J. Plasma Fusion Res. Vol.82, No.5 (2006) 282-286

総合解説 ニューラルネットワークの応用と今後の発展

2．ニューラルネットワークの基礎と理論的に重要な課題

萩原克幸
三重大学　教育学部

（原稿受付：2006年1月25日）

階層型ニューラルネットワークは，与えられた入出力データに基づき，そこに内在する規則性を学習により獲得することを目的とした数理モデルであり，画像・音声の処理・認識，制御，時系列解析・予測からデータ解析まで，その応用は多岐にわたる．本稿は，階層型ニューラルネットワークに関する基礎を関数近似・学習・汎化性の観点からまとめるとともに，最近の理論的な課題について特異モデルの観点から述べたものである．

Keywords:
layered neural network, back propagation, generalization, early stopping, regularization, singular model

2.1　はじめに

ニューラルネットワークとは，脳神経系の情報処理機構を模倣した数理モデルであり，与えられたデータに基づく学習を通して，必要とされる情報処理を実現するものである．本稿では，ニューラルネットワークの中でも，パターン分類，規則性の抽出，時系列解析・予測，データ解析など広範な応用分野をもつことから，階層型ニューラルネットワークである多層パーセプトロンに焦点を当てる．多層パーセプトロンの研究領域は，主に，関数近似・学習・汎化性に分類される．本稿では，これらの観点から，多層パーセプトロンの基礎をまとめるとともに，多層パーセプトロンに関する最近の理論研究について，簡単に触れる．以下，本稿では，多層パーセプトロンを階層型ニューラルネットワークあるいは単にニューラルネットワークと呼ぶことにする．

2.2　ニューラルネットワーク

D 入力 M 出力のニューラルネットワークは，Fig. 1に示すように，ユニットと呼ばれる演算素子を層状に結合した階層構造として表される．ユニットは神経細胞のモデルであり，ニューラルネットワークは神経細胞が結合された脳神経系のモデルとなっている．入力側の D 個のユニットを配置した層を入力層，出力側の M 個のユニットを配置した層を出力層といい，その間の層を隠れ層あるいは中間層という．ここでは，入力層を第1層と定義し，出力層を第 L 層とする．第 l 層に配置されているユニットの数を K_l と書く．

各素子間の結合は重み付けされており，この重みを結合重みという．ここでは，第 $l-1$ 層 j 番目のユニットから第 l 層 k 番目のユニットへの結合重みを $w_{l-1,j,k}$ と表す．ただし，$2<l\leq L$，$1\leq j\leq K_{l-1}$，$1\leq k\leq K_l$ である．いま，入力層に入力 $\vec{x}=(x_1,\dots,x_D)\in\boldsymbol{R}^D$ が提示されたときの第 l 層 k 番目のユニットの出力を $f_{l,k}(\vec{x})$ と表すこととすれば，第 $l-1$ 層から第 l 層 k 番目のユニットに対する入力は，

$$s_{l,k}(\vec{x})=\sum_{j=1}^{K_{l-1}} w_{l-1,j,k}f_{l-1,j}(\vec{x}) \tag{1}$$

と表される．各ユニットは，活性化関数と呼ばれる一変数関数 ϕ を備えており，第 l 層 k 番目のユニットの出力は

$$f_{l,k}(\vec{x})=\phi(s_{l,k}(\vec{x})) \tag{2}$$

で与えられる．ただし，入力層 k 番目のユニットは，x_k を入力として受け取り，x_k を出力するものとする．この演算を入力層から順次出力層まで行うことにより，出力 $f_m(\vec{x})=f_{L,m}(\vec{x})$ が得られる．一般には，各層に常に1を出力する素子を配置する．これは，各層において $f_{l-1,0}(\vec{x})=-1$ と定義し，これに対する結合重み $w_{l-1,0,k}$ を用意して，式(1)の和を $j\in[0,K_{l-1}]$ についてとれば良い．特に，$w_{l-1,0,k}$ をしきい値という．

