

# 人工知能技術の 最新動向の紹介



理化学研究所／東京大学



東京大学  
THE UNIVERSITY OF TOKYO

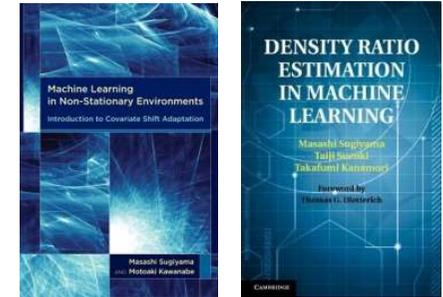
杉山 将



# 自己紹介

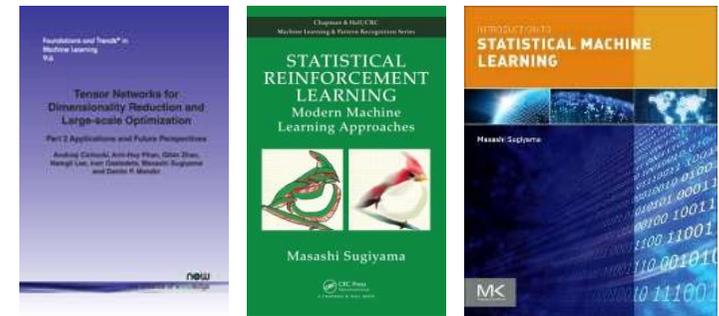
## ■ 現職:

- 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:

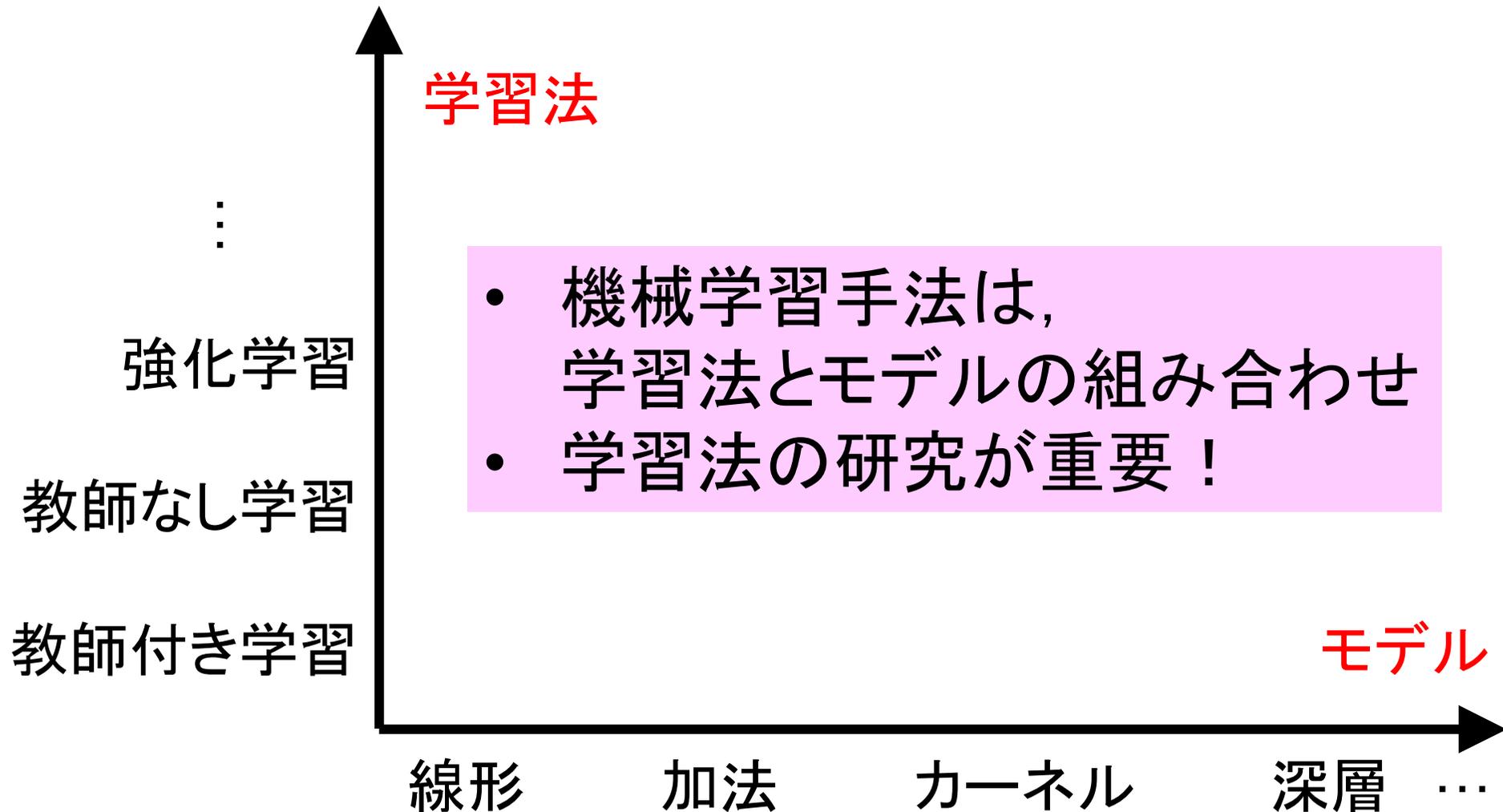
- 人工知能の数学的な基礎研究 (コンピュータ, 統計学など)
- 人工知能技術の実世界応用 (画像, 言語, 脳波, ロボット, 医療, 生命など)



