限られた情報からロバストに:
信頼できる機械学習に向けて

杉山　将

理化学研究所／東京大学

http://www.ms.k.u-tokyo.ac.jp/sugi/

自己紹介

現職：

- 理化学研究所・センター長：研究者とともに
- 東京大学・教授：学生とともに
- 企業・技術顧問：エンジニアとともに

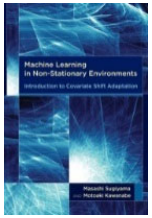
専門分野：

- 機械学習の理論とアルゴリズム
  （転移学習，密度比推定，強化学習，
  変分推論，弱教師付き学習など）
- 機械学習技術の実世界応用
  （画像，言語，音声，脳波，ロボット，
  自動運転，広告，生命，医療，教育など）
  →ぜひ皆さんと協業できれば！

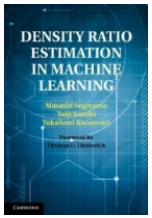
学会活動：

- プログラム委員長：NeurIPS2015, AISTATS2019, ACML2010/2020など
- 信学会IBISML研究会委員長(2022-)

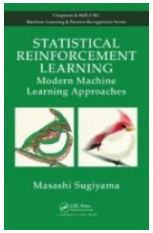
Sugiyama & Kawanabe, Machine Learning in Non-Stationary Environments, MIT Press, 2012



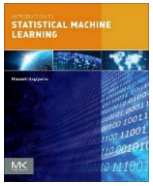
Sugiyama, Suzuki & Kanamori, Density Ratio Estimation in Machine Learning, Cambridge University Press, 2012



Sugiyama, Statistical Reinforcement Learning, Chapman and Hall/CRC, 2015

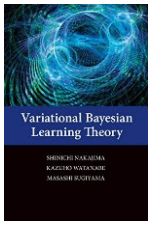


Sugiyama, Introduction to Statistical Machine Learning, Morgan Kaufmann, 2015



Cichocki, Phan, Zhao, Lee, Oseledets, Sugiyama & Mandic, Tensor Networks for Dimensionality Reduction and Large-Scale Optimizations, Now, 2017

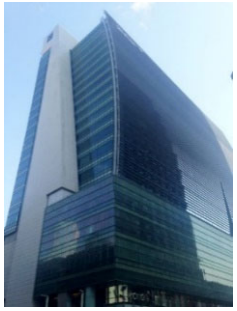
Nakajima, Watanabe & Sugiyama, Variational Bayesian Learning Theory, Cambridge University Press, 2019



理研・革新知能統合研究(AIP)センター

- 文科省AIPプロジェクト(2016~2025年度)を推進:
  - 常勤研究員130名(36%外国人, 23%女性)
  - 客員研究員200名, 学生100名
  - 延べ140名の海外インターン生

日本橋オフィス



分散拠点

- 機械学習の技術を軸足に,
  基礎から応用・社会まで一気通貫の研究体制
- 産学官で連携し, 研究成果を国際的に発信
- 国際的な高度AI人材の登竜門となることを目指す

今日の話題：ロバスト機械学習

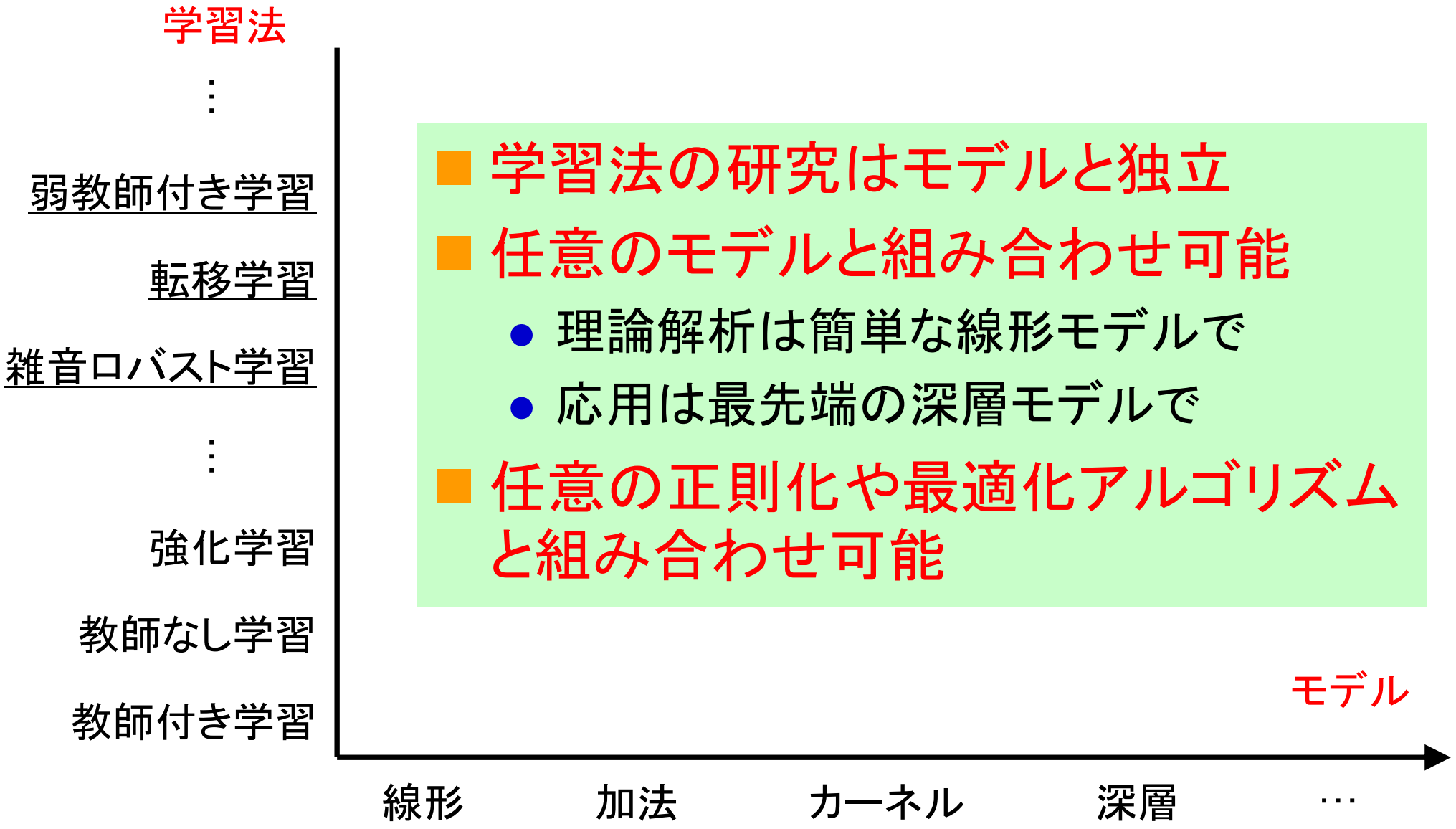
- 機械学習の実応用では，様々な不確定要因に対する信頼性向上が重要：
  - 不完全情報：弱い教師データ
  - 標本化バイアス：標本選択，環境変化
  - ラベル雑音：センサー誤差，ヒューマンエラー

- 本講演では，（手前味噌な）ロバスト機械学習の最近の研究成果をご紹介します

http://www.ms.k.u-tokyo.ac.jp/sugi/publications.html

紹介する研究の位置づけ:学習法

- 機械学習手法は，学習法とモデルの組み合わせ



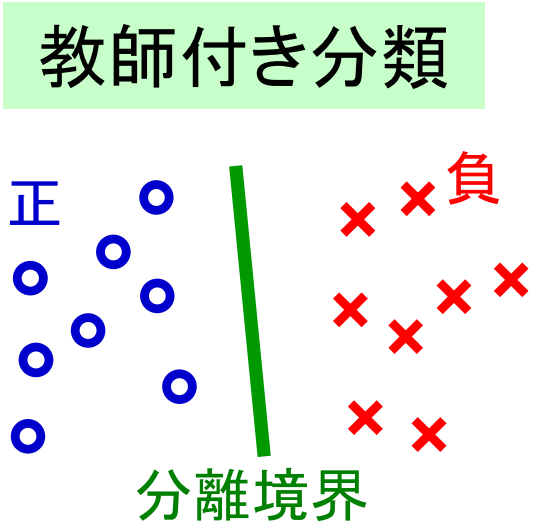
- 学習法の研究はモデルと独立
- 任意のモデルと組み合わせ可能
  - 理論解析は簡単な線形モデルで
  - 応用は最先端の深層モデルで
- 任意の正則化や最適化アルゴリズムと組み合わせ可能

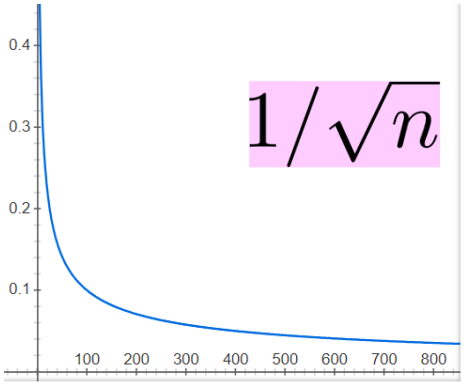
講演の流れ

1. 弱教師付き分類
   A) 正とラベルなしデータからの分類
   B) 最近の発展
2. 転移学習
3. 雑音ロバスト分類
4. 今後の展望

限られた情報からの機械学習

- 教師付き分類：大量の教師データを用いることにより，人間と同等かそれ以上の予測性能を達成：
  - 画像理解，音声認識，機械翻訳・・・
  - ラベル付きデータ数 n に対して，分離境界の推定誤差は $1/\sqrt{n}$ の速さで減っていく（最適）
- しかし，応用分野によっては，教師データを簡単に取れない：
  - 医療，自然災害，材料，プライバシ・・・
- 限られた情報からの学習が重要！

