第2回「e-サイエンスに向けた革新的アルゴリズム」シンポジウム

2011年11月22日

確率密度比を用いた機械学習アルゴリズムとその応用

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

自己紹介

- 東京工業大学　計算工学専攻　准教授
- 研究室構成員:
  - ポスドク4名，共同研究員2名
  - 博士課程8名，修士課程6名，学部生2名
  - 学生は半数以上が外国人！
- 研究テーマ:機械学習
  - 理論解析(主に統計的)
  - 汎用的なアルゴリズム開発
  - 実世界応用(主に企業と共同で)

機械学習と確率分布推定

- 目的：データの背後に潜む知識を自動的に獲得
  - パターン認識，予測，異常検出，因果推論など
- データを生成する規則（確率分布）を推定すれば，あらゆる機械学習タスクが解決できる！

データの生成規則を知る ⇔ データの全てを知る

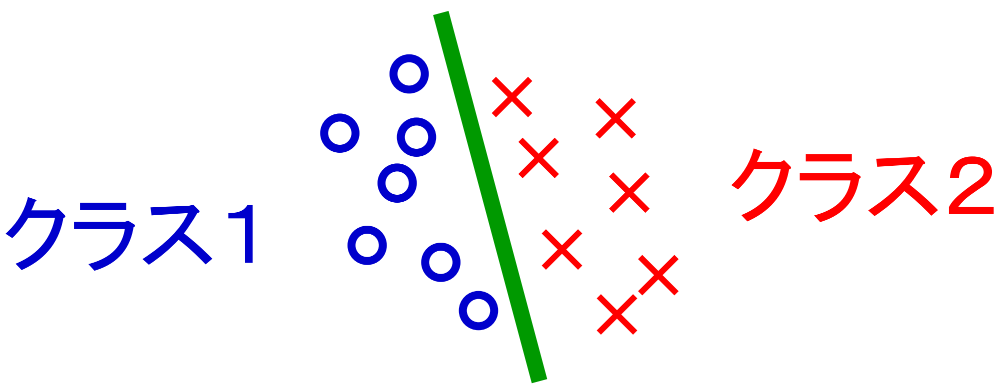
- しかし，確率分布の推定は最も難しい問題の一つ
- ひとたび確率分布の推定を行うと，推定精度が大きく低下してしまう

機械学習と確率分布推定

- 確率分布を推定することなく，解きたいタスクを解決することが望ましい

- 例：サポートベクトルマシン（パターン認識器）
  - データの生成確率分布を推定せず，パターン認識に必要な，クラス間の分離境界だけを学習する

Cortes & Vapnik (ML1995)



確率分布推定と機械学習

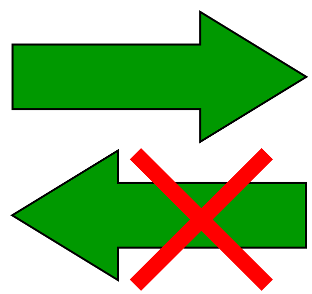
- 機械学習分野には様々なタスクがある:
  - 非定常環境下での適応学習, ドメイン適応, マルチタスク学習
  - 二標本検定, 異常値検出, 時系列の変化点検知
  - 相互情報量推定, 独立性検定, 特徴選択, 十分次元削減, 独立成分分析, 因果推論, クラスタリング, オブジェクト適合
  - 条件付き確率推定, 確率的パターン認識
- 各々のタスクに対して, 確率分布の推定を避けることのできるアルゴリズムを個別に開発することは困難

確率密度比に基づく 新たな機械学習アプローチ

- 前頁の機械学習タスクは複数の確率分布を含む
- 確率密度関数の比を推定することにする
  - その際，確率密度は推定しない
  - 密度を求めるよりも，密度比を求めるほうが易しい！

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

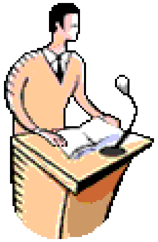
が分かる



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

が分かる

- 密度比推定により，前頁の機械学習タスクが全て解決できる！



発表の流れ

1. 密度比推定に基づく機械学習の枠組み
2. 密度比推定法
3. 密度比推定の応用事例
   A) 重点サンプリング
   B) 確率分布比較
   C) 相互情報量推定
   D) 条件付き確率推定
4. まとめ

最小二乗密度比適合(LSIF)

Kanamori, Hido & Sugiyama
(NIPS2008, JMLR2009)

