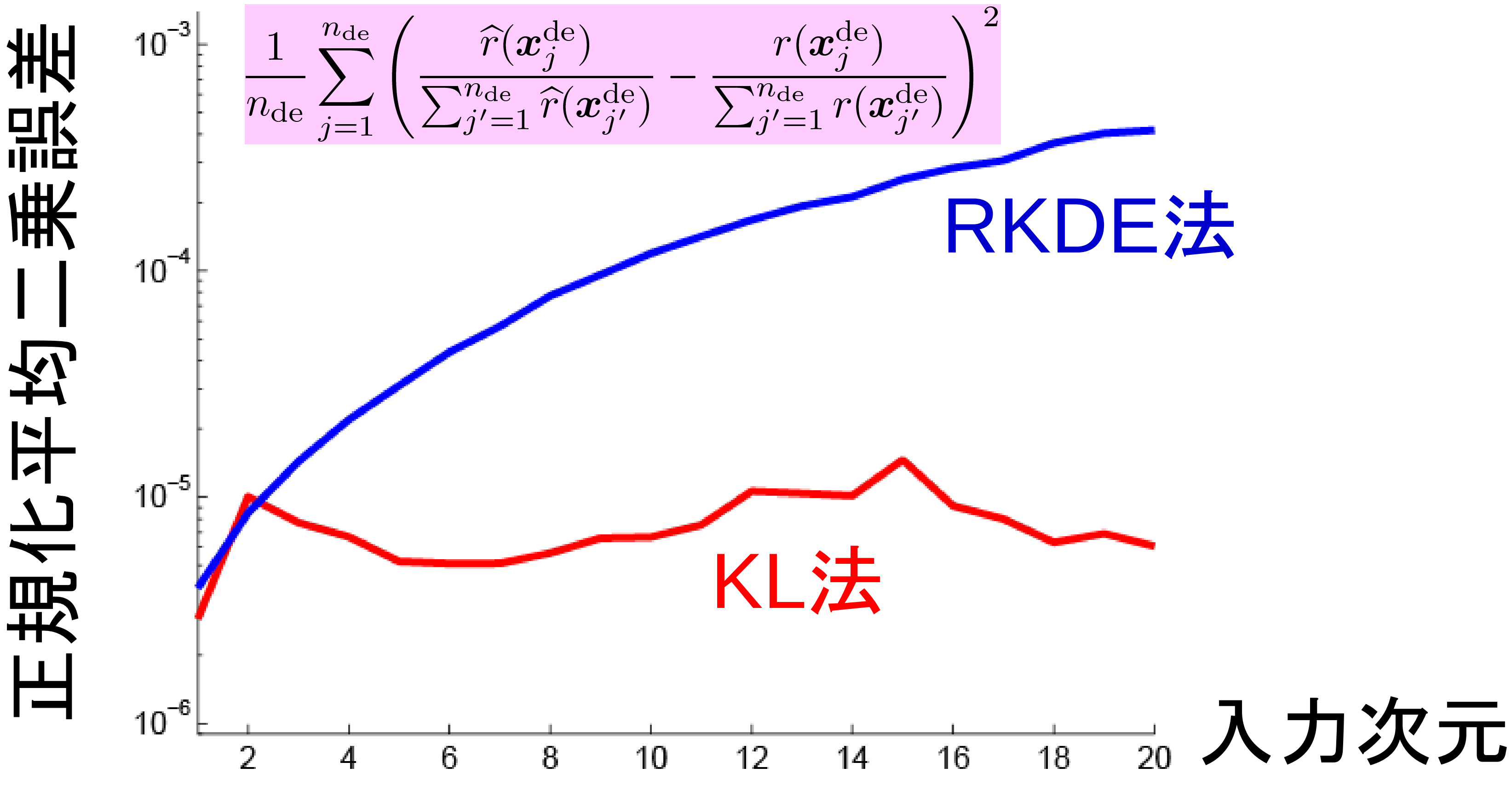
数値例（設定）

- d次元の正規分布，共分散は単位行列，平均は
  - 分母： (0,0,0,…,0)
  - 分子： (1,0,0,…,0)



- カーネル密度推定の比(RKDE法)：
  - 分母と分子の密度をそれぞれ推定し，比を取る.
  - ガウス幅は交差確認法で決定する.
- KL法：
  - 密度比を直接推定する.
  - ガウス幅は交差確認法で決定する.

