

機械学習分野の 国内外の動向と今後

東京大学 杉山 将

sugi@k.u-tokyo.ac.jp

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

国内の機械学習研究の 主要な学会

- 電子情報通信学会 情報論的学習理論と機械学習(IBISML)研究会：
 - 1998年から開催されているIBISワークショップと、人工知能学会データマイニングと統計数理(DMSM)研究会を母体とし、2010年に発足
 - 年に4回開催
- うち1回は、IBISワークショップとして継続：
 - 参加者数は316(2012), 393(2013), 475(2014), 530(2015)と大幅増加
 - 発表件数は118(2012), 93(2013), 126(2014), 134(2015)と微増傾向

機械学習研究の 主要な国際会議

- Neural Information Processing Systems (NIPS) conference
- International Conference on Machine Learning (ICML)
- International Conference on Artificial Intelligence and Statistics (AISTATS)
- これらの会議は、全て1980年代に開始された

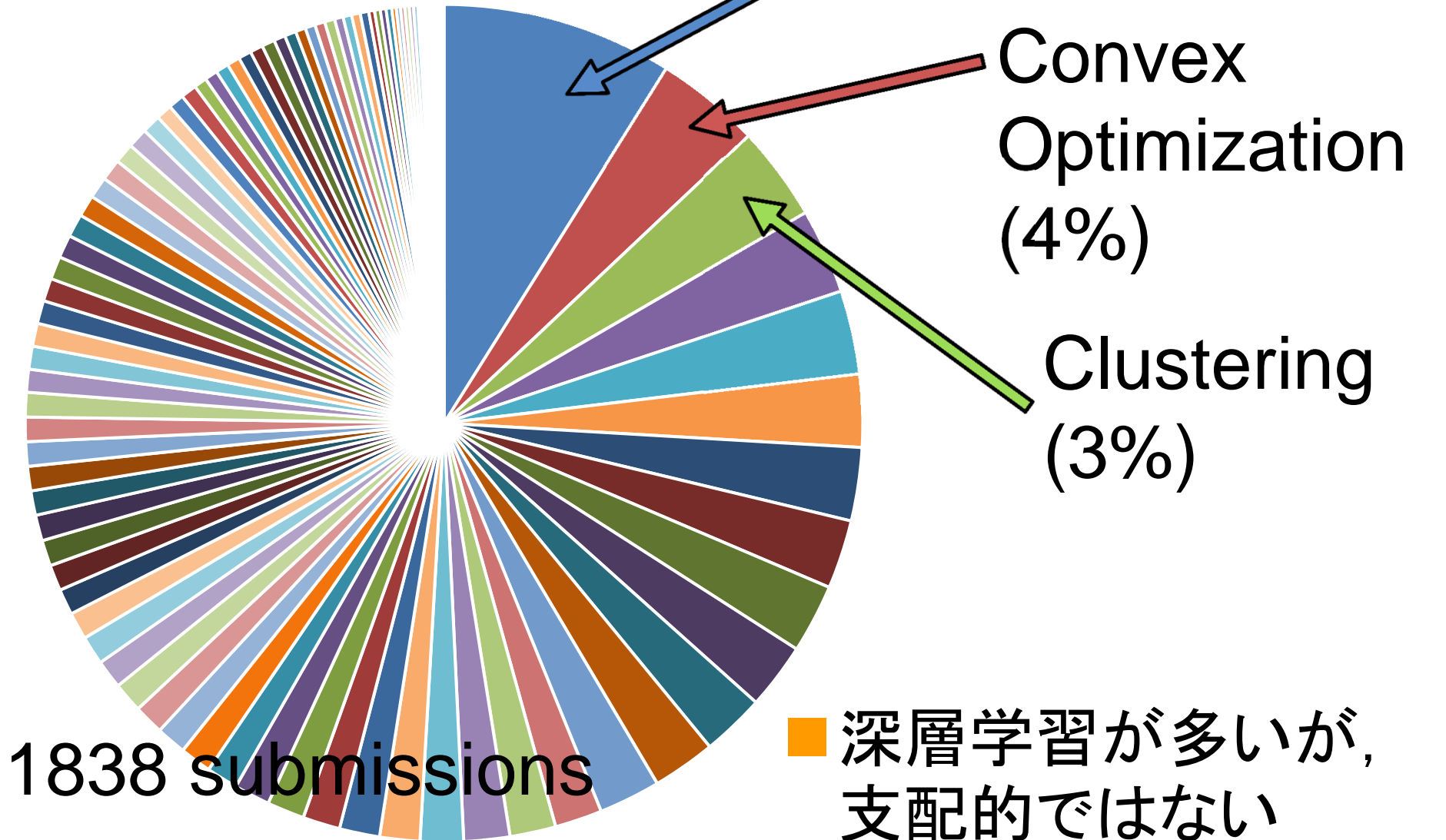
- 毎年12月に開催される機械学習分野の最大の国際会議
- 2016年で30周年を迎える
- 採択率は20~25%程度
- 全員がポスター発表(19時~24時)
- 口頭発表に選ばれるのは高々15~20件. 招待講演とともにシングルトラックで行われる.

