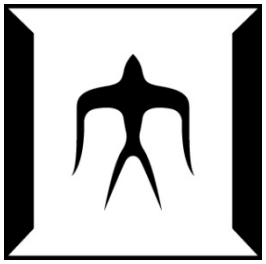
統計的機械学習



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

機械学習

- 機械学習：データの背後に潜む知識を学習する
- 様々な応用例：
  - 音声・画像・動画の認識
  - ウェブやSNSからの情報抽出
  - 商品やサービスの推薦
  - 工業製品の品質管理
  - ロボットシステムの制御
- ビッグデータ時代の到来に伴い，
  機械学習技術の重要性は
  益々高まりつつある

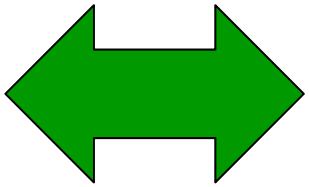
機械学習のタスク

- 機械学習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，変化点検知，クラスバランス推定
  - 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

最も汎用的なアプローチ

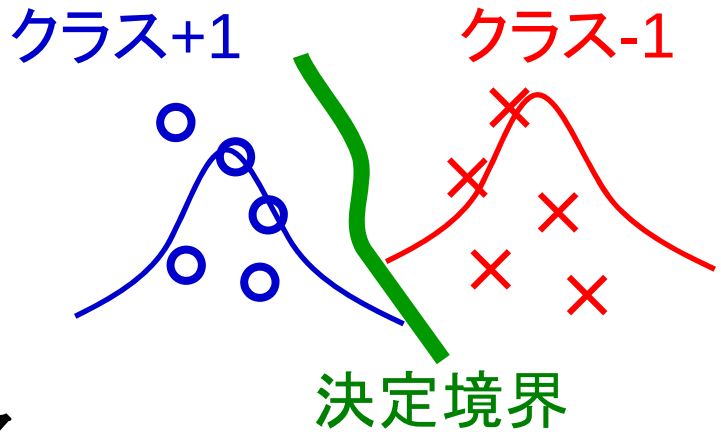
- データを生成する規則(確率分布)を推定すれば,
  あらゆる機械学習タスクが解決できる！

データの生成
規則を知る



データの
全てを知る

- 例:各クラスのデータの
  生成分布がわかれば,
  パターン認識ができる

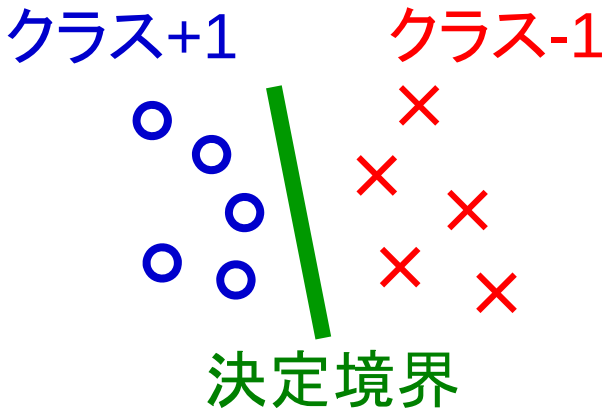


- 生成的アプローチとよばれる

各タスクに特化したアプローチ

- しかし，確率分布の推定は困難であるため，
  生成モデル推定に基づくアプローチによって，
  必ずしも高い学習精度が得られるとは限らない

- 確率分布の推定を行わず，各タスクを直接解く
  - 例：サポートベクトルマシンでは，
    各クラスのデータ生成分布
    を推定せず，パターン認識に
    必要な決定境界のみを学習
  - パターン認識に対しては，
    識別的アプローチとよばれる



