創発採択課題「深層学習の原理記述に向けた構造汎化理論スキームの開発」

AIの原理を解き明かす新理論

2021/12/15 JST

今泉允聡
東京大学 総合文化研究科 先進科学研究機構

AIと深層学習

新技術の”発見”がもたらしたもの

深層学習の"発見"

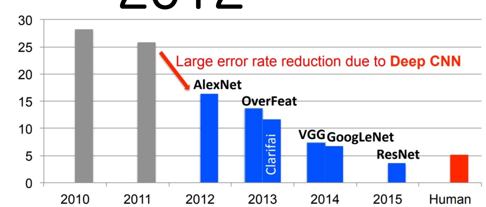
2010年代に実用化されたデータ解析技術

基礎研究 ブレイクスルー AIブーム

~2000 2012 2016~

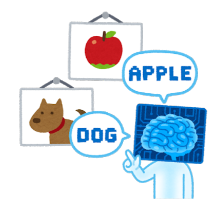
技術的課題

技術発展



高精度を発揮

画像識別精度: 75% → 96%

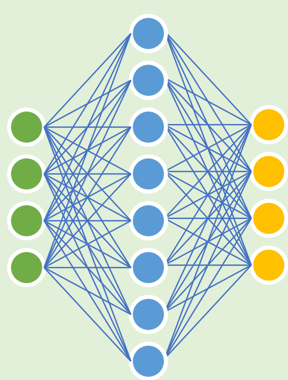


画像認識

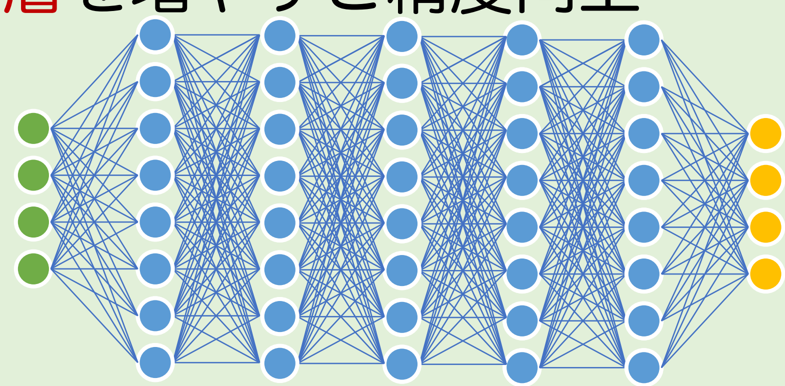
自動運転

発見：ニューラルネットワークの層を増やすと精度向上

（既存法）
層が少ない
ネットワーク



（新手法）
深層学習

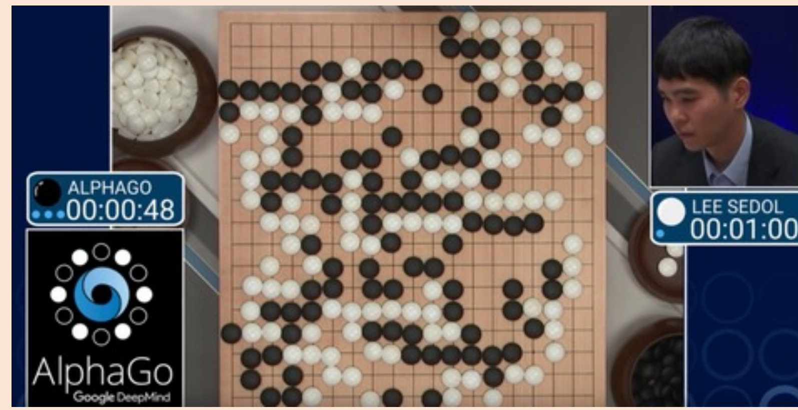


深層学習はAIの中核技術

有名な成功例

AlphaGo (DeepMind)

- 囲碁で人間超え
  - 世界トップ棋士に勝利



BERT (Google)

- 言語テストで高得点
  - 人間: 91.2%
  - BERT: 93.2%

文章

In early 2012, NFL Commissioner Roger Goodell stated that the league planned to make the 50th Super Bowl "spectacular" and that it would be "an important game for us as a league".