NIPS2015

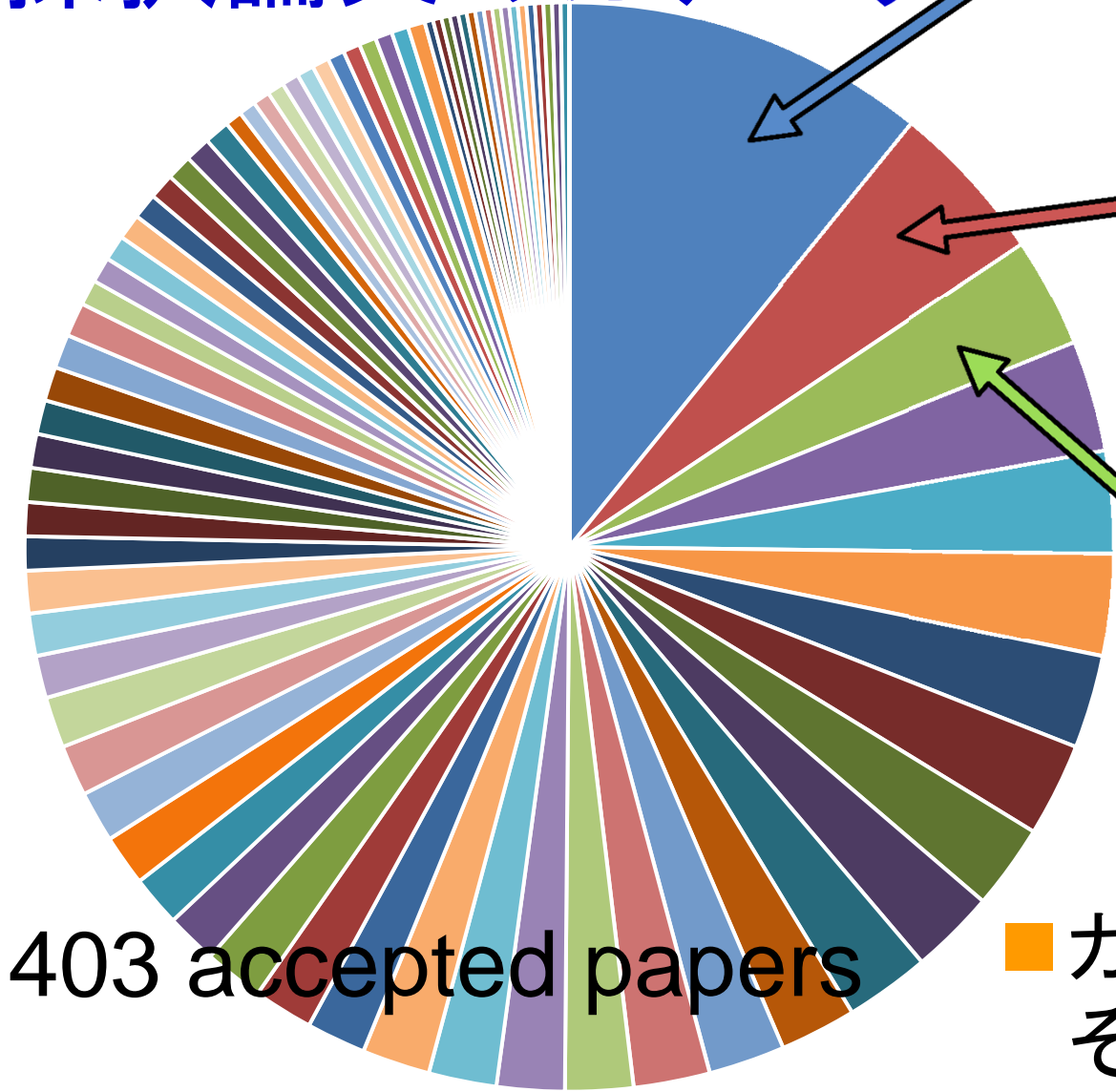
5

- 論文投稿数は1838(2014年の1678件から約10%増)
- 参加者数は3800(2014年の2500名から約50%増)
- 研究者数は微増だが、聴講者数が大幅増(国内と同傾向)

NIPS2015の 投稿論文のカテゴリ



NIPS2015の 採択論文のカテゴリ



Deep Learning (11%) 7

Convex Optimization (5%)

Statistical Learning Theory (3%)

403 accepted papers

■ カテゴリ間の差はそれほど大きくない

研究トピックの動向

- 産業界における爆発的な深層学習ブームと比較して:
 - 深層学習が一番人気のトピックであることには違いがないが、それほど支配的ではない
 - 最適化や学習理論など、長期的に研究すべき基礎のトピックを中心に、機械学習に関する様々なトピックが幅広く研究されている

NIPSの特徴

- 他の機械学習の国際会議と異なり, NIPSは神経科学者も参加する:
 - 今年はATRの川人先生が招待講演者
- 2000年ころまでは神経科学と機械学習の研究者の交流が盛んだったが, その後, NIPSは数学的な会議にシフトし, しばらくは神経科学色が薄まっていた
- 近年の深層学習ブームの到来に伴い, 脳から学ぼうという機械学習研究者が増えつつある.

