

機械学習と 統計的データ解析

杉山・横矢・石田研究室

{sugi, yokoya, ishi}@k.u-tokyo.ac.jp

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



■ 杉山 将(教授):

- 機械学習の理論とアルゴリズム
(弱教師付き学習, 雑音ロバスト学習,
転移学習など)



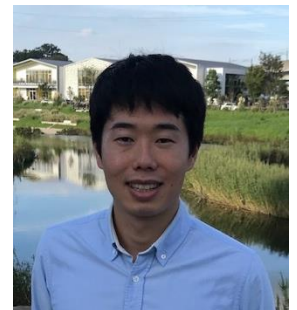
■ 横矢 直人(教授):

- 画像処理, リモートセンシングなど



■ 石田 隆(准教授):

- 弱教師付き学習,
機械学習の限界性能予測など



教員の所属

- 大学院新領域創成科学研究科
複雑理工学専攻(柏)
- 大学院情報理工学系研究科
コンピュータ科学専攻(本郷)
- 理学部情報科学科(本郷)
- Beyond AI研究推進機構(本郷)
- 理化学研究所革新知能
統合研究センター(日本橋)

所属学生数

博士課程:24名

修士課程:21名

研究生・聴講生:2名



コンピュータはどこまで賢くなれるのか？

- 我々の研究室では、**機械学習**の研究を行っています：
 - 基礎理論の構築
 - 実用的なアルゴリズムの開発
 - 実問題への応用

東京大学 杉山・横矢・石田研究室：機械学習と統計的データ解析

機械学習の基礎理論の構築と実用的なアルゴリズムの開発、及び、実問題への応用研究を行っています

[[English](#) | [Japanese](#)]

杉山研究室

横矢研究室

石田研究室

トップ

メンバー

発表論文（英語）

発表論文（日本語）

学生・研究員の

コンピュータはどこまで賢くなれるのか？

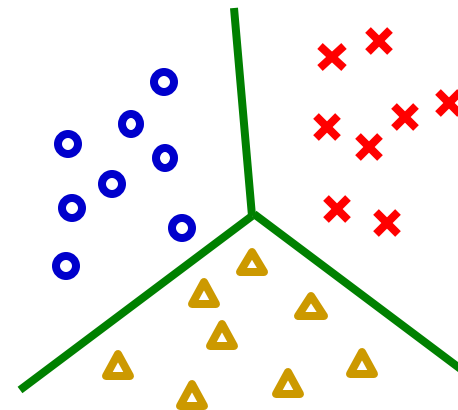
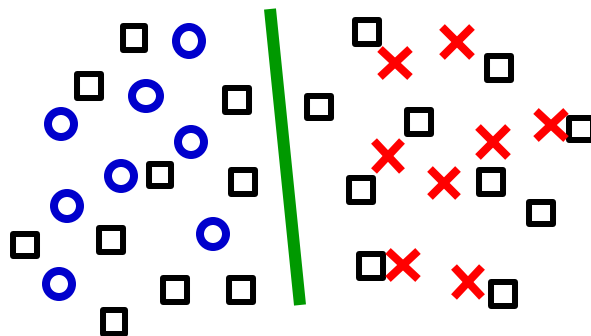
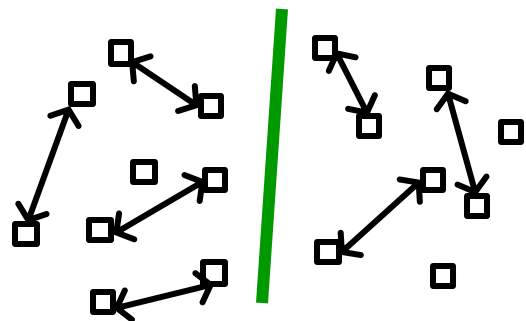
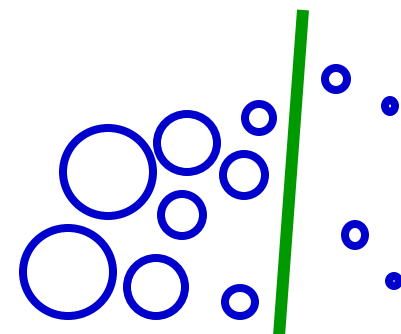
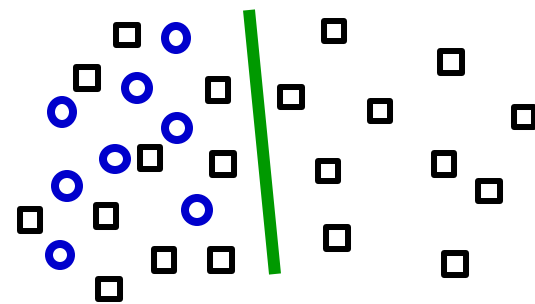
機械学習の理論構築，アルゴリズム開発，実世界応用