活性化関数としては，隠れ層において，一般に，双曲線正接関数 $\tanh(u)$，ロジスティック・シグモイド関数 $(1+\exp(-u))^{-1}$ などの非線形関数が用いられる．出力層においては，目的に応じて，パターン認識の場合には $(0,1)$ に値をとるロジスティック・シグモイド関数，関数近似・統計的回帰の場合には $\boldsymbol{R}$ に値をとる線形関数が用

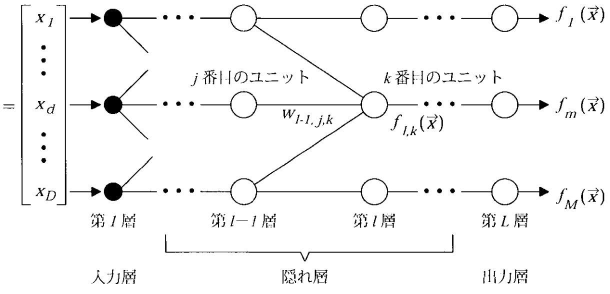

Fig. 1　階層型ニューラルネットワーク．

author's e-mail: hagi@edu.mie-u.ac.jp

いられる．

以上の構成から，ニューラルネットワークにおいて，結合重みの値を適当に変えればそれに応じて出力が変ること，また，層数・素子数を増やせばその入出力関係を複雑にできることが容易に想像できる．このことは，ニューラルネットワークが，結合の重みを適当に調整することで，対象とする現象あるいはシステムの入出力関係を近似的に実現できる可能性があることを意味する．ニューラルネットワークは，脳神経系の機能を模倣した情報処理装置として注目されたが，本質的には，パラメトリックな関数とみなせばよい．

2.3 ニューラルネットワークの基礎

2.3.1 関数近似能力

対象とする入出力関係をニューラルネットワークにより獲得させようとする場合，まず，その表現能力が問題となる．ここでは，非線形関数を活性化関数とする隠れ層を一つとし，出力層の活性化関数を線形関数とした D 入力1出力の3層構造を考える．入力 $\vec{x}\in\boldsymbol{R}^D$ に対する出力は，$\bar{x}=(-1,\vec{x})$ と定義すると，前節の議論を簡略化して，

$$f_{\vec{w}}(\vec{x})=\sum_{k=1}^{K} v_k\phi(\vec{w}_k\cdot\bar{x}) \tag{3}$$

と書ける．ただし，$\vec{w}_k=(w_{k,0},\ldots,w_{k,D})$ として，$\vec{w}$ は $\{(\vec{w}_k,v_k):1\le k\le K\}$ を並べた $\boldsymbol{R}^{K(D+2)}$ 次元ベクトルとする．また，・は内積を表すものとする．3層階層型は，最も単純な構造であるにもかかわらず，隠れ層の素子数を増やせば，ϕ に関する緩い条件の下で，有界閉集合上の任意の連続関数を任意の精度で一様に近似できることが示されている（例えば[1]）．ただし，例えば，$\vec{v}=(v_1,\ldots,v_K)\in\boldsymbol{R}^K$ とし，k のみに依存する関数を φ_k として，

$$g_{\vec{v}}(\vec{x})=\sum_{k=1}^{K} v_k\varphi_k(\vec{x}) \tag{4}$$

による関数近似を考えると，これは，古典的な巾多項式やフーリエ級数による展開であり，K を増やせばいくらでも有界閉集合上の連続関数を近似できるだろう．関数の表現(3)と(4)を基底関数の線形結合として見るときの決定的な違いは，前者がパラメータ $\vec{w}_k$ により基底関数の形状を変化させることができるのに対して，後者では基底関数 φ_k が固定されている点にある．実は，固定基底の場合には，入力次元 D が大きいとき，ある程度の近似精度を保証するために必要な基底関数の数が莫大になる．これを次元の呪いという．3層階層型の場合には，可変基底の性質により，この次元の呪いを回避できることが知られている（例えば[2]）．3層構造は，単に単純で取り扱いやすいというだけでなく，こうした保証があることから，応用上最もよく用いられる．