- 著者のメールアドレスに.jpが含まれる投稿件数
64/1838=3.5%
 - .cnは132, .krは32
- 著者のメールアドレスに.jpが含まれる採択件数
8/403=2%
 - .cnは13, .krは0
- ただし, 中国, 韓国は欧米の大学・企業に所属している学生・研究者が多数いる
- 体感的には, 日本人の存在感はほぼゼロ, 韓国は日本より少しマシ, 中国はかなり溶け込んでいるという印象

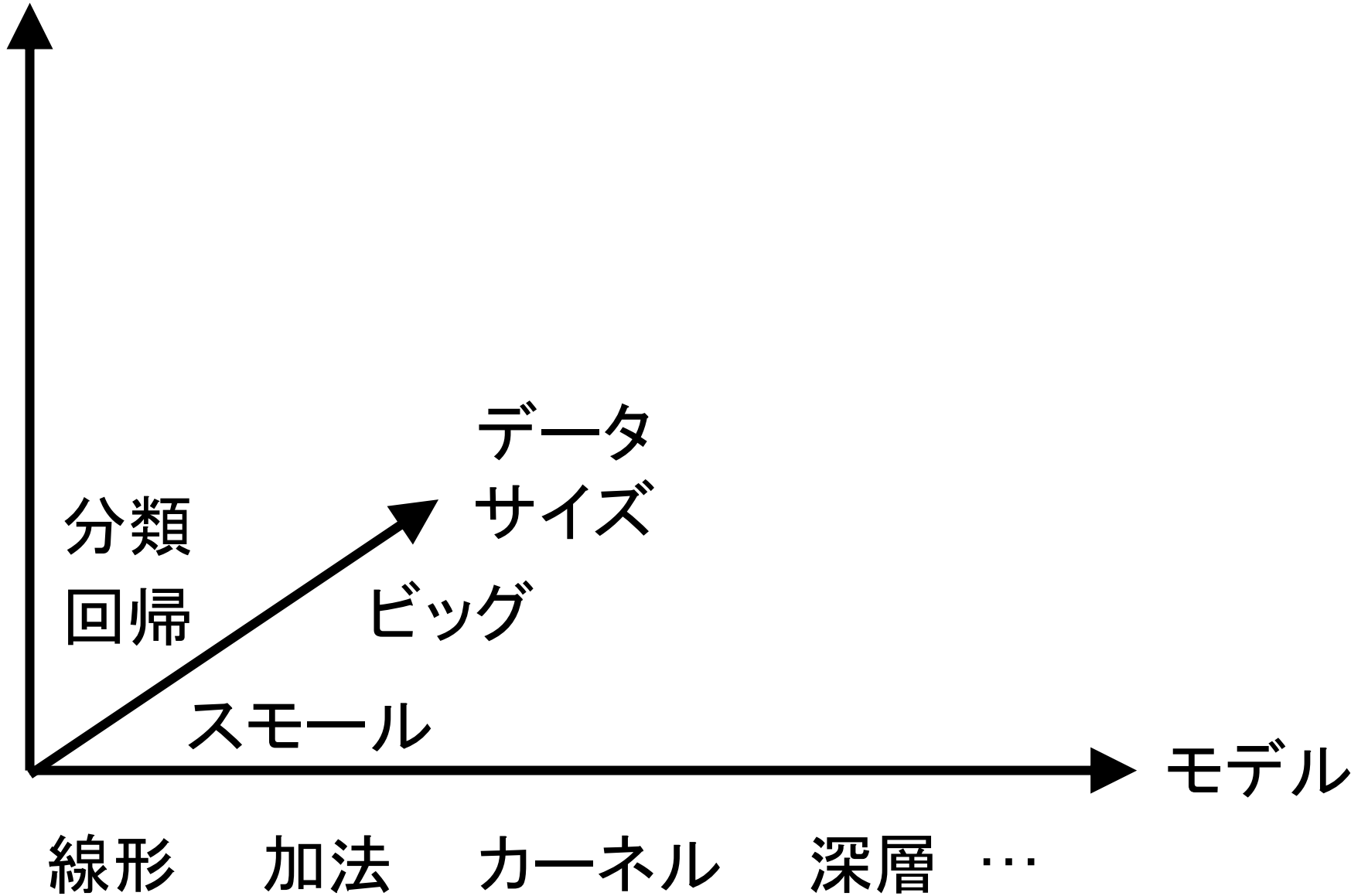
機械学習・人工知能研究の予算¹¹

- **欧米**: 巨大民間企業等が数百～千億円規模の莫大な予算を投じて研究開発を開始
 - Google, Microsoft, Facebook, Amazon, Toyota, OpenAI,...
- **国内**: 政府が中心となって, 数十億円規模の予算を幅広い分野に配分
- 欧米の研究の後追いをしても勝ち目がない
- 欧米では流行っていないが, 基礎のしっかりした理論研究をサポートし, 10年後の浮上を目指す