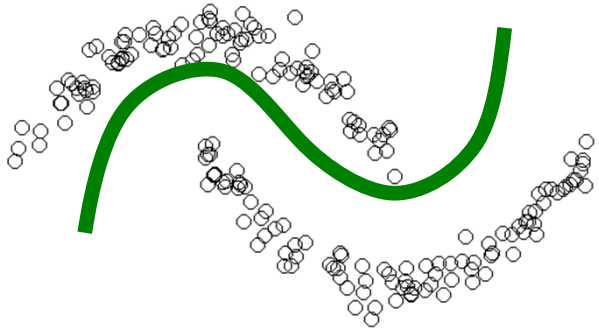




教師なし分類，半教師付き分類

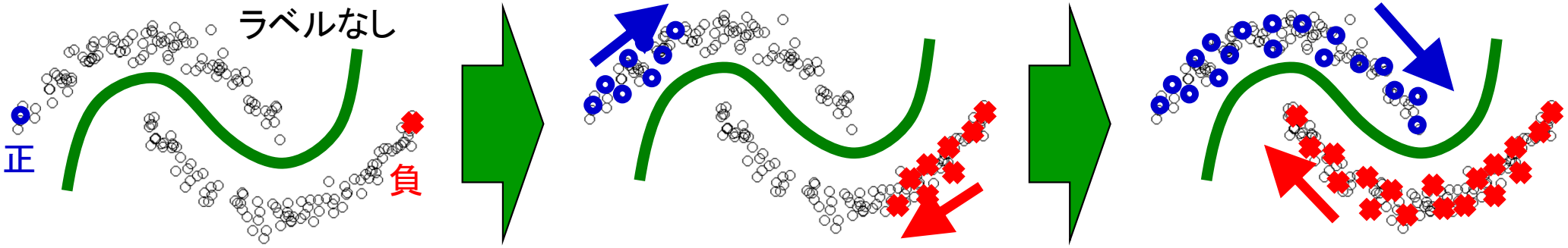
教師なし分類：

- ラベルを全く使わない
- データを塊に分けるクラスタリング
- 予測性能に関して何も保証できない

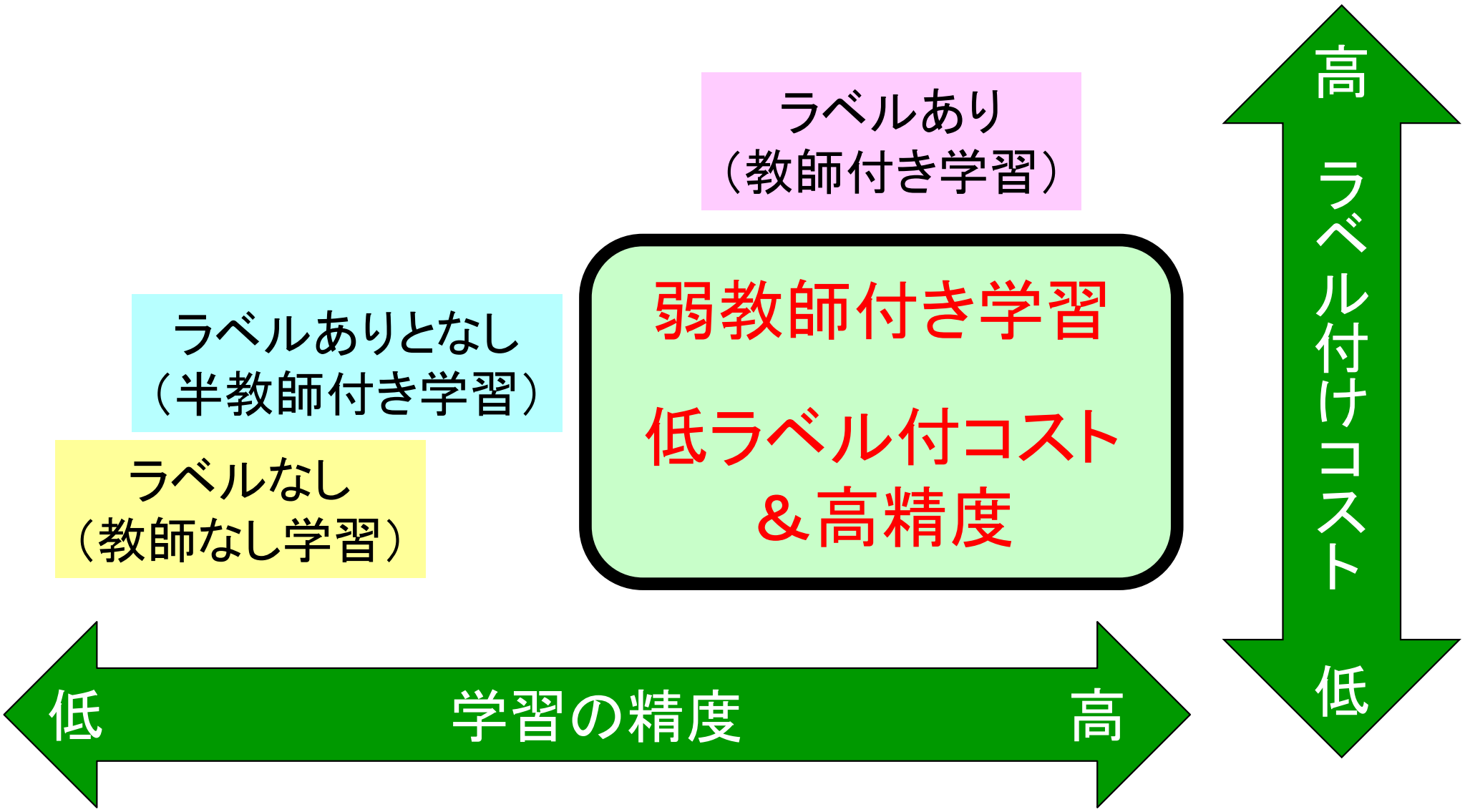


半教師付き分類：　Chapelle et al. (2006), 他多数

- 少しだけラベルを使う
- ラベルをクラスタに沿って伝播
- 予測性能に関しては，一般には何も保証できない



弱教師付き学習のねらい



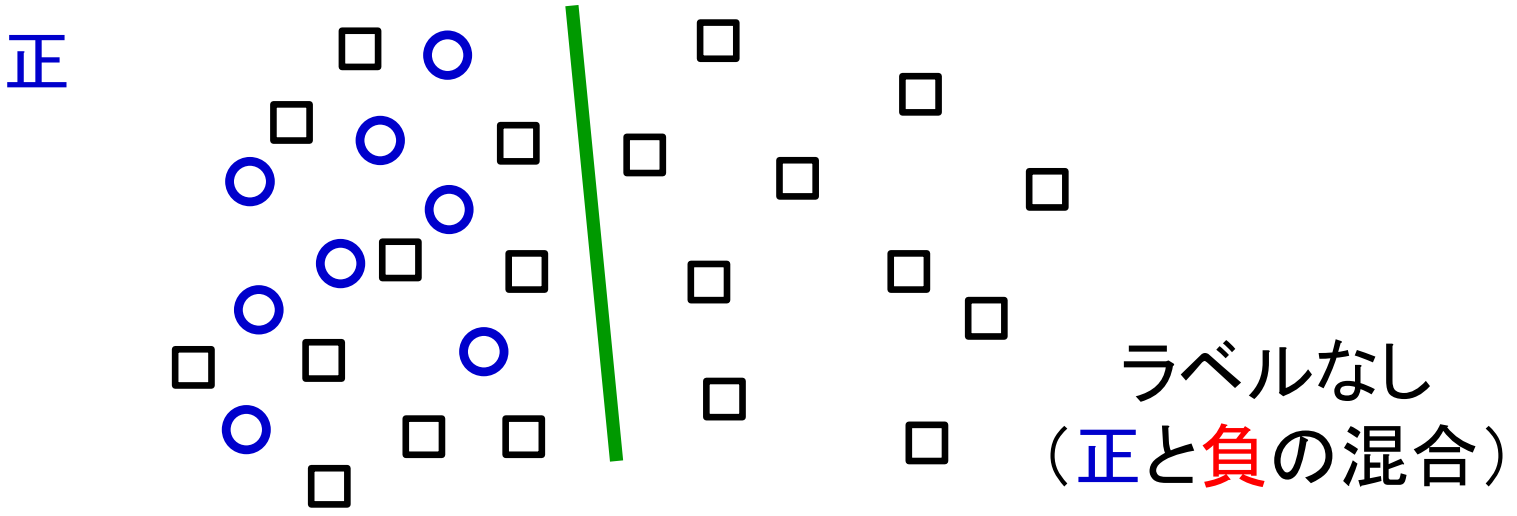
- 低コストで集められる弱教師付きデータで精度良く学習できないか？

講演の流れ

1. 弱教師付き分類
   A) 正とラベルなしデータからの分類
   B) 最近の発展
2. 転移学習
3. 雑音ロバスト分類
4. 今後の展望

正とラベルなしデータからの分類

- 正とラベルなしデータだけが与えられる:
  - 負のデータは一つも与えられない
- 例:オンライン広告配信におけるクリック予測
  - クリックされた広告:正データ
  - クリックされなかった広告:ラベルなし(負ではない!)



解き方（正ラベルなし分類）

- 与えられるデータ：正とラベルなし標本

du Plessis, Niu & Sugiyama (NIPS2014, ICML2015)

$$\{\boldsymbol{x}_i^{\mathrm{P}}\}_{i=1}^{n_{\mathrm{P}}} \overset{\text{i.i.d.}}{\sim} p(\boldsymbol{x}|y=+1) \qquad \{\boldsymbol{x}_j^{\mathrm{U}}\}_{j=1}^{n_{\mathrm{U}}} \overset{\text{i.i.d.}}{\sim} p(\boldsymbol{x})$$

- 分類リスクを分解：

$$R(f) = \mathbb{E}_{p(\boldsymbol{x},y)}\Big[\ell\Big(yf(\boldsymbol{x})\Big)\Big]$$

ℓ：損失関数

$\pi = p(y=+1)$：
クラス事前確率（既知と仮定）

$$= \pi\mathbb{E}_{p(\boldsymbol{x}|y=+1)}\Big[\ell\Big(f(\boldsymbol{x})\Big)\Big] + (1-\pi)\mathbb{E}_{p(\boldsymbol{x}|y=-1)}\Big[\ell\Big(-f(\boldsymbol{x})\Big)\Big]$$

正データに対するリスク　　負データに対するリスク

- 負データがないので，負リスクを以下のように変形：

$$\mathbb{E}_{p(\boldsymbol{x})}\Big[\ell\Big(-f(\boldsymbol{x})\Big)\Big] - \pi\mathbb{E}_{p(\boldsymbol{x}|y=+1)}\Big[\ell\Big(-f(\boldsymbol{x})\Big)\Big]$$

$$p(\boldsymbol{x}) = \pi p(\boldsymbol{x}|y=+1) + (1-\pi)p(\boldsymbol{x}|y=-1)$$

- データからの不偏リスク推定量：

$$\widehat{R}_{\mathrm{PU}}(f) = \frac{\pi}{n_{\mathrm{P}}}\sum_{i=1}^{n_{\mathrm{P}}}\ell\Big(f(\boldsymbol{x}_i^{\mathrm{P}})\Big) + \frac{1}{n_{\mathrm{U}}}\sum_{j=1}^{n_{\mathrm{U}}}\ell\Big(-f(\boldsymbol{x}_j^{\mathrm{U}})\Big) - \frac{\pi}{n_{\mathrm{P}}}\sum_{i=1}^{n_{\mathrm{P}}}\ell\Big(-f(\boldsymbol{x}_i^{\mathrm{P}})\Big)$$

理論的な性質

Niu, du Plessis, Sakai, Ma & Sugiyama (NIPS2016)

$$\widehat{R}_{\mathrm{PU}}(f) = \frac{\pi}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} \ell\Big(f(\boldsymbol{x}_i^{\mathrm{P}})\Big) + \frac{1}{n_{\mathrm{U}}} \sum_{j=1}^{n_{\mathrm{U}}} \ell\Big(-f(\boldsymbol{x}_j^{\mathrm{U}})\Big) - \frac{\pi}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} \ell\Big(-f(\boldsymbol{x}_i^{\mathrm{P}})\Big)$$

- 最適な収束率を達成：

$$R(\widehat{f}_{\mathrm{PU}}) - R(f^*) \leq C(\delta) \left(\frac{2\pi}{\sqrt{n_{\mathrm{P}}}} + \frac{1}{\sqrt{n_{\mathrm{U}}}} \right) = \mathcal{O}_p \left(\frac{1}{\sqrt{n_{\mathrm{P}}}} + \frac{1}{\sqrt{n_{\mathrm{U}}}} \right)$$

with probability $1-\delta$

$$\widehat{f}_{\mathrm{PU}} = \mathrm{argmin}_f \, \widehat{R}_{\mathrm{PU}}(f)$$

$$f^* = \mathrm{argmin}_f \, R(f)$$

$$R(f) = \mathbb{E}_{p(\boldsymbol{x},y)} \Big[\ell\Big(yf(\boldsymbol{x})\Big) \Big]$$

- リスク推定量を補正すれば更に性能が向上：

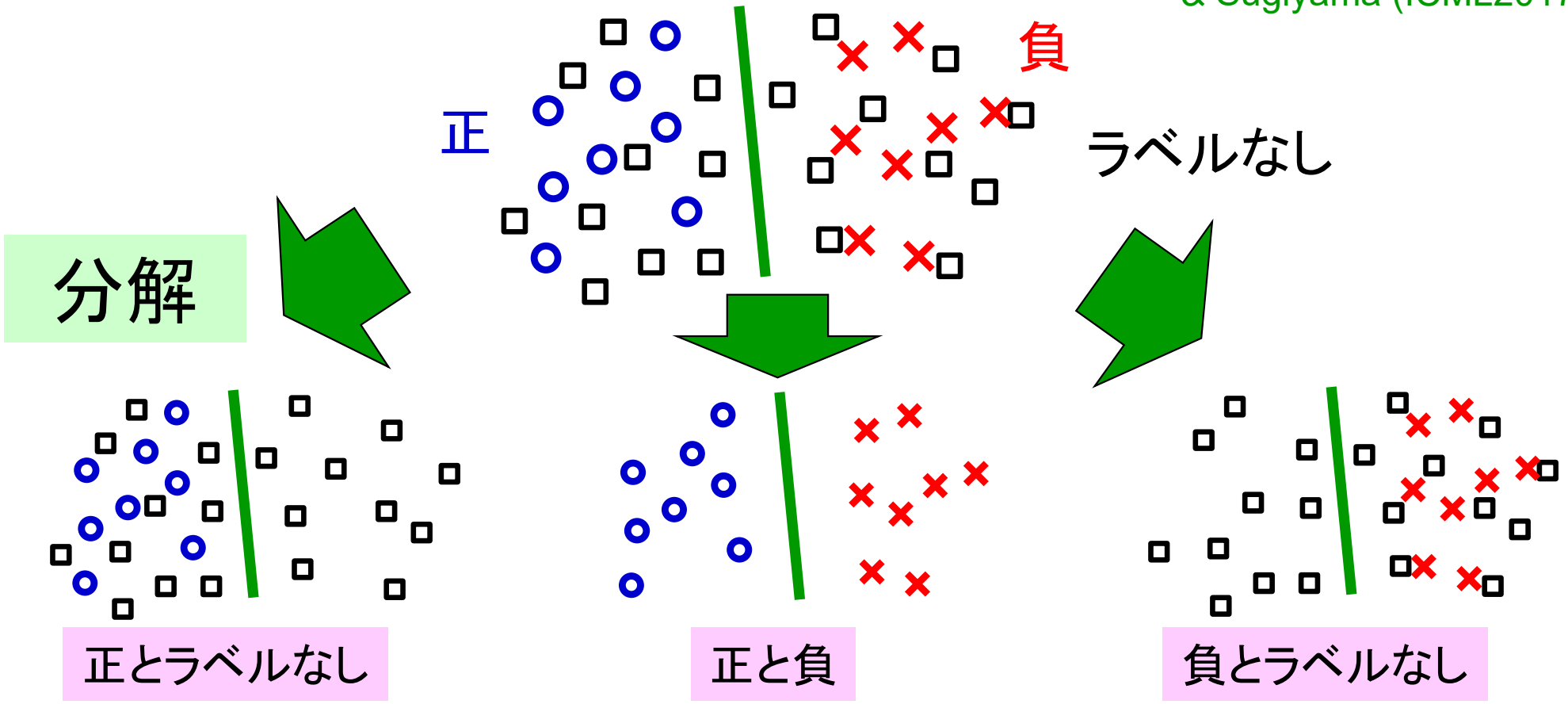
Kiryo, Niu, du Plessis & Sugiyama (NIPS2017)