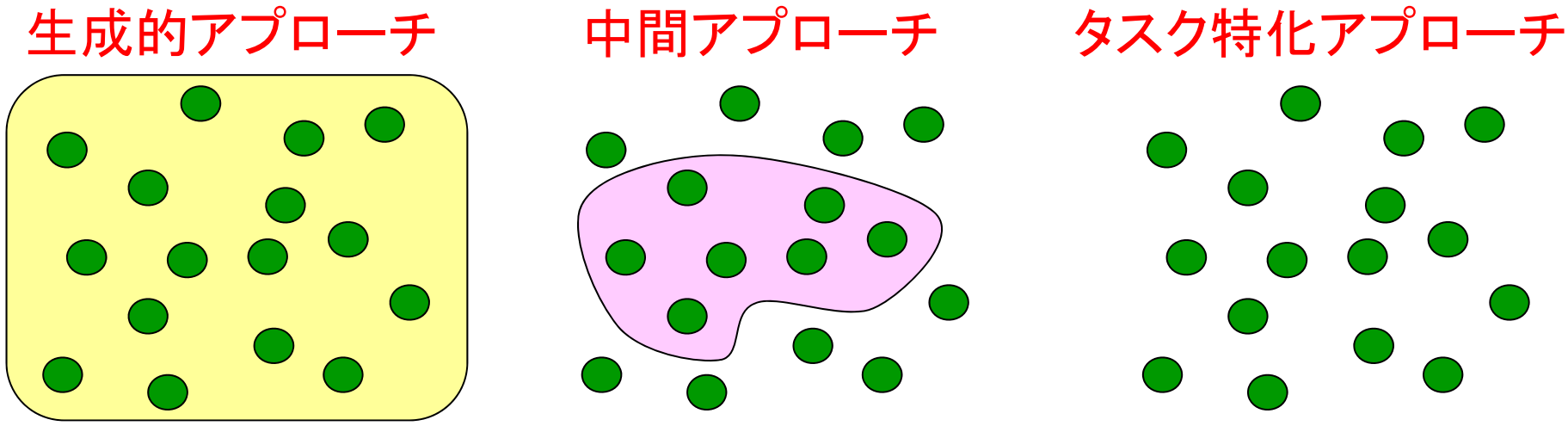
各タスクに特化したアプローチ

- 各タスクに特化したアルゴリズムを開発した方が原理的には生成的アプローチよりも性能が良い

- しかし，様々なタスクに対して個別に研究開発を行うのは大変：
  - アルゴリズム考案
  - 理論的性能評価
  - 高速かつメモリ効率の良い実装
  - エンジニアの技術習得

本日紹介するアプローチ

- 中間的なアプローチ：あるクラスのタスク群に対して，研究開発を行う
  - 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，変化点検知，クラスバランス推定
  - 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



確率密度比に基づく機械学習

- 前述の機械学習タスク群は複数の確率分布を含む

$$p(\boldsymbol{x}), q(\boldsymbol{x})$$

- しかし，これらのタスクを解くのに，それぞれの確率分布そのものは必要ない
- 確率密度関数の比が分かれば十分である

$$r(\boldsymbol{x}) = \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}$$

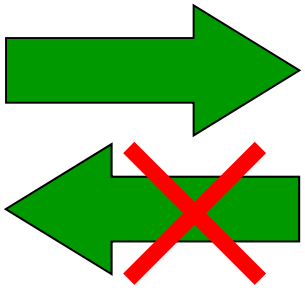
- 各確率分布は推定せず，密度比を直接推定することにする

直感的な正当化

バプニックの原理 Vapnik (1998)

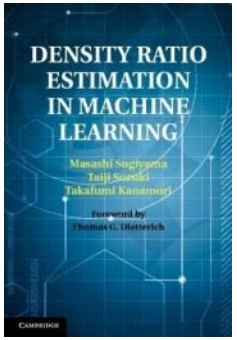
ある問題を解くとき，それより一般的な問題を途中段階で解くべきでない

$p(\boldsymbol{x}), q(\boldsymbol{x})$ が分かる

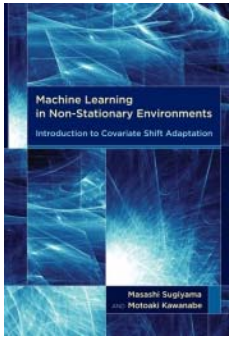


$r(\boldsymbol{x}) = \dfrac{p(\boldsymbol{x})}{q(\boldsymbol{x})}$ が分かる

■ 密度を求めるよりも，密度比を求めるほうが易しい



Sugiyama, Suzuki & Kanamori,
Density Ratio Estimation
in Machine Learning,
Cambridge University Press, 2012