機械学習の研究軸

12

学習法



我々の取り組み

学習法

⋮

密度微分

異常検知

強化学習

密度比

非定常適応

分類

回帰

データ

サイズ

ビッグ

スモール

モデル

線形

加法

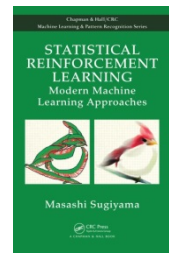
カーネル

深層

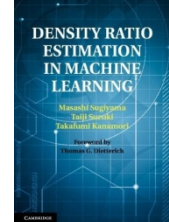
...



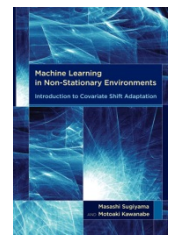
井手, 杉山,
異常検知と変化検知
講談社, 2015



Sugiyama,
Statistical Reinforcement Learning,
Taylor & Francis, 2015



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



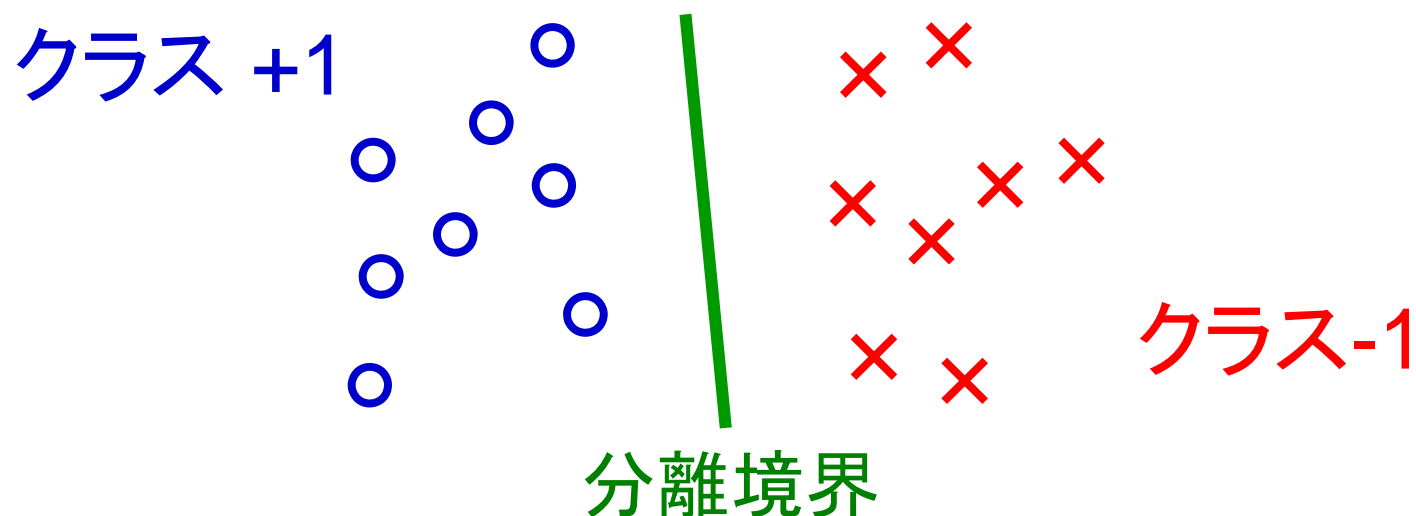
Sugiyama & Kawanabe,
Machine Learning
in Non-Stationary Environments,
MIT Press, 2012

我々の現在の取り組み： 少ない情報からの学習

- 画像や音声認識のような応用では、大量のデータが入手できる
- しかし、他の多くの場面では大量データの入手が難しい

2クラスの教師付き分類

- ラベル付きデータ: $f \gg \{ \mathbf{x}_i, y_i \}_{i=1}^f$
 - 入力 \mathbf{x}_i は q 次元の実ベクトル \mathbb{R}^q
 - 出力 y_i は2値のクラスラベル $y_i \in \{0, 1\}$



