

令和 3 年 5 月 13 日現在

機関番号：14101

研究種目：基盤研究(C)（一般）

研究期間：2018～2020

課題番号：18K11433

研究課題名（和文）縮小推定を導入した貪欲法の下でのモデル選択規準についての研究

研究課題名（英文）On model selection criteria under shrinkage estimation in greedy learning

研究代表者

萩原 克幸（Hagiwara, Katsuyuki）

三重大学・教育学部・教授

研究者番号：60273348

交付決定額（研究期間全体）：（直接経費） 2,300,000円

研究成果の概要（和文）：本研究では、階層型ニューラルネットやスパース学習に共通するモデル選択の問題を扱う。特に、ここでは、正則化・縮小推定の下でのモデル選択を考えた。まず、スパース学習において、正則化法であるLASSOのバイアス問題を解決するスケーリング法を考え、その下でのモデル選択規準を導出し、応用上の妥当性を数値的に確認した。さらに、スケーリング法を利用して、ノンパラメトリック直交回帰の下での統一的なモデリング法を与えるとともに、その汎化性を理論的に解析した。一方で、階層型ニューラルネットについては、モデル選択に関して、その深層化による学習の傾向とオーバーフィッティングの関係を調べた。

研究成果の学術的意義や社会的意義

最近、機械学習分野では、深層学習およびスパース学習という二つのキーワードが注目されており、社会的にもインパクトを与えているが、モデル選択の問題はまだ研究が続いている。これらは独立に発展しているが、深層学習の基本である階層型のニューラルネットとスパース学習は、いずれも、学習によって選択できる関数達の線形結合により構成されるモデルを考えているという共通点をもつ。これまでの研究で、こうしたモデルを貪欲法の下で学習した場合、予測誤差の推定値がモデル選択規準として応用可能な形にならないことが知られている。本研究は、この問題を解決するために、縮小推定を導入した下でのモデル選択を考えるものである。

研究成果の概要（英文）：In this project, we considered a model selection problem that is common for both of layered neural nets and sparse modeling. We considered model selection under regularization and shrinkage methods. In a sparse modeling, we derived a scaling method for LASSO, in which a bias problem is relaxed. And, we derived a risk-based model selection criterion for the estimate under the proposed method. We confirm its effectiveness through numerical experiments. Additionally, by introducing a scaling method, we derived a unified modeling method under a non-parametric orthogonal regression problem and we analyzed the generalization properties of the proposed method. On the other hand, in layered neural nets, we found that a deep structure affects over-fitting to noise.

研究分野：機械学習

キーワード：スパース学習 階層型ニューラルネット モデル選択 正則化 縮小推定

様式 C - 19、F - 19 - 1、Z - 19 (共通)

## 1. 研究開始当初の背景

最近、機械学習分野では、深層学習およびスパース学習という二つのキーワードが注目されており、社会的にもインパクトを与えている。こうした方法の応用では、常に、モデル選択の問題が生じ、正則化をはじめとして様々なヒューリスティクスが提案されてきた。一方で、それらの理論解析はそれほど進んでおらず、例えば、赤池情報量規準のような、モデル選択のための規準についてはあまり研究が進んでいない状況である。本研究は、こうした方法について、汎化性の理論解析やそれに基づくモデル選択を考える。実は、深層学習とスパース学習は独立に発展しているが、深層学習の基本である階層型ニューラルネットとスパース学習は、いずれも、可変な関数達の線形結合により構成され、その関数達を学習により選択できる方法とみなせる。特に、その学習には貪欲法を用いるのが一般的である。階層型ニューラルネットについてはモデル選択規準の問題は解決されていないが、スパース学習においては、LASSO(Least Absolute Shrinkage and Selection Operator)について、リスクの不偏推定量として、応用可能なモデル選択規準が与えられている。例えば、階層型ニューラルネットに対する最小二乗学習のように、単純な貪欲法を用いた場合、一般に、リスクの不偏推定量に真の関数の情報が現れるため、応用可能な形のモデル選択規準を構成できないことが、これまでの研究で分かっている。LASSO においてこれが可能であるのは、LASSO が貪欲法である閾値処理と同時に縮小推定を行っているためである。

## 2. 研究の目的

本研究では、貪欲法の下での学習に適切な縮小推定を導入した場合について、モデル選択の問題を理論的に考えることを目的とする。スパース学習においては、上述したように LASSO についてリスクの不偏推定量が計算されている。しかしながら、LASSO は推定量にはバイアスが生じる問題があり、スパースで汎化性の良いモデルを構成できないことが知られている。これに対して、バイアスの問題を解決する方法として、Adaptive LASSO や MCP(Minimax Concave Penalty)などの非凸な正則化法が提案されている。しかしながら、LASSO のように、ハイパーパラメータを選択するためのモデル選択規準の研究はほとんど見られない。ここでは、本研究では、スパース学習の問題において、LASSO のバイアスの問題を解決しつつも、解析的にモデル選択規準を導くことができる方法を研究する。一方、近年、深層学習の応用上の成功が深層化にあることが指摘されているが、その真偽については明らかにされていない。こうした中、関数近似や統計的学習理論の枠組みでも層数の重要性が指摘されているが、深層化のアドバンテージに対する納得のいく説明がなされていないと思えない。本研究では、特に、モデル選択に関連して、深層化と汎化性の関係について