$$\widetilde{R}_{\mathrm{PU}}(f) = \frac{\pi}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} \ell\Big(f(\boldsymbol{x}_i^{\mathrm{P}})\Big) + \max \left\{ 0, \; \frac{1}{n_{\mathrm{U}}} \sum_{i=1}^{n_{\mathrm{U}}} \ell\Big(-f(\boldsymbol{x}_i^{\mathrm{U}})\Big) - \frac{\pi}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} \ell\Big(-f(\boldsymbol{x}_i^{\mathrm{P}})\Big) \right\}$$

半教師付き分類：分解解法

半教師付き分類（正と負とラベルなし）

Sakai, du Plessis, Niu & Sugiyama (ICML2017)



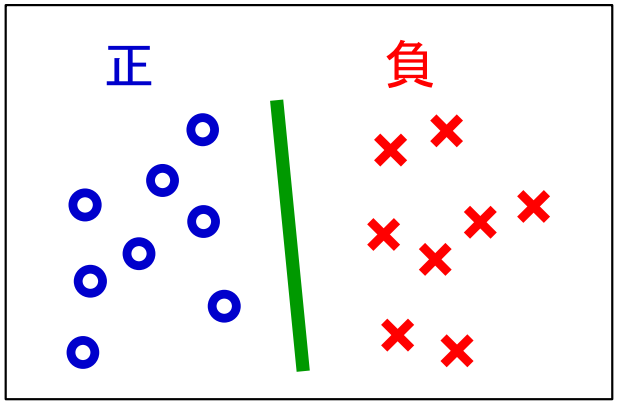
- 分解した3つの問題は，それぞれ最適に解ける
- それらを組み合わせても最適に解ける！
  （クラスタ仮定不要）

$$\mathcal{O}_p\left(1/\sqrt{n_{\mathrm{P}}}+1/\sqrt{n_{\mathrm{N}}}+1/\sqrt{n_{\mathrm{U}}}\right)$$

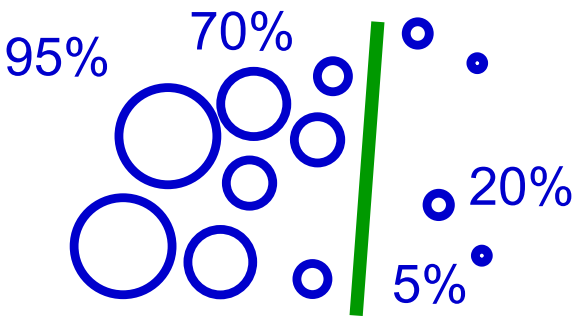
講演の流れ

1. 弱教師付き分類
   A) 正とラベルなしデータからの分類
   B) 最近の発展
2. 転移学習
3. 雑音ロバスト分類
4. 今後の展望

様々な弱教師付き分類（2クラス）



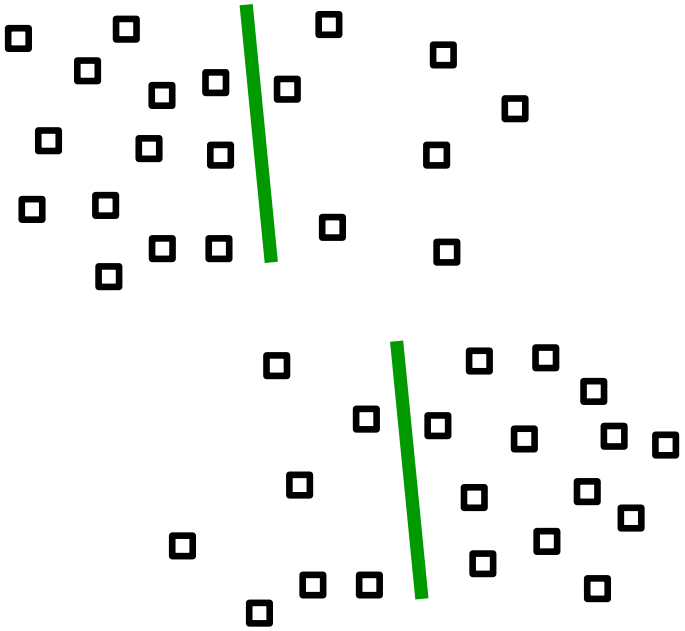
正信頼度学習



例：購買予測

Ishida et al. (NeurIPS2018)
Shinoda et al. (IJCAI2021)

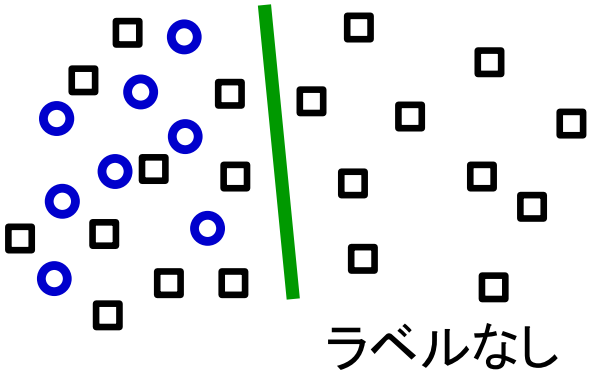
ラベルなしラベルなし分類



例：異なる母集団からの学習

du Plessis et al.,(TAAI2013)
Lu et al. (ICLR2019, AISTATS2020)
Charoenphakdee et al. (ICML2019)
Lei et al. (ICML2021)

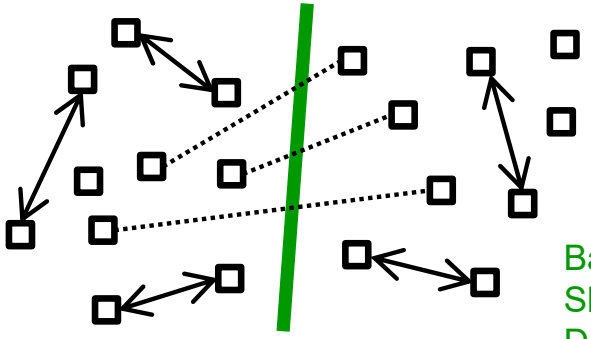
正ラベルなし分類



例：クリック予測

du Plessis et al.
(NIPS2014, ICML2015, MLJ2017)
Niu et al. (NIPS2016),
Kiryo et al. (NIPS2017)
Hsieh et al. (ICML2019)

類似非類似ラベルなし分類



例：機微情報予測

Bao et al. (ICML2018)
Shimada et al. (NeCo2021)
Dan et al. (ECMLPKDD2021)
Cao et al. (ICML2021)
Feng et al. (ICML2021)

$1/\sqrt{n}$

様々な弱教師付き分類(多クラス)

- 多数のクラスがあると，ラベル付けはますます大変
- 補ラベル：パターンが属さないクラスを示すラベル
  - 例：「クラス1に属さない」「この画像に犬はいない」

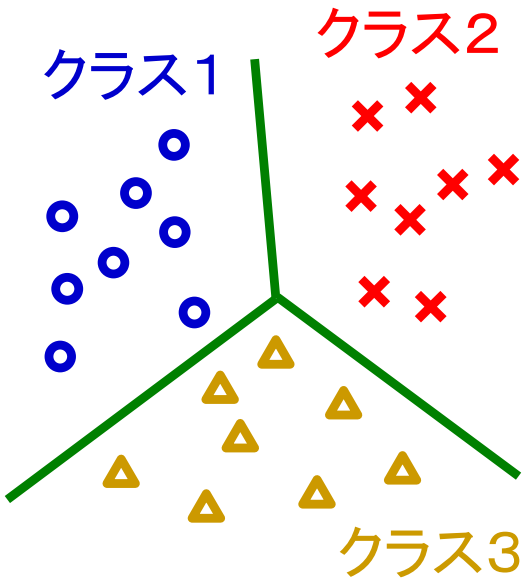
Ishida, Niu, Hu & Sugiyama (NIPS2017)
Ishida, Niu, Menon & Sugiyama (ICML2019)

- 部分ラベル：真のクラスを含むラベルのサブセット
  - 例：「クラス1か2に属する」「犬か猫か鳥のどれか」

Feng, Kaneko, Han, Niu, An & Sugiyama (ICML2020)
Feng, Lv, Han, Xu, Niu, Geng, An & Sugiyama (NeurIPS2020)

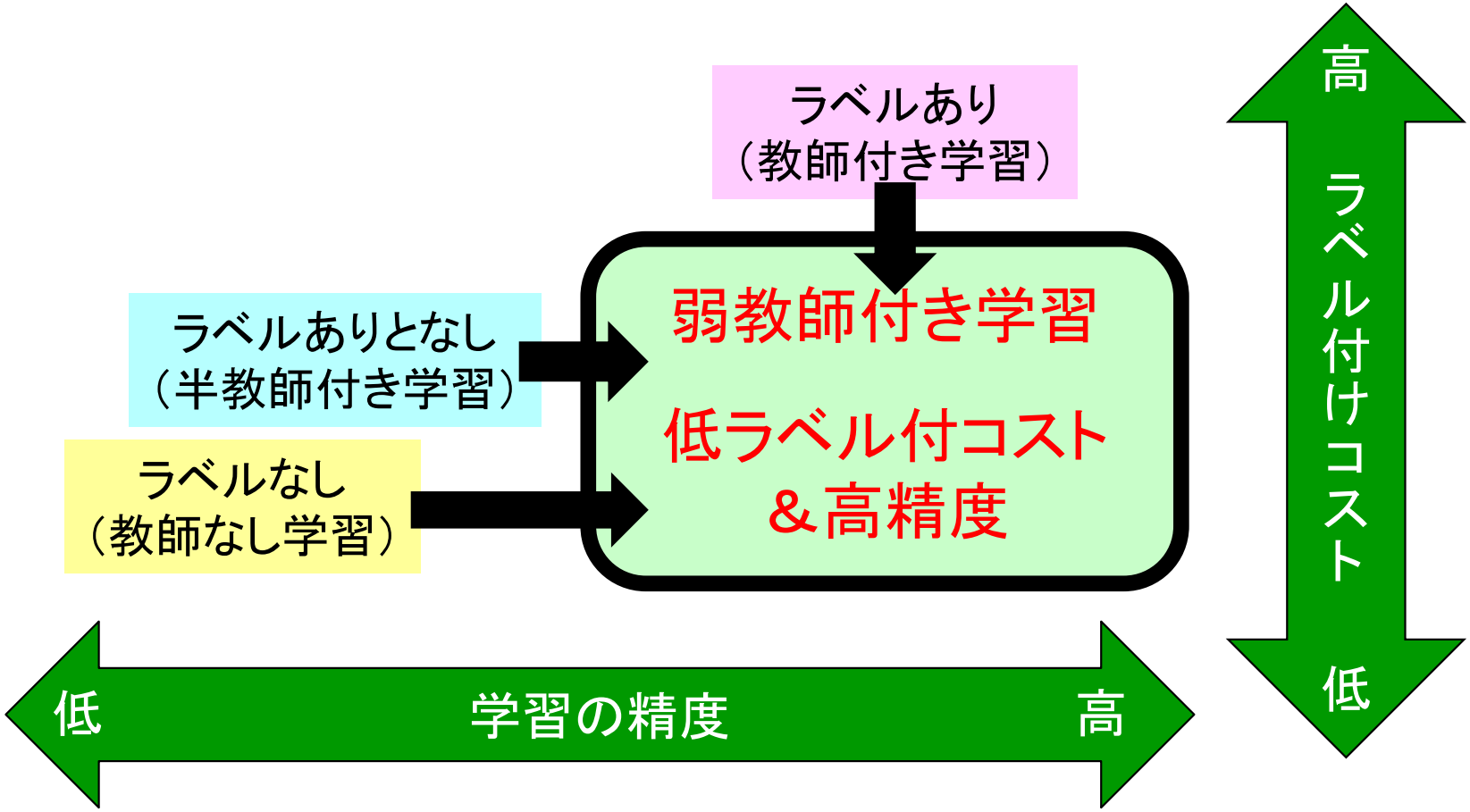
- 1クラス信頼度：信頼度データ
  - 例：「クラス1である確率が60%，
    クラス2である確率が30%，
    クラス3である確率が10%」

Cao, Feng, Shu, Xu, An, Niu & Sugiyama (arXiv2021)



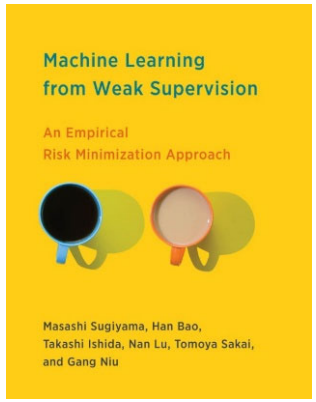
$1/\sqrt{n}$

弱教師付き学習のまとめ



- 様々な弱教師データから分類リスクを推定
  - 任意のモデル，正則化，最適化アルゴリズムと組み合わせ可能！

Sugiyama, Bao, Ishida, Lu, Sakai & Niu. Machine Learning from Weak Supervision, MIT Press, 2022.