文章に関する問い

Who was the NFL Commissioner in early 2012?

Ground Truth Answers: Roger Goodell Roger Goodell Goodell

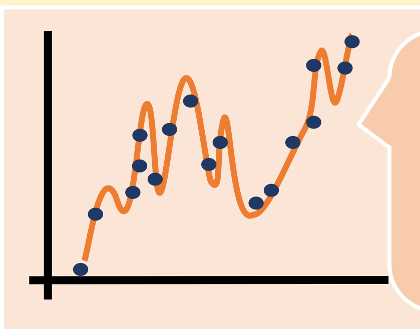
深層学習の謎

その原理はまだ理解されていない

性能をめぐる理論の謎

従来理論と深層学習は完全に食い違う

従来のデータ解析理論



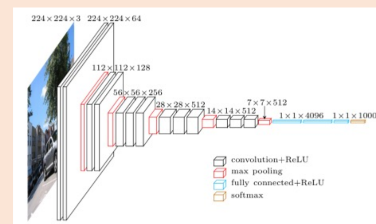
過剰な
パラメタは
過学習する
→性能悪化

誤差は $\dfrac{\text{パラメタ数}}{\text{データ数}}$ に比例

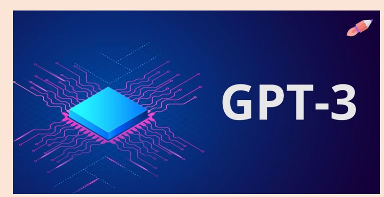
深層学習登場前の**常識**

矛盾

巨大深層学習の成功



VGG19 Net
1億パラメタ



GPT-3
千億パラメタ

パラメタを増やすほど
予測精度が向上

→ 理論再考：データ解析理論を再構築する必要性

理論がないと何が困る？

深層学習の根本的問題は、原理の不理解に依存