2.3.2 学習

前述した関数近似の結果は，目的の関数を近似するための具体的な結合重みの値を教えてくれない．したがって，次に問題となるのは，対象の入出力関係を近似的に実現するニューラルネットワークをいかにして構成するか，すなわち，その結合重みをいかにして決めるかである．多くの場合，そうした入出力関係は有限個のデータ $\{(\vec{x}_i,\vec{y}_i):\vec{x}_i=(x_{i1},\ldots,x_{iD})\in\boldsymbol{R}^D,\ \vec{y}_i=(y_{i1},\ldots,y_{iM})\in\boldsymbol{R}^M,\ 1\le i\le n\}$ として与えられる．これを学習データという．特に，対象となる $\vec{y}_i$ を教師データということがある．前述したように，結合重みの値を変えれば出力が様々に変化するから，結合重みをうまく調整すれば，与えられた教師データに出力を近づけることができるだろう．この過程を一般に学習という．これは，統計学における，パラメータ推定に対応する．

各 $\vec{y}_i$ と出力 $(f_{L,1}(\vec{x}_i),\ldots,f_{L,M}(\vec{x}_i))$ の間の近さの尺度として2乗誤差関数：

$$\ell(\vec{x}_i,\vec{y}_i,\vec{w})=\frac{1}{2}\sum_{j=1}^{m}(y_{ij}-f_{L,j}(\vec{x}_i))^2 \tag{5}$$

を考える．ここで，$\vec{w}$ は結合重みおよびしきい値をすべて並べたベクトルとする．すべての例題に対する2乗誤差の総和を $\ell(\vec{w})=\sum_{i=1}^{n}\ell(\vec{x}_i,\vec{y}_i,\vec{w})$ と書き，これを学習誤差と呼ぶことにする．学習とは，具体的には，学習誤差 $\ell(\vec{w})$ を小さくする結合重みを求める過程である．一般に，学習誤差の大域的最小値を与える結合重みを求めることは困難であり，通常は，反復法により局所最小値を与える結合重みを探索することになる．

ニューラルネットワークを含む一般の学習機械に対する学習の方法として最も基本的な方法は，$\vec{w}(0)$ を適当に決め，$t+1$ 回目の繰り返しにおいて，

$$\vec{w}(t+1)=\vec{w}(t)-\eta\Delta\vec{w}(t) \tag{6}$$

$$\Delta\vec{w}(t)=\left.\frac{\partial\ell(\vec{w})}{\partial\vec{w}}\right|_{\vec{w}=\vec{w}(t)} \tag{7}$$

により結合重みを修正する方法であり，この方法を最急降下法という．ここで，η は，学習率と呼ばれる小さな正定数である．ニューラルネットワークの場合，その階層構造より，式(7)の各成分は，

$$\Delta w_{l-1,j,k}(t)=\sum_{i=1}^{n}\delta_{i,l,k}(t)f_{l-1,j}(\vec{x}_i) \tag{8}$$

と書ける．ただし，

$$\delta_{i,l,k}(t)=\phi'(s_{l,k}(\vec{x}_i))\alpha_{i,l,k} \tag{9}$$

$$\alpha_{i,l,k}=\begin{cases}-(y_{i,k}-f_{L,k}(\vec{x}_i)) & l=L\\ \sum_{\nu=1}^{K_{l+1}}\delta_{i,l+1,\nu}w_{l,k,\nu} & l<L\end{cases} \tag{10}$$

であり，$\delta_{i,l,k}(t)$ を誤差信号という．上式は，結合重みの修正量が出力層から伝播してくる誤差信号に基づき，出力層から逆方向に順次計算できることを意味しており，この学習則を誤差逆伝播法（error backpropagation）という[3]．