数値例(結果)



- RKDE法: “次元の呪い”
- KL法: 高次元でも精度が良い

KL法の拡張

- KL法は，線形モデル以外に
  - 線形対数モデル

    Tsuboi, Kashima, Hido, Bickel & Sugiyama (JIP2009)
  - ガウス混合モデル

    Yamada & Sugiyama (IEICE-ED2009)
  - 確率的主成分分析混合モデル

    Yamada, Sugiyama, Wichern & Jaak (Submitted)

にも拡張されている.

発表の流れ

1. 密度比推定の応用例
2. 密度比推定手法
   - A) 確率的分類法
   - B) 積率適合法
   - C) 密度比適合法
     - i. カルバック･ライブラー距離
     - ii. 二乗距離
   - D) 比較
   - E) 次元削減

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

二乗距離(SQ)

Kanamori, Hido & Sugiyama (JMLR2009)

- $f(t)=(t-1)^2/2$ とおけば，BRは

$$\widehat{\mathrm{SQ}}\,(\widehat{r})=\frac{1}{2n_{\mathrm{de}}}\sum_{j=1}^{n_{\mathrm{de}}}\widehat{r}(\boldsymbol{x}_j^{\mathrm{de}})^2-\frac{1}{n_{\mathrm{nu}}}\sum_{i=1}^{n_{\mathrm{nu}}}\widehat{r}(\boldsymbol{x}_i^{\mathrm{nu}})$$

- 線形モデル $\widehat{r}(\boldsymbol{x})=\sum_{\ell=1}^{b}\alpha_\ell\phi_\ell(\boldsymbol{x})$ に対しては，

SQ最小化は解析的に行える： $\widehat{\boldsymbol{\alpha}}=(\widehat{\boldsymbol{H}}+\lambda\boldsymbol{I})^{-1}\widehat{\boldsymbol{h}}$

$$\widehat{\boldsymbol{H}}=\frac{1}{n_{\mathrm{de}}}\sum_{j=1}^{n_{\mathrm{de}}}\boldsymbol{\phi}(\boldsymbol{x}_j^{\mathrm{de}})\boldsymbol{\phi}(\boldsymbol{x}_j^{\mathrm{de}})^\top \qquad \widehat{\boldsymbol{h}}=\frac{1}{n_{\mathrm{nu}}}\sum_{i=1}^{n_{\mathrm{nu}}}\boldsymbol{\phi}(\boldsymbol{x}_i^{\mathrm{nu}})$$

- 計算効率が非常に良い！

SQ法の特徴

Kanamori, Hido & Sugiyama (JMLR2009)
Kanamori, Suzuki & Sugiyama (arXiv2009)

- パラメトリックモデル，ノンパラメトリックモデル両方に対して，最適な収束率を達成する.
- 一つ抜き交差確認が解析的に行える.
- 条件数が最小である.
- 解析解が求まることにより，相互情報量推定量の勾配が計算できる:
  - 十分次元削減
  - 独立成分分析
  - 因果推定

発表の流れ

1. 密度比推定の応用例
2. 密度比推定手法
   - A) 確率的分類法
   - B) 積率適合法
   - C) 密度比適合法
   - D) 比較
   - E) 次元削減

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

密度比推定法の比較

手法	密度推定	分母・分子の定義域	モデル選択	計算時間
RKDE	含む	違ってもよい	可能	超高速
LR	含まない	同じ	可能	遅い
KMM	含まない	同じ	不可能	遅い
KL	含まない	違ってもよい	可能	遅い
SQ	含まない	違ってもよい	可能	高速

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x})}$$

- 実用的にはSQが最も優れているであろう.

発表の流れ

1. 密度比推定の応用例
2. 密度比推定手法
   - A) 確率的分類法
   - B) 積率適合法
   - C) 密度比適合法
   - D) 比較
   - E) 次元削減

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

次元削減付き密度比直接推定 (D^3; D-cube)

- 密度推定を経由せず，密度比を直接する手法は有望である.
- しかし，高次元の場合は密度比推定はやはり困難である.
- 密度比推定と次元削減を組み合わせる.