# 研究方針

3

- 世間は「深層学習 & ビッグデータ」で賑わっているが、それが全てではない！





# 講演の流れ

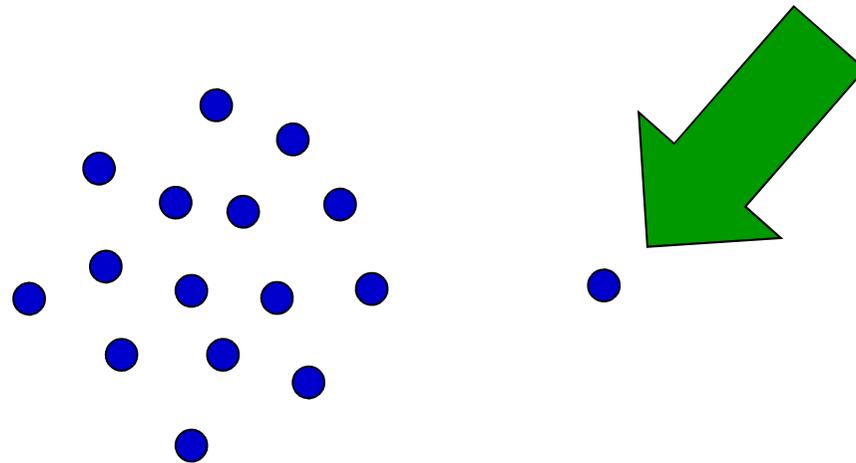
4

1. 異常検知
2. 弱教師付き分類

# 異常検出

5

- **目標**: 与えられたデータ  $\{x_i\}_{i=1}^n$  に含まれる **異常値**を見つける



# 従来アプローチ

6

## ■ 局所異常因子法

- 密度が低いとき, 異常度が高い
- 近傍点の密度が高いとき, 異常度が高い

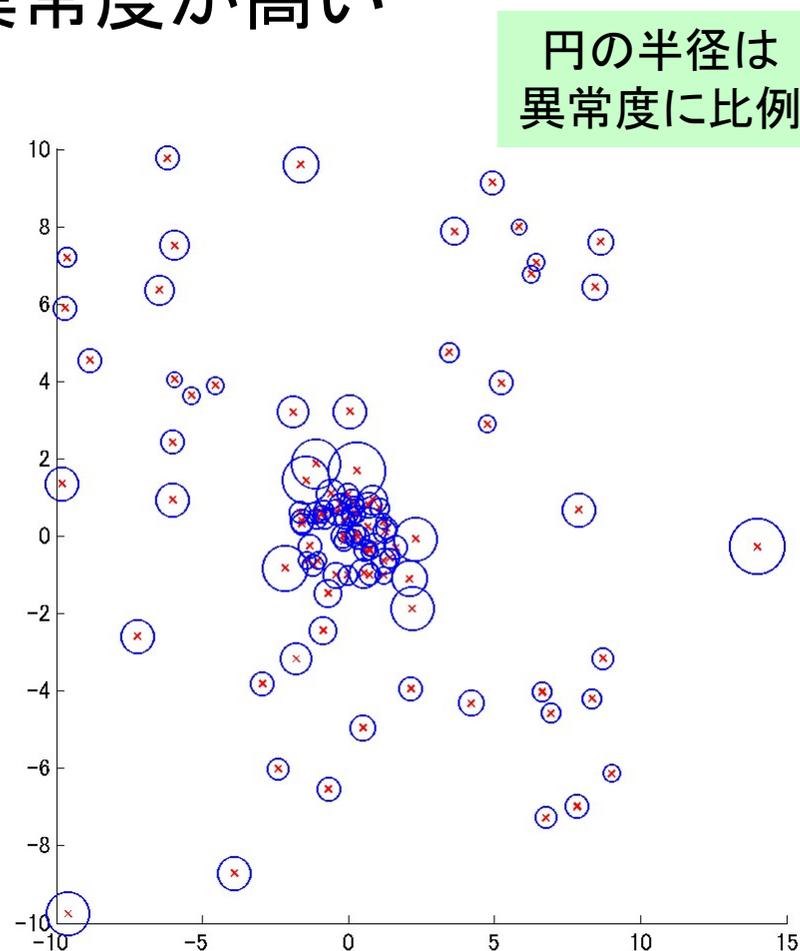
## ■ あらかじめ定められた ルールに従い異常を検知 (学習はしない)

$$\text{LOF}_k(\mathbf{x}) = \frac{\frac{1}{k} \sum_{i=1}^k \text{LRD}_k(\mathbf{x}^{(i)})}{\text{LRD}_k(\mathbf{x})}$$

$$\text{LRD}_k(\mathbf{x}) = \left( \frac{1}{k} \sum_{i=1}^k \text{RD}_k(\mathbf{x}^{(i)}, \mathbf{x}) \right)^{-1}$$

$$\text{RD}_k(\mathbf{x}, \mathbf{x}') = \max \left( \|\mathbf{x} - \mathbf{x}^{(k)}\|, \|\mathbf{x} - \mathbf{x}'\| \right)$$