Sugiyama & Kawanabe,
Machine Learning
in Non-Stationary Environments,
MIT Press, 2012

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

最小二乗密度比適合

Kanamori, Hido & Sugiyama (JMLR2009)

- データ: $\{\boldsymbol{x}_i^p\}_{i=1}^{n_p} \overset{i.i.d.}{\sim} p(\boldsymbol{x})$, $\{\boldsymbol{x}_j^q\}_{j=1}^{n_q} \overset{i.i.d.}{\sim} q(\boldsymbol{x})$
- 真の密度比$r(\boldsymbol{x})$との二乗誤差を最小にするように密度比モデル $r_{\boldsymbol{\alpha}}(\boldsymbol{x})$ を学習:

$$J(\boldsymbol{\alpha}) = \frac{1}{2}\int \Big(r_{\boldsymbol{\alpha}}(\boldsymbol{x}) - r(\boldsymbol{x})\Big)^2 q(\boldsymbol{x})d\boldsymbol{x} \qquad r(\boldsymbol{x}) = \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}$$

$$= \frac{1}{2}\int r_{\boldsymbol{\alpha}}(\boldsymbol{x})^2 q(\boldsymbol{x})d\boldsymbol{x} - \int r_{\boldsymbol{\alpha}}(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} + C$$

$$\approx \frac{1}{2n_q}\sum_{j=1}^{n_q} r_{\boldsymbol{\alpha}}(\boldsymbol{x}_j^q)^2 - \frac{1}{n_p}\sum_{i=1}^{n_p} r_{\boldsymbol{\alpha}}(\boldsymbol{x}_i^p) + C$$

アルゴリズム

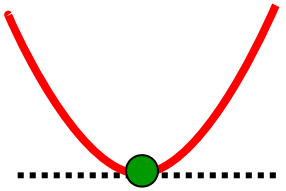
- 密度比モデル: $r_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \sum_{\ell=1}^{n_p} \alpha_\ell \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_\ell^p\|^2}{2\sigma^2}\right)$

- 最適化規準: $\min_{\boldsymbol{\alpha}} \left[\frac{1}{2}\boldsymbol{\alpha}^\top \widehat{\boldsymbol{G}}\boldsymbol{\alpha} - \widehat{\boldsymbol{h}}^\top \boldsymbol{\alpha} + \frac{\lambda}{2}\boldsymbol{\alpha}^\top \boldsymbol{\alpha}\right]$

$$\widehat{G}_{\ell,\ell'} = \frac{1}{n_q}\sum_{j=1}^{n_q} \exp\left(-\frac{\|\boldsymbol{x}_j^q - \boldsymbol{x}_\ell^p\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|\boldsymbol{x}_j^q - \boldsymbol{x}_{\ell'}^p\|^2}{2\sigma^2}\right)$$

$$\widehat{h}_\ell = \frac{1}{n_p}\sum_{i=1}^{n_p} \exp\left(-\frac{\|\boldsymbol{x}_i^p - \boldsymbol{x}_\ell^p\|^2}{2\sigma^2}\right)$$

- 大域的最適解が解析的に計算可能:



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

最小二乗密度比適合のMATLABによる実装

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

$$\widehat{G}_{\ell,\ell'} = \frac{1}{n_q} \sum_{j=1}^{n_q} \exp\left(-\frac{\|\boldsymbol{x}_j^q - \boldsymbol{x}_\ell^p\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|\boldsymbol{x}_j^q - \boldsymbol{x}_{\ell'}^p\|^2}{2\sigma^2}\right)$$

$$\widehat{h}_\ell = \frac{1}{n_p} \sum_{i=1}^{n_p} \exp\left(-\frac{\|\boldsymbol{x}_i^p - \boldsymbol{x}_\ell^p\|^2}{2\sigma^2}\right)$$

```
%人工データの生成
n=300; x=randn(n,1); y=randn(n,1)+0.5;

%密度比の推定
x2=x.^2; xx=repmat(x2,1,n)+repmat(x2',n,1)-2*x*x';
y2=y.^2; yx=repmat(y2,1,n)+repmat(x2',n,1)-2*y*x';
r=exp(-yx); s=r*((r'*r+eye(n))¥(mean(exp(-xx),2))); plot(y,s,'rx');
```