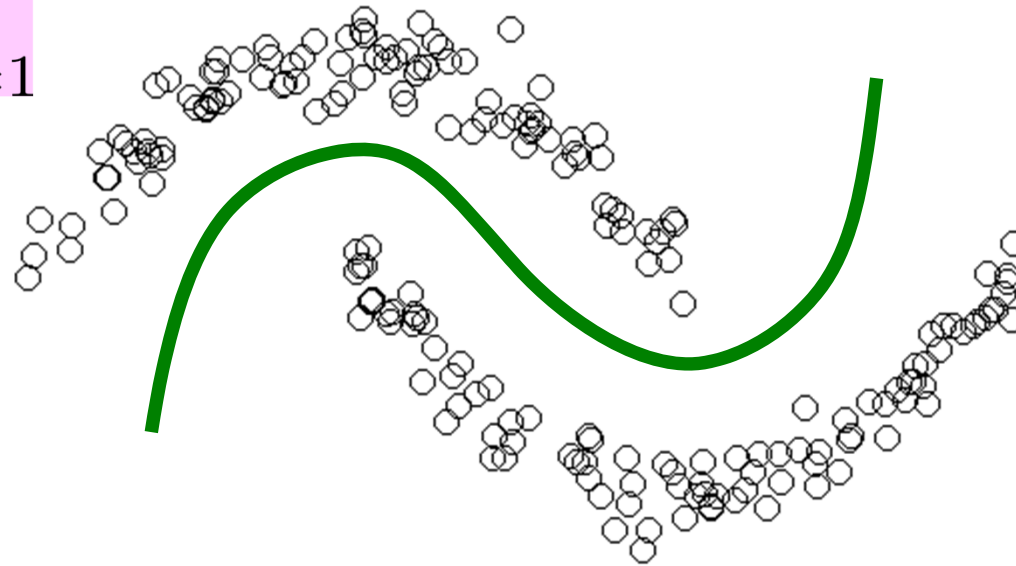
- 大量のラベル付きデータを用いれば、精度良く分類境界が学習できる

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

$$\{x'_i\}_{i=1}^{n'}$$

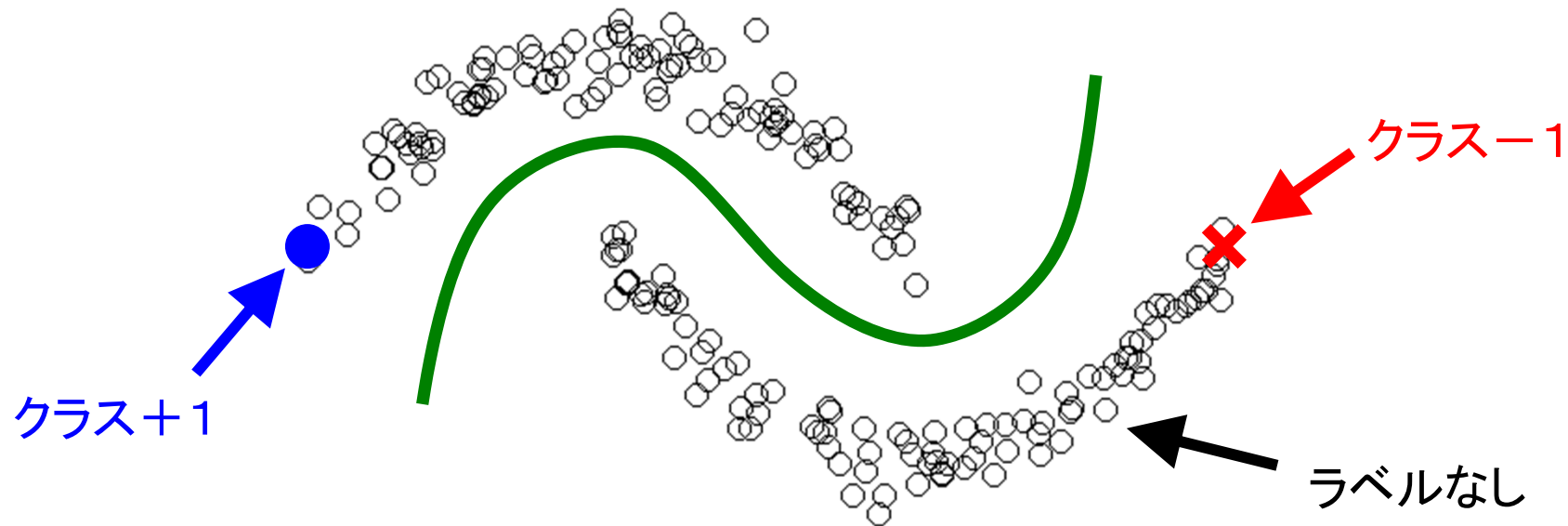


- 教師なし分類はクラスタリングともよぶ
- データがクラス毎にクラスタに分かれていないと、正しく分類できない

半教師付き分類

17

- 少量のラベル付きデータ $\{(x_i, y_i)\}_{i=1}^n$ と大量のラベルなしデータ $\{x'_i\}_{i=1}^{n'}$ を利用する
- ラベルなしデータがなすクラスタ構造に従って分類