異分布部分空間

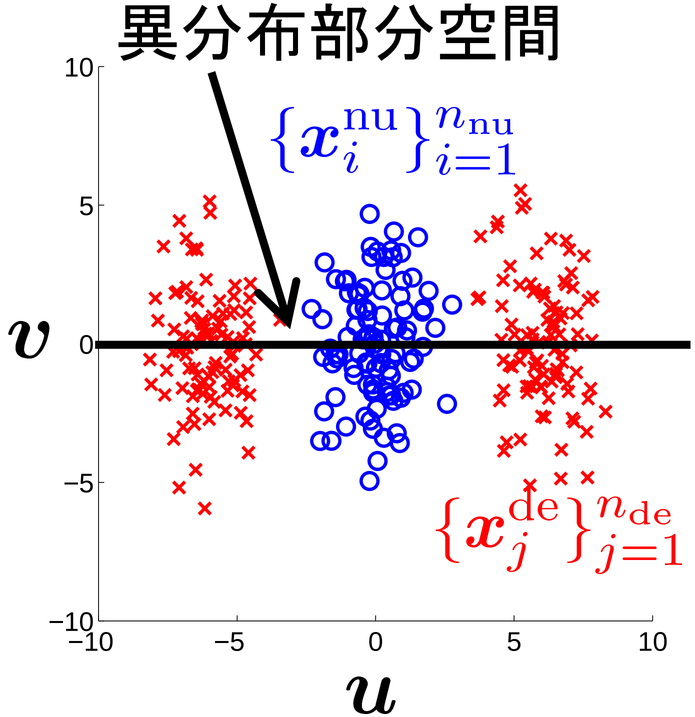
- 仮定: $p_{\mathrm{nu}}(\boldsymbol{x})$ と $p_{\mathrm{de}}(\boldsymbol{x})$ は, ある部分空間(異分布部分空間)内でのみ異なる

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{nu}}(\boldsymbol{x})}{p_{\mathrm{de}}(\boldsymbol{x})}$$

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix} \boldsymbol{x}$$

$$= \frac{p(\boldsymbol{v}|\boldsymbol{u})p_{\mathrm{nu}}(\boldsymbol{u})}{p(\boldsymbol{v}|\boldsymbol{u})p_{\mathrm{de}}(\boldsymbol{u})} = \frac{p_{\mathrm{nu}}(\boldsymbol{u})}{p_{\mathrm{de}}(\boldsymbol{u})}$$

$\begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix}$:フルランクで直交

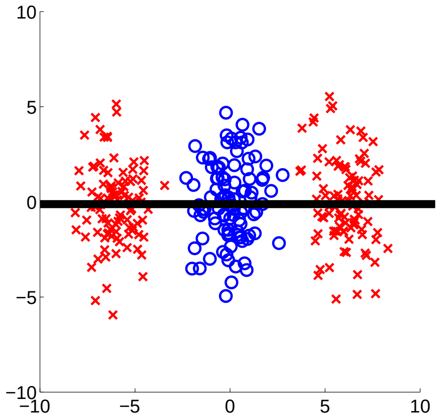


- 低次元の異分布部分空間内のみで密度比を推定できる.

教師付き次元削減による異分布部分空間の探索

Sugiyama, Kawanabe & Chui (NN2010)

- 異分布空間内では $p_{\mathrm{nu}}(\boldsymbol{x})$ と $p_{\mathrm{de}}(\boldsymbol{x})$ が異なるため，$\{\boldsymbol{x}_i^{\mathrm{nu}}\}_{i=1}^{n_{\mathrm{nu}}}$ と $\{\boldsymbol{x}_j^{\mathrm{de}}\}_{j=1}^{n_{\mathrm{de}}}$ は分離可能であろう.
- 任意の教師付き次元削減法によって，異分布部分空間を見つけることができるであろう.
- 局所フィッシャー判別分析(LFDA):
  - クラス内多峰性を扱うことができる.
  - 部分空間の次元数に上限がない.
  - 解析解が得られる.



Sugiyama (JMLR2007)

教師付き次元削減アプローチの利点と欠点

- 利点:
  - SQ+LFDAによってD^{3}の解も解析的に求められる.
- 欠点:
  - $\boldsymbol{u}$ と $\boldsymbol{v}$ が独立という理論的制限がある.

$$p(\boldsymbol{v}|\boldsymbol{u}) = p(\boldsymbol{v}) \qquad \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix} \boldsymbol{x}$$

  - 分離可能性を高めることにより異分布部分空間が見つかるとは限らない(例:重なりのある異なる分布)

分離可能 $\underset{\not\Leftarrow}{\Rightarrow}$ 分布が異なる

異分布部分空間の特徴付け

Sugiyama, Hara, von Bünau, Suzuki, Kanamori & Kawanabe (SDM2010)

