

数学と情報科学で解き明かす多様な対象の数理構造と活用
2019 年度採択研究者

2021 年度 年次報告書

二反田 篤史

九州工業大学 大学院情報工学研究院
准教授

深層学習の潜在的な正則構造の理解に基づく学習法の安定化と高速化

§ 1. 研究成果の概要

ニューラルネットワークの学習は非線形最適化問題に帰着するにも関わらず最適解へ収束することおよび高い汎化能力を持つことが経験的に知られている。本研究ではその理由を解明すべく過剰パラメトリゼーションのもとでの最適化ダイナミクスの解析を進めた。

(1) ニューラルネットワークの最適化問題はそのパラメータについて非凸最適化であり非常に困難な問題であるが、ニューラルネットワークそのものを変数として、その最適化ダイナミクスを適切に捉えると目的関数の凸性を活かすことができる。そのための理論的枠組みの一つにニューラルタンジェントカーネル (NTK) がある。NTK 理論ではニューラルネットワークの学習ダイナミクスを NTK に付随する再生核ヒルベルト空間で記述する。それにより高次元ニューラルネットワークの最適化をカーネル法の理論を援用して議論できる。実際、NTK を用いた既存研究においてニューラルネットワークの最適化法に対する大域的収束性および汎化性が示されたが、その理論には改善の余地があった。本研究では二乗損失を用いた回帰問題において NTK の固有値減衰を考慮にいたれた精緻な汎化誤差解析を与え、平均化確率的勾配降下法が NTK 理論を適用できる条件下でミニマックス最適な汎化誤差を達成することを示した³⁾。

(2) ニューラルネットワークの学習を凸最適化に帰着するもう一つのアプローチとしてニューラルネットワークの平均場理論がある。この理論は個人研究者が開拓したものであり、本年度の研究でさらに深化させることに成功した。本理論が対象とする平均場ニューラルネットワークは表現学習という重要な特徴を備えたモデルであるが、同時に最適化が困難なモデルであるとも考えられていた。本年度の研究では負エントロピー項を正則化として加えると効率的な最適化が可能であることを有限次元の凸最適化手法を無限次元の設定に拡張することで証明した²⁾。これは有限サイズの平均場ニューラルネットワークが効率的に学習できることを示した初の研究成果である。さらに標準的な最適化法であるノイズ付き勾配降下法に対しても新しい理論解析方針を与えた¹⁾。これは平均場ニューラルネットワークに対する平均場ランジュバンダイナミクスの理論に他ならず、今後さらなる発展が期待される。

【代表的な原著論文情報】

- 1) Atsushi Nitanda, Denny Wu, and Taiji Suzuki. Convex Analysis of the Mean Field Langevin Dynamics. The 25th International Conference on Artificial Intelligence and Statistics (AISTATS2022), Proceedings of Machine Learning Research, 151:9741--9757, 2022.
- 2) Atsushi Nitanda, Denny Wu, and Taiji Suzuki. Particle Dual Averaging: Optimization of Mean Field Neural Networks with Global Convergence Rate Analysis. In Advances in Neural Information Processing Systems 34 (NeurIPS2021), pp.19608--19621, 2021.
- 3) Atsushi Nitanda and Taiji Suzuki. Optimal Rates for Averaged Stochastic Gradient Descent under Neural Tangent Kernel Regime. The 9th International Conference on Learning Representations (ICLR), 2021.