- 真の密度比 $r(\boldsymbol{x})$ とそのモデル $\widehat{r}(\boldsymbol{x})$ の
  二乗誤差を最小にするように学習：

$$\min_{\widehat{r}} \left[J(\widehat{r}) \right]$$

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

$$J(\widehat{r}) = \frac{1}{2} \int \left(\widehat{r}(\boldsymbol{x}) - r(\boldsymbol{x}) \right)^2 p_{\mathrm{de}}(\boldsymbol{x}) d\boldsymbol{x}$$

$$= \frac{1}{2} \int \left(\widehat{r}(\boldsymbol{x}) \right)^2 p_{\mathrm{de}}(\boldsymbol{x}) d\boldsymbol{x} - \int \widehat{r}(\boldsymbol{x}) p_{\mathrm{nu}}(\boldsymbol{x}) d\boldsymbol{x} + C$$

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

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

アルゴリズム

- 線形密度比モデルを用いる:

$$\widehat{r}(\boldsymbol{x}) = \sum_{\ell=1}^{b} \alpha_\ell \phi_\ell(\boldsymbol{x}) = \boldsymbol{\alpha}^\top \boldsymbol{\phi}(\boldsymbol{x})$$

$\phi_\ell(\boldsymbol{x})$: 基底関数
（例えばガウスカーネル）

- 最適化規準:

$$\min_{\boldsymbol{\alpha}} \left[\frac{1}{2} \boldsymbol{\alpha}^\top \widehat{\boldsymbol{H}} \boldsymbol{\alpha} - \widehat{\boldsymbol{h}}^\top \boldsymbol{\alpha} + \frac{\lambda}{2} \boldsymbol{\alpha}^\top \boldsymbol{\alpha} \right]$$

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

$$\widehat{\boldsymbol{h}} = \frac{1}{n_{\mathrm{nu}}} \sum_{i=1}^{n_{\mathrm{nu}}} \boldsymbol{\phi}(\boldsymbol{x}_i^{\mathrm{nu}})$$

- 大域的最適解が解析的に計算可能:

$$\widehat{\boldsymbol{\alpha}} = (\widehat{\boldsymbol{H}} + \lambda \boldsymbol{I})^{-1} \widehat{\boldsymbol{h}}$$

  - 大規模な問題に対しても適用可能！

理論解析

■ 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

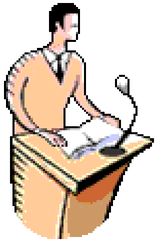
$n = \min(n_{\mathrm{nu}}, n_{\mathrm{de}})$

Kanamori, Hido & Sugiyama (JMLR2009)

■ 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

Kanamori, Suzuki & Sugiyama (ML2012)