- 異分布部分空間はピアソン距離(PE)の最大解:

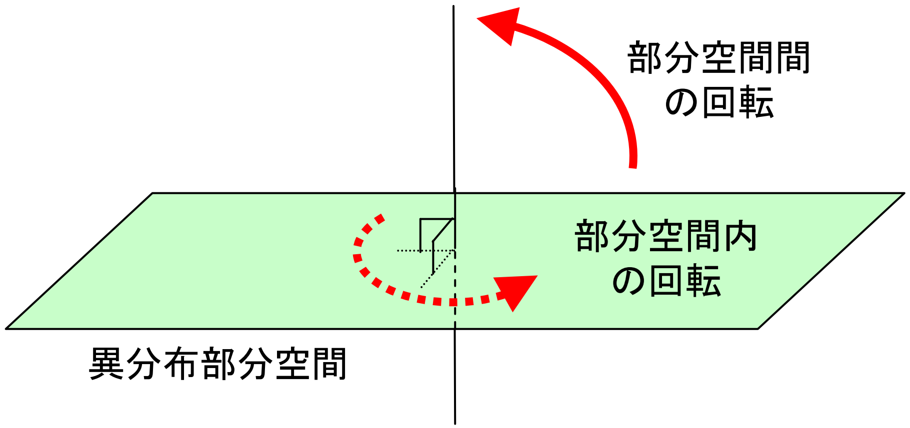
$$\mathrm{PE}[p_{\mathrm{nu}}(\boldsymbol{u}), p_{\mathrm{de}}(\boldsymbol{u})] = \int \left(\frac{p_{\mathrm{nu}}(\boldsymbol{u})}{p_{\mathrm{de}}(\boldsymbol{u})} - 1 \right)^2 p_{\mathrm{de}}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$$

- 従って, PEを $\boldsymbol{U}$ に関して最大化すれば, 異分布部分空間が求まる.
- PEはSQ法によって解析的に近似できる.

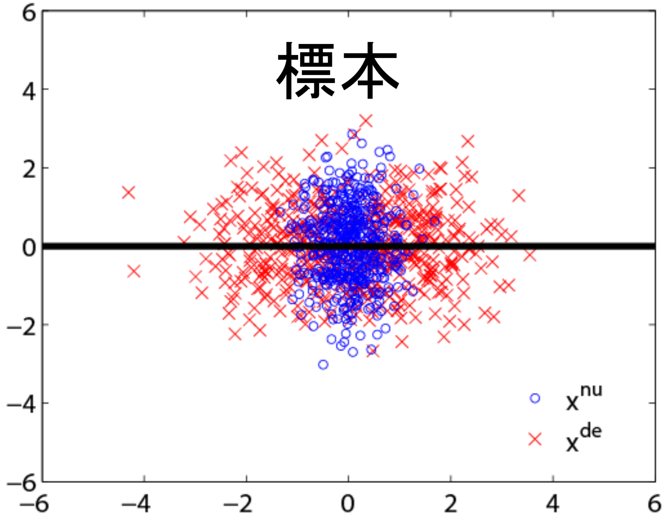
$$\widehat{\mathrm{PE}} = \frac{1}{n_{\mathrm{de}}} \sum_{j=1}^{n_{\mathrm{de}}} \left(\widehat{r}(\boldsymbol{u}_j^{\mathrm{de}}) - 1 \right)^2 \qquad \boldsymbol{u}_j^{\mathrm{de}} = \boldsymbol{U}\boldsymbol{x}_j^{\mathrm{de}}$$