$\mathbf{x}^{(k)} : \{\mathbf{x}_i\}_{i=1}^n$ の中で $\mathbf{x}$ に $k$ 番目に近い標本



# 従来アプローチ(続き)

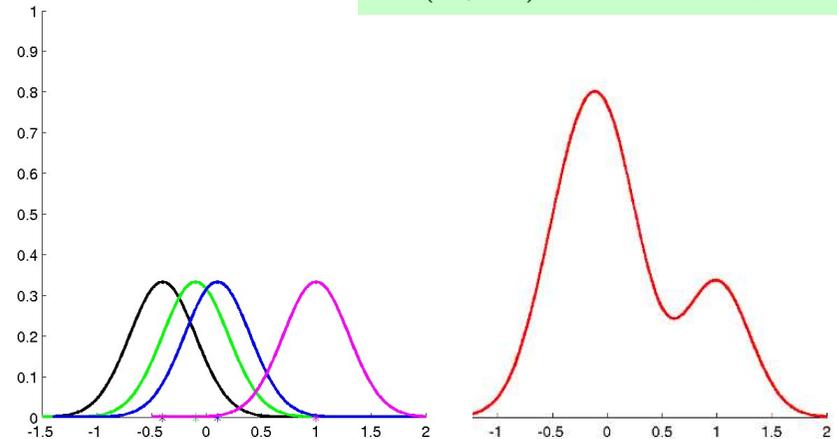
7

## ■カーネル密度推定

- 密度が低いとき, 異常とみなす
- 高次元データに対して精度が悪い

$$\hat{p}(x) = \frac{1}{n} \sum_{i=1}^n K(x, x_i)$$

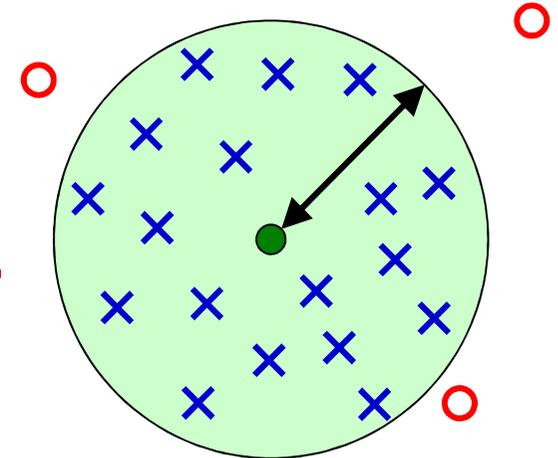
$K(x, x')$ :カーネル関数



## ■サポートベクトル異常検知

- 大多数のデータを含む超球を求める
- チューニングパラメータを客観的に決めるすべがない

## ■データの分布や超球は学習するが, どの点が異常かは学習しない



# 弱教師付き異常検知

8

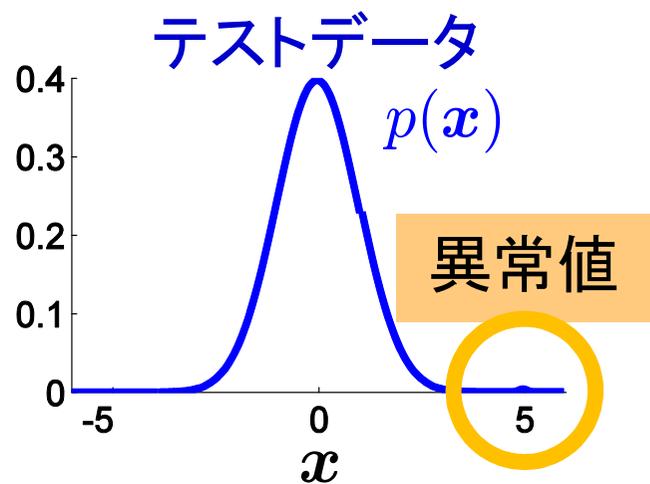
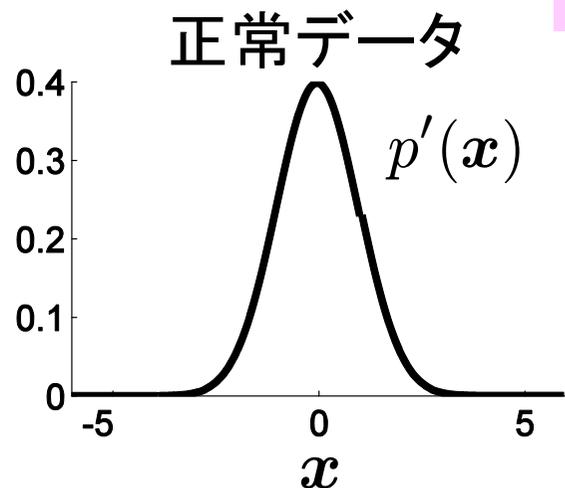
- ラベルなしの標本  $\{x_i\}_{i=1}^n$  だけでは、まともに異常検知はできない
- **正常とわかっている標本**  $\{x'_{i'}\}_{i'=1}^{n'}$  が別途与えられると仮定する
- **ナイーブな方法**: 正常標本の尤度が大きくなるようにパラメータを調整

$$\prod_{i'=1}^{n'} \hat{p}(x'_{i'})$$

- 確率密度の推定を避けたい

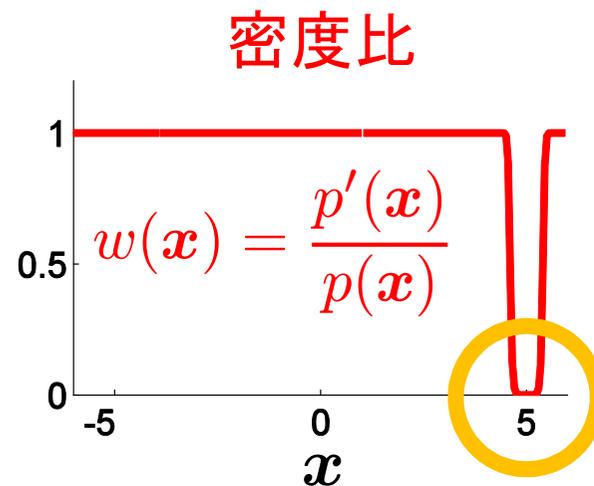
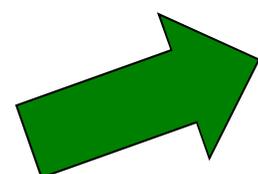
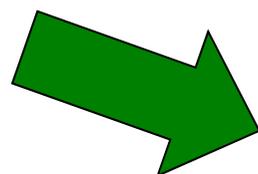
# 確率密度比の推定

■ 異常値は密度比  $w(\boldsymbol{x}) = \frac{p'(\boldsymbol{x})}{p(\boldsymbol{x})}$  の値が小さい

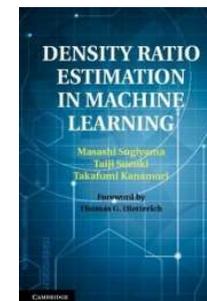


$$\{\boldsymbol{x}'_{i'}\}_{i'=1}^{n'} \stackrel{\text{i.i.d.}}{\sim} p'(\boldsymbol{x})$$

$$\{\boldsymbol{x}_i\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} p(\boldsymbol{x})$$



各密度を推定せず、  
密度比を直接推定！



# 実世界応用例

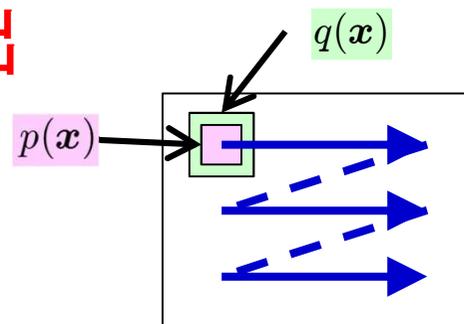
10

- 製鉄プロセスの異常診断
- 光学部品の品質検査
- ローン顧客の審査

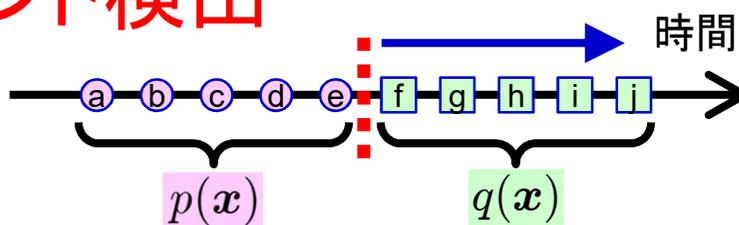
# 同様な技術で変化検知も可能

11

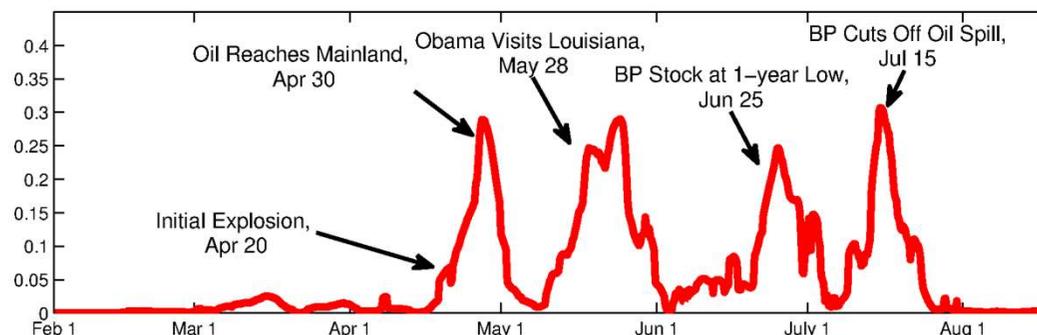
## ■ 画像中の注目領域抽出



## ■ 動画からのイベント検出



## ■ ツイッターデータ解析



# 密度比に基づく異常検出:まとめ<sup>12</sup>

- 一つの標本集合だけから異常は学習できない
- 異常を見つけないたいテスト標本に加えて、正常標本の集合が与えられていると仮定する
- 正常標本とテスト標本の確率密度を求めずに、それらの**確率密度の比**を直接推定する
- 交差確認により**調整パラメータの値**を**客観的に決定**できる



