

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 5 月 24 日現在

機関番号：14101

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21500215

研究課題名（和文）多層パーセプトロンのモデル選択に関する研究

研究課題名（英文）Research on model selection of multi-layer perceptron

研究代表者

萩原 克幸（HAGIWARA KATSUYUKI）

三重大学・教育学部・准教授

研究者番号：60273348

研究成果の概要（和文）：本研究では、多層パーセプトロンのモデル選択を考える上で、基底関数の可変性に着目し、データ数と同じ個数の関数集合の中を基底関数が動く場合のモデル選択の方法を確立した。その方法では、まず、集合内の関数に対して直交変換を行い、直交化された基底関数に対して、その係数を推定する。その係数に対する閾値処理により、不必要な係数を零にする。その後、逆変換により、元の基底関数の係数を求める。この際、閾値は、基底関数の可変性の性質を反映して導出しており、理論的に妥当なものである。この方法では、直交化の方法に応じて、縮小推定となる場合やスパースな表現が得られる。この方法は、ある学習法の下で、多層パーセプトロンに対しても応用可能である。

研究成果の概要（英文）：In this research, we focus on a variability of basis functions in multi-layer perceptron and establish a model selection method for the case where basis functions are selected from a finite set of functions, in which the number of candidates is equal to the number of data. In the proposed method, functions in the set are orthogonalized and their coefficients are estimated. Inefficient coefficients are set to zero by thresholding. Then inverse transform yields the weights of original basis functions. The threshold level in this method is theoretically reasonable since it derived based on the variability of basis functions. We can obtain a smooth output and/or sparse representation depending on the method of orthogonalization. The method is applicable to multi-layer perceptron under a certain training algorithm.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,200,000	360,000	1,560,000
2010年度	900,000	270,000	1,170,000
2011年度	600,000	180,000	780,000
年度			
年度			
総計	2,700,000	810,000	3,510,000

研究分野：機械学習・統計科学

科研費の分科・細目：情報学・感性情報学・ソフトコンピューティング

キーワード：多層パーセプトロン・特異モデル・モデル選択・可変基底・ノンパラメトリック回帰

1. 研究開始当初の背景

多層パーセプトロンは、予測・パターン識別の問題に対して有効であることから、工学・

医療・経済など様々な分野での応用が試みられている。多層パーセプトロンの応用においては、未学習のデータに対して誤差を小さくするという、汎化性が求められる。このための一つの方法として、多層パーセプトロンのサイズを決めるモデル選択の問題がある。統計学では、モデル選択のための規準が古くから提案されている（例えば、Akaike(1972)）。しかしながら、多層パーセプトロンの場合、同じ出力を与える重みが一意に決まらない場合があることから、従来のモデル選択規準の理論的妥当性が失われることが知られている(Hagiwara et. al. (1993))。こうしたモデルは、近年、特異モデルと呼ばれており、実は、機械学習の分野において重要なモデルの多くは特異モデルであることが知られている。こうした問題点があるため、応用上のモデル選択は扱いが難しく、これまでに発見的な方法のみが提案され、理論的根拠をもつ方法は見られない。一方で、多層パーセプトロンの汎化誤差などの理論解析は、統計的学習理論の枠組みで古くからおこなわれているが（例えば Haussler(1992)など）、実用的なモデル選択の方法には至っていない。これに対して、申請者は、多層パーセプトロンを可変な基底関数をもつモデルとして捉えることにより、その学習誤差・汎化誤差の解析が、ある確率変数列の最大値の評価に帰着されることを示しており、学習誤差の上界を導いている(Hagiwara and Fukumizu(2008))。しかしながら、そうした理論研究に基づき、具体的なモデル選択の方法を与えることは未だにできていない。

2. 研究の目的

本研究では、多層パーセプトロンの統計的性質を調べ、理論的根拠のあるモデル選択方法を確立することを目的とする。

3. 研究の方法

これまでの研究で、可変基底の観点から、学習誤差などの評価が、ある確率変数列の最大値の評価の問題に帰着されることが分かっている。この観点から、申請者は、ノイズに対するオーバーフィッティングの度合いの上界のオーダーを与えているが、これに基づく具体的なモデル選択規準の構成には至っていない。これは、多層パーセプトロンの場合、確率変数の相関構造が複雑であることに起因する。一方で、関数の有限集合から基底関数を選ぶ場合で、特に、その関数達が直交性を満たす場合には、独立な正規確率変数列の最大値の評価の問題に帰着され、この場合については、学習誤差・汎化誤差の実用的な評価がなされており、そのモデル選択の