問題：膨大な計算コスト

適切な設計が不明
大量に試験しよう

膨大な計算コスト
（大手による技術独占）

問題: ブラックボックスな挙動

失敗したが
原因は不明！

信頼できる
製品が作れない
（社会への技術浸透の遅れ）

深層学習の原理を説明できる理論がない

➡ 高コスト・非効率な運用が主流に

深層学習と基礎理論の関係

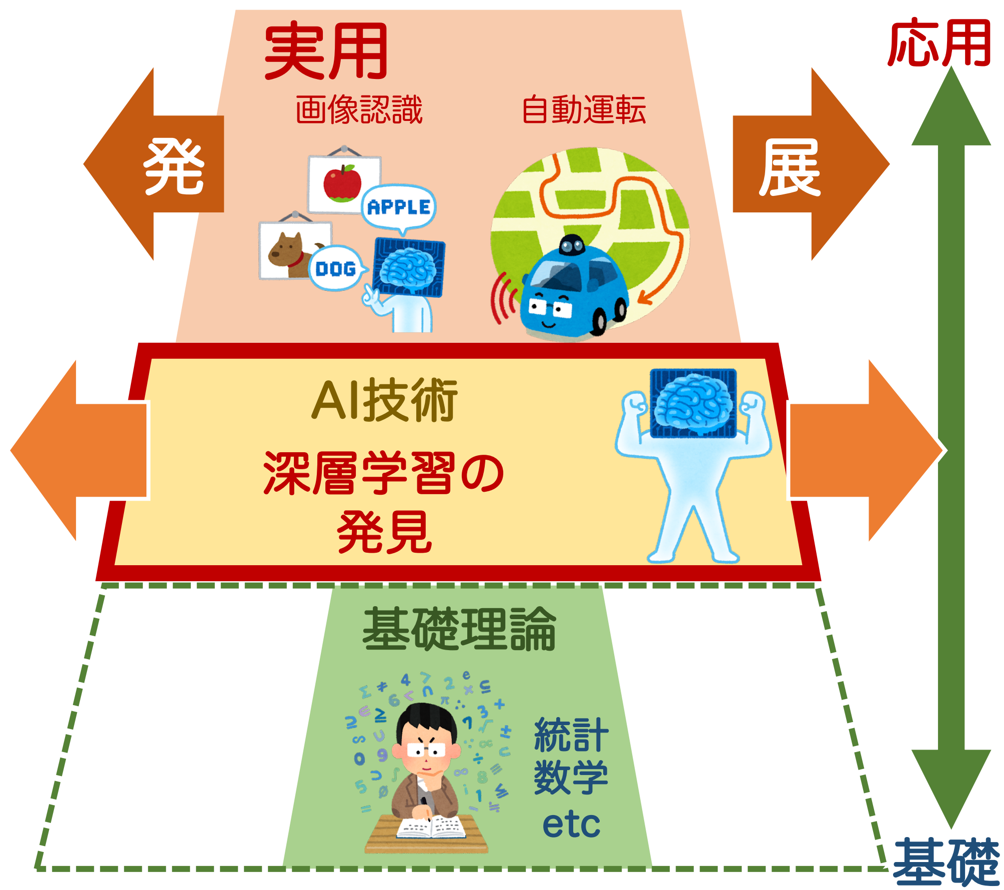
“発見”された手法の理解は発展途上

実用の世界

- 深層学習の恩恵で発展

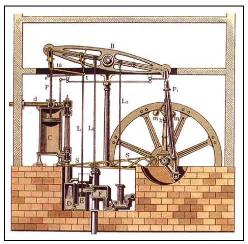
理論の世界

- 理解・支えるべき
- 実用に追いついていない

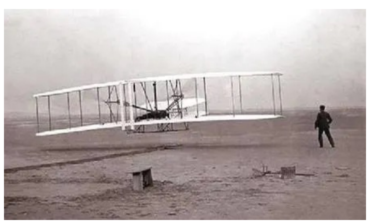


“発見”を理論で記述すること

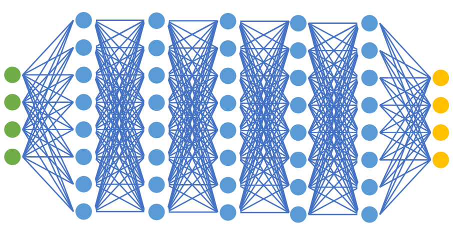
- 歴史的には共通の現象



蒸気機関の発明
(1769年) → 熱力学の成立



飛行機の発明
(1903年) → 航空力学の成立



深層学習の発明
(2012年) → ?