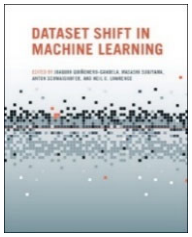
講演の流れ

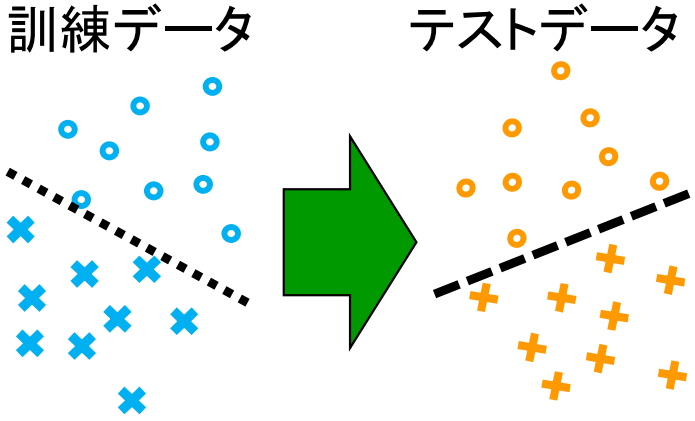
1. 弱教師付き分類
2. 転移学習
   A) 共変量シフト適応
   B) 最近の発展
3. 雑音ロバスト分類
4. 今後の展望

訓練データのバイアス

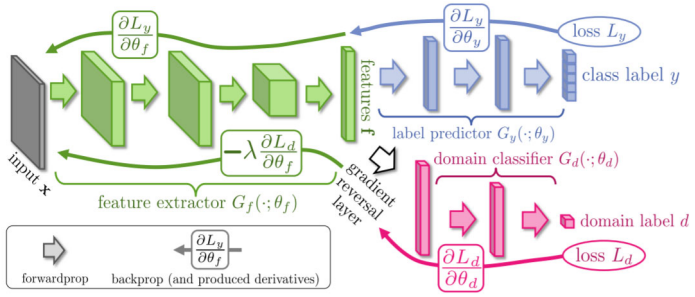
- 訓練データとテストデータの分布が異なると，標準的な機械学習法はうまくいかない：
  - 環境の変化
  - 標本選択バイアス

Quiñonero-Candela, Sugiyama, Schwaighofer & Lawrence (MIT Press 2009)



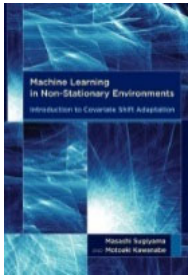


- 転移学習（ドメイン適応）：
  訓練データをテストデータに何らかの方法で適応させる
  - 訓練データを変換する
  - 訓練データに重みを付ける



Ben-David et al. (NIPS2006)
Ganin & Lempitsky (ICML2015)

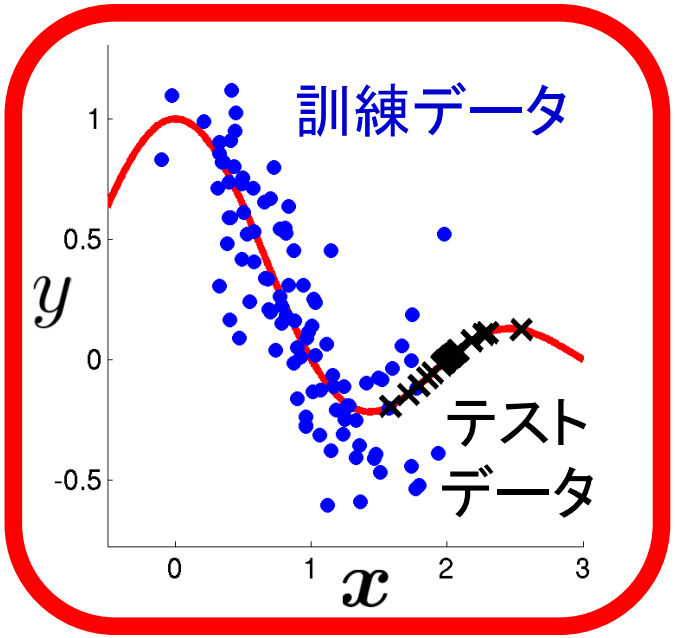
Sugiyama & Kawanabe (MIT Press 2012)

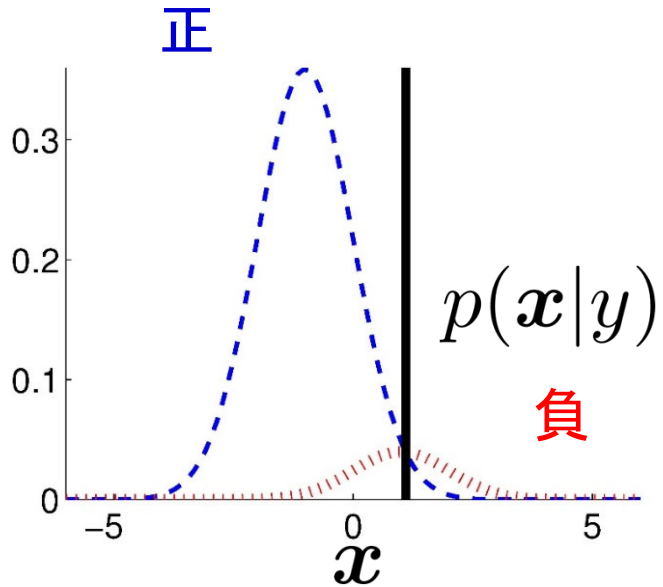


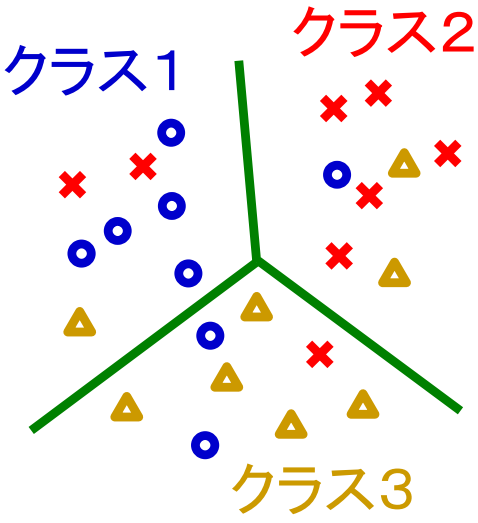
いろいろな問題設定

$\boldsymbol{x}$：入力　　y：出力

- 任意の分布変化：$p_{\mathrm{tr}}(\boldsymbol{x}, y) \neq p_{\mathrm{te}}(\boldsymbol{x}, y)$
- 共変量シフト：$p_{\mathrm{tr}}(\boldsymbol{x}) \neq p_{\mathrm{te}}(\boldsymbol{x})$
- クラス事前確率シフト：$p_{\mathrm{tr}}(y) \neq p_{\mathrm{te}}(y)$
- 出力雑音：$p_{\mathrm{tr}}(y|\boldsymbol{x}) \neq p_{\mathrm{te}}(y|\boldsymbol{x})$
- クラス条件付き分布シフト：$p_{\mathrm{tr}}(\boldsymbol{x}|y) \neq p_{\mathrm{te}}(\boldsymbol{x}|y)$







講演の流れ

1. 弱教師付き分類
2. 転移学習
   A) 共変量シフト適応
   B) 最近の発展
3. 雑音ロバスト分類
4. 今後の展望

共変量シフト下での回帰

共変量シフト： Shimodaira (JSPI2000)

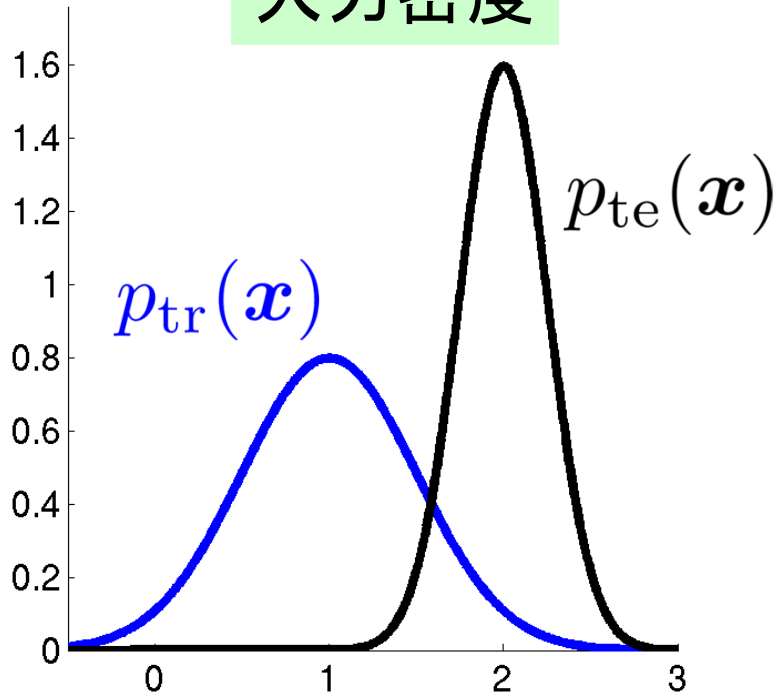
- 訓練データとテストデータの入力分布が異なる

$$p_{\mathrm{tr}}(\boldsymbol{x}) \neq p_{\mathrm{te}}(\boldsymbol{x})$$

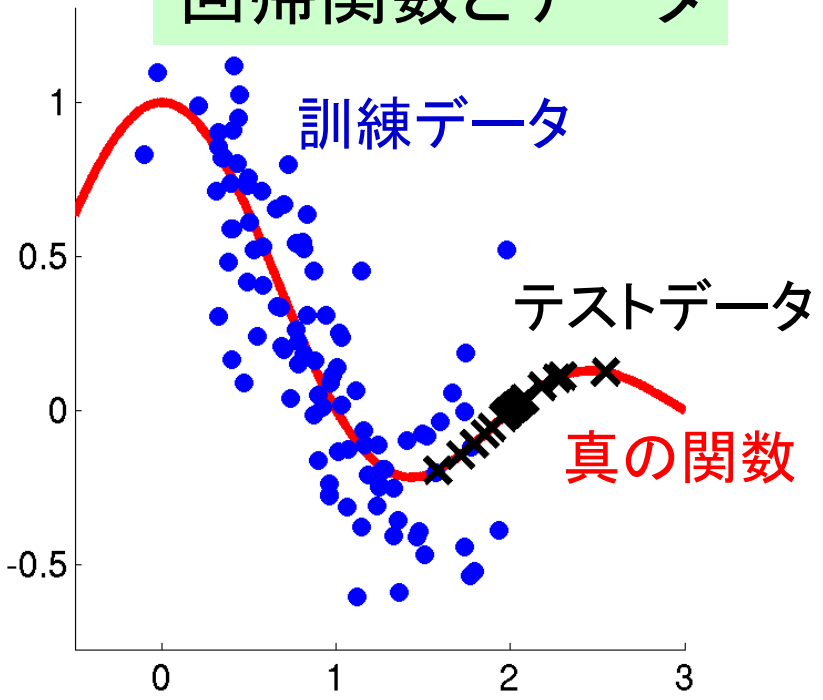
- 入出力関係は変わらない

$$p_{\mathrm{tr}}(y|\boldsymbol{x}) = p_{\mathrm{te}}(y|\boldsymbol{x}) = p(y|\boldsymbol{x})$$

