

令和 4 年 6 月 13 日現在

機関番号：17104

研究種目：若手研究

研究期間：2019～2021

課題番号：19K20337

研究課題名（和文）超高次元機械学習モデルの学習ダイナミクスの究明と効率的学習法の開発

研究課題名（英文）Study on learning dynamics of high-dimensional machine learning models and development of efficient learning methods

研究代表者

二反田 篤史（Nitanda, Atsushi）

九州工業大学・大学院情報工学研究院・准教授

研究者番号：60838811

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：深層学習に代表される高次元機械学習モデルが機能する要因の究明および効率的学習法の開発を目指し、学習ダイナミクスの研究を推進した。特に代表的学習法である（確率的）勾配降下法に対し次の成果を得た。

(1)線形モデルを用いた判別問題に対し判別誤差が低ノイズ条件下で指数収束することを証明。(2)確率的勾配降下法で学習された二層ニューラルネットワークの汎化性が最適効率を達成することをNTK理論を精緻化し証明。(3)輸送写像の関数勾配法理論に基づくニューラルネットワークの解析方法を考案し、新たな学習法を提案。(4)平均場ニューラルネットワークの最適化ダイナミクスの考案と収束性を証明。

研究成果の学術的意義や社会的意義

深層学習の原理解明に向けた二種の最適化理論：NTK理論および平均場ニューラルネットワーク理論の進展に寄与した。具体的にはNTK理論を精緻化しニューラルネットワークを理論上最適な効率で学習可能であることを初めて証明し、またデータへの適応性に優れた平均場ニューラルネットワークの最適化ダイナミクスを解析する新たな研究の流れを創出した。これらの成果は深層学習の最適化ダイナミクスの基礎を与えるもので、深層学習の効率化への重要なステップである。

研究成果の概要（英文）：We study learning dynamics of machine learning models, aiming to understand why high-dimensional models such as deep learning work well and to develop efficient learning methods. In particular, we obtained the following results for the (stochastic) gradient descent method, which is a representative learning method.

(1) We proved that the classification error converges exponentially under low noise conditions for classification problems using linear models. (2) We proved that the generalization ability of the two-layer neural network trained by the stochastic gradient descent method achieves optimal efficiency by refining the NTK theory. (3) We developed a way for analyzing neural networks based on the functional gradient theory of transport mapping and proposed a new learning method. (4) We developed an optimization dynamics of mean-field neural networks and proved its convergence.

研究分野：機械学習

キーワード：機械学習 深層学習 ニューラルネットワーク 確率的勾配降下法 ランジュバンダイナミクス

1. 研究開始当初の背景

近年、物体・音声認識、自然言語処理、データ生成技術等の高度な人工知能技術が高い精度で実現され始めた事で、その根幹技術である機械学習に対する注目と期待が産業界・学术界において急速に高まっている。このパラダイムシフトは深層学習という名のもとに引き起こされたが、その実態である所の深層ニューラルネットを始めとする超高次元モデルの優れた性能を裏付ける理論的理解は未だ不十分であり、モデリング・チューニングの理論的指針が無いまま闇雲な試行錯誤が要求される実状にある。巨大モデルの学習には膨大な計算コストを必要とする為、超高次元モデルの理論的究明及び効率的学習法の開発は急務である。

2. 研究の目的

深層ニューラルネットは非常に多くのパラメータを備える超高次元非凸モデルであるが、正則化無しでも高い汎化性を示す事が経験的に知られている。しかしながら、この現象に対する理論的理解は未だ不十分であり解明が求められている。また超高次元非凸モデルであるため最適化が困難であり、その為のパラメータチューニングに多大なコストを要するという問題もある。