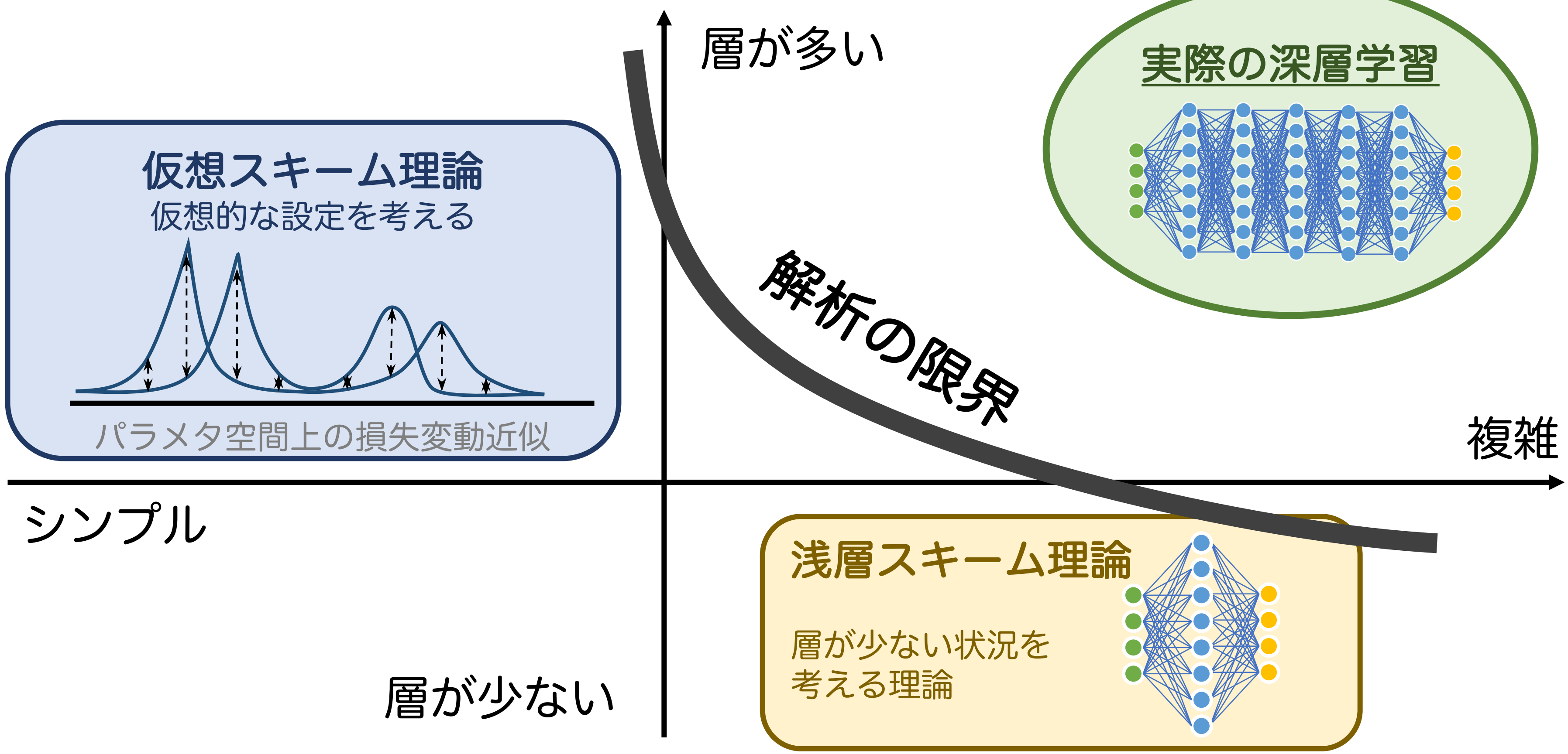
問：深層学習は記述・理解できる理論は構築できるか？

本研究プロジェクト

深層学習を記述する”構造汎化スキーム”