# 講演の流れ

13

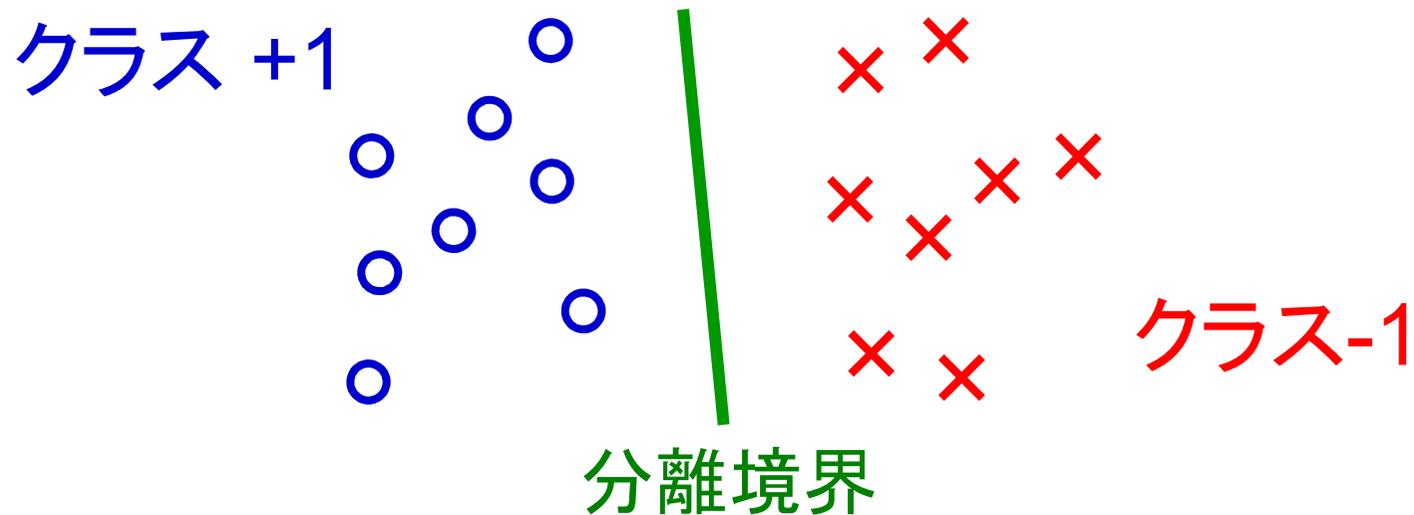
1. 異常検知
2. 弱教師付き分類

# ビッグデータを用いた機械学習 14

- 画像認識, 音声認識, 機械翻訳などで, 人間と同等かそれ以上の性能を達成
- しかし, 応用分野によっては, 教師付きビッグデータを簡単に取れない
  - 医療データ解析
  - インフラの管理
  - 自然災害の防災・減災
  - 機能材料の開発
- 限られた情報からの学習が重要！

# 2クラスの教師付き分類

15

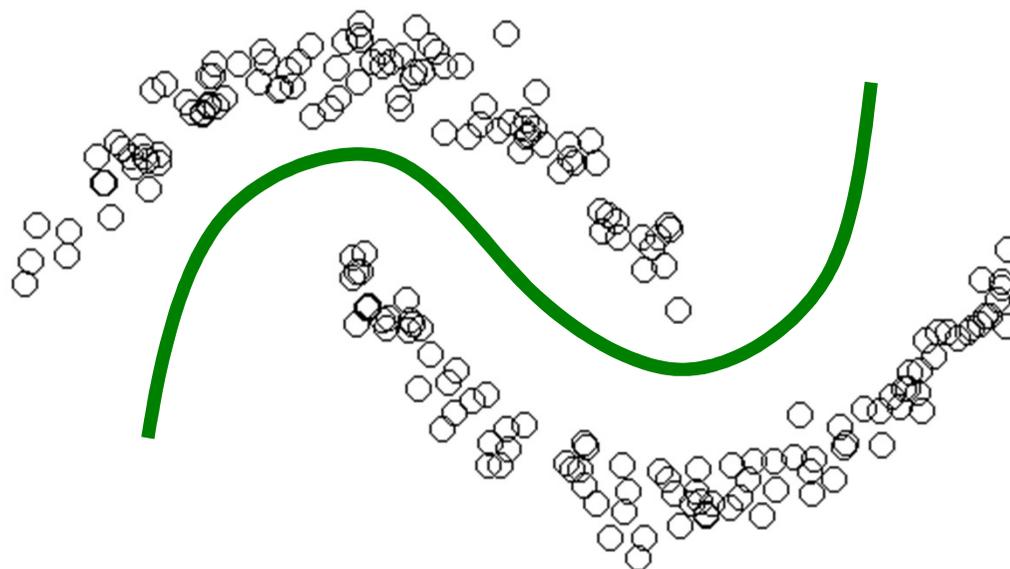


- 大量のラベル付きデータを用いれば、精度良く分類境界が学習できる
- ラベル付きデータ数 $n$ に対して、分離境界の推定誤差は $1/\sqrt{n}$ の速さで減っていく