理論解析

■ パラメトリックモデルの場合：

$$r_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \sum_{\ell=1}^{b} \alpha_\ell \phi_\ell(\boldsymbol{x})$$

- 学習したパラメータは $n^{-\frac{1}{2}}$ の速さで最適値に収束
- 最適な収束率を達成している $n = \min(n_p, n_q)$

Kanamori, Hido & Sugiyama (JMLR2009)

■ ノンパラメトリックモデルの場合：

$$r_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \sum_{\ell=1}^{n_p} \alpha_\ell \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_\ell^p\|^2}{2\sigma^2}\right)$$

- 学習した関数は $n^{-\frac{1}{2+\gamma}}$ の速さで真の関数に収束
  （関数空間の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 $0 < \gamma < 2$

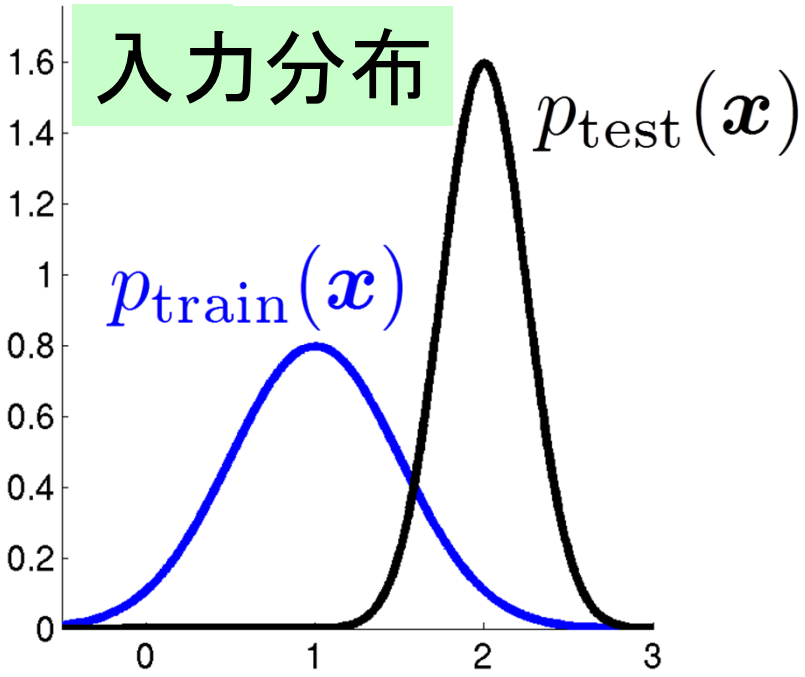
Kanamori, Suzuki & Sugiyama (ML2012)

