

# 汎用基盤技術研究グループの 活動紹介

杉山 将  
汎用基盤技術研究グループ  
グループディレクター

2021年度  
AIPシンポジウム成果報告会

# 汎用グループ

- **目標**: 機械学習の理論構築・アルゴリズム開発において、国際的なイニシアチブを取る

実応用

学習・推論・探索・検定の  
アルゴリズム開発

理論保証

学習理論・最適化の  
理論構築

数学的知見

基礎数理

# 汎用グループ

- **目標**: 機械学習の理論構築・アルゴリズム開発において, 国際的なイニシアチブを取る

実応用

高信頼学習

因果推論

仮説検定

類似検索

理論保証

深層学習

学習理論・最適化  
理論構築

非凸最適化

数学的知見

最適輸送

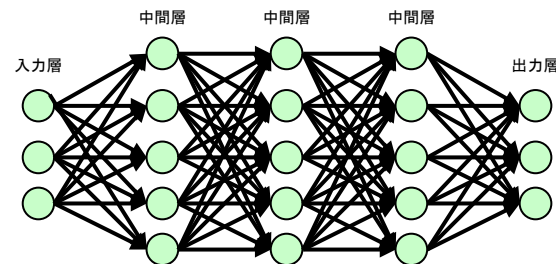
構造データ

# 深層学習理論

## 成果1: 深層学習が浅層学習よりも優れた予測性能を持つことを証明

- Neural Tangent Kernelに基づく解析 (無限幅のモデルを凸的に解析できる)

Nitanda & Suzuki, ICLR2021 Outstanding Paper Award



$$\mathbb{E}[\|f_T - f^*\|_{L_2}^2] \leq \epsilon_M + O\left(T^{-\frac{2r\beta}{2r\beta+1}}\right)$$

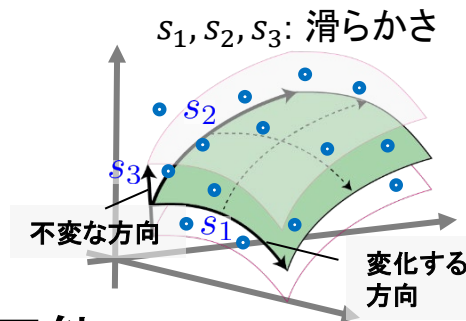
$\epsilon_M$ : 横幅  $M \rightarrow \infty$  で0に収束する項

速い学習レート ( $O(1/\sqrt{T})$  より速い)  
→ Minimax最適レート

## 成果2: 深層学習は次元の呪いを受けない

- 一般に滑らかな関数は簡単に学習できる
- 浅層学習の誤差は、関数の最も滑らかでない方向が支配 (次元の呪い)
- 深層学習は、方向ごとにモデルを自動調整 (次元の呪いを受けない)

Suzuki & Nitanda, NeurIPS2021 Spotlight.



深層  $n^{-\frac{\tilde{s}}{\tilde{s}+1}}$   
 $\tilde{s} = (s_1^{-1} + s_2^{-1} + s_3^{-1})^{-1}$

浅層  $n^{-\frac{s_1}{s_1+d}}$   
 (次元の呪いを受ける)

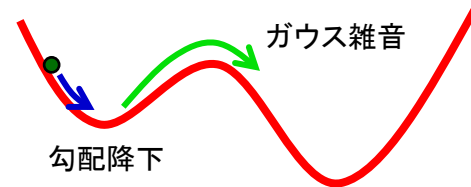
## 成果3: 深層学習は大域的最適化が可能

- 確率的勾配法に適切なガウス雑音を加えれば、大域的最適解に収束

Suzuki & Akiyama, ICLR2021 Spotlight.

$$\hat{L}(X_k) - \int \hat{L}(x) d\pi_\infty(x) \lesssim \exp(-\Lambda_\eta^* k \eta) + \frac{c_\beta}{\Lambda_0^*} \eta^{1/2-\kappa}$$

$$dX_t = -\nabla \left( \hat{L}(X_t) + \frac{\lambda}{2} \|X_t\|_{\mathcal{H}_K}^2 \right) dt + \sqrt{\frac{2}{\beta}} d\xi_t$$



# 多レベル最適化理論

## ■ 多レベル最適化: 上位の最適化の解が下位の最適化に含まれる

- 機械学習におけるハイパーパラメータ最適化
- 敵対的学習における  
設計者, 攻撃者, 操業者のせめぎ合い

$$\begin{aligned} \min_{\lambda} & \|y_{val} - X_{val}w_{\lambda}\|^2 \\ \text{s.t. } & w_{\lambda} = \arg \min_w \|y_{tr} - X_{tr}w\|^2 + \lambda\|w\|_1 \end{aligned}$$