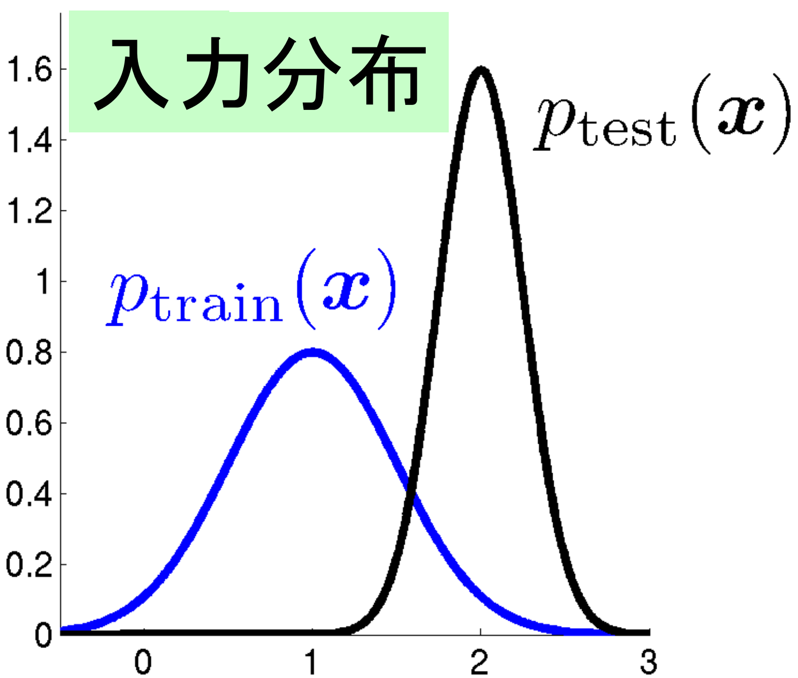


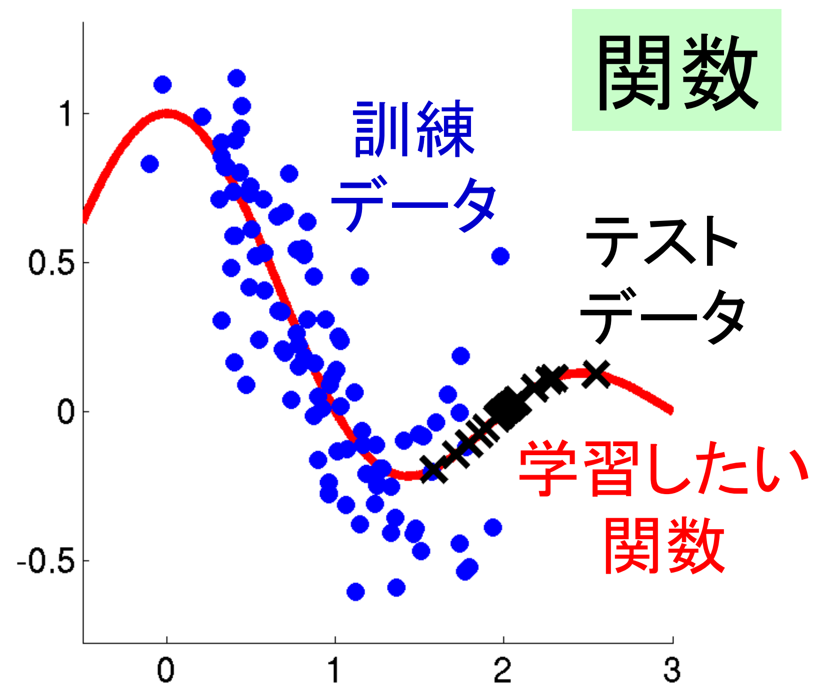
発表の流れ

1. 密度比推定に基づく機械学習の枠組み
2. 密度比推定法
3. 密度比推定の応用事例
   - A) 重点サンプリング
   - B) 確率分布比較
   - C) 相互情報量推定
   - D) 条件付き確率推定
4. まとめ