入力密度

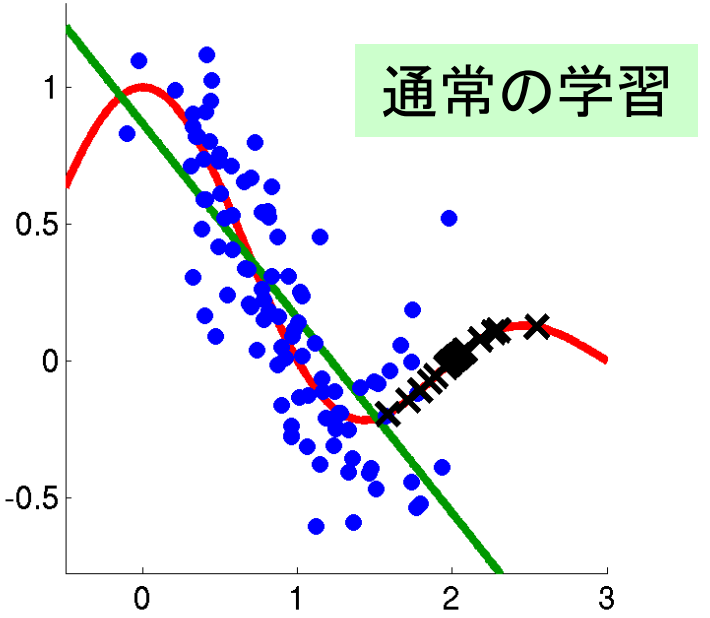


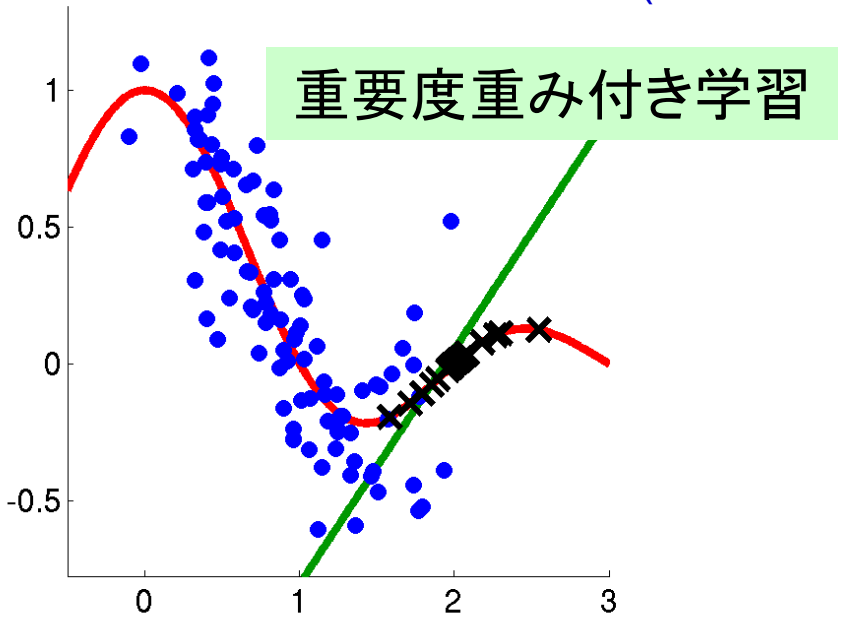
回帰関数とデータ



重要度重み付き学習

Shimodaira (JSPI2000)





$$\underset{f\in\mathcal{F}}{\operatorname{argmin}}\left[\sum_{i=1}^{n_{\mathrm{tr}}}\ell(f(\boldsymbol{x}_i^{\mathrm{tr}}),y_i^{\mathrm{tr}})\right]$$

$$\underset{f\in\mathcal{F}}{\operatorname{argmin}}\left[\sum_{i=1}^{n_{\mathrm{tr}}}\frac{p_{\mathrm{te}}(\boldsymbol{x}_i^{\mathrm{tr}})}{p_{\mathrm{tr}}(\boldsymbol{x}_i^{\mathrm{tr}})}\ell(f(\boldsymbol{x}_i^{\mathrm{tr}}),y_i^{\mathrm{tr}})\right]$$

重要度

$$\{(\boldsymbol{x}_i^{\mathrm{tr}},y_i^{\mathrm{tr}})\}_{i=1}^{n_{\mathrm{tr}}}\overset{\text{i.i.d.}}{\sim}p_{\mathrm{tr}}(\boldsymbol{x},y)$$

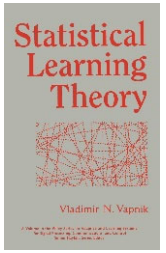
- 共変量シフト下では，通常の学習は一致性がない
- 重要度で重み付けすると，一致性を持つ

■ 重要度はどうやって推定するか？

重要度推定

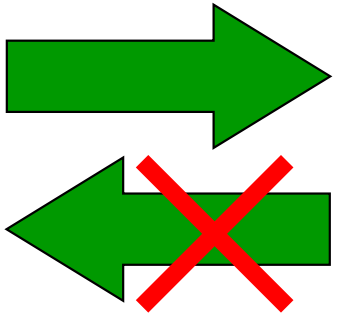
ヴァプニックの原理：Vapnik (Wiley, 1998)

ある問題を解くときに，それより
一般的な問題を途中で解くべきでない



密度を知る

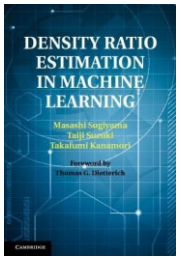
$$p_{\mathrm{te}}(\boldsymbol{x}), p_{\mathrm{tr}}(\boldsymbol{x})$$



密度比を知る

$$r^*(\boldsymbol{x}) = \frac{p_{\mathrm{te}}(\boldsymbol{x})}{p_{\mathrm{tr}}(\boldsymbol{x})}$$

- 密度比の推定より密度の推定の方が一般的
- 密度を推定せず密度比を推定すべき：
  - 直接密度比推定



Sugiyama, Suzuki & Kanamori
(Cambridge University Press, 2012)

最小二乗重要度推定

Kanamori, Hido & Sugiyama (JMLR2009)

- 与えられるデータ: 訓練とテストの入力標本

$$\{\boldsymbol{x}_i^{\mathrm{tr}}\}_{i=1}^{n_{\mathrm{tr}}} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{tr}}(\boldsymbol{x}) \quad \{\boldsymbol{x}_j^{\mathrm{te}}\}_{j=1}^{n_{\mathrm{te}}} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{te}}(\boldsymbol{x})$$

- 重要度モデルrを最小二乗法で直接学習:

$$\min_r Q(r) \qquad Q(r) = \int \Big(r(\boldsymbol{x}) - r^*(\boldsymbol{x})\Big)^2 p_{\mathrm{tr}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

$$r^*(\boldsymbol{x}) = \frac{p_{\mathrm{te}}(\boldsymbol{x})}{p_{\mathrm{tr}}(\boldsymbol{x})}$$

  - データによる経験近似:

$$Q(r) = \int r(\boldsymbol{x})^2 p_{\mathrm{tr}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} - 2\int r(\boldsymbol{x}) p_{\mathrm{te}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + C$$

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

講演の流れ

1. 弱教師付き分類
2. 転移学習
   A) 共変量シフト適応
   B) 最近の発展
3. 雑音ロバスト分類
4. 今後の展望

重要度と予測モデルの同時推定

- 従来は二段階推定:
  1. 重要度を推定

$$\widehat{r} = \operatorname*{argmin}_{r} \mathbb{E}_{p_{\mathrm{tr}}(\boldsymbol{x})}[(r(\boldsymbol{x}) - r^*(\boldsymbol{x}))^2]$$

  2. 推定した重要度を使って予測モデルを重み付き学習

$$\widehat{f} = \operatorname*{argmin}_{f} \mathbb{E}_{p_{\mathrm{tr}}(\boldsymbol{x},y)}[\widehat{r}(\boldsymbol{x})\ell(f(\boldsymbol{x}), y)]$$

- 1.での推定誤差が2.で増幅される恐れ
- 重要度と予測モデルを同時推定したい

上界同時最小化

Zhang, Yamane, Lu & Sugiyama (ACML2020, SNCS2021)

■ 与えられるデータ：訓練入出力とテスト入力の標本

$$\{(\boldsymbol{x}_i^{\mathrm{tr}}, y_i^{\mathrm{tr}})\}_{i=1}^{n_{\mathrm{tr}}} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{tr}}(\boldsymbol{x}, y) \qquad \{\boldsymbol{x}_j^{\mathrm{te}}\}_{j=1}^{n_{\mathrm{te}}} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{te}}(\boldsymbol{x})$$

■ 予測誤差の上界の同時最小化：

$$\min_{r,f} J_{\ell'}(r, f) \qquad J_{\ell'}(r, f) \geq \tfrac{1}{2} R_\ell(f)^2 \qquad R_\ell(f) = \mathbb{E}_{p_{\mathrm{te}}(\boldsymbol{x},y)}[\ell(f(\boldsymbol{x}), y)]$$

$$\ell \leq 1, \ell' \geq \ell, r \geq 0$$

$$J_{\ell'}(r, f) = \mathbb{E}_{p_{\mathrm{tr}}(\boldsymbol{x})}[(r(\boldsymbol{x}) - r^*(\boldsymbol{x}))^2] \quad \leftarrow \text{最小二乗重要度推定}$$

$$+(\mathbb{E}_{p_{\mathrm{tr}}(\boldsymbol{x},y)}[r(\boldsymbol{x})\ell'(f(\boldsymbol{x}), y)])^2 \quad \leftarrow \text{重要度重み付き学習}$$

- 従来法は上界の二段階最小化に相当
- 収束性を理論保証：

$$\widehat{f} = \operatorname*{argmin}_{f \in \mathcal{F}} \min_r \widehat{J}_{\ell'}(r, f)$$

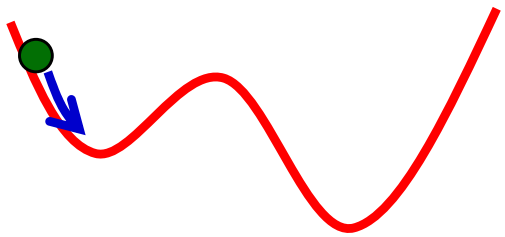
$$R_\ell(\widehat{f}) \leq \sqrt{2} \min_{f \in \mathcal{F}} R_{\ell'}(f) + \mathcal{O}_p(n_{\mathrm{tr}}^{-1/4} + n_{\mathrm{te}}^{-1/4})$$

動的推定

Fang, Lu, Niu & Sugiyama (NeurIPS2020)

- ニューラルネットの確率的勾配学習：

$$f \leftarrow f - \eta \nabla \widehat{R}(f) \qquad \eta > 0 : \text{学習率}$$



- 与えられるデータ：訓練とテストの入出力標本

$$\{(\boldsymbol{x}_i^{\mathrm{tr}}, y_i^{\mathrm{tr}})\}_{i=1}^{n_{\mathrm{tr}}} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{tr}}(\boldsymbol{x}, y) \qquad \{(\boldsymbol{x}_j^{\mathrm{te}}, y_j^{\mathrm{te}})\}_{j=1}^{n_{\mathrm{te}}} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{te}}(\boldsymbol{x}, y)$$

- 各ミニバッチ $\{(\bar{\boldsymbol{x}}_i^{\mathrm{tr}}, \bar{y}_i^{\mathrm{tr}})\}_{i=1}^{\bar{n}_{\mathrm{tr}}}, \{(\bar{\boldsymbol{x}}_j^{\mathrm{te}}, \bar{y}_j^{\mathrm{te}})\}_{j=1}^{\bar{n}_{\mathrm{te}}}$ に対して，重要度をカーネル平均適合で推定：

Huang, et al. (NeurIPS2007)