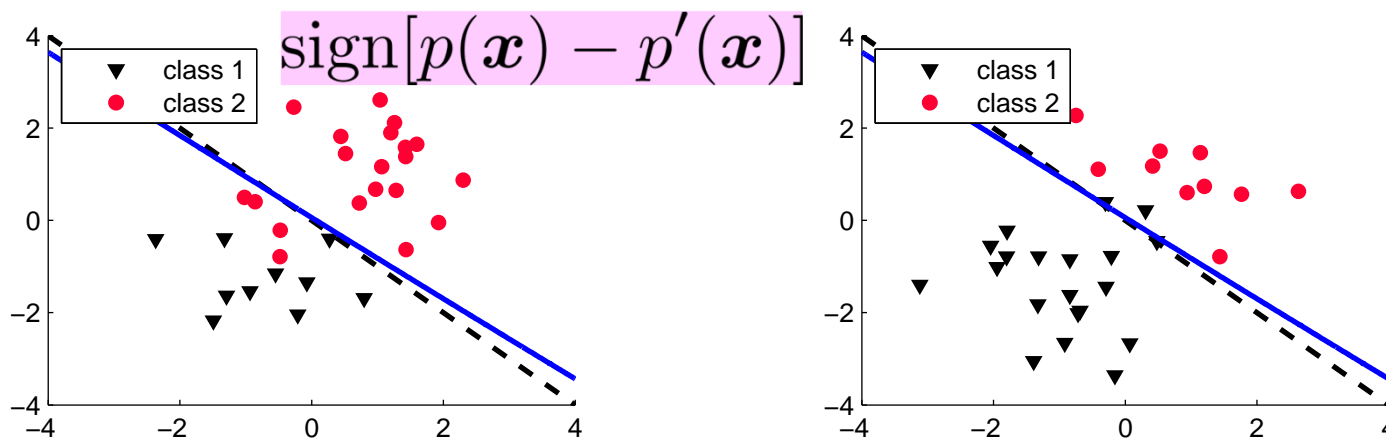
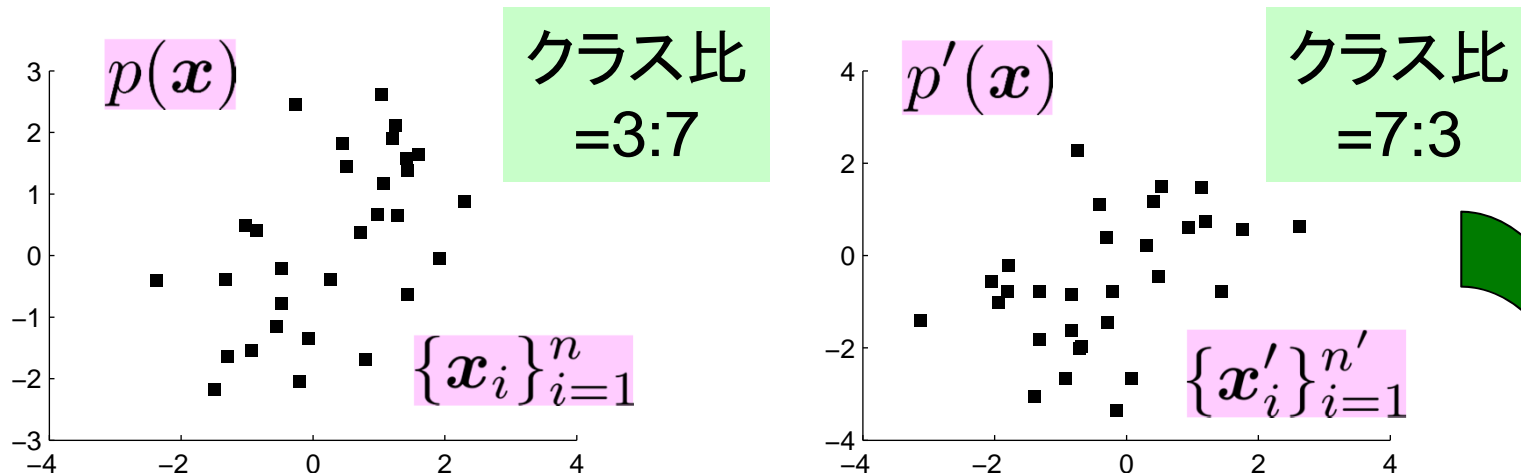
- 同じクラスタに属するデータが同じラベルを持つとき、うまく分類できる
- そのような仮定が常に成り立つとは限らない

提案法1: 教師なし分類(2)

18

du Plessis, Niu & Sugiyama (TAAI2013)

