

機械学習の停滞に関する新たなメカニズムを解明

～入力データゆらぎに誘導される停滞現象の発見により効率の良い機械学習系の構成へ期待～

ポイント

- ・オンライン機械学習の停滞現象のメカニズムをランダム力学系理論で解明。
- ・入力データゆらぎが大きくなると、より強い学習の停滞が生じることを発見。
- ・様々な機械学習系における勾配消失や過学習の問題を分析する有力な方法として期待。

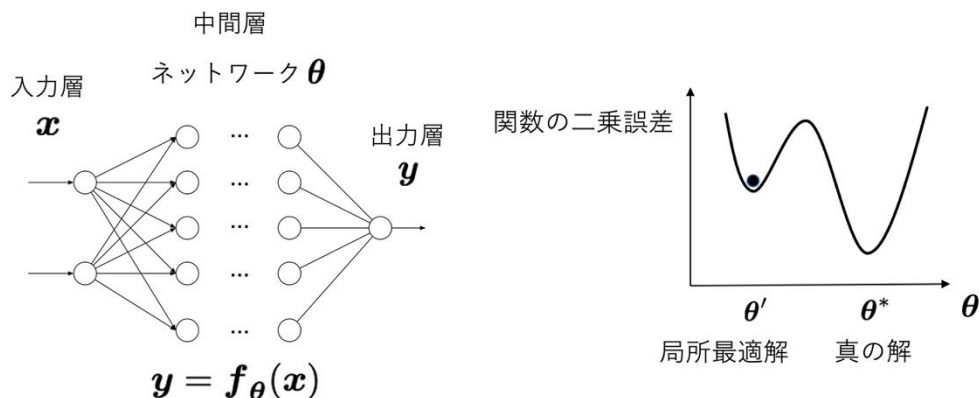
概要

北海道大学電子科学研究所の佐藤 譲准教授は、大阪大学大学院理学研究科の藤原彰夫教授、同博士後期課程筒井大二氏と共同で、オンライン機械学習の停滞に関する新たなメカニズムを、ランダム力学系*1理論に基づいて解析し、入力データゆらぎがこの停滞現象をむしろ強化するという新たな現象を発見しました。さらに、学習を効率化する最適な入力データゆらぎの存在を示しました。

機械学習がうまくいかなくなる原因として、勾配消失、と過学習の二つの問題が知られています。佐藤准教授らは、機械学習の基本要素の一つである三層パーセプトロンモデルのオンライン学習に焦点を当て、この問題を考察しました。結果として、対応する学習の動力学をランダム力学系でモデル化して安定性解析を行うことにより、オンライン学習に固有の雑音によって誘起される停滞現象を新規に発見しました。

本研究成果により提案された理論的枠組みと用いられた評価法は、深層学習を含む様々な機械学習系における勾配消失や過学習の問題を分析する有力な方法となることが期待されます。さらに、入力データの揺らぎの大きさを制御することによって、効率の良い学習系を構成できる可能性が示唆されました。

なお、本研究成果は、2021年11月22日（月）公開のPhysica誌「機械学習と力学系」特集号にオンライン掲載されました。



(左)多層パーセプトロン。パラメーター θ で与えられるネットワーク構造を最適化する。

(右)勾配降下法。局所最適解に陥ると、より誤差の小さい大域最適解(真の解)に到達できない。

【背景】

近年、急速に発達している機械学習技術において、最もよく用いられるニューラルネットワークは多層パーセプトロンです (p.1 図左)。特に階層数の多い多層パーセプトロン (深層ネットワーク) を用いた学習は深層学習とよばれ、多くの現実的問題を効率よく解決する方法として盛んに研究されています。多層パーセプトロンで近似された関数の二乗誤差^{*2}を最小化するアルゴリズムとしては勾配降下法^{*3}がしばしば用いられます。勾配降下法で学習がうまくいなくなる原因として、局所最適解へのトラップ (p.1 図右) に加えて、勾配消失と過学習という二つの非自明な問題が知られています。