そこで本研究では学習ダイナミクスそのものに超高次元モデルの成功理由があると考え相補的な関係にある以下の2つの課題に取り組む。

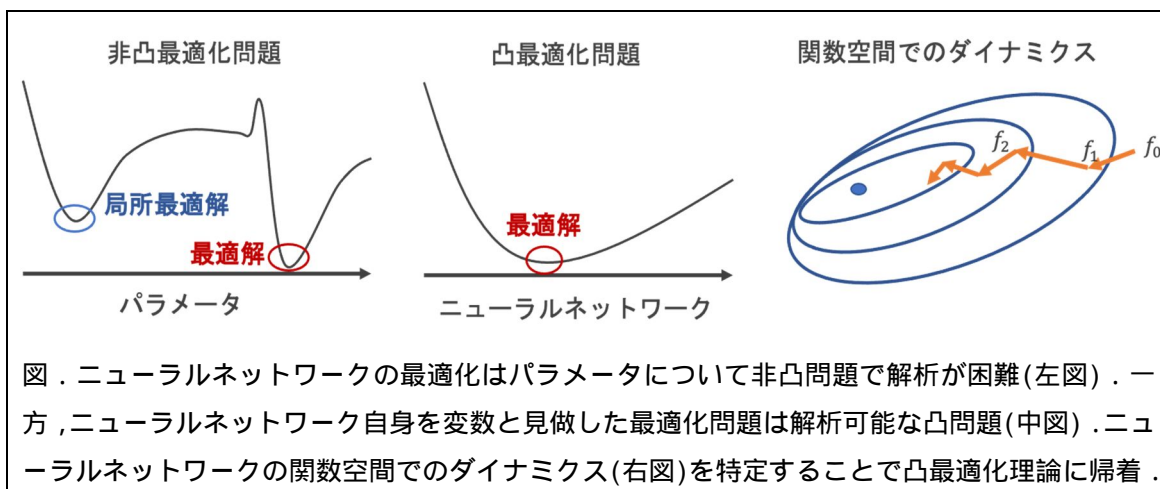
課題 1: 超高次元モデルの学習ダイナミクスに基づく汎化理論の構築。

課題 2: 理論に裏付けされた超高次元モデル用の効率的学習手法の開発。

3. 研究の方法

深層学習で用いられるような高次元機械学習モデルの汎化性能は使用する学習法の性質に依存することが知られている。そこで代表的学習法である確率的勾配降下法あるいはその派生手法の収束解析を通じて高次元モデルの研究を推進する。学習問題は凸最適化問題と非凸最適化問題のいずれかに帰着する。本研究ではいずれのケースも取り扱う。凸最適化問題に帰着する場合は従来の機械学習理論と統計的学習理論を駆使する。

ニューラルネットワークの場合は非凸最適化になるため解析が一般に困難である。しかし、これはニューラルネットワークのパラメータを変数と見做した際の問題であり、ニューラルネットワーク自身を変数と見做せば凸最適化問題という解析可能なクラスに帰着する。そこでニューラルネットワークの最適化ダイナミクスを適当な関数空間で記述し凸最適化理論に帰着させる理論：ニューラルタンジェントカーネル(NTK)理論と輸送写像の関数勾配法理論・平均場理論を駆使し研究を進める。



4. 研究成果

(1) 凸経験損失最小化問題に対する確率的最適化手法の研究

ミニバッチを用いた確率的最適化手法は深層学習を含めた機械学習モデルの代表的学習手法である。ミニバッチは確率的勾配（経験損失関数の勾配の確率的近似）の分散を縮小させ収束に掛かる反復数を軽減する目的で用いられる。本研究では、ミニバッチ法をさらに効率化させた分散縮小法を研究し、収束に掛かる最小反復数およびその際必要となる最小ミニバッチサイズを具体的な確率的最適化手法を提案するとともに示した。分散縮小法は正則化付き経験損失最小化法の高速化を実現する技術であるため、課題2の解決に貢献する成果である。

(2) 関数勾配法に基づく逐次的な残差ネットワークの学習法

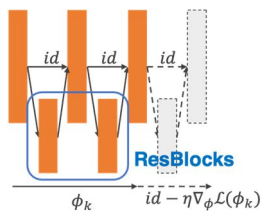


図1: ResNet 最適化

と本手法は残差ブロックの凸結合で分類可能なデータセットが学習可能であることが事前に保証される。提案手法の最適化過程は入力層のノード追加および層の追加で実現されるため（図1）、データに適応的なネットワーク構造最適化手法ともみなすことができ、課題2に貢献するものである。