- クラスタ構造が無くても, **クラス比の異なるラベルなしデータが2セット**あれば, 正しく分類可能

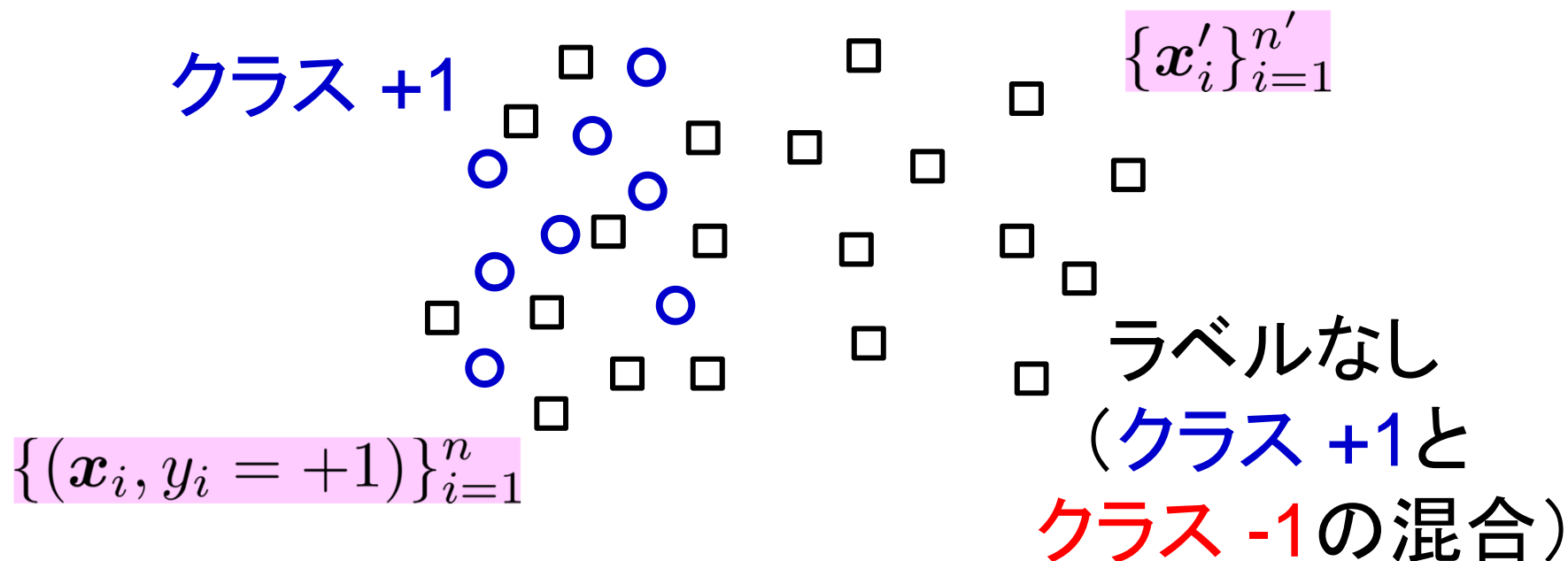


提案法2:

正例とラベルなしデータからの分類

du Plessis, Niu & Sugiyama (NIPS2014, ICML2015)