# 教師なし分類

16

- ラベル付きデータの収集にはコストがかかるため、容易に入手できるラベルなしデータを用いる

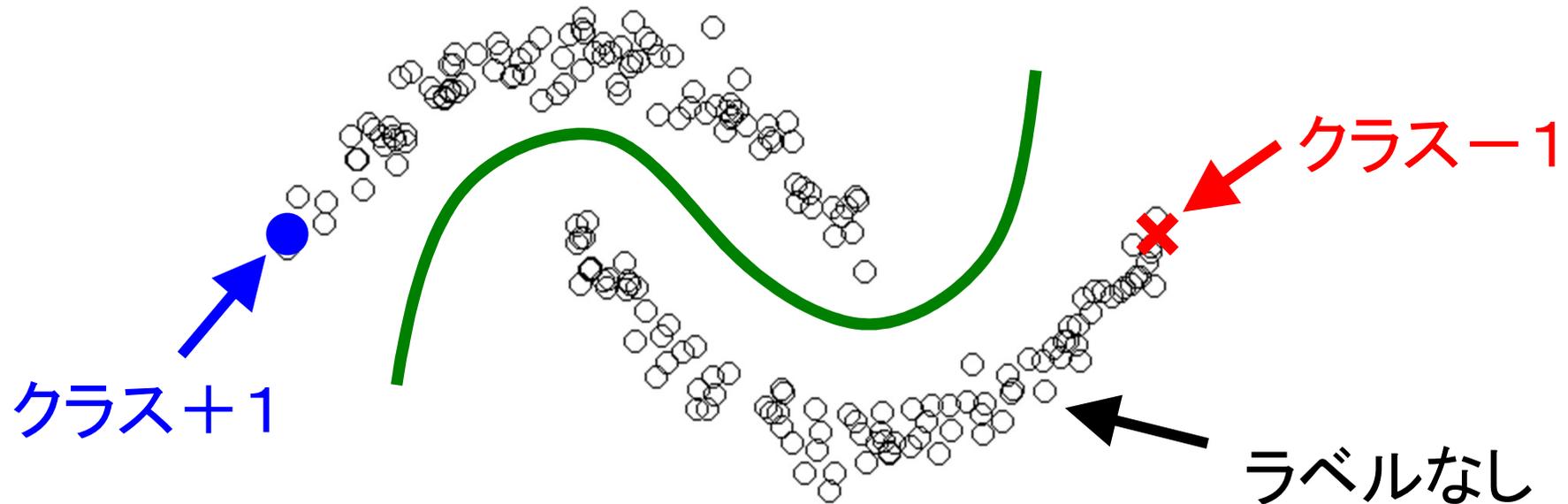


- 教師なし分類はただのクラスタリングに過ぎない
- データがクラス毎にクラスタに分かれていないと、正しく分類できない

# 半教師付き分類

17

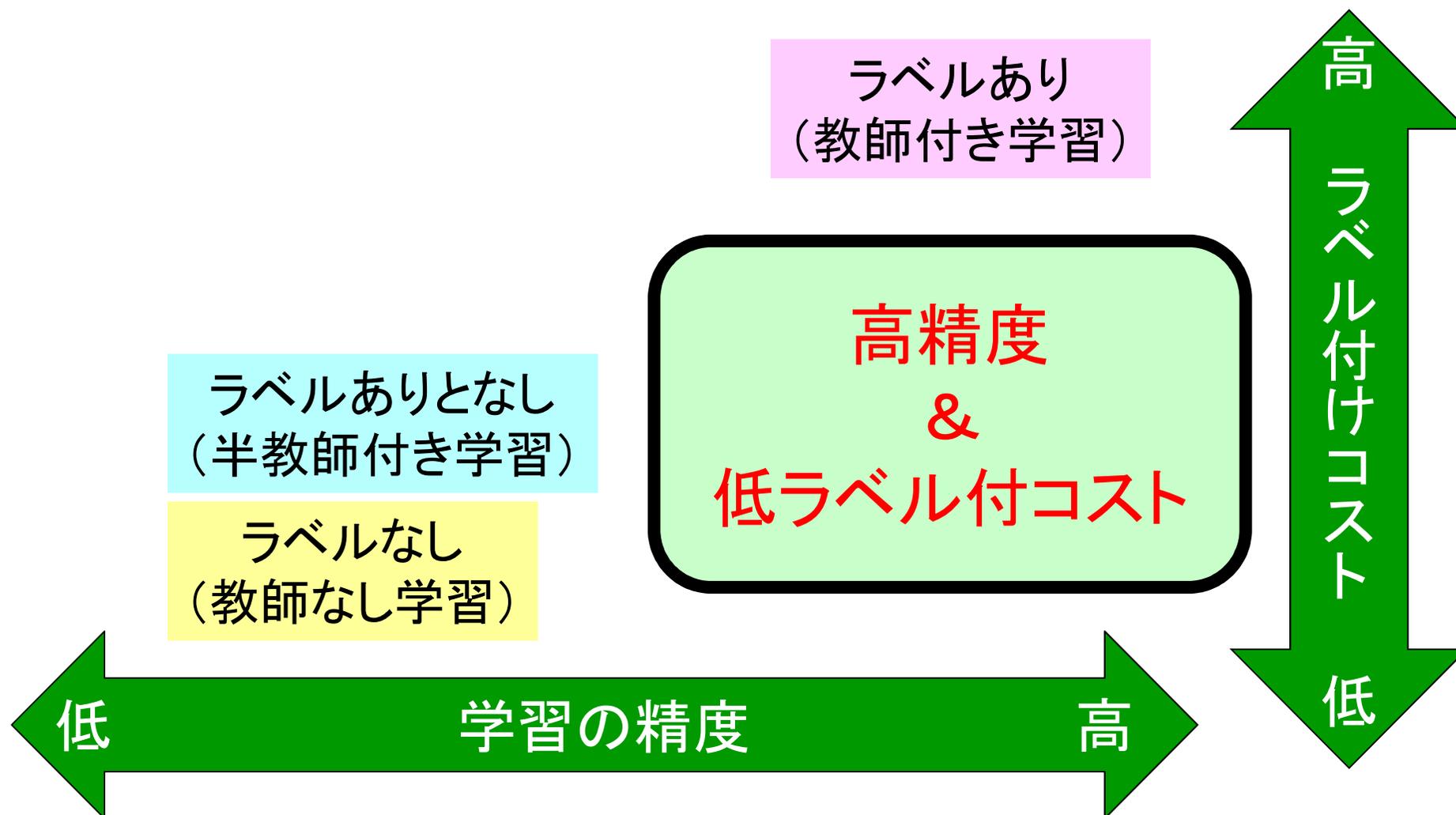
- 少量のラベル付きデータと大量のラベルなしデータを利用
- ラベルなしデータがなすクラスタ構造に従って分類