問題の一つ目は勾配消失と呼ばれる、学習過程で勾配が0に近くなり、長時間の停滞が生じる現象です。多層パーセプトロンの勾配降下学習において、しばしば勾配消失(vanishing gradient)が生じ、学習がなかなか進展しなくなることがあります。この現象は、損失関数を時間の関数として描いた学習曲線に平坦な部分(プラトー)が生じることから、プラトー現象とよばれています。このときネットワーク内では、(1)中間層の複数のパラメーターが対称的な値をとる (2)ダイナミクスの有効自由度が小さくなる (3)勾配が0に近づき学習が停滞する、という現象が起きています。この停滞状態から回復するためにはネットワークダイナミクスの有効自由度を大きくする必要があります。

問題の二つ目は過学習という入力データに過剰適合した汎用性のない関数(一対一の対応表)が学習される現象です。ニューラルネットワークの関数近似能力を活かすためには、ある程度のネットワークサイズが必要です。しかし逆に大きすぎるネットワークでデータに適合する関数を表現しようとするとう過学習 (overfitting) が生じます。例えば極端な場合、データ点をすべて結んでできる関数、つまり入出力関係の完全な一対一の対応表が構成されてしまい、新しいデータに対する出力が不適切になる、という意味で、汎化に失敗してしまいます。この過学習状態から回復するためにはネットワークダイナミクスの有効自由度を小さくする必要があります。

本研究では、単純なモデルを構成し、ランダム力学系理論でこれら二つの問題を分析しました。

【研究手法】

中間層に二つのニューロンを持つ3層パーセプトロンが十分なデータに基づく平均勾配降下法でプラトー現象を示す最小のモデルであることがわかっています。この最小モデルにおけるオンライン学習を考え、対応する確率勾配降下法の動力学を考察します。このとき学習方程式は4次元のランダム力学系で与えられるので、その局所最適解近傍の確率的ダイナミクスを数値解析と安定性理論に基づいて解析しました。

【研究成果】

オンライン学習系の確率勾配降下法では停滞現象が入力データ揺らぎによりむしろ強化されることがわかりました。この現象は「勾配系にノイズを加える」という古典的な描像 (図1左) では捉えることのできない非自明な現象で、ランダム力学系理論におけるノイズによる安定化現象 (図1右) という形で、はじめてそのメカニズムが明らかになったものです。発見されたオンライン学習に固有の学習の停滞現象は、雑音によって誘起される停滞現象であることから、雑音誘起縮退現象(noise-induced degeneration)と名付けました。

また、数値実験により、非常に小さい入力データ揺らぎでは停滞領域での滞在時間は長く、中間スケールの揺らぎでは短く、さらに大きな揺らぎでは雑音誘起縮退現象により停滞領域での滞在時間は長くなることわかりました。つまり「学習が停滞する時間を最小にする最適な揺らぎサイズ」が存在することが示され、入力データ揺らぎの大きさを制御することによって、効率の良い学習系を構成できる

可能性が示唆されました（図2）。

【今後への期待】

本研究成果により、入力データゆらぎがこの停滞現象をむしろ強化する現象が発見され、学習が停滞する時間を最小にするような最適揺らぎサイズが存在することが示されたことで、オンライン学習の停滞のメカニズムがランダム力学系理論に基づいて明らかになりました。また、提案された理論的枠組みと用いられた評価法は、深層学習を含む様々な機械学習系における勾配消失や過学習の問題を分析する有力な方法となることが期待されます。さらに、入力データの揺らぎの大きさを制御することによって、効率の良い学習系を構成できる可能性が示唆されました。

【謝辞】

本研究はLondon Mathematical Laboratory External Fellowship, 日本学術振興会科学研究費補助金基盤研究 (C)「ランダム力学系理論に基づく確率カオスの現象論とその応用」、基盤研究 (B)「非線形確率微分方程式系における確率カオスの定量解析とその応用」、基盤研究 (B)「古典および量子統計的システムにおける新規な情報幾何構造の探究」の支援を受けて実施されました。