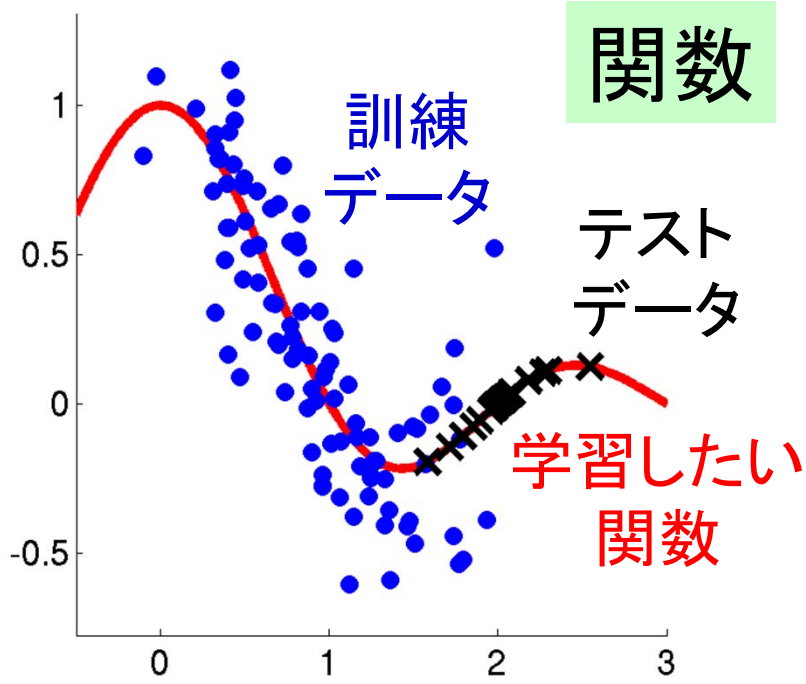
発表の流れ

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

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

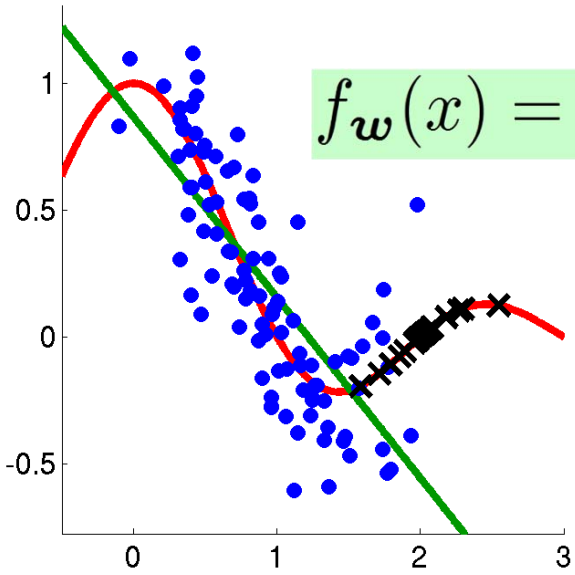




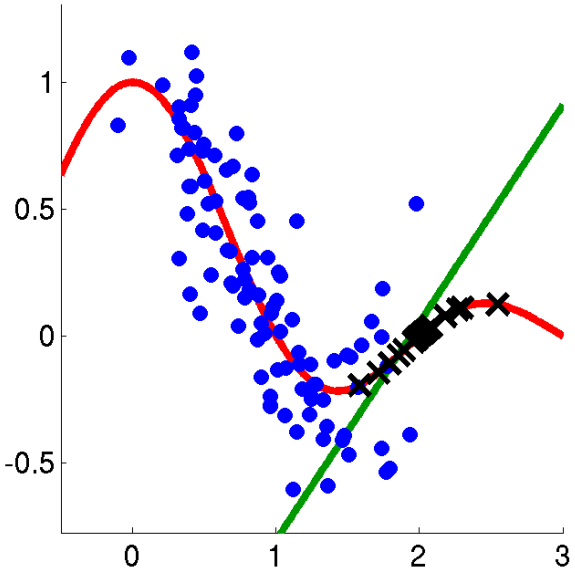
重要度重み付き最小二乗学習

$$\min_{\boldsymbol{w}} \sum_{i=1}^{n} \left(f_{\boldsymbol{w}}(\boldsymbol{x}_i) - y_i \right)^2$$

$$\min_{\boldsymbol{w}} \sum_{i=1}^{n} \frac{p_{\mathrm{test}}(\boldsymbol{x}_i)}{p_{\mathrm{train}}(\boldsymbol{x}_i)} \left(f_{\boldsymbol{w}}(\boldsymbol{x}_i) - y_i \right)^2$$



$$f_{\boldsymbol{w}}(x) = w_1 + w_2 x$$



- 共変量シフト下では，通常の最小二乗学習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)