残差ネットワークは残差ブロック（あるいはスキップ接続）という層構造を備える深層学習モデルで、とりわけ高いパフォーマンスを発揮することが知られている。研究代表者は残差ブロックを関数空間における最適化のステップとみなすことで関数空間におけるフランク・ウルフ型最適化法が残差ネットワークの構築に適用可能であることを示し、同時に汎化性能保証もあたえた。学習後に大域的収束性のための条件の充足性が判定される既存手法に比べる

(3) NTK 理論下の高次元二層ニューラルネットワークの最適収束率

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すれば、凸最適化問題と見なせる。すなわち、パラメータについての最適化法から関数としての更新則を特定出来れば、この凸性を最大限に活かした理論解析が可能となる。そのための道具としてニューラルタンジェン

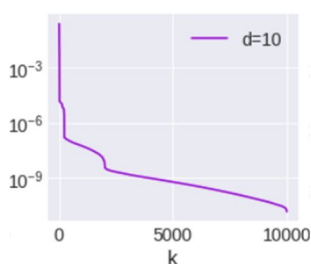


図2: NTK の固有値減衰

ミニマックス最適レートを達成することを示した。これは NTK 理論を精緻化するものであり、課題1に貢献するものである。また NTK 理論の研究を通じて、高次元ニューラルネットワークの学習がより効率化されるための条件が見えつつあり、課題2に関して重要な知見をもたらすと考えている。

トカーネル（NTK）および研究代表者が提案した平均場ニューラルネットワークの二つの理論がある。NTKの理論ではニューラルネットワークの横幅を無限大に大きくした極限での勾配法の挙動を NTK が定める再生核ヒルベルト空間内で記述する。本研究では NTK の固有値減衰を考慮に入れた解析により特定条件下で平均化確率的勾配降下法が大域的収束することと汎化誤差についての

(4) 平均場ニューラルネットワークの効率的最適化手法と収束率解析

大域収束性を持つモデルとして平均場ニューラルネットワークがある。このモデルは NTK 理論が対象と

するニューラルネットワークに比べて、表現学習の性質を備えることから重要な研究対象になっている。簡易的な実験で平均場ニューラルネットワークのパラメータが真の関数に効率的に収束することが確認できる(図 3)。しかし同時に最適化が難しいモデルでもあり効率的な収束性の担保には強い条件が必要と考えられていた。実際、一般的な条件下における収束効率については不明な部分が多い。本研究では、KL 情報量による正則化を課せば平均場ニューラルネットワークを劣線形の効率性で最適化できることを有限次元の凸最適化法を自然に拡張することで示した。さらにより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平均場ランジュバンダイナミクスの理論を通じて収束率解析を与えた。これらは課題 1,2 に貢献する成果である。