方法も考えられている(Hagiwara(2006))。そこで、本研究では、直交化された基底関数に対する結果を一般的な場合に適用することを考える。

4. 研究成果

上記の目的を達成するために、本研究では、以下のようなモデル選択の方法を考えた。まず、データ数と同じ数の線形独立な関数の有限集合を与えられているものとする。これは、基底関数の候補となる集合であり、基底関数が有限集合の中を動くという意味で可変基底の特別な場合と見なせる。本方法では、その集合の中の関数を直交化し、その直交化された基底関数の係数を推定する。この場合、直交化された基底関数の数はデータ数と等しい。係数の値が予め決められた閾値より小さい場合、その値を零とし、それ以上の場合、係数の値はそのままにしておく。すなわち、直交化された基底関数の上でスパースな表現を得る。こうして得られた係数を逆変換することで、元の関数の重みが得られる。この方法は、関数形に関して一般的な方法であり、統計学的観点からは、ノンパラメトリック回帰について、予測誤差の小さいモデルを構成する方法として位置づけることができる。本研究の重要な点は、基底関数を直交化することにより、その係数がデータ中の雑音のみに影響されている場合の確率的上界・下界が理論的に明らかにされた点である。ここで導出された閾値は、係数の二乗に対して、 $2S \log(n)$ で与えられる。ここで、 S はノイズ分散、 n はデータ数である。これは、係数がノイズのみに影響される不必要なものである場合の確率的上界であるとともに、ほぼ、確率的下界に等しい量である点が重要である。この意味で、ここで導出した閾値はタイトであり、理論的な妥当性をもつ。従来のモデル選択規準では、係数の二乗を定数で評価するのと等価であるのに対して、可変性をもつ基底の場合には、データ数の増加関数、特に、 $\log(n)$ のオーダーの量で評価する必要がある点は重要である。

ここで得られた方法は、直交化の方法として固有値分解を用いた場合、集合の中の関数をすべて基底関数として用いるが、その係数については縮小推定量を与えることになる。この方法は、想定した集合内の関数に対する主成分分析を行う方法と等価である。一方、直交化の方法としてGram-Schmidtの直交化を用いた場合には、元の基底関数の上でもスパースな表現を与える。すなわち、集合の中の関数で、ノイズのみに影響される不必要な基底関数を除いた表現が得られる。特に、スパースな表現が得られる点は、多層パーセプトロンの場合の隠れ素子数の決定に応用で

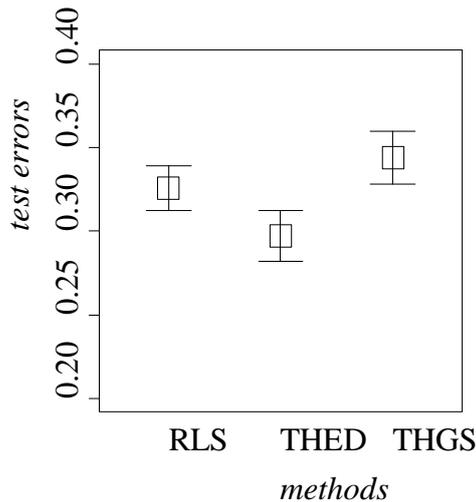


図1 テストデータに対する平均誤差

きる。また、この方法は、逐次的に直交化できる点で、計算量が大幅に削減できる。

本方法をいくつかのベンチマークデータに対して、適用した結果を述べる。データとしては、<http://www.liaad.up.pt/~ltorgo> より取得した Ailerons・Delta_ailerons・Elevators・Delta_elevators・Kin8nm・Puma8NH・Puma32H・Auto の8つのデータセットを用いた。これらのデータセットはいずれも多次元入力の特徴データである。適用した方法は、二乗誤差和に重みの二乗ノルムを付加して構成されたコスト関数を最小化する正則化法(Regularized Least Squares(RLS))、固有値分解を用いた提案法(THED)および Gram-Schmidt 法を用いた提案法(THGS)である。RLS については、正則化パラメータを leave one out cross validation により選んだ。ここで、提案法における閾値の計算に必要なノイズ分散の推定値としては、Carter and Eagleson(1992)により提案されているものを用いた。各データセットを学習データとテストデータに分割し、学習データに対して各方法を用いて推定したモデルをテストデータに対して評価する。これらを50回行い、テストデータに対する誤差(テスト誤差)の平均を計算する。図1は、すべてのデータセットに対するテスト誤差の平均とその標準偏差の平均を合わせて示してある。標準誤差を含めて考えれば、汎化性の点で、RLS に対して、THED は同等あるいは優れており、THGS は同等あるいはやや劣る結果となった。