共変量シフト適応

- 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

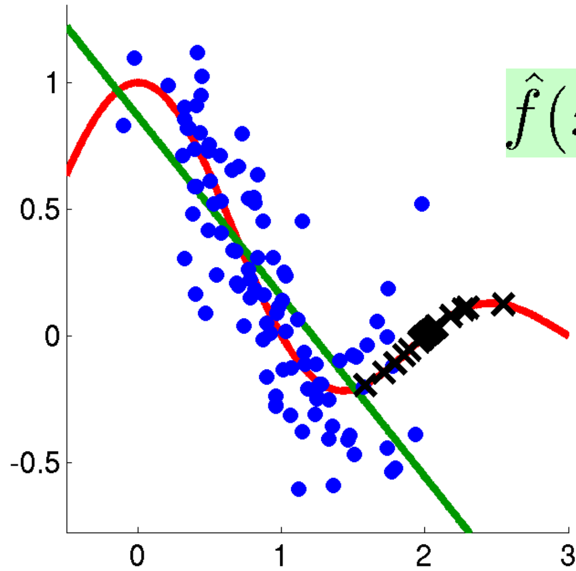




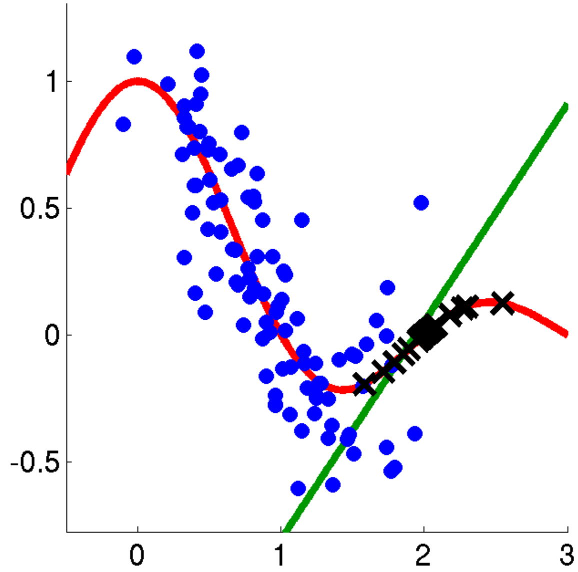
重要度重み付き最小二乗法

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

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



$$\hat{f}(x) = \alpha_1 + \alpha_2 x$$



- 共変量シフト下では，通常の最小二乗法は一致性を持たない（ $n \to \infty$ でも最適解に収束しない）

- 共変量シフト下でも一致性を持つ
- 様々な学習法に適用可能：
  - サポートベクトルマシン，ロジスティック回帰，条件付き確率場など

実世界応用例

- 顔画像からの年齢予測:
  - 照明環境の変化

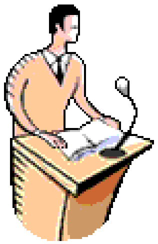
    Ueki, Sugiyama & Ihara (IEICE-ED2011)
- 話者認識:
  - 声質の変化

    Yamada, Sugiyama & Matsui (SigPro2010)
- テキスト分割:
  - ドメイン適応

    Tsuboi, Kashima, Hido, Bickel & Sugiyama (JIP2008)
- ブレイン・コンピュータインターフェース:
  - 心理状態の変化

    Sugiyama, Krauledat & Müller (JMLR2007)

    Li, Kambara, Koike & Sugiyama (IEEE-TBE2010)