近年の理論研究の背景

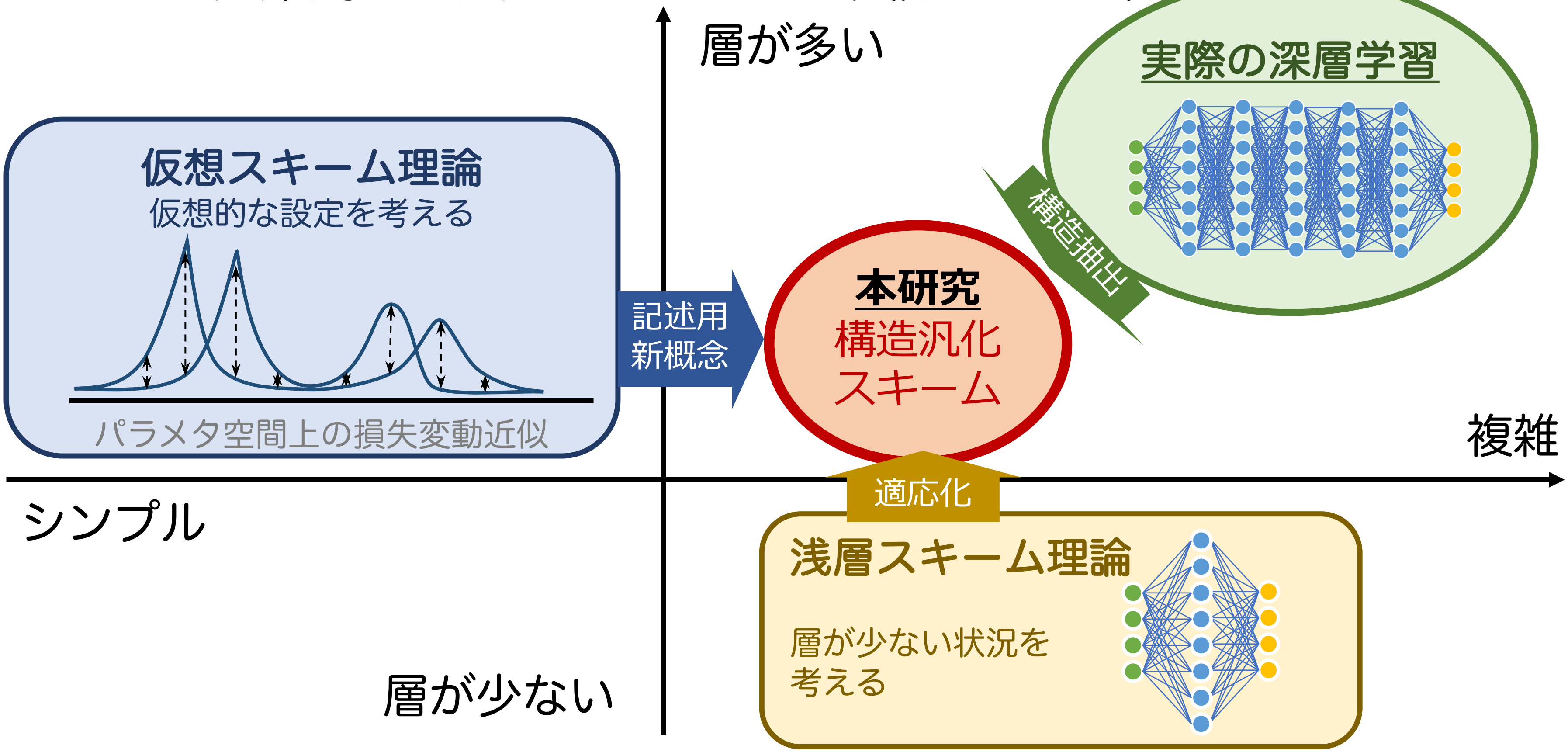
仮想的・限定的な設定下の理論が多い



本研究の枠組み

構造汎化スキーム

- 実現象と既存スキームを架橋する理論



深層学習を記述する理論スキームを創出

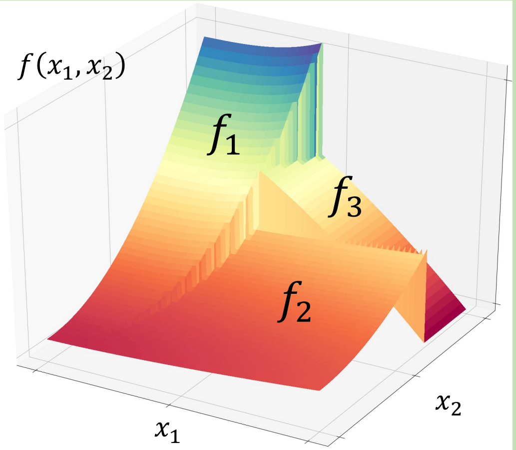
これまでの研究

今泉研による理論成果例（3点）

1/3: データの非滑らかさ

データ構造が非滑らかな時、深層学習が優越することを証明

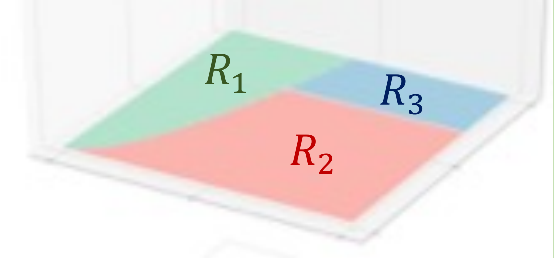
対象：区分上でのみ滑らかな関数



例：物理的相転移現象

$$\mathcal{F}^{PS} := \{f = \Sigma_m f_m 1_{R_m}\}$$

f_m: ヘルダー級関数 , 1_{R_m}: 区分R_mの指示関数



関数の台を区分に分割

主定理：深層学習の優越 (ミニマックス汎化誤差)