結合重みの空間上での$\ell(\bar{w})$の形状すなわち誤差曲面は，一般に複雑であるため，最急降下学習，したがって，誤差逆伝播法の収束は遅いことが指摘されている．このため，例えば，結合重みの不必要な振動を抑えるために前回の更新方向を現在の更新方向に慣性項として加える方法[3]を始め，結合重みの履歴に関する平滑化微分と現在の更新の方向の比較による学習率の増減法[4]，勾配の増減に依存した慣性係数の増減法[5]，など数多くの発見的な方法が提案されてきた．こうした発見的な方法の有効性は経験的に示されているが，各方法において導入されている学習のためのパラメータの系統的な設定規準がないという問題が残されている．一方，学習の問題は，非線形最適化の問題であり，そこではニュートン法・準ニュートン法・共役勾配法など最急降下法より収束の速いアルゴリズムが多数存在する．そうしたアルゴリズムあるいはそれを改良したものをニューラルネットワークに適用した研究も多く見られる(例えば[6])．こうした学習の加速化の方法については，例えば[7]を参照してほしい．

2.3.3 汎化性

我々は，学習後のニューラルネットワークに対して，データに内在する規則性を獲得していること，あるいは，同じ生成構造をもつ未学習のデータに対して当てはまりが良いことを期待する．この性質がなければ，パターン認識において未知パターンを適切に分類できない，予測・解析結果が信頼できないなどの問題が生じる．一般に，データに潜む規則性を的確に獲得していれば未学習データに対してはある程度当てはまりが良いし，その逆も言える．未学習データに対してニューラルネットワークの当てはまりが良いとき，それは汎化性が良いと言われる．応用においては，汎化性の良いニューラルネットワークをいかにして構成するかが問題となる．汎化性は汎化誤差により測られる．汎化誤差は，学習後のニューラルネットワークが未学習データに対して与える誤差の期待値として定義される．すなわち，未学習データの色々な出現の仕方に関する平均的な誤差が汎化誤差である．

ここでは，直感的な理解のために，Fig. 2 に示すように，何らかの決定的な関数の出力(薄線)に雑音が重畳してデータ（丸印）が観測される場合，すなわち，統計的回帰の問題を考える．このとき，我々の目的は，データの背後にある決定的な関数の出力をニューラルネットワークの出力(Fig. 2 実線）により獲得することにある．この決定的な関数を真の関数と呼ぶことにする．ニューラルネットワークとして3層構造を仮定したとき，隠れ素子数が少ないと，真の関数を表現できないため，学習後のニューラルネットワークの出力は真の関数の出力から偏る(Fig. 2(a))．一方，前述したように，隠れ素子数を増やすと，より複雑な関数を表現できる．したがって，多くの隠れ素子を用いて学習を行えば，2乗誤差を小さくするような学習においては，与えられた学習データを忠実に表現する出力が獲得される(Fig. 2(c))．しかしながら，データには雑音成分が混入していることを考えると，学習後の出力は，真の関数の出力からは隔たったものとなる．これは，いわゆる，オーバーフィッティングの状態である．すなわち，隠れ素子数が過度に少ない場合や過度に多い場合は，学習後の出力が真の関数から隔たり，このとき，汎化誤差は大きくなる．このことは，適切な隠れ素子数を決める必要があることを意味している（Fig. 2(b)）．一般に，この問題は，統計的モデル選択と呼ばれており，そのための規準として，例えば，AIC（Akaike Information Criterion）[8]などが知られている．しかしながら，後述するように，ニューラルネットワークに対してAICを直接適用しても，適切なモデル選択を行えないことが知られており[9]，現在のところ，ニューラルネットワークに対して理論的な妥当性をもつ実用的なモデル選択規準は与えられていない．