- 同じクラスタに属するデータが同じラベルを持つとき, うまく分類できる
- そのような仮定が常に成り立つとは限らない

# 分類問題の分類

- 高精度でラベル付コストの低い分類手法が重要！

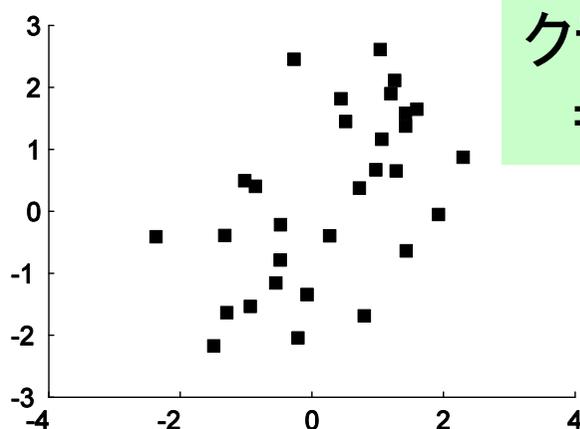


# 新手法1: 教師なし分類

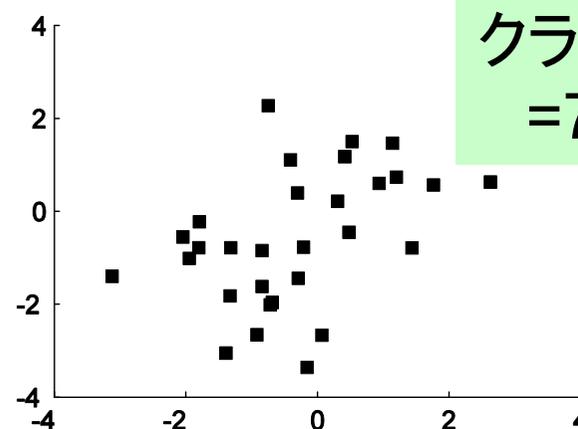
19

du Plessis, Niu & Sugiyama (TAAI2013)

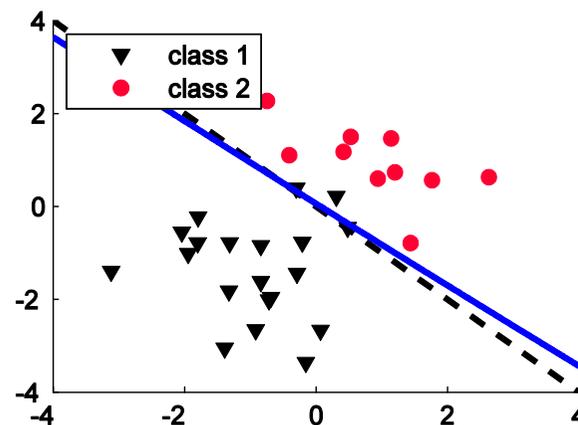
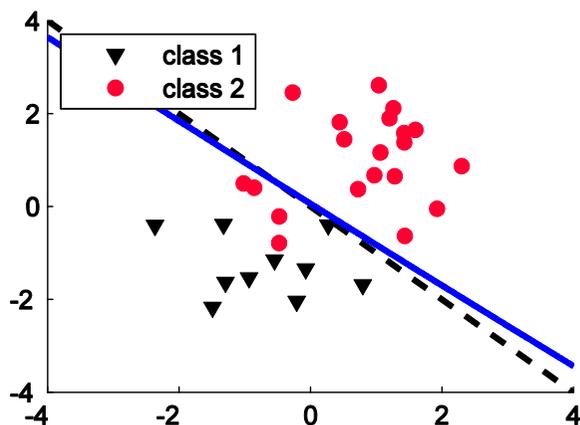
- クラスタ構造がない場合でも, **クラス比の異なる** ラベルなしデータが2セットあれば, 教師付き学習と同じ収束率を達成可能



クラス比  
=3:7



クラス比  
=7:3



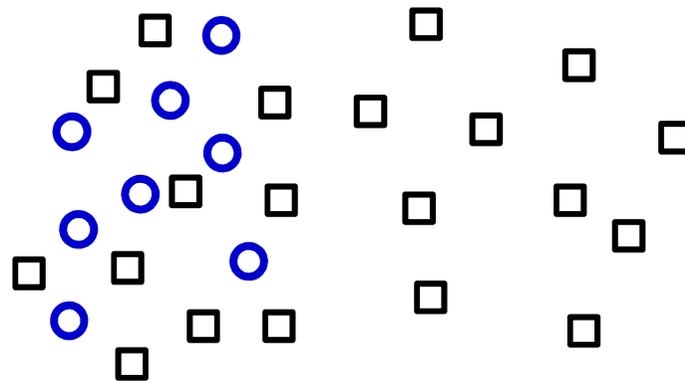
# 新手法2: 正例とラベルなし データからの分類

20

Niu, du Plessis, Sakai, Ma & Sugiyama (NIPS2016)  
du Plessis, Niu & Sugiyama (NIPS2014, ICML2015)

- 負例が全くななくても, 正例とラベルなしデータだけから, 教師付き学習と同じ収束率を達成可能

クラス +1



ラベルなし

(クラス +1と

クラス -1の混合)

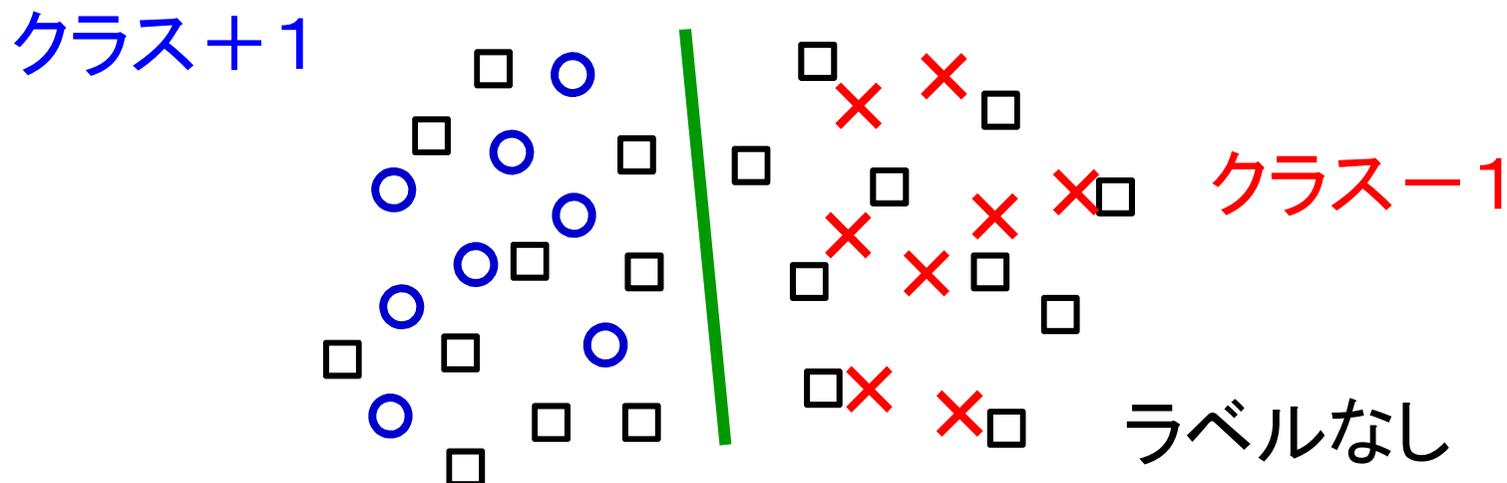
- 例:

- クリック vs. 非クリック
- 友達 vs. 非友達

# 新手法3: 半教師付き分類

21

Sakai, du Plessis, Niu & Sugiyama (ICML2017)



- 「正例とラベルなしデータからの分類」と「正例と負例からの分類」を組み合わせる

$$R_{\text{PU+PN}}^{\gamma}(f) = \gamma R_{\text{PU}}(f) + (1 - \gamma) R_{\text{PN}}(f) \quad 0 \leq \gamma \leq 1$$

- 提案法では、ラベルなしデータからもラベルの情報が抽出できる！

$$R_{\text{PU+PN}}^\gamma(f) = \gamma R_{\text{PU}}(f) + (1 - \gamma) R_{\text{PN}}(f) \quad 0 \leq \gamma \leq 1$$

## ■ 汎化誤差上界:

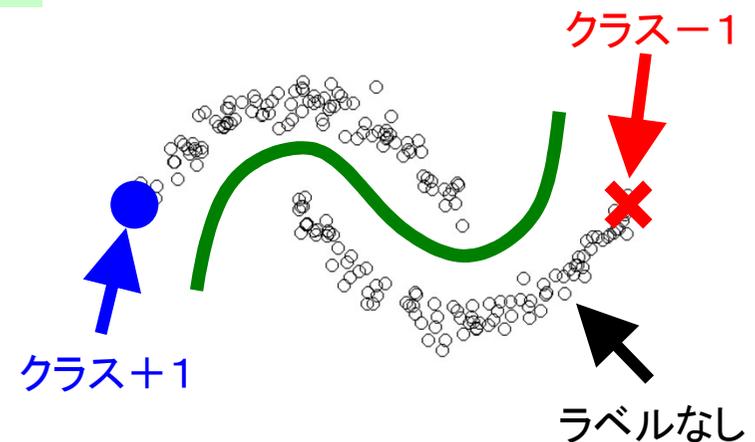
$$R_{\ell_{0/1}}(f) \leq 2\hat{R}_{\text{PU+PN}}^\gamma(f) + C(\delta) \left( \frac{(2 - \gamma)\pi}{\sqrt{n_P}} + \frac{\gamma(1 - \pi)}{\sqrt{n_N}} + \frac{(1 - \gamma)}{\sqrt{n_U}} \right)$$

with probability  $1 - \delta$

$n_P, n_N, n_U$  : 正, 負,  
ラベルなしデータの数

$\hat{R}_{\text{PU+PN}}^\gamma$  :  $R_{\text{PU+PN}}^\gamma$  の経験近似

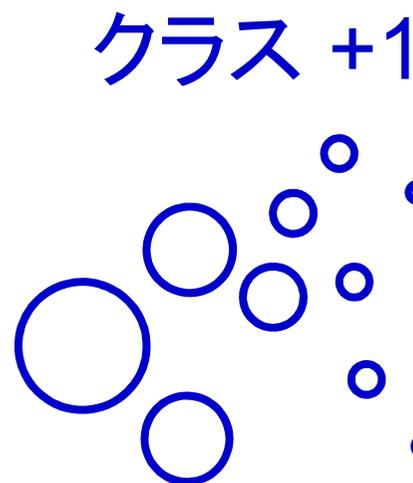
## ■ クラスタ仮定がなくても ラベルなしデータが 活用できる!



# 新手法4: 正信頼度からの分類 23

Ishida, Niu & Sugiyama (arXiv2017)

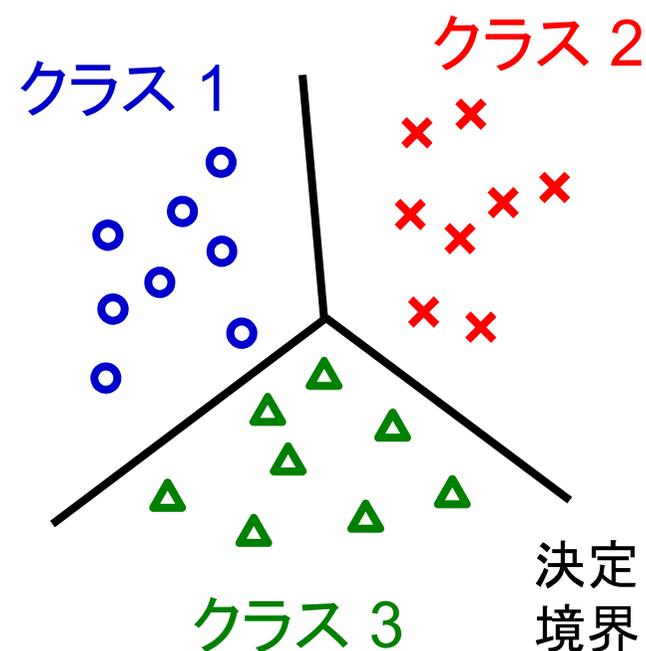
- 正クラスのデータしか取れない:
  - 他社のデータは取れず自社のデータしか取れない
  - 成功例は入手できても失敗例は入手できない
- **信頼度**さえ分かれば,  
正しく分類器を学習できる！