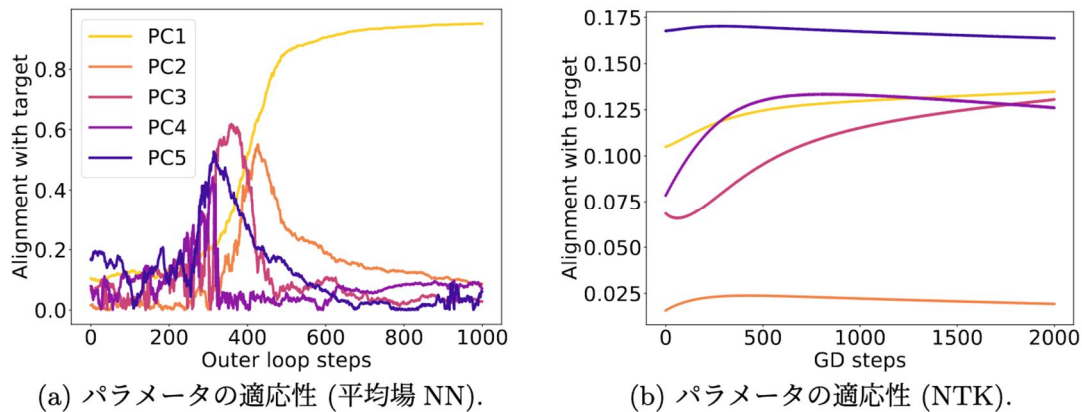


図 3. 真の関数とパラメータの類似度の推移。

(5) 低ノイズ条件下でランダム特徴を用いた確率的勾配降下法の線形収束性

二層ニューラルネットワークの最適化は入力層パラメータの学習を本質的には対象としている。

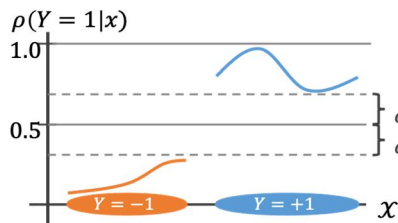


図 3 : ラベルの低ノイズ性

一方で出力層の学習はランダム特徴近似を用いたカーネル法と同一視される。カーネル法における確率的勾配降下法に関しては、研究代表者の既存研究においてラベルの低ノイズ性が識別誤差の収束速度を指数的に加速することを示している。さらに本研究ではこの理論をランダム特徴近似を用いたカーネル法へと拡張した。この理論と NTK

理論はカーネルで記述可能なニューラルネットワークの学習の研究を進展させるものである。本成果は課題 1 の解決に貢献する。

(6) その他

上記研究に並行し深層学習モデルがその高い適応性によってカーネル法を優越することを示す研究やグラフ埋め込みの汎化性能を調べる研究に取り組んだ。

【引用文献】

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.

Atsushi Suzuki, Atsushi Nitanda, Jing Wang, Linchuan Xu, Kenji Yamanishi, and Marc Cavazza. Generalization Bounds for Graph Embedding Using Negative Sampling: Linear vs Hyperbolic. In Advances in Neural Information Processing Systems 34 (NeurIPS2021), pp.1243--1255, 2021.

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.

Taiji Suzuki and Atsushi Nitanda. Deep learning is adaptive to intrinsic dimensionality of model smoothness in anisotropic Besov space. In *Advances in Neural Information Processing Systems 34 (NeurIPS2021)*, pp.3609--3621, 2021.

Shingo Yashima, Atsushi Nitanda, and Taiji Suzuki. Exponential Convergence Rates of Classification Errors on Learning with SGD and Random Features. *The 24th International Conference on Artificial Intelligence and Statistics (AISTATS), Proceedings of Machine Learning Research*, 130:1954–1962, 2021.

Atsushi Nitanda, Tomoya Murata, and Taiji Suzuki. Sharp Characterization of Optimal Minibatch Size for Stochastic Finite Sum Convex Optimization. *Knowledge and Information Systems (KAIS)*, 63(9):2513--2539, 2021. (Journal version of ICDM2019 paper)

Atsushi Suzuki, Atsushi Nitanda, Jing Wang, Linchuan Xu, Kenji Yamanishi, and Marc Cavazza. Generalization Error Bound for Hyperbolic Ordinal Embedding. *The 38th International Conference on Machine Learning (ICML2021), Proceedings of Machine Learning Research*, 139:10011--10021, 2021.

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.

Atsushi Nitanda and Taiji Suzuki. Functional Gradient Boosting for Learning Residual-like Networks with Statistical Guarantees. *The 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), Proceedings of Machine Learning Research*, 108:2981–2991, 2020.

Atsushi Nitanda, Tomoya Murata, and Taiji Suzuki. Sharp Characterization of Optimal Minibatch Size for Stochastic Finite Sum Convex Optimization. *2019 IEEE International Conference on Data Mining (ICDM2019)*, pp. 488–497. 2019