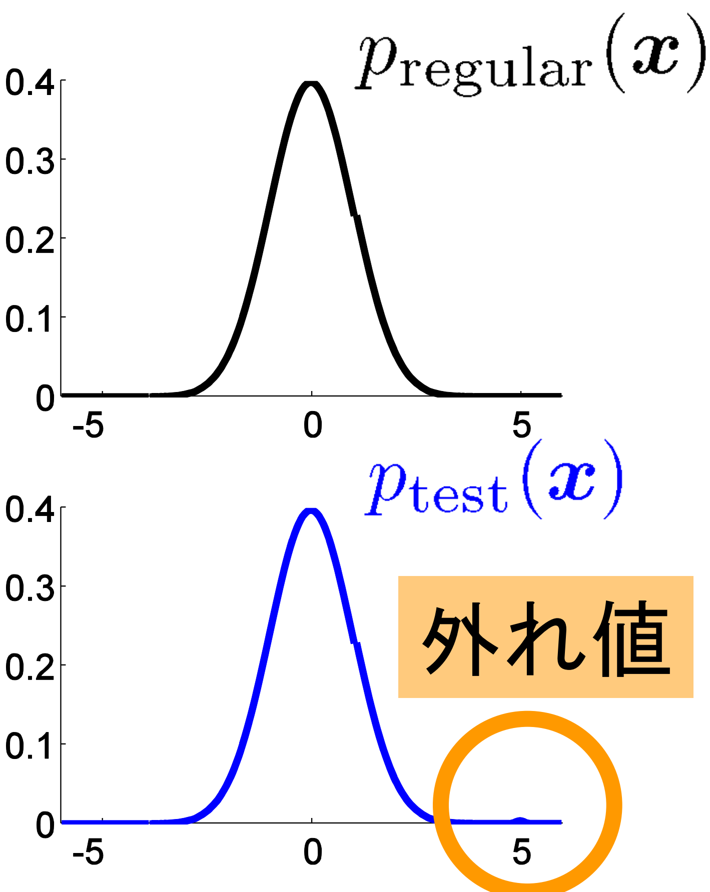
発表の流れ

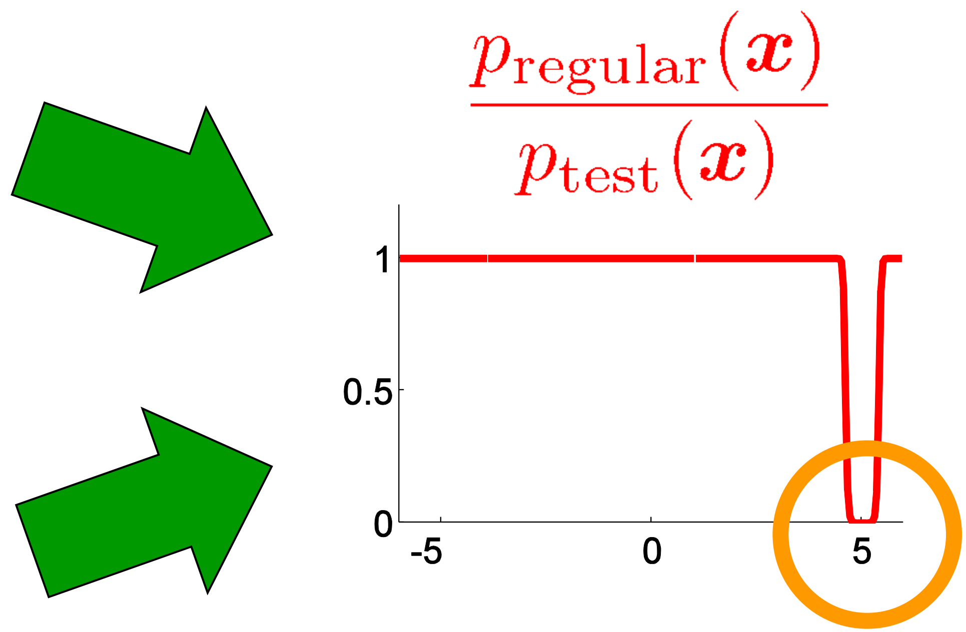
1. 密度比推定に基づく機械学習の枠組み
2. 密度比推定法
3. 密度比推定の応用事例
   A) 重点サンプリング
   B) 確率分布比較
   C) 相互情報量推定
   D) 条件付き確率推定
4. まとめ

正常値に基づく異常値検出

Hido, Tsuboi, Kashima, Sugiyama & Kanamori (ICDM2008, KAIS2011)
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．





正常データを有効活用することにより，高精度な解が得られる

実世界応用例

- 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 (KAIS2011)

- 生体データからの睡眠診断

  Kawahara & Sugiyama (SADM2011)

二標本検定

Sugiyama, Suzuki, Ito, Kanamori & Kimura (NN2011)

- 目的: 二つのデータセットの背後の確率分布が同じかどうかを検定する

$$\{\boldsymbol{x}_i\}_{i=1}^{n} \overset{i.i.d.}{\sim} P \qquad \{\boldsymbol{x}'_j\}_{j=1}^{n'} \overset{i.i.d.}{\sim} P'$$

- アプローチ: 密度比を用いて分布間の距離を推定する
  - カルバック・ライブラー距離: $\int p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{p'(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}$
  - ピアソン距離: $\int p'(\boldsymbol{x}) \left(\frac{p(\boldsymbol{x})}{p'(\boldsymbol{x})} - 1 \right)^2 \mathrm{d}\boldsymbol{x}$

実世界応用例

- 画像中の注目領域抽出:

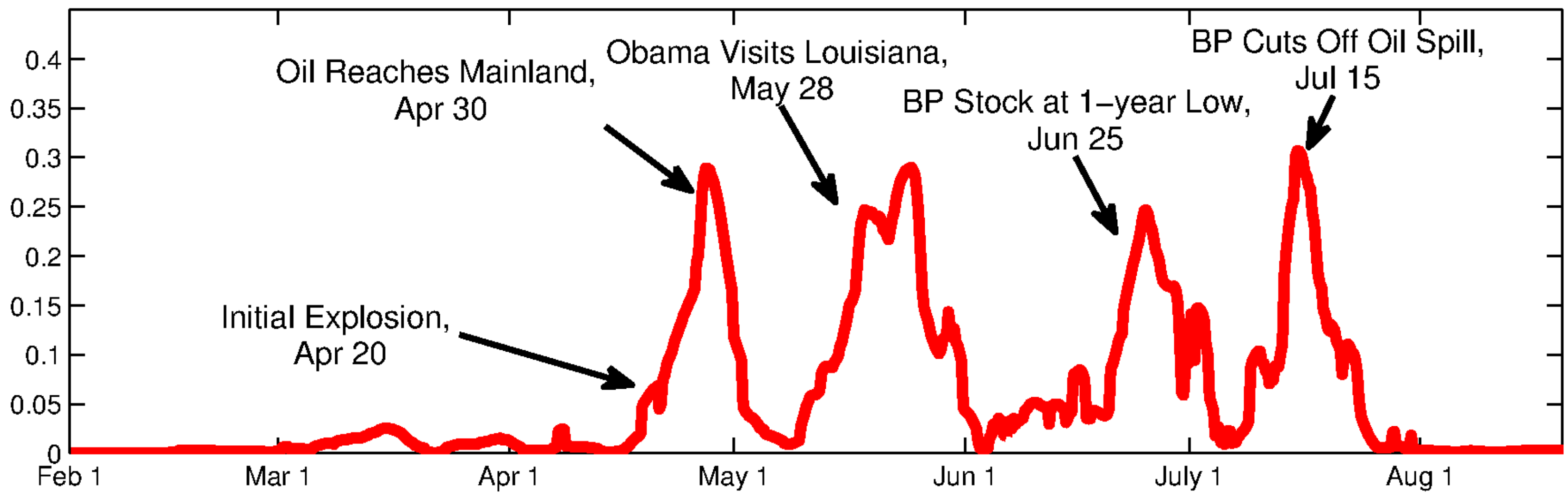
  Yamanaka, Matsugu & Sugiyama (MIRU2010)