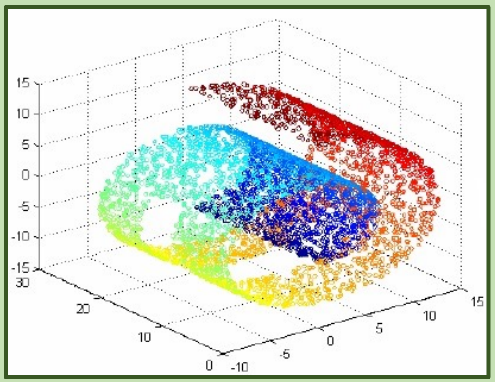
$$\inf_{\hat{f}^{DL}} \sup_{f^* \in \mathcal{F}^{PS}} E\left[\left\|f^* - \hat{f}^{DL}\right\|_{L^2(\mu)}^2\right] < \inf_{\hat{f}^{\mathrm{lin}}} \sup_{f^* \in \mathcal{F}^{PS}} E\left[\left\|f^* - \hat{f}^{\mathrm{lin}}\right\|_{L^2(\mu)}^2\right]$$

深層学習の予測誤差　　　　従来法による予測誤差

2/3: データの隠れた低次元性

データの隠れ低次元性が深層学習の性能を向上

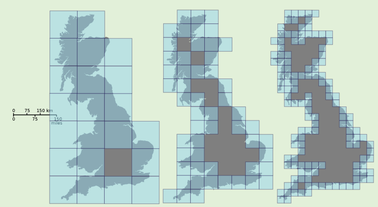
対象：隠れた低次元性を持つデータ



例：低次元多様体に集中するデータ

次元の定義：Minkowski 次元

$$\inf\left\{ d > 0 \,\middle|\, \limsup_{\varepsilon \to 0} \mathcal{N}_\varepsilon(\Omega)\varepsilon^d = 0 \right\}$$



$\mathcal{N}_\varepsilon(\Omega)$：$\Omega$を$\|\cdot\|_\infty$で被覆する$\varepsilon$球の数.

主定理：隠れ次元に依存する誤差収束

n: データ数
β: データの滑らかさ

D次元データが隠れ次元d ($d < D$)を持つとき：

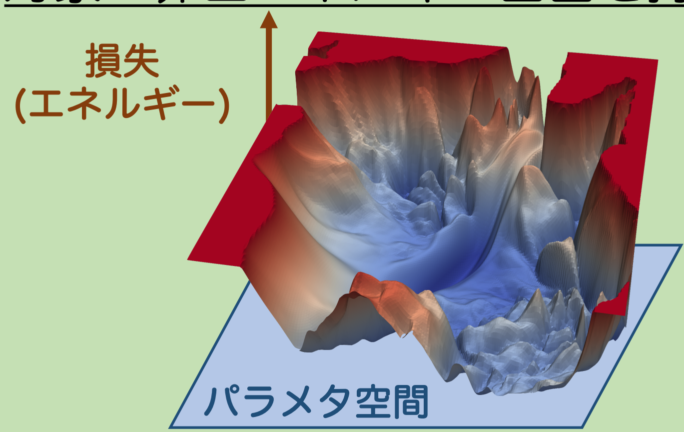
$$\left\|\hat{f}^{DL} - f^*\right\|_{L^2(\mu)}^2 \le C_1 n^{\frac{-2\beta}{2\beta+d}}(1 + \log n)^4 + \frac{16}{n}$$