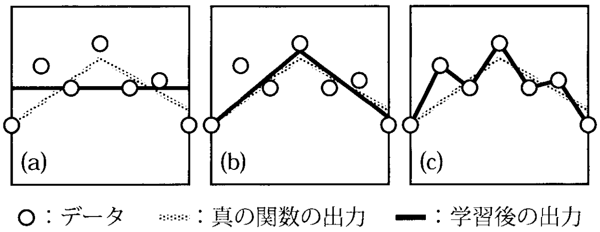

Fig. 2　統計的回帰の場合における汎化性の説明．

応用上は，与えられたデータを学習用と評価用に分割し，学習用データに対して学習したニューラルネットワークの評価用データに対する誤差を計算し，それをサイズの異なるニューラルネットワーク間で比較する方法が用いられる．この方法は，統計学で知られている leave-one-out cross validation を簡略化した方法で，2-fold cross validation という[10]．この場合の評価用データは未学習データであるから，汎化誤差を近似的に比較しているという意味で，ある程度妥当な選択が可能である．ただし，公平なモデル選択のためには，評価用データ数を増やす必要があるが，これにより学習用データ数が減るため，学習の精度が落ちる．したがって，与えられた全データ数の下で，分割の割合を適切に決める必要がある．また，学習の段階で，素子および結合をダイナミックに追加・削除する発見的な方法も多数提案されている．詳細については，文献[11]を参照してほしい．

一方，見方を変えると，モデル選択の問題は，2乗誤差のようなデータとの近さしか見ない規準に関して学習を行うことに起因していると考えることができる．そこで，その規準に関して学習し過ぎないように適当なところで学習を止める方法や学習の規準自体を修正する方法が考えられる．前者をearly stoppingといい，後者としては正則化法がある．

一般に，比較的大きいサイズのニューラルネットワークの学習では，学習が進行するにつれて学習誤差は減少するが，汎化誤差はある時点から増加する場合がある．この現象を過学習という．もし，過学習が生じるならば，汎化誤差が増加し始める時点で学習を停止させることにより，ある程度汎化性の良いニューラルネットワークが構成できる．これを early stopping という．early stopping の方法としては，2-fold cross validation を用いる方法が一般的であ

る．すなわち，学習の各段階において，評価用データに対する誤差をモニターしながら，それが増加し始めた時点で学習を止める方法である．

一方，$r(\vec{w})$ をオーバーフィッティングに対するペナルティを与える関数として，

$$c(\vec{w})=\ell(\vec{w})+\lambda r(\vec{w}) \tag{11}$$

に関して学習を行う方法を正則化法といい，$r(\vec{w})$ を正則化項という．$\lambda>0$ は，学習誤差 $\ell(\vec{w})$ とペナルティである正則化項 $r(\vec{w})$ のバランスを決めるパラメータであり，正則化パラメータと呼ばれている．正則化法は，大きめのサイズの下での学習において，比較的簡単に汎化性の良いニューラルネットワークを構成できることから，応用においてもよく用いられる．一般に，オーバーフィッティングの状態にあるニューラルネットワークの結合重みは大きな値となっていることから，正則化項としては，結合重みの値に対するペナルティを採用するのが一般的である．正則化項の種類としては，結合重みベクトルを $\vec{w}=(w_1,\ldots,w_m)$ と書くと，代表的なものとして，$r_1(\vec{w})=\sum_{j=1}^{m}|w_j|$[12]や $r_2(\vec{w})=\sum_{j=1}^{m}|w_j|^2$[13]などが提案されている．$r_2$ は単純に結合重みの値を制限するのに対して[7]，r_1 は正則化パラメータの値により決まるレベル以下の値の結合重みを削除してニューラルネットワークを構造化する効果がある[12]．正則化法において応用上重要なのは，正則化パラメータをいかにして決めるかという問題であるが，これに対する一つの解決策としては，これまでに見た2-fold cross-validation を用いる方法が考えられる．