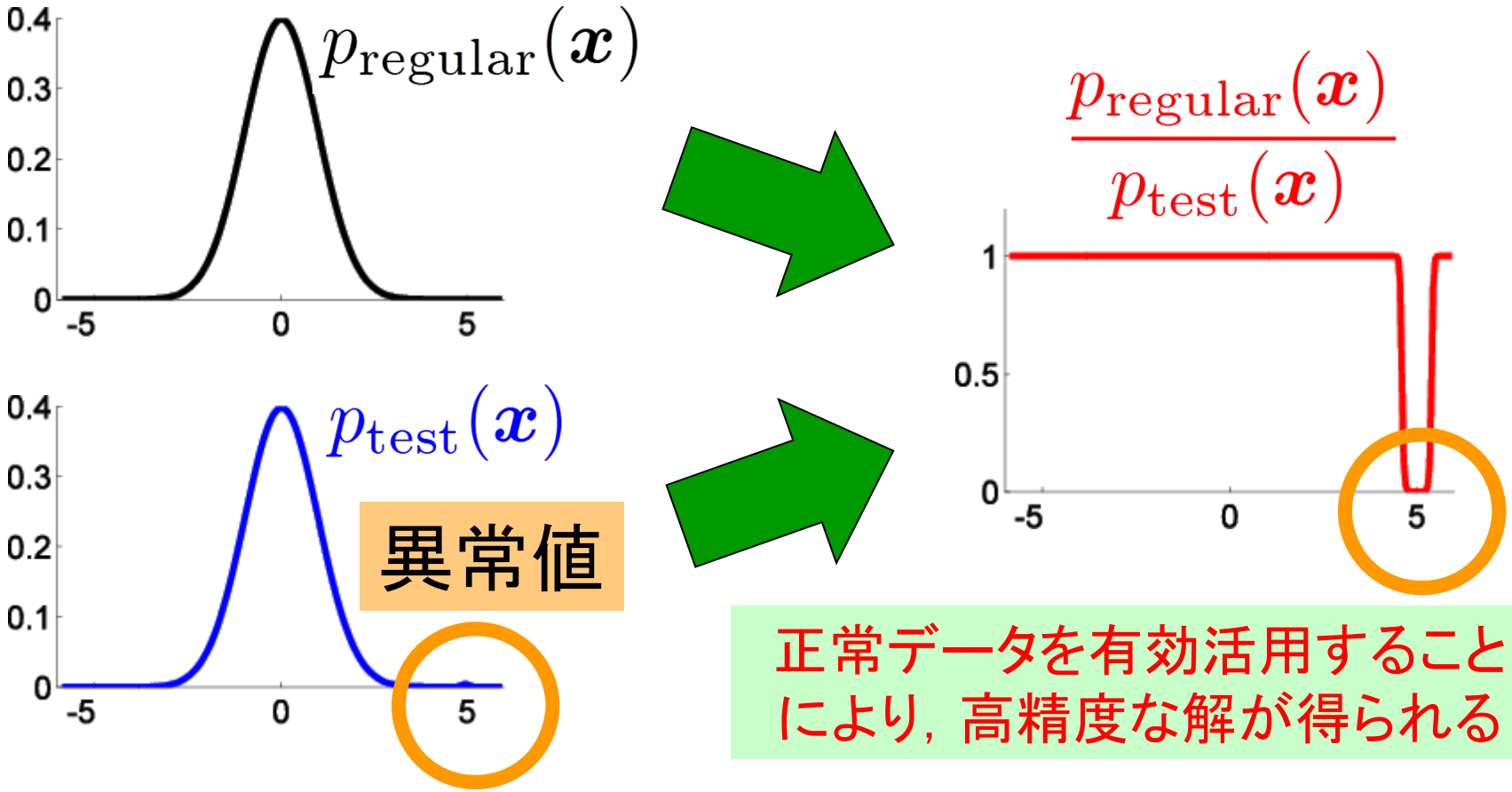
発表の流れ

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

正常値に基づく異常値検出

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

- 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

- 製鉄プロセスの異常診断

  Hirata, Kawahara & Sugiyama (Patent2010)

- 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)

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

$$\{\boldsymbol{x}_i^p\}_{i=1}^{n_p} \overset{i.i.d.}{\sim} p(\boldsymbol{x})$$

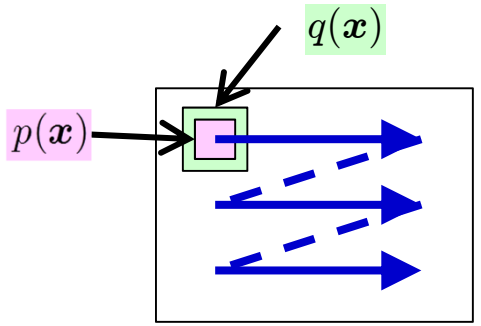
$$\{\boldsymbol{x}_j^q\}_{j=1}^{n_q} \overset{i.i.d.}{\sim} q(\boldsymbol{x})$$