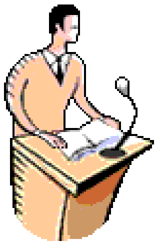
- 動画からのイベント検出:

  Yamanaka, Matsugu & Sugiyama (VECTaR2011)

- ツイッターデータ解析:

  Liu, Yamada & Sugiyama (IBIS2011)





発表の流れ

1. 密度比推定に基づく機械学習の枠組み
2. 密度比推定法
3. 密度比推定の応用事例
   A) 重点サンプリング
   B) 確率分布比較
   C) 相互情報量推定
   D) 条件付き確率推定
4. まとめ

相互情報量推定

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

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

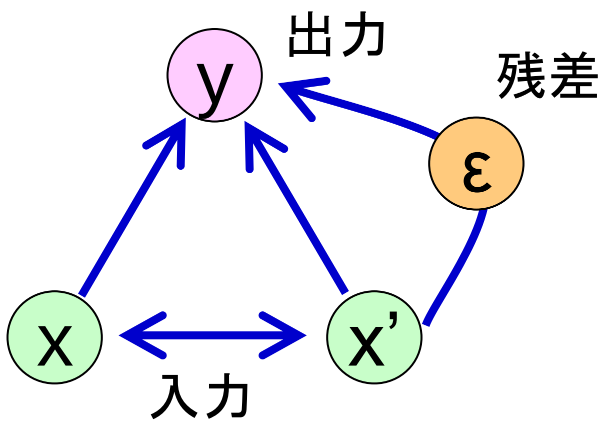
- 相互情報量は密度比を用いて計算できる

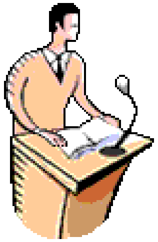
$$r(\boldsymbol{x}, \boldsymbol{y}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})} \qquad \widehat{\mathrm{MI}} = \frac{1}{n} \sum_{i=1}^{n} \log \widehat{r}(\boldsymbol{x}_i, \boldsymbol{y}_i)$$

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)
  - クラスタリング
    Kimura & Sugiyama (JACIII2011)
    Sugiyama, Yamada, Kimura & Hachiya (ICML2011)
- 入力間の独立性判定:
  - 独立成分分析
    Suzuki & Sugiyama (NeCo2011)
  - オブジェクト適合
    Yamada & Sugiyama (AISTATS2011)
- 入力と残差との独立性判定:
  - 因果推定
    Yamada & Sugiyama (AAAI2010)





発表の流れ

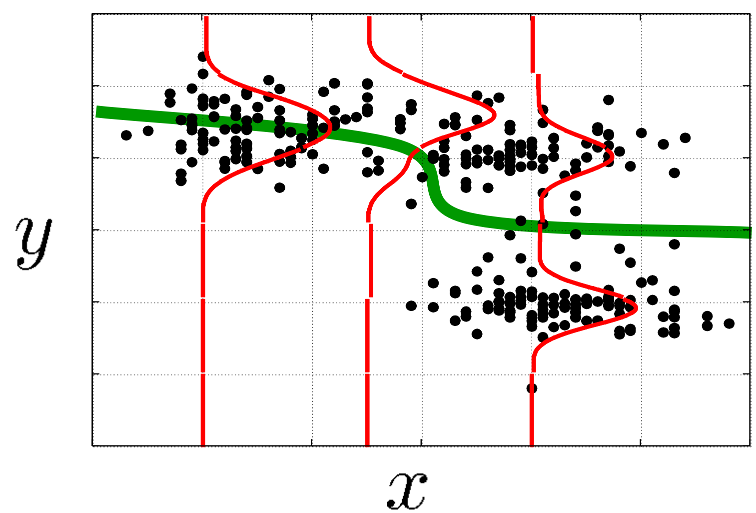
1. 密度比推定に基づく機械学習の枠組み
2. 密度比推定法
3. 密度比推定の応用事例
   A) 重点サンプリング
   B) 確率分布比較
   C) 相互情報量推定
   D) 条件付き確率推定
4. まとめ