5. 主な発表論文等

〔雑誌論文〕 計10件（うち査読付論文 10件 / うち国際共著 2件 / うちオープンアクセス 9件）

1. 著者名 Nitanda Atsushi, Murata Tomoya, Suzuki Taiji	4. 巻 63
2. 論文標題 Sharp characterization of optimal minibatch size for stochastic finite sum convex optimization	5. 発行年 2021年
3. 雑誌名 Knowledge and Information Systems	6. 最初と最後の頁 2513-2539
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10115-021-01593-1	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Taiji Suzuki, Atsushi Nitanda	4. 巻 34
2. 論文標題 Deep learning is adaptive to intrinsic dimensionality of model smoothness in anisotropic Besov space	5. 発行年 2021年
3. 雑誌名 Advances in Neural Information Processing Systems (NeurIPS2021)	6. 最初と最後の頁 3609-3621
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Atsushi Nitanda, Denny Wu, Taiji Suzuki	4. 巻 34
2. 論文標題 Particle Dual Averaging: Optimization of Mean Field Neural Networks with Global Convergence Rate Analysis	5. 発行年 2021年
3. 雑誌名 Advances in Neural Information Processing Systems (NeurIPS2021)	6. 最初と最後の頁 19608-19621
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1. 著者名 Atsushi Nitanda, Denny Wu, Taiji Suzuki	4. 巻 151
2. 論文標題 Convex Analysis of the Mean Field Langevin Dynamics	5. 発行年 2022年
3. 雑誌名 Proceedings of Machine Learning Research (AISTATS2022)	6. 最初と最後の頁 9741-9757
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1. 著者名 Atsushi Nitanda and Taiji Suzuki	4. 巻 108
2. 論文標題 Functional Gradient Boosting for Learning Residual-like Networks with Statistical Guarantees	5. 発行年 2020年
3. 雑誌名 Proceedings of Machine Learning Research (AISTATS2020)	6. 最初と最後の頁 2981-2991
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Shingo Yashima, Atsushi Nitanda, and Taiji Suzuki	4. 巻 130
2. 論文標題 Exponential Convergence Rates of Classification Errors on Learning with SGD and Random Features	5. 発行年 2021年
3. 雑誌名 Proceedings of Machine Learning Research (AISTATS2021)	6. 最初と最後の頁 1954-1962
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Atsushi Nitanda and Taiji Suzuki	4. 巻 89
2. 論文標題 Stochastic Gradient Descent with Exponential Convergence Rates of Expected Classification Errors	5. 発行年 2019年
3. 雑誌名 Proceedings of Machine Learning Research (AISTATS2019)	6. 最初と最後の頁 1417-1426
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Nitanda Atsushi, Murata Tomoya, and Suzuki Taiji	4. 巻 -
2. 論文標題 Sharp Characterization of Optimal Minibatch Size for Stochastic Finite Sum Convex Optimization	5. 発行年 2019年
3. 雑誌名 In Proceedings of IEEE International Conference on Data Mining	6. 最初と最後の頁 488-497
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Satoshi Hara, Atsushi Nitanda, and Takanori Maehara	4. 巻 32
2. 論文標題 Data Cleansing for Models Trained with SGD	5. 発行年 2019年
3. 雑誌名 Advances in Neural Information Processing Systems	6. 最初と最後の頁 4215-4224
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Atsushi Suzuki, Jing Wang, Feng Tian, Atsushi Nitanda, and Kenji Yamanishi	4. 巻 101
2. 論文標題 Hyperbolic Ordinal Embedding	5. 発行年 2019年
3. 雑誌名 In Proceedings of Machine Learning Research (ACML2019)	6. 最初と最後の頁 1065-1080
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

[学会発表] 計19件 (うち招待講演 4件 / うち国際学会 5件)

1. 発表者名 Shun-ichi Amari, Jimmy Ba, Roger Grosse, Xuechen Li, Atsushi Nitanda, Taiji Suzuki, Denny Wu, and Ji Xu
2. 発表標題 When Does Preconditioning Help or Hurt Generalization?
3. 学会等名 International Conference on Learning Representation (ICLR2021) (国際学会)
4. 発表年 2020年 ~ 2021年

1. 発表者名 Atsushi Nitanda and Taiji Suzuki
2. 発表標題 Optimal Rates for Averaged Stochastic Gradient Descent under Neural Tangent Kernel Regime
3. 学会等名 International Conference on Learning Representation (ICLR2021) (国際学会)
4. 発表年 2020年 ~ 2021年