## ■ 成果1: 微分不可能な 2レベル最適化の理論保証

- スパース学習など微分不可能な問題が対象
- 従来は劣微分を使っていたが, 理論保証ができない
- 最適性必要条件 (scaled bilevel-KKT条件) を導入し,  
収束保証付き解法を初めて構築

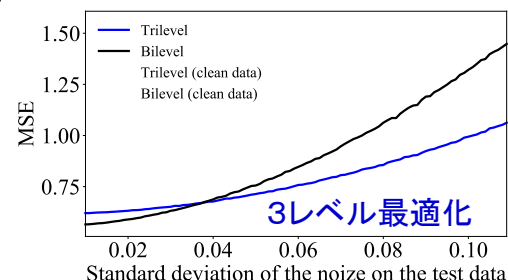
Okuno, Takeda, Kawana & Watanabe, [JMLR2021](#)

## ■ 成果2: 一般の多レベル最適化の理論保証

- 多レベル最適化は90年代から研究されているが,  
理論保証は未開拓
- 漸近的理論保証付き解法を初めて構築

Sato, Tanaka & Takeda, [NeurIPS2021](#)

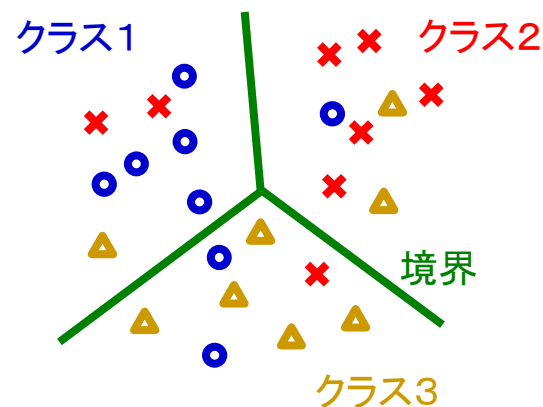
$$\begin{aligned} \min_{x,y,z} & f(x,y,z) \\ \text{s.t. } & y = \arg \min_{y',z'} g(x,y',z') \\ & z' = \arg \min_{z''} h(x,y',z'') \end{aligned}$$



# ラベル雑音に対するロバスト学習理論 6

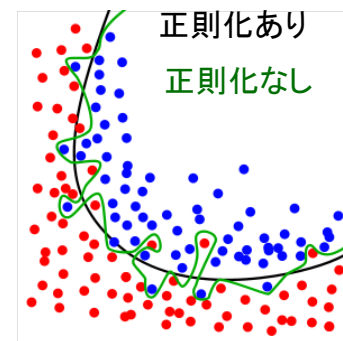
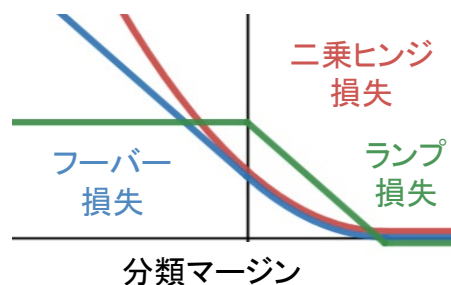
## 背景:

- ラベル雑音が存在するとき, 大量の訓練データを用いても学習はうまくいかない
- 雑音を除去する明示的な機構が必要



## 従来のアプローチ:

- 教師なし異常値除去:  
分類よりも難しく信頼性が低い
- ロバスト損失, 正則化による平滑化:  
回帰(加法雑音)には有効だが,  
分類(ラベル変化)には不十分



## 一連の成果: ラベル雑音学習の理論体系を構築

- ラベルの変化確率(雑音遷移行列)の明示的なモデル化
- 雑音の無いデータがあれば, 雑音遷移行列は推定可能
- 雑音のあるデータだけからは, 数学的には同定不可能
- 雑音遷移行列と分類器の同時学習
- 入力依存ラベル雑音への拡張