# 新手法5: 補ラベルからの分類 24

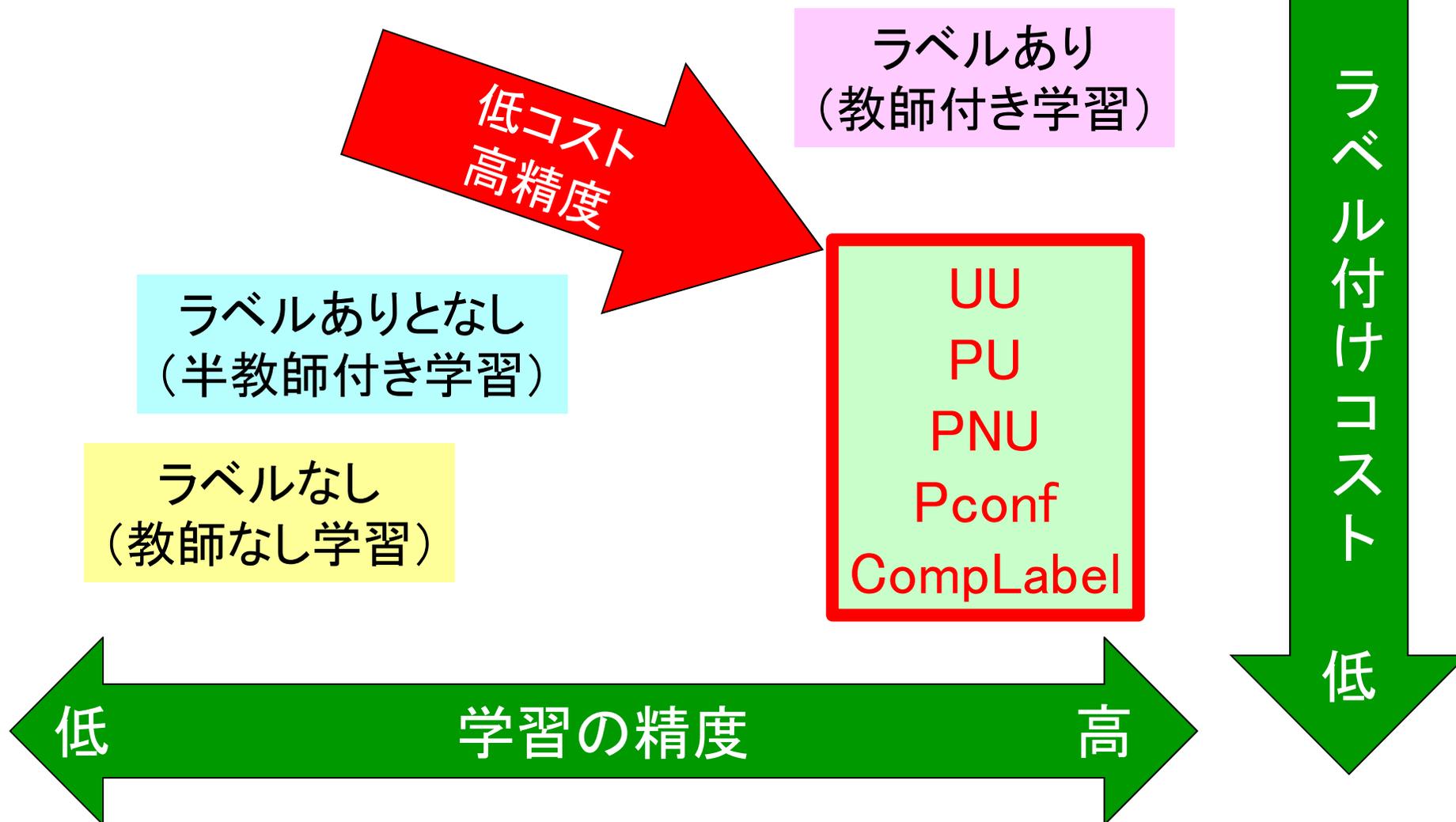
Ishida, Niu & Sugiyama (NIPS2017)

- 多クラスの訓練データのラベル付けは高コスト
  - 多数の候補クラスから正しいものを選ぶ必要がある
- 補ラベル:
  - パターンが属さないクラスのラベル(補ラベル)を与える
  - 補ラベルをつけるのは低コスト
- 「間違っただラベル」だけから、分類器を正しく学習できる！



# まとめ

- ラベル付けコストが低く、精度が高い機械学習手法が有用！



# 今後の展望: AI人材育成

26

## ■ AI研究者・技術者は世界的に不足している:

- 日本ではここ数年でAI技術者の雇用が活発化
- 米国の巨大IT企業は, 2000年ころからAI人材を積極的に雇用し, 人材育成を行っていた
- 米国の大学では, **情報科学(Computer Science)**の学生数が大きく増大
- 日本では, AIに関する高度な知識・技術を持つ人材が限られている

## ■ 人工知能の技術が理解できる人(会社)とできない人(会社)の差が拡大!

[http://www.mext.go.jp/a\\_menu/koutou/itaku/\\_icsFiles/afieldfile/2015/06/29/1357552\\_1.pdf](http://www.mext.go.jp/a_menu/koutou/itaku/_icsFiles/afieldfile/2015/06/29/1357552_1.pdf)

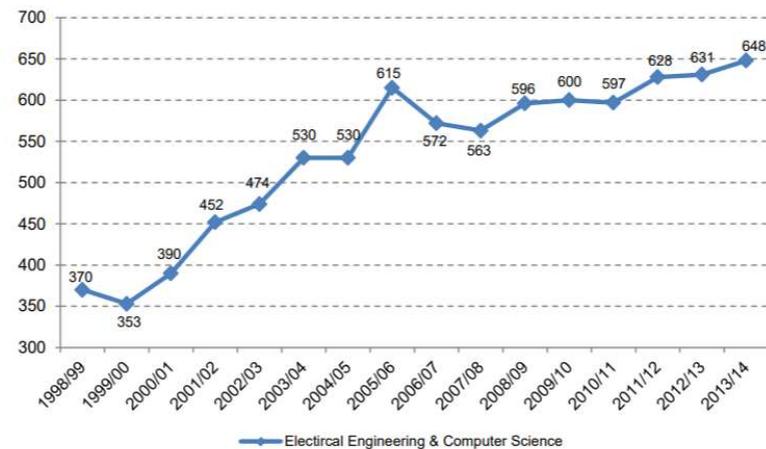
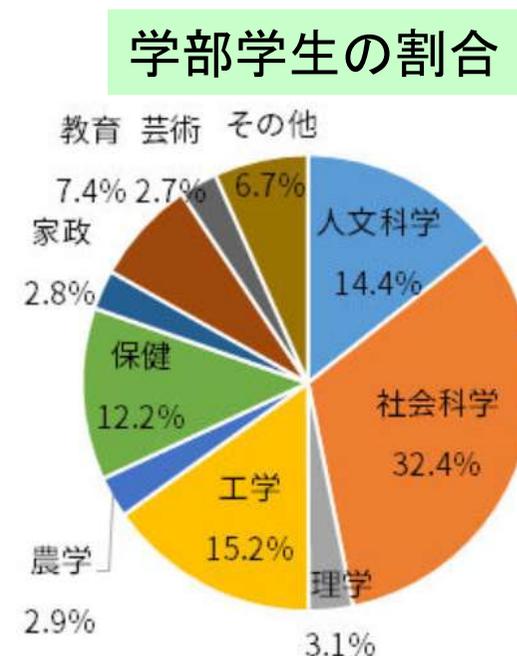


図 7.4 MIT 博士課程学生数の推移【機械工学・計算機科学】(1998/99-2013/14)

# 今後の展望：AI人材育成（続き） 27

- **理工学系（2割）**：  
AI技術の発展のため、  
プログラミングだけでなく数学も
- **その他理系（2割）**：  
AI活用のため、  
データ解析に対する素養を
- **政治・経済・ビジネス・社会系（6割）**：  
AI利用のため、  
最低限の技術リテラシを



<https://matome.naver.jp/odai/2146937280793973301>

受け皿となる情報科学系の  
教員と学生を増やす必要！