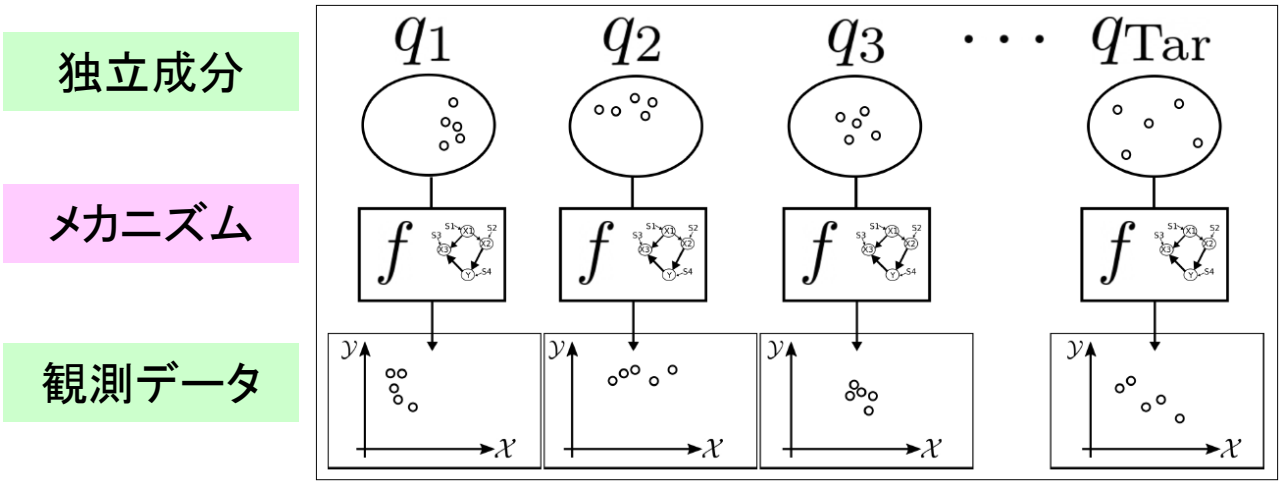
$$\frac{1}{\bar{n}_{\mathrm{tr}}} \sum_{i=1}^{\bar{n}_{\mathrm{tr}}} r_i \ell(f(\bar{\boldsymbol{x}}_i^{\mathrm{tr}}), \bar{y}_i^{\mathrm{tr}}) \approx \frac{1}{\bar{n}_{\mathrm{te}}} \sum_{j=1}^{\bar{n}_{\mathrm{te}}} \ell(f(\bar{\boldsymbol{x}}_j^{\mathrm{te}}), \bar{y}_j^{\mathrm{te}})$$

- 共変量シフトの仮定は不要！

転移学習のまとめ

- 重要度重み付き学習において，
  重要度と予測モデルの同時学習を実現：
  - 共変量シフト：上界同時最小化
  - 任意の分布変化：動的推定
- 訓練・テストデータの分布が大きく異なる場合は？
  - メカニズム転移

Teshima, Sato & Sugiyama (ICML2020)



Bai, Zhang, Zhao, Sugiyama & Zhou (arXiv2202)

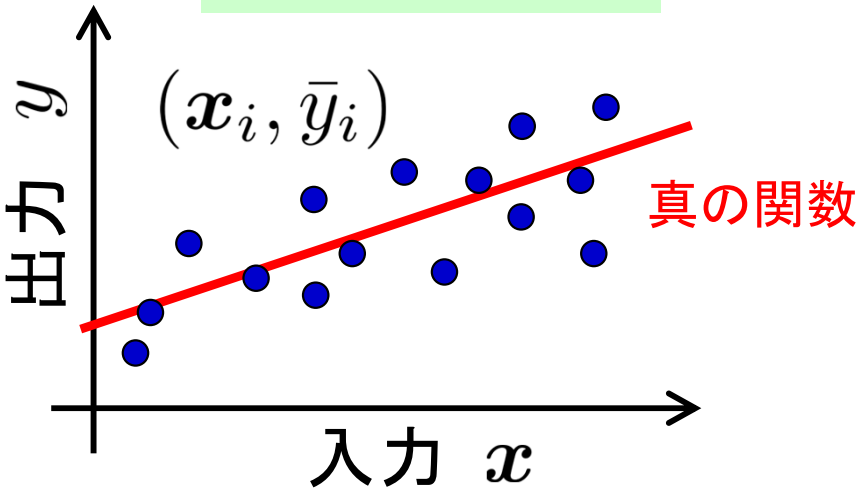
- 更なる発展：連続的な分布変化への対応

講演の流れ

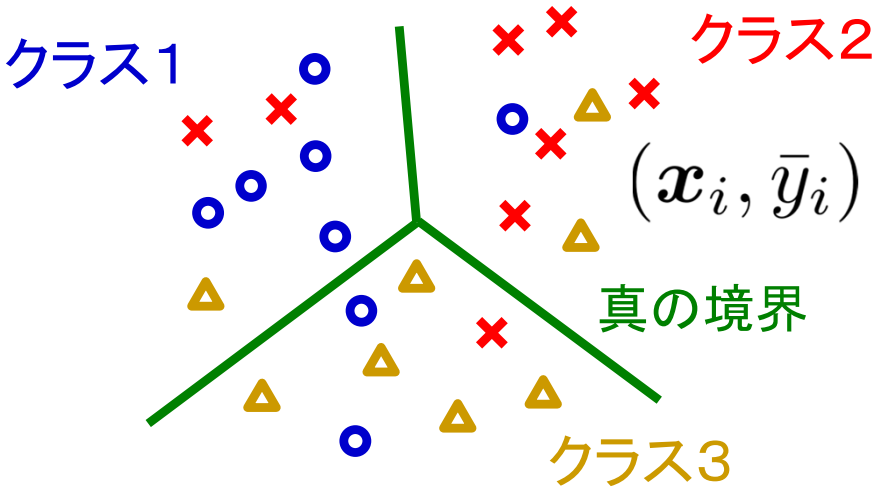
1. 弱教師付き分類
2. 転移学習
3. 雑音ロバスト分類
   A) 雑音遷移行列による補正
   B) 最近の発展
4. 今後の展望

雑音を含む出力からの教師付き学習

回帰(加法雑音)



分類(ラベル反転)



$$\min_g \sum_{i=1}^{n} \ell\left(\bar{y}_i, g(\boldsymbol{x}_i)\right)$$

ℓ : 損失　　g : 予測モデル

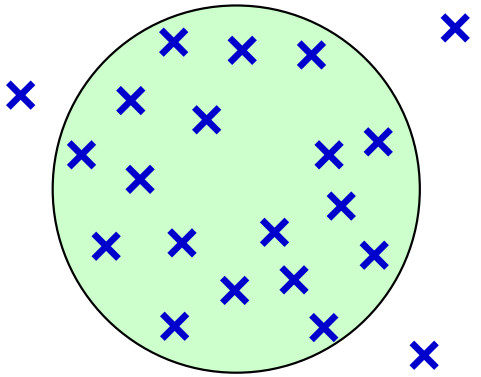
$\bar{y}$: 雑音を含む出力

- このような古典的な問題は，解決済み？
  - 回帰 : 単にデータを増やせばOK(一致性がある)
  - 分類 : 単にデータを増やすだけではダメ(一致性がない)．
    ラベル雑音を抑制する明示的な機構が必要！

ラベル雑音への古典的対処法

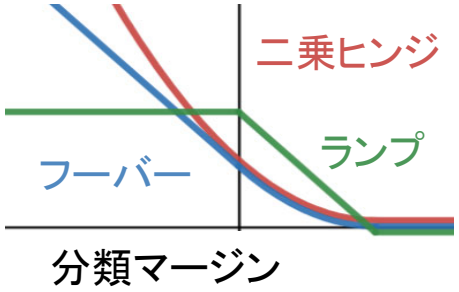
教師なし外れ値除去:

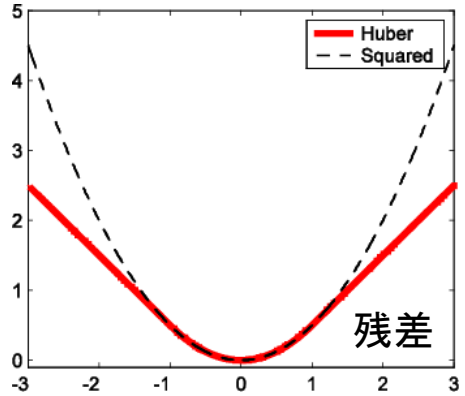
- そもそも教師付き分類よりも難しい



ロバスト損失:

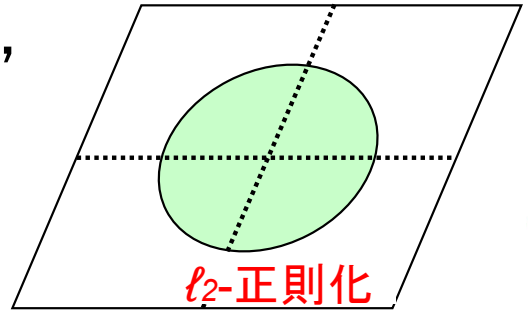
- 回帰に対しては有効だが,
  分類に対する効果は限定的

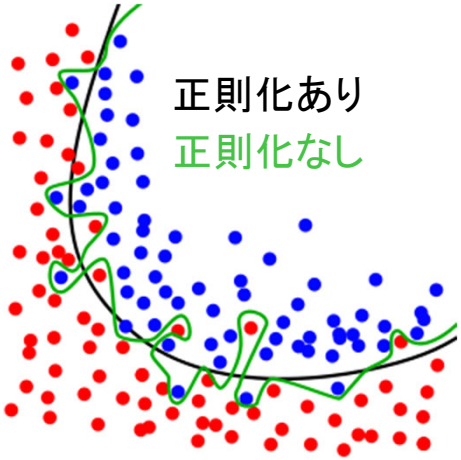




正則化:

- ある程度過適合を防げるが,
  強い雑音に対しては不十分





https://en.wikipedia.org/wiki/Overfitting

新しい手法が必要！

講演の流れ

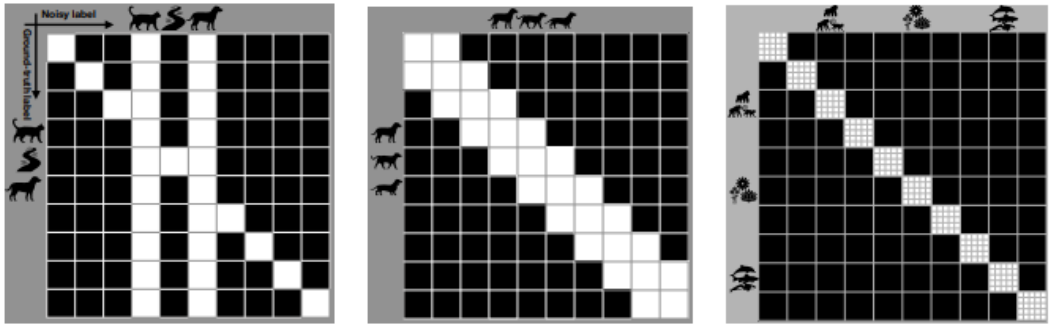
1. 弱教師付き分類
2. 転移学習
3. 雑音ロバスト分類
   A) 雑音遷移行列による補正
   B) 最近の発展
4. 今後の展望

ラベル雑音のモデル化

- 雑音遷移行列: $T_{y,\bar{y}} = \bar{p}(\bar{y}|y)$
  - ラベル y が $\bar{y}$ に反転する確率

y \ $\bar{y}$			
	1	0	0
	0.1	0.8	0.1
	0.5	0.5	0

- 人間の認知バイアスを組み込める:

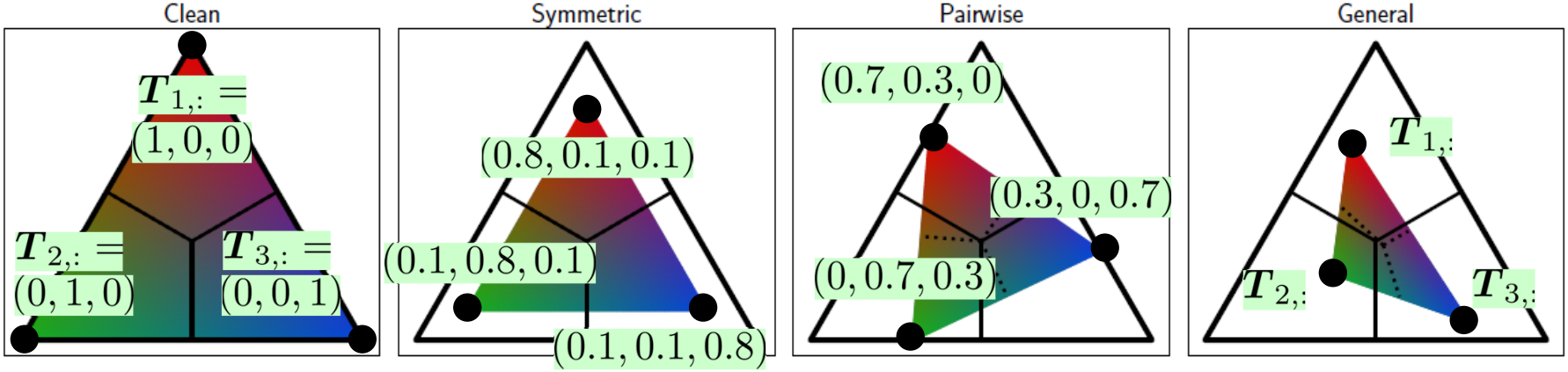


(a) Column-diagonal (b) Tri-diagonal (c) Block-diagonal

Han, Yao, Niu, Zhou, Tsang, Zhang & Sugiyama (NeurIPS2018)