- 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：
  $$\int p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}$$
  - ピアソン距離：
  $$\int q(\boldsymbol{x}) \left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} - 1 \right)^2 \mathrm{d}\boldsymbol{x}$$

実世界応用例

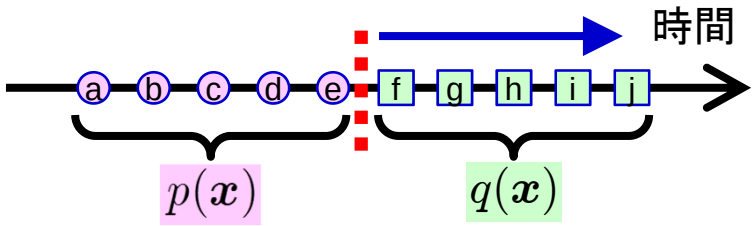
■画像中の注目領域抽出

Yamanaka, Matsugu
& Sugiyama (IPSJ-TOM2013)



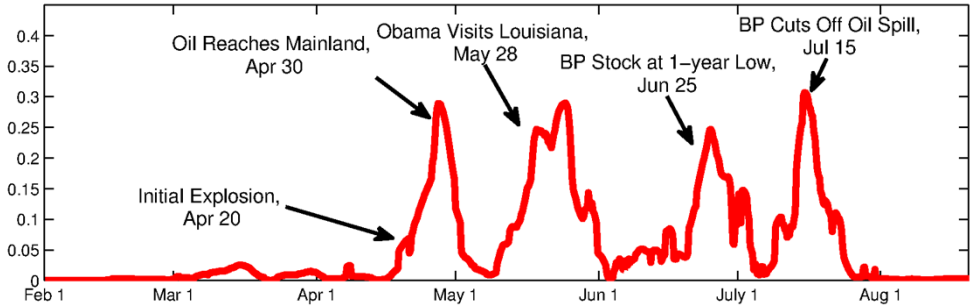
■動画からのイベント検出

Yamanaka, Matsugu
& Sugiyama (IPSJ-TOM2013)



■ツイッターデータ解析

Liu, Yamada
& Sugiyama (NN2013)



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

相互情報量推定

Suzuki, Sugiyama, Sese & Kanamori (FSDM2008), Sugiyama (Entropy2013)

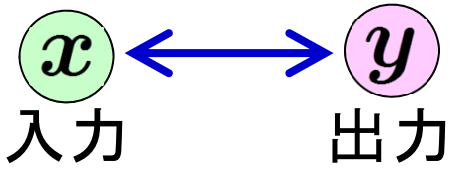
- 相互情報量: $$\mathrm{MI} = \int p(\boldsymbol{x}, \boldsymbol{y}) \log \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})} d\boldsymbol{x} d\boldsymbol{y}$$

$\mathrm{MI} = 0$ ⟺ $\boldsymbol{x}$ と $\boldsymbol{y}$ は統計的に独立

- 相互情報量は密度比を用いて計算できる
- 最小二乗密度比推定には，
  二乗損失相互情報量が自然:

$$\mathrm{SMI} = \int p(\boldsymbol{x})p(\boldsymbol{y}) \left(\frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})} - 1 \right)^2 d\boldsymbol{x} d\boldsymbol{y}$$

相互情報量に基づく機械学習



- 入出力間の独立性判定:
  - 特徴選択
  - クラスタリング

Suzuki, Sugiyama, Sese & Kanamori (BMC-Bioinfo2009)

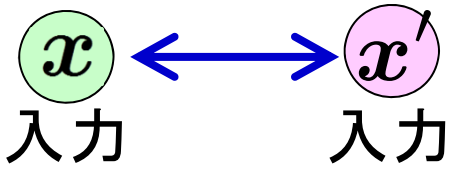
Suzuki & Sugiyama (NeCo2012)

Sugiyama, Niu, Yamada, Kimura & Hachiya (NeCo2013)

- 実世界応用例:
  - 遺伝子解析
  - 画像認識
  - 音響認識

相互情報量に基づく機械学習

- 入力間の独立性判定:
  - 独立成分分析
  - オブジェクト適合



Suzuki & Sugiyama (NeCo2011)

Karasuyama & Sugiyama (NN2012)