条件付き確率密度の推定

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

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

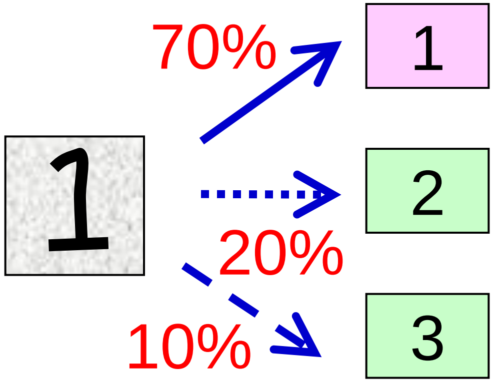
- 回帰分析：条件付き期待値の推定
- 非対称なノイズや多峰性を持つようなデータに対しては，回帰分析では不十分
- 応用例：
  - 移動ロボットの状態遷移予測



確率的パターン認識

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

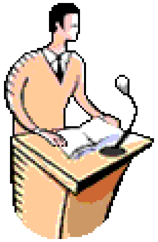
Sugiyama (IEICE-ED2010)



- 出力 y がカテゴリのとき,
  条件付き確率の推定は
  確率的なパターン認識に対応
- 応用例:
  - 顔画像からの年齢推定

    Ueki, Sugiyama, Ihara & Fujita (ACPR2011)
  - 加速度データからの行動認識

    Hachiya, Sugiyama & Ueda
    (Neurocomputing, to appear)



発表の流れ

1. 密度比推定に基づく機械学習の枠組み
2. 密度比推定法
3. 密度比推定の応用事例
   A) 重点サンプリング
   B) 確率分布比較
   C) 相互情報量推定
   D) 条件付き確率推定
4. まとめ

密度比の世界

実問題応用例:
ブレイン・コンピューターインターフェース，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

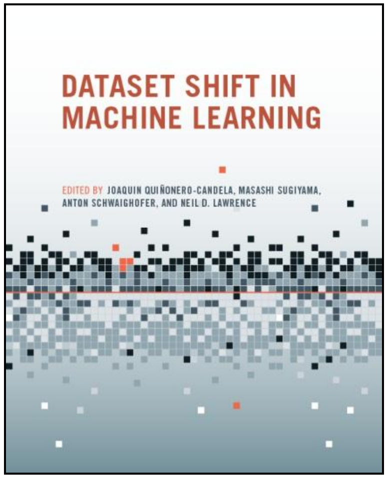
機械学習アルゴリズム:
重点サンプリング（共変量シフト適応，ドメイン適応，多タスク学習），
二標本問題（二標本検定，外れ値検出，変化点検知），
相互情報量推定（独立性検定，変数選択，独立成分分析，
次元削減，因果推定，クラスタリング，オブジェクト適合）
条件付き確率推定（可視化，状態遷移推定，確率的パターン認識），

密度比推定法:
基本アルゴリズム（LR，KMM，KLIEP，LSIF），
大規模対応，高次元対応，安定化，ロバスト化，統一化

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）

関連書籍

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



- Sugiyama & Kawanabe Machine Learning in Non-Stationary Environments, MIT Press, 2012年4月.

- Sugiyama, Suzuki & Kanamori, Density Ratio Estimation in Machine Learning, Cambridge University Press, 2012年2月.

現在進行中のプロジェクト

- 科研費若手(A)
  - 非定常環境下での機械学習
- さきがけ
  - 密度比推定に基づく機械学習
- 科研費新学術領域(代表:銅谷賢治先生)
  - 機械学習の脳科学への応用
- 最先端研究開発支援プログラム(代表:喜連川優先生)
  - 機械学習のサイバーフィジカルシステムへの応用
- 民間企業との応用研究
  - 画像,音声,自然言語,ウェブ,医療,制御,光学計測など

機械学習に関するポスドク研究員募集中!