情報通信技術の飛躍的な性能向上に伴い、これまで人間にしかできなかった知的な情報処理が、コ

最近の主な研究テーマ

■ 弱教師付き学習:

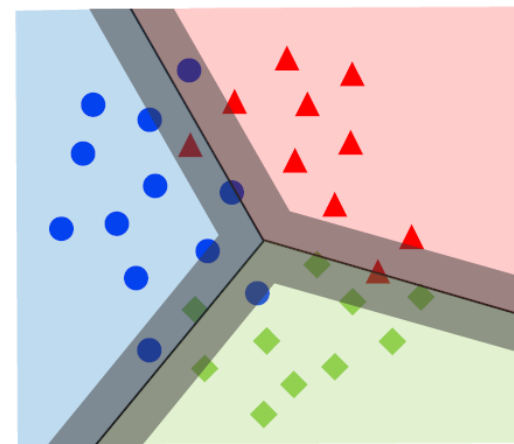
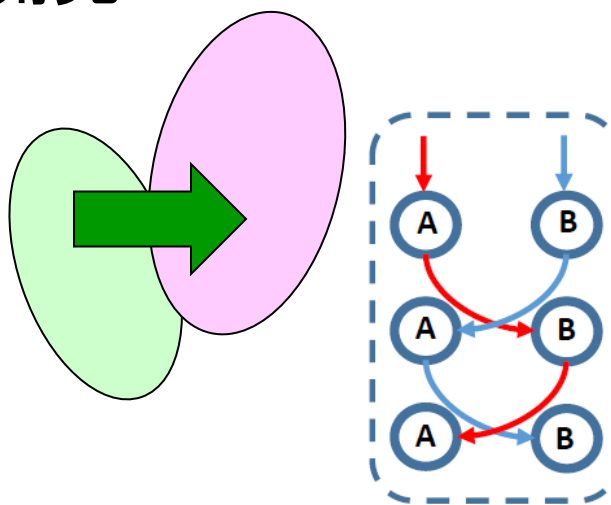
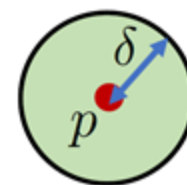
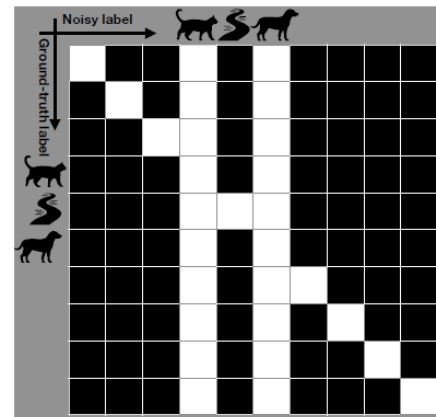
- データのラベル付けは高コスト
- しかし, 小データからの学習は一般には難しい
- 容易に入手できる**弱いデータ**を**活用**する新しい枠組みを開発
- **例**: PU, PNU, Pconf, UU, SDU, 補ラベル学習など



最近の主な研究テーマ

■ ロバストで信頼性の高い学習:

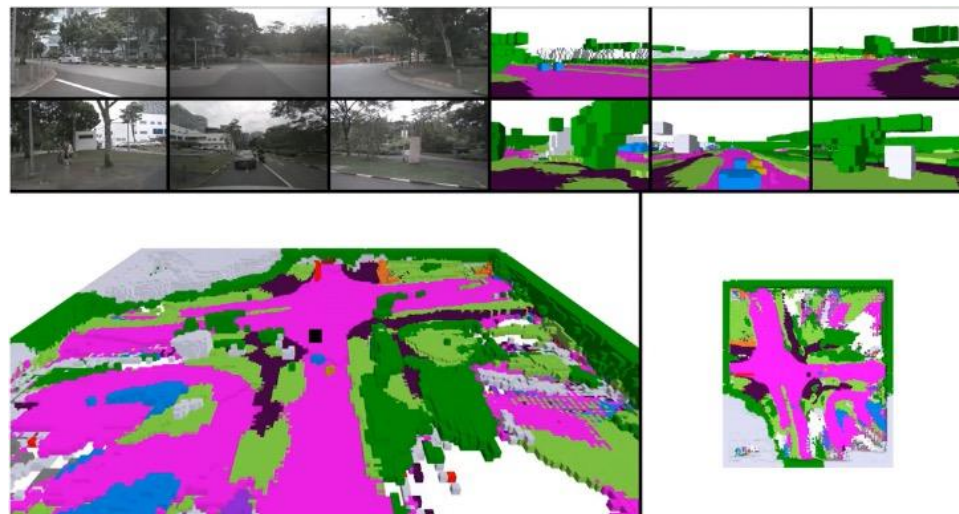
- 実世界のデータには、**雑音**、**異常値**、**非定常性**などが含まれることが多い。
- 標準的な機械学習技術では、そのようなデータからうまく学習できない
- そのようなデータに対しても有効な、新しい学習技術を開発
- **例**: 共教示,
ラベルノイズ学習,
棄却付き分類,
逐次的意思決定,
転移学習など



最近の主な研究テーマ

■ 複雑な実世界シーンの視覚的理解

- 環境の変動性やラベルと計算資源の不足が課題
- 多様なデータの融合と、資源の効率化により、実用性の高いコンピュータビジョン技術を開発
- **例**: マルチモーダル学習, 軽量モデル設計, 少量データ学習など



最近の主な研究テーマ

■ リモートセンシング画像処理

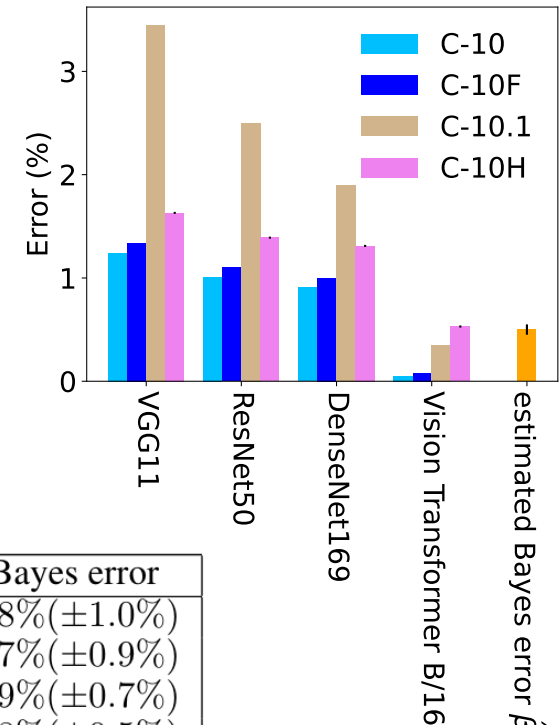
- 解像度の限界と、画素毎の教師データコストが課題
- 超解像や地図作成において、教師データコストの低い新しい技術を開発
- **例**: 画像融合超解像, 合成データからの学習, 異なる種類の教師データからの学習など



最近の主な研究テーマ

■ 機械学習における最良性能の推定

- 最良性能を推定することができれば, **SOTAとの比較, テスト過適合の検知, データセットの難易度の把握**等に役立つ
- 分類問題の場合の最良性能「ベイズ誤差」を推定する技術を開発



ICLR's Bayes error	
2017	6.8%(±1.0%)
2018	8.7%(±0.9%)
2019	7.9%(±0.7%)
2020	8.8%(±0.5%)
2021	9.3%(±0.5%)
2022	9.6%(±0.5%)
2023	8.0%(±0.4%)

Table 1: Parenthesis is 95% confidence interval.

杉山・横矢・石田研究室：

機械学習と統計的データ解析

10

- **目標**：人間のように学習するコンピュータを作る
 - **教師付き学習**：人間が直接コンピュータの学習を手伝う
 - **教師なし学習**：人間はコンピュータの学習を手伝わない
 - **強化学習**：人間が間接的にコンピュータの学習を手伝う
- **研究の観点**：
 - **理論**：確率論，統計学，最適化，情報理論など
 - **アルゴリズム**：有効性，効率性，実用性など
 - **応用**：信号，画像，センサ，ウェブ，言語，ロボット，生命，脳，医療など