考える。

### 3 . 研究の方法

本研究はスパース学習と深層学習についての研究に分けられる。

スパース学習については、LASSO 推定量をスケーリングにより改良する方法を考える。これまでの研究で、直交回帰の下では、soft thresholding 推定量 (LASSO 推定量の特殊な場合) の経験的スケーリングによる改良が Non-negative Garrote 推定量と一致することを示している。これは、soft thresholding 推定量のバイアス問題を解決するとともに、そのリスクの不偏推定量である SURE(Stein's Unbiased Risk Estimate)が計算できるものとなっている。そこで、本研究では、一般の線形回帰について、LASSO 推定量に対して単一の経験的スケーリング値を導入し、その SURE を計算する。さらに、性能のよい推定量を開発するために、係数ごとにスケーリング値を導入する方法を考える。このために、まずは、ノンパラメトリック直交回帰の下で、スケーリングの方法を拡張し、その下での SURE を計算することを考える。

深層学習については、特に、全結合層 (階層型ニューラルネット) の部分について、深層化が学習にどのように影響するかを調べ、それと汎化性との関係を明らかにする。

### 4 . 研究成果

3.のそれぞれについての研究成果を述べる。

リスクの意味で最適な単一のスケーリング値に対して、LASSO 推定量をプラグインした形の推定値を考え、この推定値を経験的スケーリング値として採用することにした。このスケーリング値は LASSO 推定量を最小二乗推定量に近づける働きをし、正則化パラメータが大きいとき (スパースな表現のとき) に LASSO 推定量のバイアス問題をよく改良するものとなっていることが示された。このスケーリング値によりスケーリングされた LASSO 推定量は Stein's lemma が適用できることから、SURE を導出した。簡単な数値実験により、この推定値が LASSO よりリスクを改善するとともに、モデル選択規準としてみたときに、LASSO と比較して、適切にスパースな表現を選択できる規準となっていることを確認した。一方で、係数ごとにスケーリングを導入する方法については、ノンパラメトリック直交回帰の下で考えた。ここでは、soft thresholding 推定量の経験的スケーリングとして、hard thresholding 推定量を与えるスケーリング値を Taylor 展開し、展開次数をハイパーパラメータとするスケーリング値を考えた。こ

のスケーリング推定量はハイパーパラメータにより、soft thresholding と hard thresholding をつなぐ bridge 推定量となっており、soft thresholding を特殊な場合として含む。さらに、この方法は Non-negative Garrote を含むとともに、adaptive ALSSO の別の表現を与えている。この推定量については SURE が計算でき、さらに、それを簡単な数値実験により確かめた。hard thresholding は前述した貪欲法であり、それに対して、リスクに基づく応用可能なモデル選択規準を構成することはできない。この bridge 推定量は、hard thresholding 推定量と異なり、応用可能なモデル選択規準が構成できるというメリットがある。この bridge 推定量は、展開次数が大きいつきに hard thresholding に近いため、この研究では、さらに、hard thresholding のオーバーフィッティングの状況について解析した。その結果、特に、スパースかつノンパラメトリックのコンテキストでは、真値が零である係数の推定値が、soft thresholding(一般には LASSO)の場合と異なった仕方で、オーバーフィッティング(モデルの自由度)に大きく関係することなどを明らかにした。

階層型ニューラルネットについては、ReLU(Rectified Linear Unit)を隠れ素子の活性化関数、ロジステックシグモイド関数を出力層の活性化関数として用いる場合について、層数とネット出力との関係性を調べた。この場合、勾配の爆発で知られるように、層数の増加に従い重みの更新量は大きくなる。しかしながら、対象が不連続な関数の場合、それを実現するために重みの発散が必要であり、この場合、層数が多いほど、その収束は早いことが示される。一方、これまでの研究から、階層型ニューラルネットの場合、ノイズに対するオーバーフィッティングの仕方として、不連続な関数を実現することで対応する傾向にあることが示されている。このことから、深層化によりオーバーフィッティングの問題はシリアスになる可能性が高いことが示唆される。

## 5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件/うち国際共著 2件/うちオープンアクセス 1件）

1. 著者名 Katsuyuki Hagiwara	4. 巻 arXiv:2104.09703
2. 論文標題 Bridging between soft and hard thresholding by scaling	5. 発行年 2021年
3. 雑誌名 arXiv:2104.09703	6. 最初と最後の頁 1-10
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 無
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 Hagiwara Katsuyuki	4. 巻 II
2. 論文標題 A Model Selection Criterion for LASSO Estimate with Scaling	5. 発行年 2019年
3. 雑誌名 ICONIP2019, LNCS 11945	6. 最初と最後の頁 248 ~ 259
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-36711-4_22	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 該当する

1. 著者名 Hagiwara Katsuyuki	4. 巻 I
2. 論文標題 On a Fitting of a Heaviside Function by Deep ReLU Neural Networks	5. 発行年 2018年
3. 雑誌名 ICONIP 2018, LNCS 11301	6. 最初と最後の頁 59 ~ 69
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-04167-0_6	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 該当する

〔学会発表〕 計3件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 Katsuyuki Hagiwara
2. 発表標題 A model selection criterion for LASSO estimate with scaling
3. 学会等名 電子情報通信学会 情報論的学習理論と機械学習研究会
4. 発表年 2019年

1. 発表者名 Katsuyuki Hagiwara
2. 発表標題 A Model Selection Criterion for LASSO Estimate with Scaling
3. 学会等名 ICONIP2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Katsuyuki Hagiwara
2. 発表標題 On a Fitting of a Heaviside Function by Deep ReLU Neural Networks
3. 学会等名 ICONIP2018 (国際学会)
4. 発表年 2018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