- 正例とラベルなしデータだけから正しく分類可能

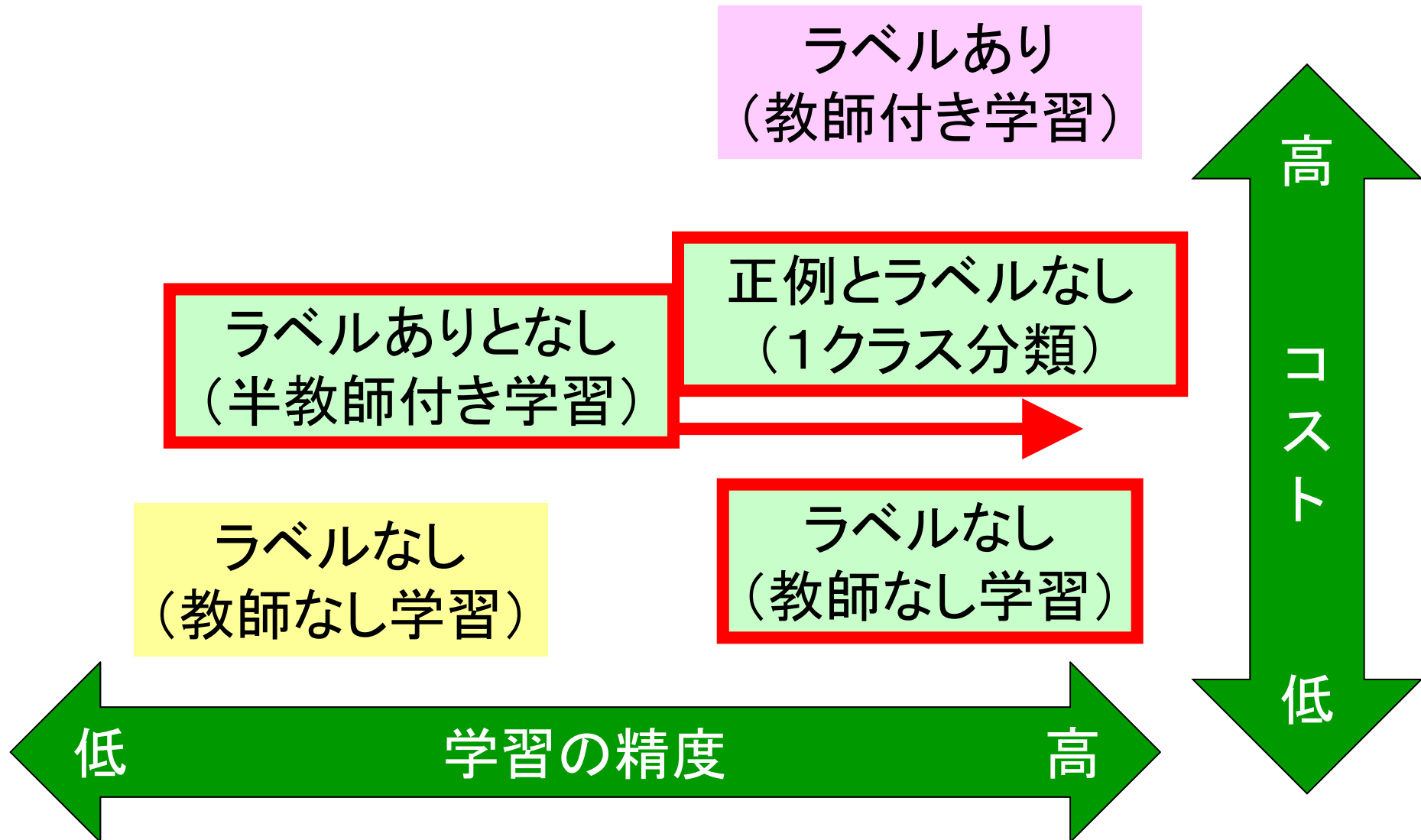


- ロバストな分類器を使用
- 正例とラベル無しに異なる損失を使用

少ない情報からの分類

20

- データ収集のコストに合わせて手法を選ぶ



強化学習

- 教師付き学習: 入力と出力のペアから学習
- 教師なし学習: 入力だけから学習
- 強化学習: 入力とその評価(報酬)から学習

- ビッグデータ
+ 深層学習で,
コンピュータゲーム
の戦略を学習

Mnih et al. (Nature2015)

- しかし実世界では
大量にデータを
集められない

動画

■ 少ない標本からの強化学習

- 高次元データの次元削減
- 過去のデータの再利用
- 分散爆発の抑制



動画

動画

Sugimoto, Tangkaratt, Wensveen,
Zhao, Sugiyama & Morimoto
(IEEE-RAM2016)

Xie, Hachiya & Sugiyama (ICML2012)
Xie, Zhao, Tian, Zhang & Sugiyama (IJCAI2015)

国内の研究費配分スキーム

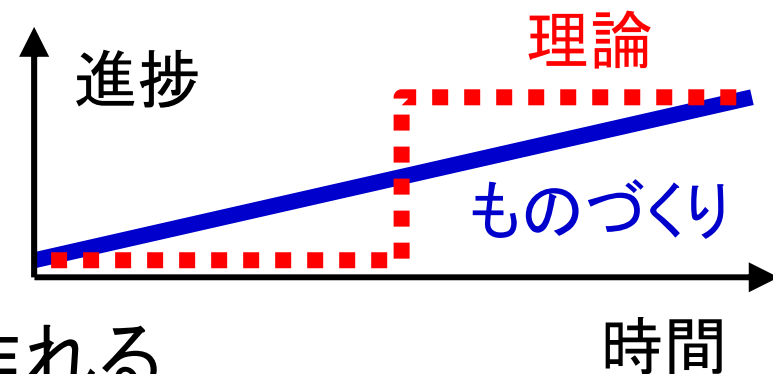
23

■ これまでの日本の研究予算は、詳細な研究計画に基づいて審査されている

- 1年後はこれをして、2年後はこれをして、...
- 1年後の何月にはこの会議で発表して、
2年後の何月にはこの会議で発表して、...

■ このような方式はものづくりには向くが、機械学習の理論研究にはそぐわない:

- アイデアは前触れなく突然やってくる
- アイデアが出れば即座にアルゴリズムが作れる



国内の研究費配分スキーム

24

- 研究計画ベースの研究費配分スキームでは、コンセプトを生み出す理論研究者は育ちにくい
 - 研究がどのように発展するか予測不可能
 - 5~10年先が見えている確実性の高い研究計画に大きな予算を配分するスタイルは合わない
- 若手研究者が中期的に理論研究を継続できる安定的なポストを提供し、理論研究に挑戦しやすい環境づくりが重要
 - 新しいコンセプトを生み出す研究は極めて挑戦的であり、予算を与えた全てのプロジェクトから成果が得られることは期待しない。起業支援と同様、積極的にリスクを取るべき。

- 世の中は機械学習(の応用)ブーム
- すでに欧米で流行っている技術の後追いしても、研究規模で到底かなわない
- 日本には機械学習の理論研究者が少ない
- 10年後を見据え、機械学習理論の若手研究者の育成が重要