Yamada & Sugiyama (AISTATS2011)

- 実世界応用例:
  - モーションキャプチャデータの解析
  - 医療画像の位置合わせ
  - 写真の自動レイアウト

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

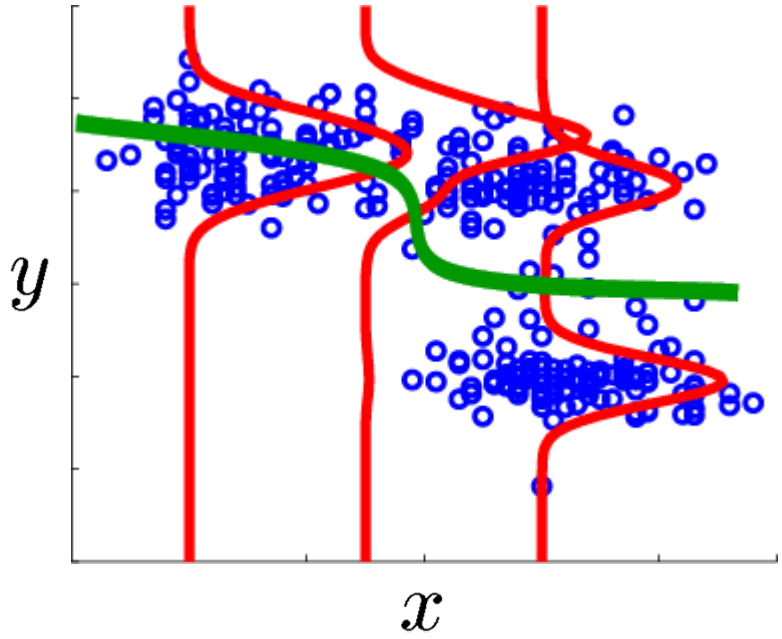
条件付き確率密度の推定

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

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
- 実世界応用例：
  - ヒューマノイドロボット制御（ATR）

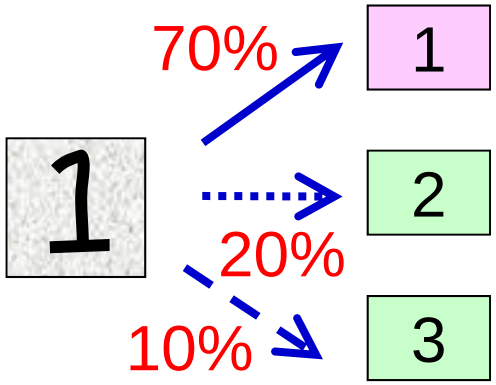
Sugimoto, Tangkaratt, Wensveen, Zhao, Sugiyama & Morimoto (submitted)



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)



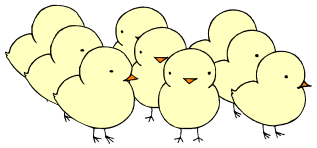
- 出力 y がカテゴリのとき，
  条件付き確率の推定は
  確率的なパターン認識に対応
- 実世界応用例:
  - 顔画像からの年齢推定

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

    Hachiya, Sugiyama & Ueda
    (Neurocomputing2012)

まとめ

- 密度比は，単純な最小二乗法で精度・効率良く推定できる
- 多くの学習タスクが実は最小二乗法で解ける：
  - 重点サンプリング：$\sum_{i=1}^{n} \frac{p_{\mathrm{test}}(\boldsymbol{x}_i)}{p_{\mathrm{train}}(\boldsymbol{x}_i)} \mathrm{loss}(\boldsymbol{x}_i)$



  - ダイバージェンス推定：$\int p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}$

  - 相互情報量推定：$\iint p(\boldsymbol{x}, \boldsymbol{y}) \log \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})} \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}$
  - 条件付き確率推定：$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x})}$

試験について

- 日時：8月1日（金）3,4限
- 場所：H121（普段の講義と異なるので注意！）
- 試験内容：
  - 専門用語の英語名を答えよ
  - 次の（a），（b），（c）からテーマを二つ選び，自由に論ぜよ.

    （a）積率母関数，（b）中心極限定理

    （c）任意の分布に従う乱数を生成する方法
- 教科書，ノートは持ち込み不可！