2.4 理論的に重要な課題

以上，学習の高速化・汎化性の向上を目的とした応用上の重要な方法について述べてきた．これらの方法は，発見的ではあるが，経験的にその有効性が示されている．しかしながら，ニューラルネットワークの性質はそれほど明確にされているわけではなく，そうした方法に対する理論的なサポートは極めて少ない．以下では，ニューラルネットワークの性質に関する理論解析の最近の話題について簡単に触れる．

今，式(3)において $K=1$ の場合を $f_1(\vec{x})=v_1\phi(\vec{w}_1\cdot\vec{x})$，$K=2$ の場合を $f_2(\vec{x})=v_{2,1}\phi(\vec{w}_{2,1}\cdot\vec{x})+v_{2,2}\phi(\vec{w}_{2,2}\cdot\vec{x})$ と書き，f_1 を f_2 で表現することを考える．その表現の仕方には，(1)$v_{2,2}=0$ として $\vec{w}_{2,1}=\vec{w}_1$，(2)$v_{2,1}=0$ として $\vec{w}_{2,2}=\vec{w}_1$，(3)$\vec{w}_{2,1}=\vec{w}_{2,2}=\vec{w}_1$ として $v_{2,1}+v_{2,2}=v_1$ がある．このとき，(1)の場合 $\vec{w}_{2,2}$，(2)の場合 $\vec{w}_{2,1}$，(3)の場合 $v_{2,1}$，$v_{2,2}$ に任意性が残る．すなわち，f_1 を f_2 により表現する場合の結合重みは一意ではない[14,9]．このことは，$K>2$ の場合，常に，それより小さいサイズの表現に対応する結合重みの集合があって，その上では誤差は同じ値となることを意味しており，これは，与えられた学習データによらない性質である．この性質は，関数近似のところで述べた基底の可変性によりもたらされるものであり，学習および汎化性の理論考察を行う上で極めて重要である．

文献[15]において，同じ出力を与える結合重みの集合が

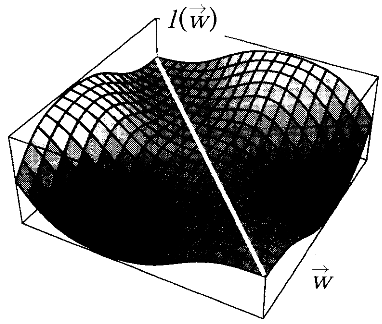

Fig. 3　誤差が等しい直線状の領域付近の誤差曲面の概念図．

直線状の領域を形成しており，その領域には極小点となる部分と鞍点となる部分があるということが示されている．結合重み $\vec{w}$ の空間上のこの付近での誤差曲面 $\ell(\vec{w})$ の概念図を Fig. 3 に示した．図中の白線が誤差が等しい直線状の領域である．この直線状の領域付近での勾配は極めて小さいため，学習の停滞が生じる．最急降下学習において，学習の長い停滞の後に学習誤差が減少する場合があるが，これは，極小点となる部分において直線状の領域に近づくことで誤差が減少するが，勾配の小さい直線状の領域において学習の停滞が生じ，鞍点部分において，再び誤差が減少する場合として説明されるだろう．また，学習の停滞が生じる直線状の領域は，谷構造と見ることもでき，そうした谷領域の存在は古くから経験的に知られていた[16]．したがって，谷構造を想定した発見的な加速化法[17]や比較的条件数の悪い二次形式の形をした誤差曲面に有効な準ニュートン法などは，ニューラルネットワークの学習法として有効であると考えられる．