深層学習による予測誤差　　隠れ次元dのみに依存する誤差レート

3/3: 複雑なエネルギーの影響

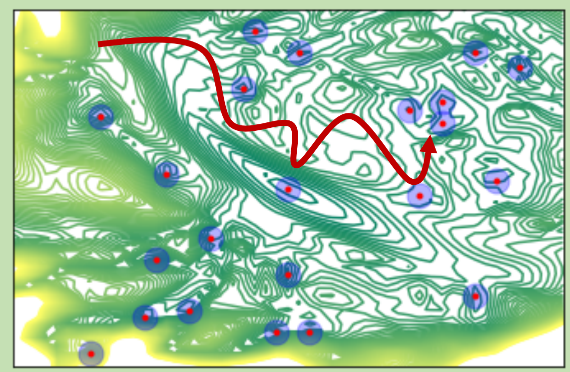
深層学習のエネルギー曲面の複雑性が予測精度を説明

対象：非凸エネルギー曲面を持つ深層学習



図はLi et al. (2018)より

複雑な曲面とパラメタ探索アルゴリズム



主定理：曲面形状による予測誤差

正の確率で以下が成立：

$$R(\hat{f}) \leq \inf_{f \in B_\delta} R(f) + O\left(\frac{\sqrt{d \log K} + SL\delta}{\sqrt{n}}\right)$$

深層学習の予測誤差

曲面形状で決まる誤差（パラメタ数と独立）

Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2018). Visualizing the loss landscape of neural nets. NeurIPS.

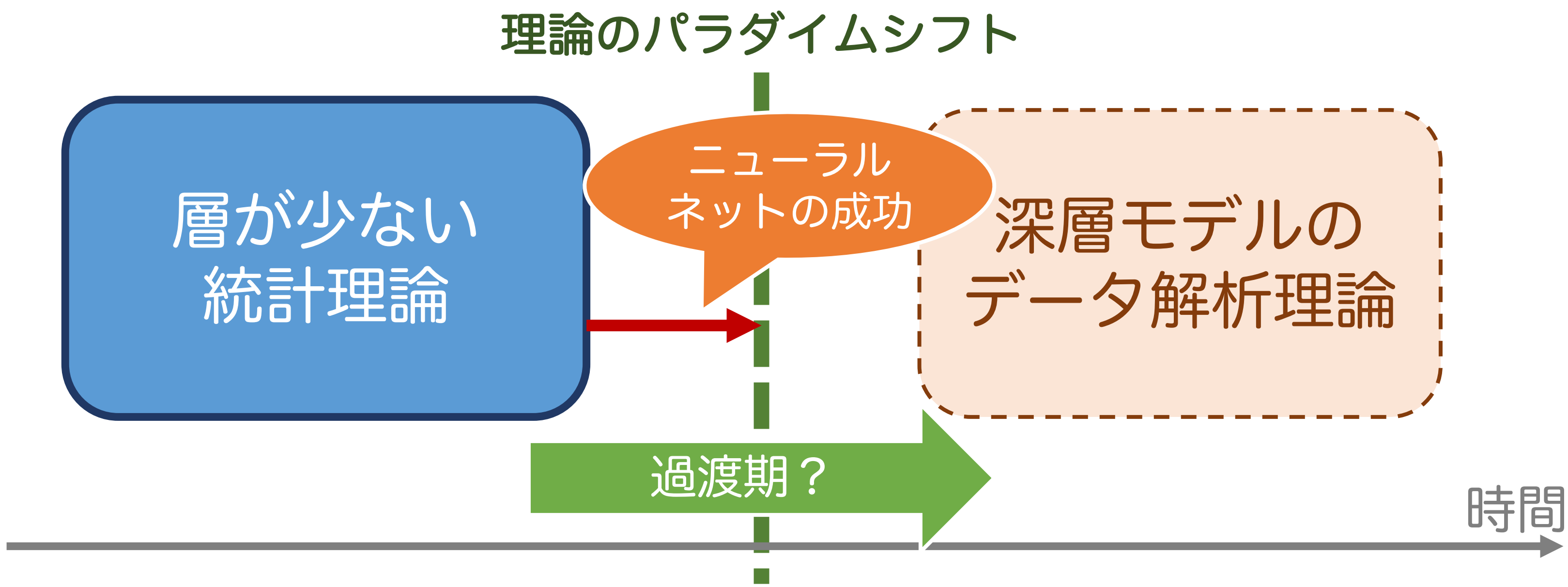
本研究の展望

理論のパラダイムと社会の発展

新理論のパラダイムを目指して

AI・データ解析理論の基盤創出

- 試金石としての深層ニューラルネットワーク



- 現在は新しい理論が定着するかの分水嶺

社会への影響

- 理論・理解に基づいた新技術開発
  - 深層学習の問題を根底から改善