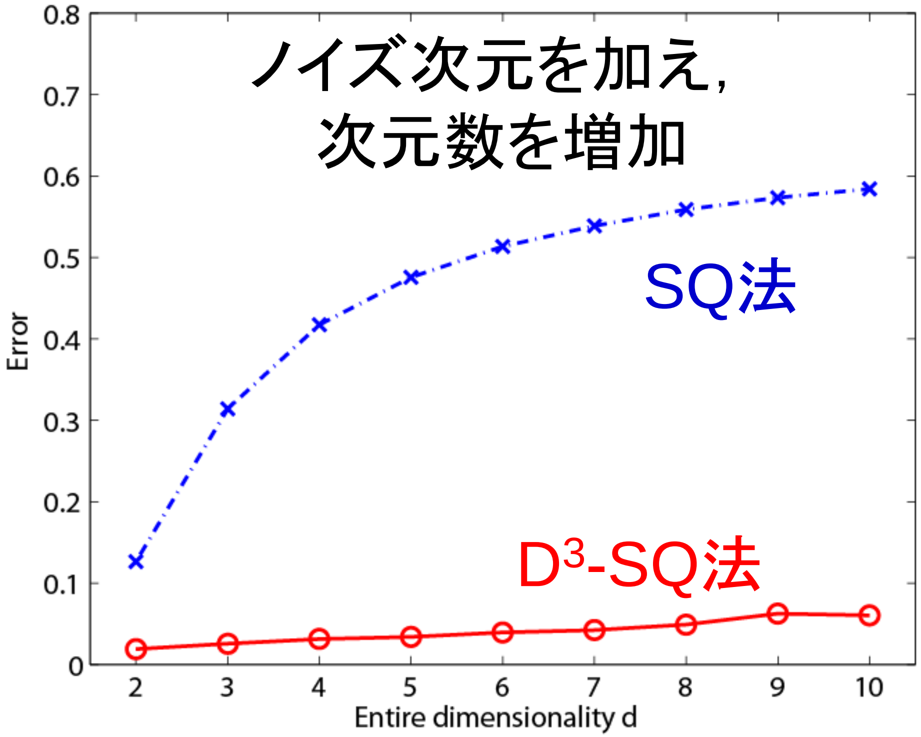
異分布部分空間の直接探索

- 勾配法:
- 自然勾配法: Amari (NeuralComp1998)
  - スティーフェル多様体の構造を利用して収束性を改善
- ギブンズ回転:
  - 二つの軸を選び，その平面内で部分空間を回転
- 部分空間回転: Plumbley (NeuroComp2005)
  - 部分空間内での回転を無視することにより，計算効率を改善

