

創発的研究支援事業

終了報告書

研究担当者	杉山 磨人
研究機関名	国立情報学研究所
所属部署名	情報学プリンシプル研究系
役職名	准教授
研究課題名	過剰パラメータ化が導く学習原理の再設計
研究実施期間	2022年4月1日～2022年9月30日

研究成果の概要

現代的な機械学習手法が持つ高い汎化性能を達成するためには、過剰パラメータ化と呼ばれるデータ数をも超える大量パラメータを使用するモデルが重要であるが、一旦汎化誤差が減少した後に増加して再び減少する二重降下現象など、従来のモデル選択基準やバイアスバリエンストレードオフでは適切な取り扱いが困難であることが知られており、今後の機械学習の発展には理論と実践双方からの研究推進が欠かせない。そこで本研究では、空間の体積などを利用した幾何学的なアプローチを創出することで、過剰パラメータ化のもとでの汎化性能を理論的に解析し、理論をもとに過剰パラメータ化を活用する機械学習アルゴリズムを構築することを目的として研究を実施した。特に、線形モデル、決定木アンサンブルモデル、生成モデルといった代表的な機械学習モデルについて、過剰パラメータ化の解析と検証を実施した。

特筆すべき成果として、決定木アンサンブルにおける過剰パラメータ化の解析が挙げられる。決定木アンサンブルは、機械学習応用で頻繁に用いられている手法であり、幅広い状況で良い性能を示す。この手法における過剰パラメータ化を解析するために、木の本数の無限大に増やしたときの振る舞いを Neural Tangent Kernel (NTK) を用いて解析した。NTK を用いることで、学習過程を関数空間で記述することができ、無限本の木のアンサンブルモデルを直接解析的に扱うことができるようになる。さらに、無限のパラメータを持つ過剰パラメータ化モデルと等価な振る舞いを、NTK を用いたカーネルリッジ回帰でシンプルに実現することができる。この結果を理論的に示すとともに、木の深さ方向や、パラメータの共有などについての解析を進め、パラメータ数無限大では、oblivious tree と呼ばれる同じ深さでのパラメータ共有をおこない効率的に学習を進めるモデルと、そうでないモデルの学習能力が等価であることを理論的に示すことに成功した。この結果は、機械学習・深層学習分野でのトップ会議 The 10th International Conference on Learning Representations (ICLR 2022) で報告した。