- 単体(三角形)として可視化できる:

Zhang, Niu & Sugiyama (ICML2021)



雑音遷移行列による損失補正

- 基本となる損失補正理論： Patrini et al. (CVPR2017)
  - 前向き補正：T で分類モデルの出力に雑音を付加

$$\ell(T^\top g(x))$$

  - 後向き補正：T^{-1} で損失から雑音を除去

$$T^{-1}\ell(g(x))$$

ℓ：損失　g：分類モデル　$T_{y,\bar{y}} = \bar{p}(\bar{y}|y)$

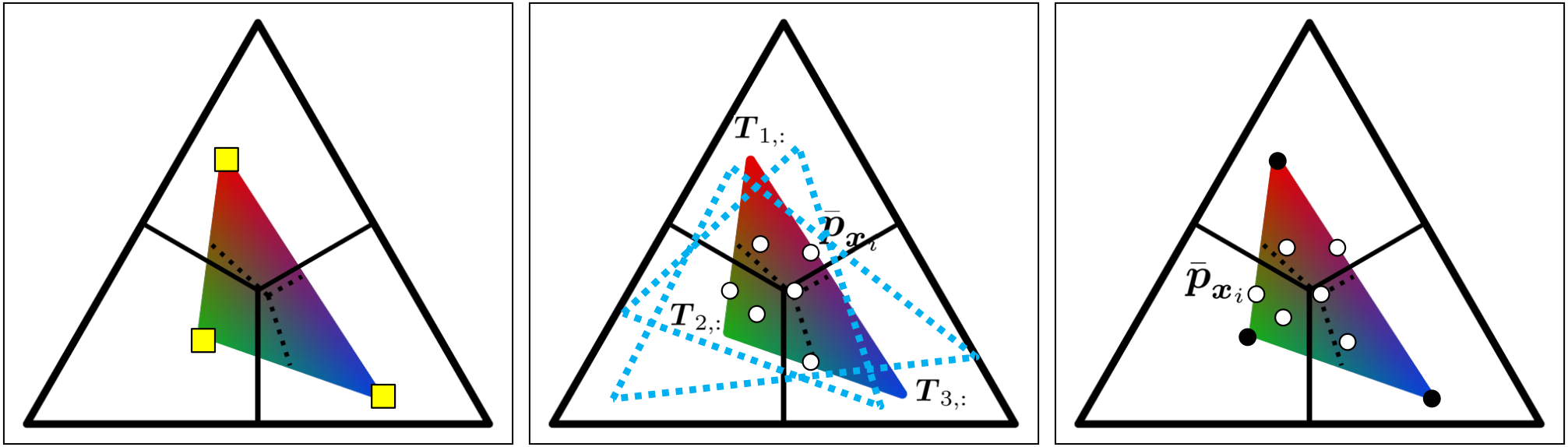
$\ell_y(g(x)) = \ell(y, g(x))$：損失のベクトル表現

- 雑音遷移行列が与えられれば一致性を保証できる

- 雑音遷移行列がわからない場合はどうするか？

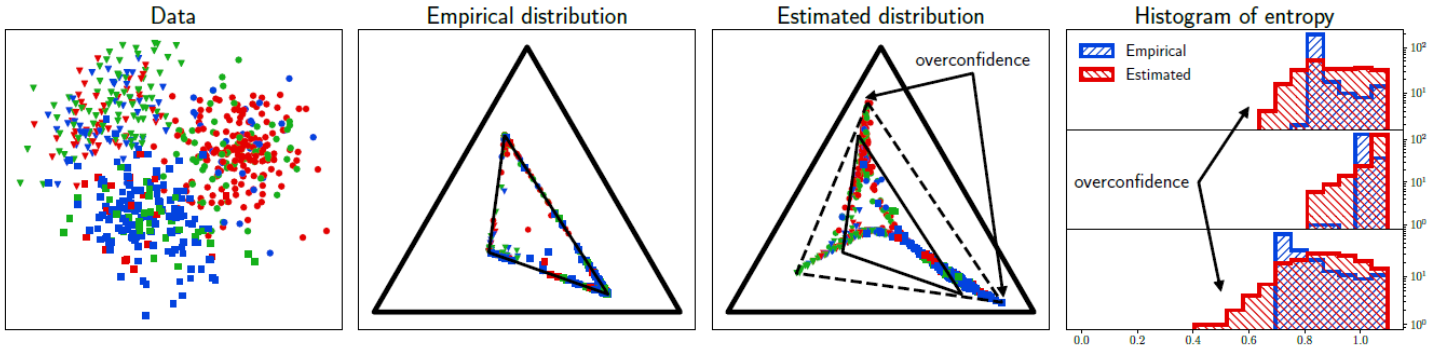
雑音遷移行列の同定不能性

- 雑音なしデータがあれば雑音遷移行列は推定可
  - 三角形の頂点がわかる
- 雑音ありデータだけから推定できないか？
  - 一般には雑音遷移行列は同定不可能
- 雑音なしデータが訓練データ中に存在すると仮定:
  - それらを雑音ありデータを使って見つける



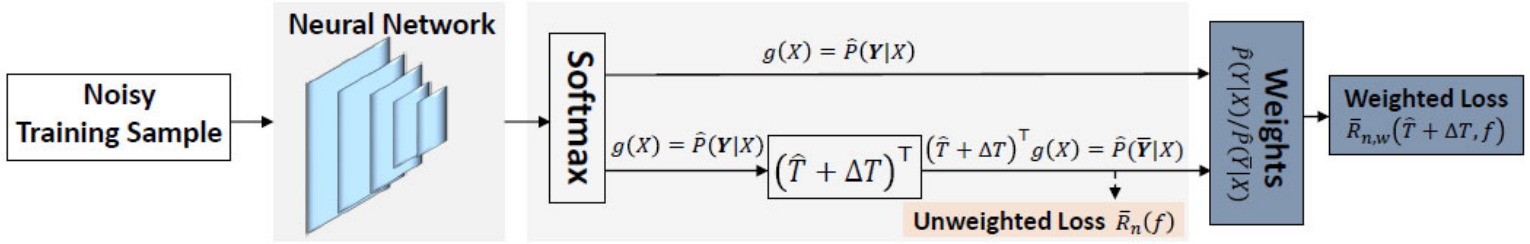
雑音遷移行列の推定

- 雑音を含むデータから確率的分類モデルを学習:
  - 分類の信頼度が最大のものを雑音なしデータとみなす
- しかし，ニューラルネットでは信頼度を過剰推定:



Zhang, Niu & Sugiyama (ICML2021)

- 推定した雑音遷移行列を分類モデル学習中に補正



Xia, Liu, Wang, Han, Gong, Niu & Sugiyama (NeurIPS2019)

- 雑音なしデータを明示的に使わず雑音遷移行列を推定

Yao, Liu, Han, Gong, Deng, Niu, Sugiyama & Tao (NeurIPS2020)

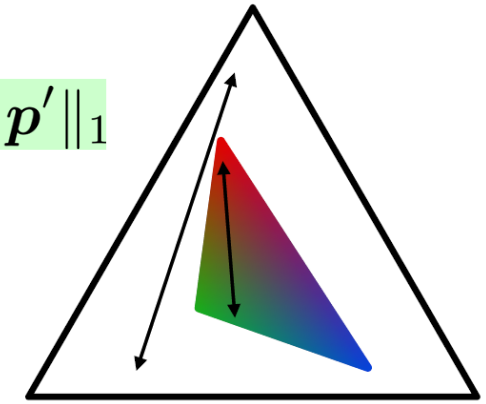
講演の流れ

1. 弱教師付き分類
2. 転移学習
3. 雑音ロバスト分類
   A) 雑音遷移行列による補正
   B) 最近の発展
4. 今後の展望

雑音遷移行列と分類モデルの同時推定

Zhang, Niu & Sugiyama (ICML2021)

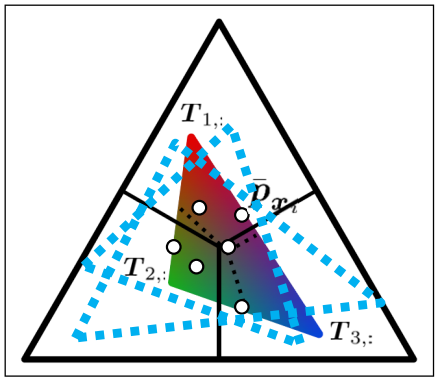
- 従来は二段階推定:
  1. 雑音遷移行列を推定
  2. 推定した雑音遷移行列を使って分類モデルを学習
- 1.での推定誤差が2.で増幅される恐れ
- 雑音遷移行列と分類モデルを同時推定:
  - ナイーブに同時推定すると解が一意に定まらない
  - 雑音遷移の縮小性に基づく正則化により，一致性を理論保証 $\|T^\top p - T^\top p'\|_1 \leq \|p - p'\|_1$
- 雑音なしデータが手元になくても，その存在を暗に仮定できればOK

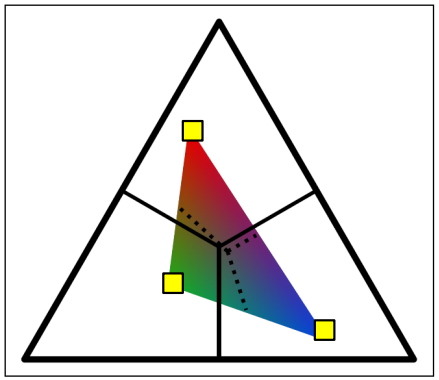


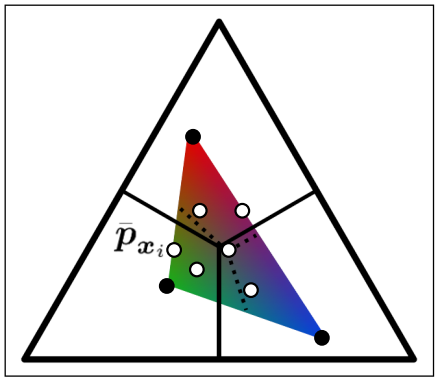
同定可能条件の一般化

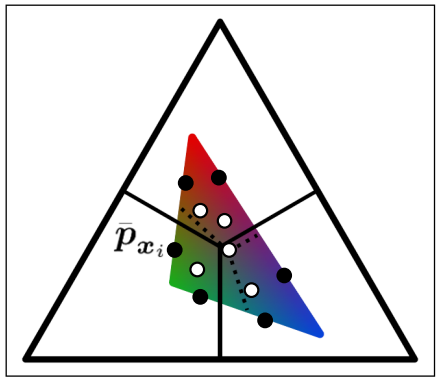
- 一般には雑音遷移行列は同定不可能:
  - 雑音なしデータが明示的に与えられれば同定可能
  - 雑音なしデータが訓練データ中に存在すれば同定可能
  - 雑音なしデータが分布中に存在すれば同定可能
- 雑音なしデータがなくても，訓練データが「十分に散らばって」いれば同定可能:
  - 一辺に二つの点が存在
  - 体積最小化正則化により一致性を理論保証

Li, Liu, Han, Niu & Sugiyama (ICML2021)



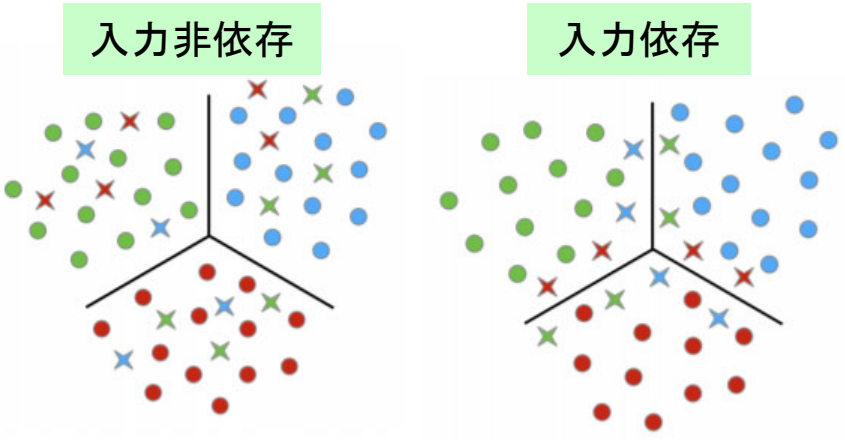






入力依存ラベル雑音への拡張

- 現実世界ではラベル雑音は入力に依存する？
  - 例：境界近くは雑音が大きい



- 雑音遷移行列関数：$T_{y,\bar{y}}(\boldsymbol{x}) = \bar{p}(\bar{y}|y,\boldsymbol{x})$
  - いかに $\boldsymbol{T}(\boldsymbol{x})$ を（近似的に）推定するか？