1. 発表者名 二反田篤史
2. 発表標題 確率的最適化法の収束解析
3. 学会等名 RAMP数理最適化シンポジウム (招待講演)
4. 発表年 2020年

1. 発表者名 二反田篤史, 鈴木大慈
2. 発表標題 確率的勾配降下法のNTK理論による最適収束率
3. 学会等名 統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 二反田篤史, Denny Wu, 鈴木大慈
2. 発表標題 粒子双対平均化法: 平均場ニューラルネットワークの大域的収束保証付最適化法
3. 学会等名 情報論的学習理論ワークショップ (IBIS)
4. 発表年 2020年

1. 発表者名 森雄人, 二反田篤史, 武田朗子
2. 発表標題 二段階最適化によるモデル抽出攻撃に対する防御
3. 学会等名 情報論的学習理論ワークショップ (IBIS)
4. 発表年 2020年

1. 発表者名 Shun-ichi Amari, Jimmy Ba, Roger Grosse, Xuechen Li, Atsushi Nitanda, Taiji Suzuki, Denny Wu, and Ji Xu
2. 発表標題 When Does Preconditioning Help or Hurt Generalization?
3. 学会等名 The 12th OPT Workshop on Optimization for Machine Learning (国際学会)
4. 発表年 2020年

1. 発表者名 二反田篤史
2. 発表標題 二層ニューラルネットワークの最適化理論
3. 学会等名 第2回若手数学者交流会
4. 発表年 2021年

1. 発表者名 Atsushi Nitanda
2. 発表標題 Stochastic Gradient Descent with Exponential Convergence Rates for Classification Problems
3. 学会等名 Summer School 2019 on Transfer Learning (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 二反田篤史
2. 発表標題 高次元ニューラルネットに対する勾配法の大域収束性と汎化性能解析
3. 学会等名 日本オペレーションズ・リサーチ学会 研究部会 最適化とその応用 (OPTA) (招待講演)
4. 発表年 2019年

1. 発表者名 二反田篤史
2. 発表標題 学習アルゴリズムの大域収束性と帰納的バイアス
3. 学会等名 情報論的学習理論ワークショップ (IBIS) (招待講演)
4. 発表年 2019年

1. 発表者名 八嶋晋吾, 二反田篤史, 鈴木大慈
2. 発表標題 Random Featureを用いた確率的勾配法の期待識別誤差の収束解析
3. 学会等名 情報論的学習理論ワークショップ (IBIS)
4. 発表年 2019年

1. 発表者名 原聡, 二反田篤史, 前原貴憲
2. 発表標題 SGDの挙動解析に基づくデータクレンジング
3. 学会等名 情報論的学習理論ワークショップ (IBIS)
4. 発表年 2019年

1. 発表者名 二反田篤史, 鈴木大慈
2. 発表標題 高次元二層ニューラルネットに対する勾配降下法による識別誤差の大域収束性と汎化性能解析
3. 学会等名 情報論的学習理論ワークショップ (IBIS)
4. 発表年 2019年

1. 発表者名 二反田篤史
2. 発表標題 識別問題に対する高次元二層ニューラルネットの大域収束性と汎化性能解析
3. 学会等名 情報系 WINTER FESTA Episode 5
4. 発表年 2019年

1. 発表者名 Atsushi Nitanda, Taiji Suzuki
2. 発表標題 Exponential convergence of stochastic gradient descent for binary classification problems
3. 学会等名 The Conference of Data Science, Statistics & Visualisation (DSSV) (国際学会)
4. 発表年 2019年

1. 発表者名 二反田篤史, 鈴木大慈
2. 発表標題 識別問題に対する高次元ニューラルネットの勾配降下法の大域収束性と汎化性能解析
3. 学会等名 日本応用数理学会年会
4. 発表年 2019年

1. 発表者名 二反田篤史, 鈴木大慈
2. 発表標題 識別問題に対する高次元二層ニューラルネットの勾配法による汎化性能解析
3. 学会等名 統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 八嶋晋吾, 二反田篤史, 鈴木大慈
2. 発表標題 カーネル法における random featureを用いた確率的勾配法の期待識別誤差の線形収束性
3. 学会等名 統計関連学会連合大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関