一方、表1に、データセットごとのデータ数および、HTGSにより得られた基底関数の数(50回の試行における平均)を示す。準備した基底関数の数はデータ数と同じである。表1より、HTGSにより得られた基底関数の数は非常に少なく、スパースな表現が得られている。このとき、RLS および THED については、

推定されたモデルの基底関数の数はデータ数に等しいことに注意する。また、図1より、テスト誤差の観点からは、RLS と同等あるいはやや劣るものであり、スパースかつある程度汎化性の良いモデルが推定されていると言える。重要な点は、近年よく研究されている support vector machine(SVM)やここで述べた RLS など標準的な手法は、調整すべき学習パラメータを含むのに対して、ここで提案した方法は、そうしたパラメータを含まない点にある。既存手法においては、そうした学習パラメータがモデルの汎化性やスパース性に大きく影響するため、その設定の方法自体が問題となるが、その系統的な方法は得られていないのが現状である。このため、学習パラメータの設定には、理論解析を必要としないが時間のかかるクロス・バリデーションなどの方法を用いる場合が多い。本研究で提案する方法は、学習パラメータを必要とせず、理論的に妥当な方法であるという点で、高速で実用的な方法である。

表1 HTGSにおけるスパース性

データ	データ数	基底数
Ailerons	4000	10.54
Delta_ailerons	4000	53.20
Elevators	4000	101.42
Delta_Elevators	4000	42.00
Kin8nm	4000	612.28
Puma8NH	3000	254.82
Puma32H	3000	31.96
auto	300	24.52

一方で、多層パーセプトロンの効率的な学習方法として Extreme Learning Machine(ELM)が提案されている(Huang, Zhu and Siew(2006))。これは、ランダムに生成した入力重みで決まる隠れ素子から適当な方法で選んだものをネットワークに組み込んでいく方法である。隠れ素子の選択方法としては、残差を小さくするように中間素子を選ぶ方法などが提案されている(Lan, Soh and Huang(2010))。多層パーセプトロンの場合の隠れ素子は基底関数に対応するため、ランダムに選ばれた隠れ素子の集まりに対して、本研究で得た HTGS を適用することで、多層パーセプトロンのモデル選択規準としても有効になるものと考えられる。さらには、HTGS は、Gram-Schmidt 法を用いた学習法である Orthogonal Least Squares(OLS)にも適用できる点で今後様々な拡張が期待できる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕（計 2 件）

① K.Hagiwara, “Nonparametric regression method based on orthogonalization and thresholding”, IEICE Trans. INF. & SYST., E94-D, 1610-1619, 2011. （査読有）

② K.Hagiwara, “Orthogonalization and thresholding method for a nonparametric regression problem”, Advances in Neuro-Information Processing, Lecture Notes in Computer Science, 5507, 187-194, Springer-Verlag, Berlin Heidelberg, 2009. （査読有）

〔学会発表〕（計 2 件）

① 萩原 克幸, “Orthogonalization and thresholding method for a nonparameteric regression problem”, 電子情報通信学会技術研究報告, IBISML-2010-8, 41-48, 6月14日-15日, 2010.

② K.Hagiwara, “On a training scheme based on orthogonalization and thresholding for a nonparametric regression problem”, Proc. of IJCNN2010, Barcelona, Spain, 18-23, July, 3131-3138, 2010.

〔その他〕

ホームページ等

<http://sunflower.edu.mie-u.ac.jp/~hagi>

6. 研究組織

(1) 研究代表者

萩原 克幸 (HAGIWARA KATSUYUKI)
三重大学・教育学部・准教授
研究者番号：60273348

(2) 研究分担者

なし

(3) 連携研究者

なし