

論文要旨

射影追跡回帰は、高次元データを低次元空間に射影することによって低次元空間における特徴的な非線形写像を発見しようとするものである。その射影方向はランダムな探索と勾配法により求められているが、本研究ではRBFネットワークとファジィ集合の考え方を採用することにより、ランダムな探索によらずに大域的な最小値を発見しやすい固有値法と勾配法の併用法を提案する。提案法は、データの応答変量を大小二つのファジィクラスに分割し、それぞれのファジィクラスのデータが射影された低次元空間においてデータの下にある関数の形状がある程度現れるような射影方向を最初に求めるものである。固有値法により得られた結果からヒトのパターン認識能力を活用して近似関数であるRBFネットワークの学習におけるパラメータの初期値を適切に決めることができる。提案法の有効性を数値実験により示すとともに、FisherのIrisデータに適用し、判別分析にも応用できることを示す。

ニューロ・ファジィ射影追跡による回帰と判別*

三好 哲也・市橋 秀友・長坂 一徳・中尾 圭介**

Regression and Classification by Neuro Fuzzy Projection Pursuit*

Tetsuya MIYOSHI, Hidetomo ICHIHASHI, Kazunori NAGASAKA, Keisuke NAKAO**

The important feature of the projection pursuit(PP) is that it is one of the multivariate methods able to bypass the “curse of dimensionality”. The aim of PP is to find an interesting or characteristic structure by working in low-dimensional linear projections. PP for regression was originally proposed by Friedman and Stuetzle. In this paper, a neuro-fuzzy approach to the projection pursuit regression (PPR) is proposed for nonparametric regression and nonparametric classification. Our proposed method is based on the membership function and the eigenvector of the covariance matrix to avoid being trapped by a local minimum of the projection indices. The dimensionality of predictors is reduced to one or two in order for good use of human ability of instantaneous pattern discovery. The radial basis function neural network is applied to the function approximation in a projected low-dimensional space. The projection direction is also changed by the adaptive learning (steepest descent) method.

1. はじめに

近年, ニューラルネットワークを用いた非線形回帰手法が広く用いられており, RBF(Radial Basis Function) ネットワーク¹⁾⁻⁴⁾による非線形回帰もその一種である. RBF ネットワークは中間層にガウス関数などのラジアル基底関数を用いた3層ニューラルネットワークの一種であり, 各ニューロンからの出力の重み付き線形和が出力となる. RBF ネットワークは, 従来のシグモイド関数を用いたニューラルネットワークに比べ, 関数近似能力に優れ, 学習の収束も速いと報告されている¹⁾⁻³⁾. また, RBF ネットワークは結論部を実数とする簡略ファジィ推論モデルともみなすことができ, ファジィルールをニューラルネットワーク的に自動的に獲得することから, ニューロ・ファジィ・モデリングともよばれている⁴⁾⁻⁶⁾. RBF ネットワークによる非線形回帰手法では, 説明変数が高次元の場合にモデルを構成するパラメータの数が指数的に増大する.

本研究は, 射影追跡回帰(projection pursuit regression)⁷⁾を用いて説明変量の次元を低減し, モデルを構成するパラメータ数を小さくすることを目的としている. FriedmanとTukey⁸⁾により始められた射影追跡の目的は, 高次元空間におけるデータ点を低次元空間に射影することによってデータの興味深い構造を発見することである. そのためにある目的関数, すなわち射影指標を最大化する線形射影をコンピュータを用いて自動探索する. 射影指標は, 通常多くの極値を持つ関数であるので, 未知パラメータの初期値のランダムな探索とニュートン法などの勾配法によるパラメータ最適化が用いられている⁹⁾⁻¹¹⁾.

多次元空間におけるデータを低次元空間に射影するという考え方は, 多変量解析で一般に用いられているもので, 主成分分析や判別分析などは射影追跡の特別な場合であるとみなされている^{9),10)}. FriedmanとStuetzle⁷⁾によって拡張された射影追跡回帰は, 低次元空間における非線形写像を発見するものである. 線形和とシグモイド関数による3層のニューラルネットワークとの構造上の類似性から射影追跡学習ネットワーク¹²⁾という呼び方もされている. 射影追跡回帰の解法もDFP法やニュートン法などの勾配法であり, 固有値法による多変量解析の手法