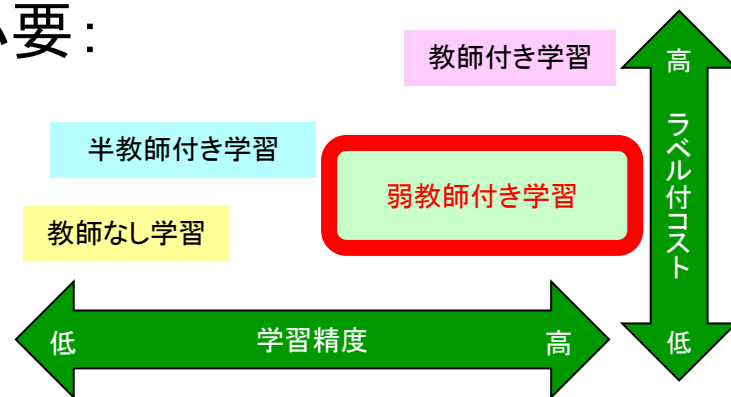
1	0.1	0.5
0	0.8	0.5
0	0.1	0

# 弱教師付き学習理論

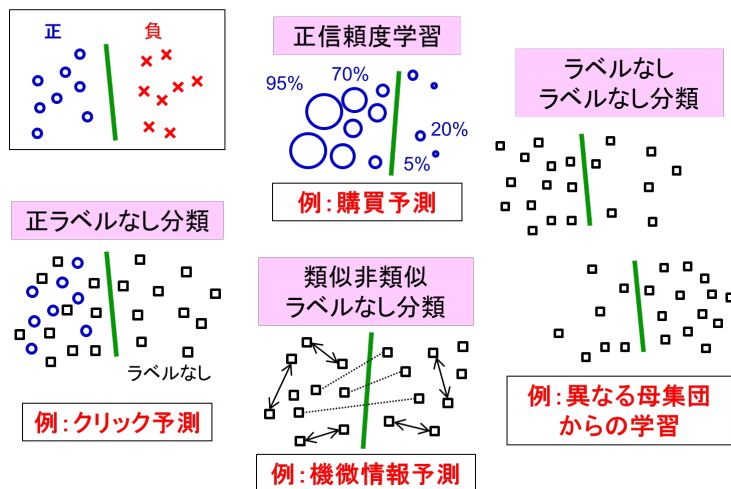
## ■ 今の機械学習は大量の教師データが必要:

- 容易に収集できる「弱い」教師情報を活用

## ■ 様々な弱教師付き学習問題に適用できる汎用的な学習理論とアルゴリズムを構築

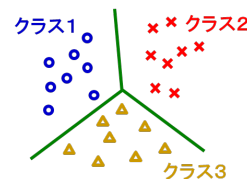


### 弱教師付き分類(2クラス)



### 弱教師付き分類(多クラス)

- **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」
- **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」
- **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%」



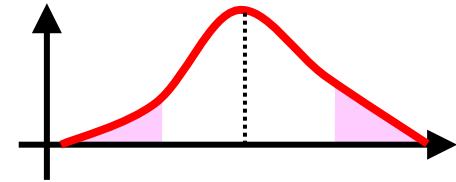
## ■ 英語専門書を出版:

Sugiyama, M., Bao, H., Ishida, T., Lu, N., Sakai, T., & Niu, G.  
**Machine Learning from Weak Supervision: An Empirical Risk Minimization Approach**,  
 320 pages, MIT Press, 2022.

# 更なる研究成果

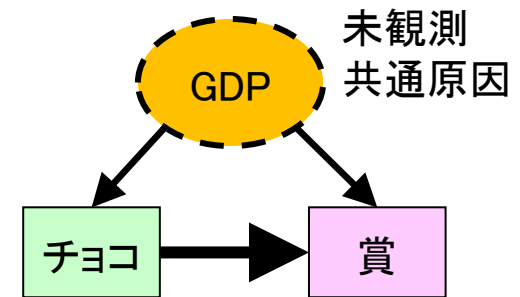
## ■ 仮説検定:

- 従来法はデータ駆動科学では偽陽性が多い
- 観測データで条件づけした選択的推論



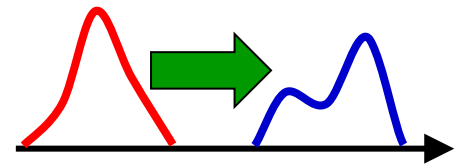
## ■ 因果推論:

- 従来の統計的機械学習は相関に基づく予測
- 因果構造は確率分布に含まれない



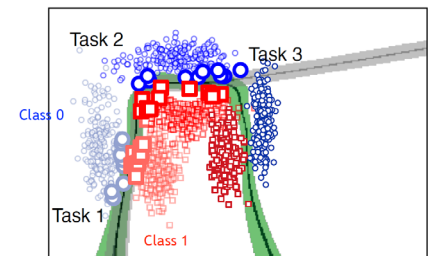
## ■ 無限次元関数空間の理論:

- 深層モデル(汎化解析), 最適輸送(ロバスト距離),  
クープマン作用素(構造データ解析),  
半無限計画(巨大最適化)...



## ■ 継続学習:

- 長期的に様々なタスクを学習するシステム
- 強化学習, 転移学習, メタ学習などの発展型





# 今後の展望

## ■ 機械学習研究技の更なる促進：

- 深層学習，最適化などの理論
- ロバスト，弱教師付き，継続，因果などの学習手法

## ■ 研究成果の普及：

- 国際チュートリアル講演を実施
- 英語専門書を出版し新学問分野を創出

## ■ リーダーシップ：

- 国際ワークショップを主催
- 国際的研究拠点としての人材育成