- ヒューリスティックな解法：
  - パーツ分解に基づく近似
  - 追加の信頼度情報の活用
  - 多様体正則化の活用

Xia, Liu, Han, Wang, Gong, Liu, Niu, Tao & Sugiyama (NeurIPS2020)

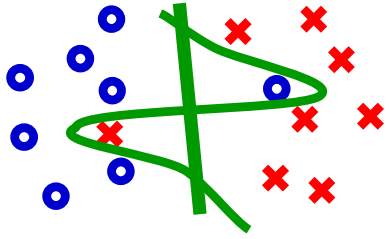
Berthon, Han,Niu, Liu & Sugiyama (ICML2021)

Cheng, Liu, Ning, Wang, Han, Niu, Gao & Sugiyama (CVPR2022)

共教示：雑音ロバスト分類の別研究

ニューラルネットの記憶能力：

- 確率的降下学習は雑音なしデータを早く記憶
- しかし，単純な早期終了ではうまくいかない



2つのニューラルネットを用いた共教示：

- 誤差の小さいデータを選んで教え合う

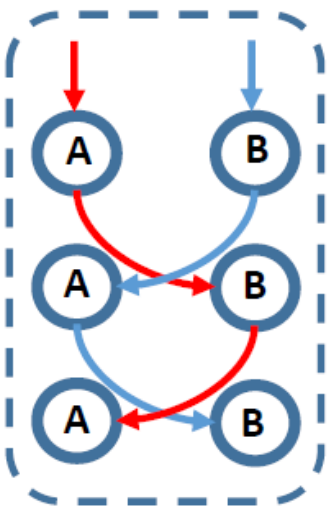
  Han, Yao, Yu, Niu, Xu, Hu, Tsang & Sugiyama (NeurIPS2018)

- 出力が合致しないデータだけを教える

  Yu, Han, Yao, Niu, Tsang & Sugiyama (ICML2019)

- 誤差の大きいデータに対して勾配上昇

  Han, Niu, Yu, Yao, Xu, Tsang & Sugiyama (ICML2020)



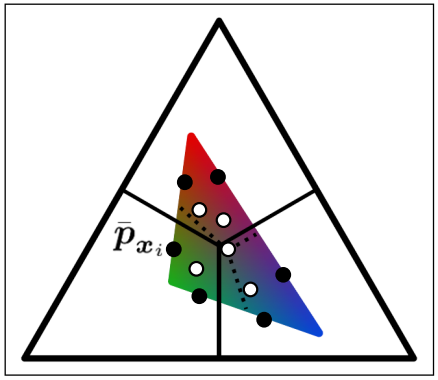
理論はないが，実験的には超ロバスト：

- 50%のラベルをランダムに変えても大丈夫！

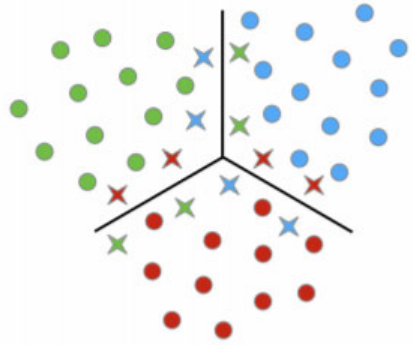
雑音ロバスト分類のまとめ

- 回帰と違い，分類ではラベル雑音の影響を明示的に考慮する必要がある
- しかし，雑音遷移行列は一般に同定不可能：
  - 雑音遷移行列の推定が課題
  - ある程度弱い条件のもとで，一致性を持つ学習法の開発に成功

$$T_{y,\bar{y}} = \bar{p}(\bar{y}|y)$$



- 実問題では雑音遷移行列が入力依存：
  - ヒューリスティックな解法をいくつか開発



- 共教示による超ロバスト学習：
  - ヒューリスティックな解法をいくつか開発

講演の流れ

1. 弱教師付き分類
2. 転移学習
3. 雑音ロバスト分類
4. 今後の展望

技術的な着目点

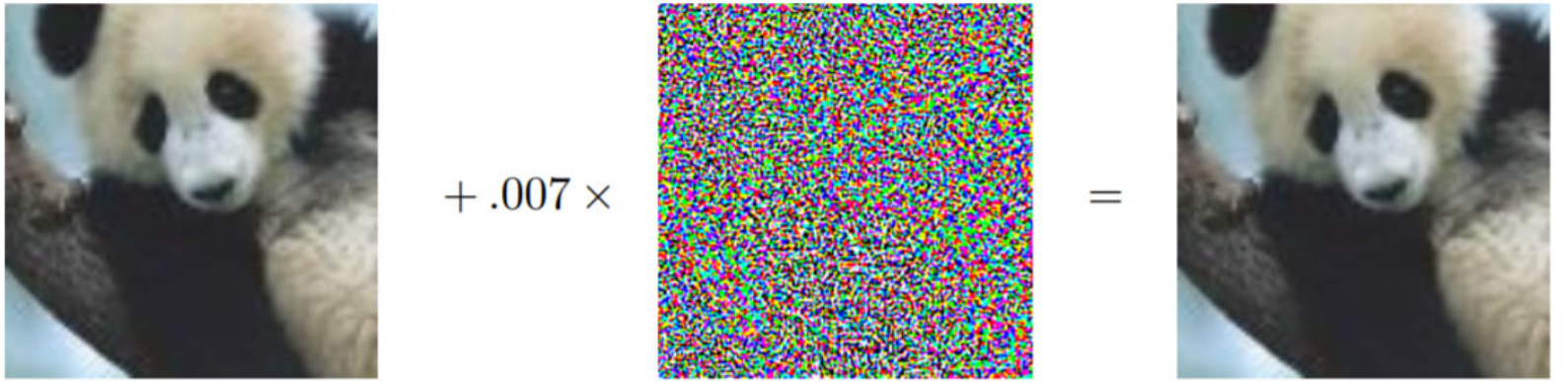
- カーネル法などの従来の学習法では，
  最終的な学習結果をきちんと求められるため，
  その良し悪しを議論していた
- 深層学習は，最終的な学習結果がきちんと
  求められないため，徐々に学習していく
  - 学習の途中結果を利用するという新しい概念が登場

- これにより，従来手法の限界を超える
  超ロバスト機械学習手法が開発されつつある

重要課題：敵対的攻撃への対応

- ニューラルネットはテスト入力の微小な摂動に弱い

Goodfellow et al. (ICLR2015)



- 多数の攻撃・防御の研究が行われているが，未だ決定的な解決策は見つかっていない

Tsuzuku, Sato & Sugiyama (NeurIPS2018)
Hu, Niu, Sato & Sugiyama (ICML2018)
Ni, Charoenphakdee, Honda & Sugiyama (NeurIPS2019)
Bao, Scott & Sugiyama (COLT2020)
Zhang, Xu, Han, Niu, Cui, Sugiyama & Kankanhalli (ICML2020)
Zhang, Zhu, Niu, Han, Sugiyama & Kankanhalli (ICLR2021)
Du, Zhang, Han, Liu, Rong, Niu, Huang, & Sugiyama (ICML2021)
Gao, Liu, Zhang, Han, Liu, Niu & Sugiyama (ICML2021)
Yan, Zhang, Niu, Feng, Tan & Sugiyama (ICML2021)
Wang, Liu, Han, Liu, Gong, Niu, Zhou & Sugiyama (NeurIPS2021)
Yan, Zhang, Feng, Sugiyama & Tan (IJCAI2022)
Xu, Zhang, Liu, Sugiyama & Kankanhalli (ICML2022)
Zhang, Xu, Han, Liu, Niu, Cui & Sugiyama (TMLR2022)

まとめ

- 機械学習システムの信頼性向上は不可欠:
  - 想定できる悪状況に対するロバスト性:
    - 悪状況をモデル化し悪影響を補正:
      悪状況のモデル化誤差をどう抑えるかが課題
  - 想定できない悪状況に対するロバスト性:
    - 最悪ケースを考えて補正:
      保守的になりすぎないようにすることが課題
    - 人間のサポートに頼る:
      自動運転など実時間応用では使えない？
  - 実用的にはその中間くらいが重要そう？
- ヒトとの相性が良くなればAIの信頼性は増す？
  - 脳の学習機構，認知バイアスなどを取り込む
  - 社会常識，文化などをAIに反映

共同研究者

■東大：機械学習と統計的データ解析研究室

- 講師
  - 横矢 直人（複雑理工学，コンピュータ科学，情報科学，理研）
  - 石田 隆（複雑理工学，コンピュータ科学，情報科学）
- 特任講師
  - 伊藤 信貴（複雑理工学）
- 准教授 （2020年4月より佐藤研究室へ異動）
  - 佐藤 一誠（コンピュータ科学，情報科学，複雑理工学）
- 特任助教
  - 蒋 兆凱（複雑理工学）
- 特任研究員（ポスドク）
  - 長野 祥大（複雑理工学）
  - 吴 栋贤（複雑理工学）
  - 鲁 楠（複雑理工学）
- 特任専門職員
  - 川島 優子（複雑理工学）
  - 横井 創磨（複雑理工学）
  - 佐藤 芙実（複雑理工学）
- 博士学生
  - 得居 誠也（コンピュータ科学）※佐藤研究室所属
  - 中台慎二（コンピュータ科学）
  - 野沢 健人（複雑理工学）※佐藤研究室所属
  - 木了 龍一（コンピュータ科学）
  - 藤澤 将広（複雑理工学）※佐藤研究室所属
  - 李 鐘瑛（コンピュータ科学）
  - 張 天绎（複雑理工学）
  - 張 一凡（コンピュータ科学）
  - チョッカリンガム ヴァリアッパ（コンピュータ科学）
  - リュ シャルル（コンピュータ科学）
  - 房 彤彤（複雑理工学）
  - 陳 柏佑（複雑理工学）
  - 董 暁宇（複雑理工学）
  - 張 宇傑（複雑理工学）
  - 蔡 欣強（複雑理工学）
  - 宋 健（複雑理工学）
  - 甘 万水（複雑理工学）
  - 唐 玉亭（複雑理工学）
  - 中村 紳太郎（複雑理工学）
  - ラヴェー オア（複雑理工学）
  - アッカーマン ヨハネス（コンピュータ科学）
- 修士学生
  - 梶塚 時央（コンピュータ科学）
  - 朴 炯奎（複雑理工学）※佐藤研究室所属
  - 王 旭杰（複雑理工学）
  - 小寺 俊希（コンピュータ科学）
  - 青木 勇磨（コンピュータ科学）
  - 李 佳歡（コンピュータ科学）
  - 梁 程偉（コンピュータ科学）
  - 周 煥健（複雑理工学）
  - 楊 坤（複雑理工学）
  - 白坂 貴規（複雑理工学）
  - 飯塚 玲夫（複雑理工学）
  - 侯 驍謀（複雑理工学）
  - メータセート アナン（コンピュータ科学）
  - エラート セマル（コンピュータ科学）
  - 山本 健斗（コンピュータ科学）
  - 太田 一毅（コンピュータ科学）
  - 八尋 一宇（コンピュータ科学）
  - 藤田 光（コンピュータ科学）
  - 姚 禹（複雑理工学）
  - 中野 嘉文（複雑理工学）
  - 西森 創一朗（複雑理工学）
  - 牛尾 凌太（複雑理工学）
  - 賢 天奎（複雑理工学）

東京大学 THE UNIVERSITY OF TOKYO

RIKEN

研究員 Gang Niu	特別研究員 Jingfeng Zhang
特別研究員 Jiaqi Lyu	特別研究員 Shuo Chen
客員主管研究員 中島 伸一	客員研究員 長 隆之
客員研究員 Florian Baptiste Yger	客員研究員 黒木 祐子
客員研究員 Miao Xu	客員研究員 Feng Liu
客員研究員 Bo Han	客員研究員 Tongliang Liu
客員研究員 Lei Feng	

＋多数のインターン生

■理研AIP:不完全情報学習チーム