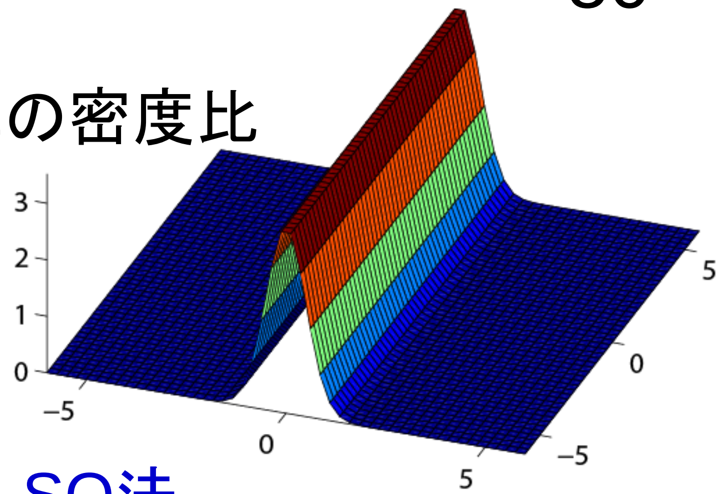


数値例

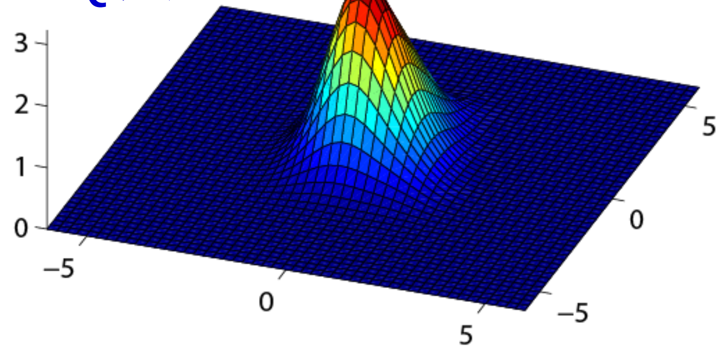




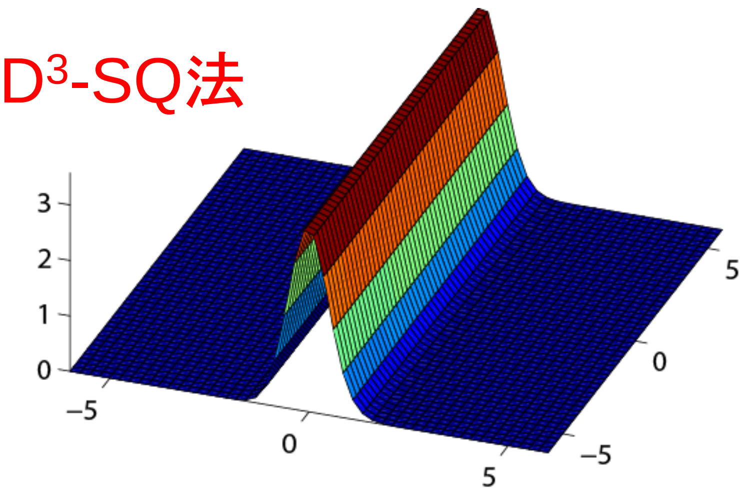
真の密度比



SQ法



D^3-SQ法



謝辞

■ 共変量シフト適応：

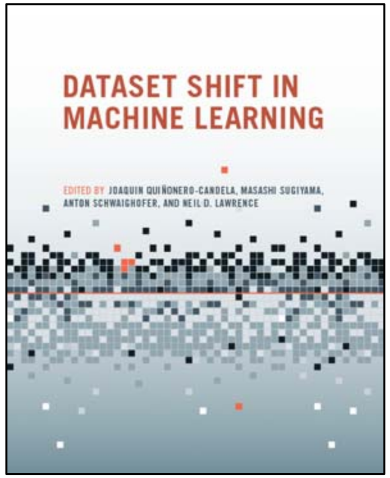
- 科研費若手(B), 基盤(B), 若手(A)
- ドイツ･Alexander von Humboldt財団
- 財団法人 大川情報通信基金
- ドイツ･Mathematisches Forshungsinstitut Oberwolfach (MFO)
- マイクロソフト産学連携研究機構CORE連携研究プロジェクト
- 米国･IBM Faculty Award

■ 密度比推定：

- JFE21世紀財団
- 米国･アジア宇宙航空研究開発事務所(AOARD)
- 財団法人テレコム先端技術研究支援センター(SCAT)
- 科学技術振興機構 さきがけ研究（知の創生と情報社会）

書籍

- Quiñonero-Candela, Sugiyama, Schwaighofer & Lawrence (Eds.), Dataset Shift in Machine Learning, MIT Press, 2009.



- Sugiyama, von Bünau, Kawanabe & Müller, Covariate Shift Adaptation in Machine Learning, MIT Press（近日出版予定）

- Sugiyama, Suzuki & Kanamori, Density Ratio Estimation in Machine Learning, Cambridge University Press（近日出版予定）

解説論文

- 密度比推定の枠組み:
  - 杉山，密度比に基づく機械学習の新たなアプローチ，統計数理，2010.
  - Sugiyama, Kanamori, Suzuki, Hido, Sese, Takeuchi & Wang. A density-ratio framework for statistical data processing. IPSJ Transactions on Computer Vision and Applications, vol.1, pp.183-208, 2009.
- 密度比推定法:
  - Sugiyama, Suzuki & Kanamori. Density ratio estimation: A comprehensive review. 京都大学数理解析研究所講究録 “Statistical Experiment and Its Related Topics”, 2010.
- 共変量シフト適応:
  - 杉山，非定常環境下での教師付き学習:データの入力分布が変化する場合，画像ラボ, vol.18, no.10, pp.1-6, 2007.
  - 杉山，共変量シフト下での教師付き学習，日本神経回路学会誌, vol.13, no.3, pp.111-118, 2006.
- 全ての論文は以下のホームページからダウンロードできます:

http://sugiyama-www.cs.titech.ac.jp/~sugi

まとめ：密度比の世界

実問題応用例：
ブレイン・コンピュータインターフェース，ロボット制御，音声認識，
画像認識，自然言語処理，バイオインフォマティクス，データマイニング

機械学習アルゴリズム：
重点サンプリング（共変量シフト適応，ドメイン適応，多タスク学習），
二標本問題（二標本検定，外れ値検出，変化点検知），
相互情報量推定（独立性検定，変数選択，独立成分分析，
次元削減，因果推定）
条件付き確率推定（可視化，状態遷移推定，確率的パターン認識），

密度比推定法：
基本アルゴリズム（LR法，KMM法，KL法，SQ法），
大規模対応，高次元対応，安定化，ロバスト化

理論解析：
収束性解析（確率論），情報量規準（統計学），安定性解析（最適化）