論文発表の概要

論文名	Noise-induced degeneration in online learning (オンライン学習における雑音誘起退化)
著者名	佐藤 譲 ^{1,2} , 筒井大二 ³ , 藤原彰夫 ³ (¹ 北海道大学電子科学研究所, ² ロンドン数学研究所, ³ 大阪大学大学院理学研究科)
雑誌名	Physica D (非線形物理学の専門誌)
DOI	10.1016/j.physd.2021.133095
公表日	2021年11月22日(月)(オンライン公開)

お問い合わせ先

北海道大学電子科学研究所 准教授 佐藤 譲 (さとうゆずる)

T E L 011-706-9453 F A X 011-706-9453

メール ysato@math.sci.hokudai.ac.jp

U R L <http://www.math.sci.hokudai.ac.jp/~ysato/>

配信元

北海道大学総務企画部広報課 (〒060-0808 札幌市北区北8条西5丁目)

T E L 011-706-2610 F A X 011-706-2092 メール jp-press@general.hokudai.ac.jp

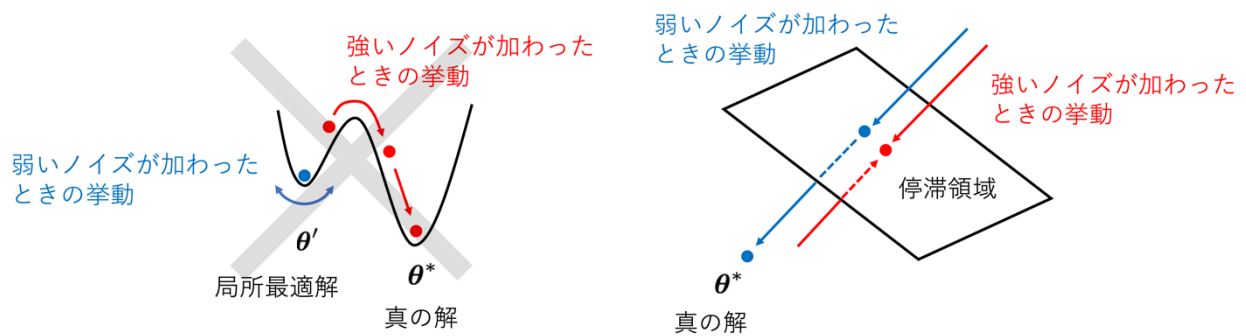


図1. (左)決定論的な平均勾配系に加法的にノイズを加えたときの、局所最適解からの脱出。
 (右)確率勾配法に従うランダム力学系におけるノイズによる安定化。

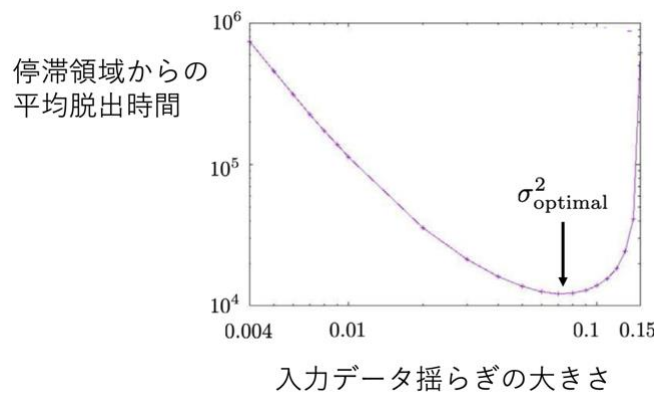


図2. 停滞領域からの平均脱出時間と入力データ揺らぎサイズの関係。最適揺らぎが存在する。

【用語解説】

- *1 ランダム力学系 … 駆動力としてノイズや外力が加わっている力学系。その運動方程式は確率微分方程式やランダム写像で与えられる。
- *2 二乗誤差 … ある θ に対して $E(\theta) = \|\mathbf{f}_\theta - \mathbf{f}_{\theta^*}\|^2$ で与えられる関数の近似誤差。
- *3 勾配降下法 … ある θ に対して $\Delta\theta = -\eta \frac{\partial E(\theta)}{\partial \theta}$ に基づいて θ の値を更新することにより最適解 θ^* を得るアルゴリズム。確率変数を含む更新式に従う方法をとくに確率勾配降下法という。