一方，一般的なパラメトリック・モデルの統計的推定において，データ数が十分大きい場合，適当な仮定の下で，パラメータの推定量が真値の周りで正規分布に従うことが示されており，これを漸近正規性という[18]．こうした漸近論を用いた解析は，統計的パラメトリック・モデルを扱う際の常套手段である．実際，汎化性の議論で触れたモデル選択規準 AIC は，漸近正規性に基づき導かれている[8]．しかしながら，ニューラルネットワークの場合，前述したように，同じ出力を与える結合重みが一意性をもたない場合があることから，学習後の結合重みがある真値の周りに集中するような状況が生じない場合がある．特に，オーバーフィッティングが生じるような状況を理論的に取り扱う場合には，このことが本質的となる[9]．すなわち，統計学における漸近論の議論をニューラルネットワークに直接適用することはできない．最近の研究により，この場合の学習誤差・汎化誤差などの振る舞いが通常の統計的漸近論が成り立つ状況とは全く異なるものとなることが示されている[9,19]．例えば，最小2乗規範の下では，雑音成分に対する学習誤差が統計的漸近論が成り立つ状況より小さくなる[9]．関数近似のところで，ニューラルネットワークは基底の可変性によって効率的に関数近似を行えることを述べたが，上記の結果も基底の可変性によってもたらされるものである．パラメータ値と出力が一対一に対応しないこうしたモデルは特異モデルとして位置付けられており，現

在，研究が進展しつつある．

2.5 おわりに

本稿では，ニューラルネットワークの基礎を関数近似・学習・汎化性の観点からまとめるとともに，特異モデルの観点からの最近の理論研究について簡単に触れた．ニューラルネットワークについては，応用上の学習・汎化性の問題を解決するための発見的な方法が数多く提案され，その有効性が経験的に示されており，実に様々な分野で応用が試みられている．しかしながら，それらをサポートする理論解析は極めて少ないため，今後の理論研究の発展が望まれる．

謝辞

本稿の執筆の機会を与えていただいた電気通信大学　竹田辰興教授に感謝いたします．

参考文献

[1] K. Funahashi, Neural Netw. **2**, 183 (1989).
[2] A.R. Barron, IEEE Trans. Inf. Theory **39**, 930 (1993).
[3] D.E. Rumelhart *et al.*, *Parallel Distributed Processing* **1** (MIT Press. 1986).
[4] R.A. Jacobs, Neural Netw. **1**, 295 (1988).
[5] S.E. Fahlman, *Proc. of the 1988 Connectionist Models Summer School* (1988) p.38.
[6] R. Battiti, Neural Comput. **4**, 141 (1992).
[7] C.M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, New York, 1995).
[8] H. Akaike, *2nd InternationalSymposium on Information Theory*, B.N. Petrov and F. Csaki eds., (Budapest, Akademiai Kiado, 1973) p.267.
[9] K. Hagiwara *et al.*, Neural Netw. **14**, 1419 (2001).
[10] P. Burman, Biometrika **76**, 503 (1989).
[11] R. Reed, IEEE Trans. Neural Netw. **4**, 740 (1993).
[12] P.M. Williams, Neural Comput. **7**, 117 (1995).
[13] D. Plaut *et al.*, **CMU-CS-86-126**, Carnegie-Melon, Department of ComputerScience, Pittsburgh, PA (1986).
[14] H.J. Sussmann, Neural Netw. **5**, 589 (1992).
[15] K. Fukumizu and S. Amari, Neural Netw. **13**, 317 (2000).
[16] R.S. Sutton, *Proc. of Eightth Annual Conference of the Cognitive Science Society* (1986) p.823.
[17] K. Ochiai *et al.*, Neural Netw. **7**, 797 (1994).
[18] A.W. van der Varrt, *Asysmptotic statistics* (Cambridge University Press, 1998).
[19] S. Watanabe, Neural Comput. **13**, 899 (2001).

はぎ わら かつ ゆき
萩 原 克 幸

1995年豊橋技術科学大学大学院工学研究科システム情報工学専攻修了．同年三重大学工学部助手，2003年名古屋工業大学助教授を経て，2004年より三重大学教育学部助教授．博士(工学)．統計的学習理論，モデル選択・縮小推定の研究に従事．特に，多層パーセプトロンなど，Fisher 情報行列が縮退する特異モデルの統計的性質を調べている．