* 原稿受付 1996年?月?日

** 大阪府立大学 工学部 College of Engineering, Osaka Prefecture University; Gakuen-cho 1-1, Sakai city, Osaka 593, JAPAN

Key Words: Projection pursuit regression, RBF network, eigenvalue method

は用いられていない。本研究では、射影追跡回帰にファジィ集合のメンバシップ関数の考え方を採用する。そして、人間のパターン認識能力を活用することによりランダムな探索によらずに大域的な最小値を発見しやすとした固有値法と勾配法の併用法を提案する。

従来の射影追跡回帰では、残差に対する回帰を階層的に繰り返す。しかし、残差に対する射影データから低次元空間における特徴的な構造を発見することは難しい。そこで固有値法を利用した2次元平面への射影追跡回帰を併せて提案する。また、提案の射影追跡回帰を判別分析にも応用できることをFisherのIrisデータ¹³⁾に適用することにより示す。

2. 固有値法による射影追跡

n 次元の説明変量 \mathbf{x}_j と1次元の応答変量 z_j からなる J 個のデータがあり、この説明変量と応答変量との関数関係を求めることを考える。射影追跡回帰は「次元の呪い」を回避する手法である。「次元の呪い」とは、高次元空間においてデータ間の距離が大きくなり、解析が困難になる状況をいう。一般的に回帰手法はデータの平滑化による手法であり、決定すべきパラメータ数に対して十分なデータが必要である。しかし高次元空間ではデータ間の距離が大きいため、平滑化による回帰手法は実行が困難となる。そこで、射影追跡回帰を用いて説明変量の次元を低減する。まず、1次元（直線）への射影追跡回帰について考える。

$\mathbf{x}_j (j=1, \dots, J)$ のデータの集合を二つのファジィ集合に分割する。ファジィ集合（ファジィクラス）を C と D の二つとし、それぞれ $\{(\mathbf{x}_j, \mu(z_j)) ; j=1, \dots, J\}$ と $\{(\mathbf{x}_j, 1-\mu(z_j)) ; j=1, \dots, J\}$ とする。ただし、 \mathbf{x}_j は n 次元空間におけるデータ点で、 $\mu(z_j)$ は j 番目のデータがクラス C に属する度合いを表すメンバシップである。 $\mu(z_j)$ は、

$$\mu(z_j) = \frac{z_j - z_{\min}}{z_{\max} - z_{\min}} \quad (1)$$

と定める。ただし、 z_{\max} , z_{\min} はそれぞれ応答変量 z_j の最大値および最小値である。 z_j の値を単位区間 $[0, 1]$ に基準化したものをメンバシップ値 $\mu(z_j)$ としているので、クラス C は応答変量の値の大きいデータのファジィ集合を意味し、 D は応答変量の値の小さいデータのファジィ集合である。また、 \mathbf{y} を単位ベクトル $\mathbf{p} = (p_1, \dots, p_n)^T$ により定まる直線への \mathbf{x} の正射影とすると

$$\mathbf{y} = \mathbf{p}^T \mathbf{x} \quad (2)$$

となる。ただし、 T は転置を意味する。ファジィクラス C 内のデータの y の標本分散を以下のように定義する。

$$\begin{aligned} \sigma_C^2 &= \frac{1}{J_C} \sum_{j=1}^J \mu(z_j)(y_j - \bar{y})^2 \\ &= \mathbf{p}^T \sum_C \mathbf{p} \end{aligned} \quad (3)$$

ただし、

$$J_C = \sum_{j=1}^J \mu(z_j) \quad (4)$$

$$\bar{y} = \frac{\sum_{j=1}^J \mu(z_j) \cdot y_j}{J_C} \quad (5)$$

$$\sum_C = \frac{1}{J_C} \sum_{j=1}^J \mu(z_j)(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T \quad (6)$$

$$\bar{\mathbf{x}} = \frac{\sum_{j=1}^J \mu(z_j) \cdot \mathbf{x}_j}{J_C} \quad (7)$$

である。一方、ファジィクラス D の \bar{y} （クラス C 内の平均）からの変動を以下のように定義する。

$$\begin{aligned} \sigma_D^2 &= \frac{1}{J - J_C} \sum_{j=1}^J (1 - \mu(z_j))(y_j - \bar{y})^2 \\ &= \mathbf{p}^T \sum_D \mathbf{p} \end{aligned} \quad (8)$$

ただし、

$$\sum_D = \frac{1}{J - J_C} \sum_{j=1}^J (1 - \mu(z_j)) \cdot (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T \quad (9)$$

である。クラス C 内の分散(σ_C^2)を小さくし、一方でクラス D の y_j ができるだけ \bar{y} （クラス C 内の平均）から離れるような \mathbf{p} を求める。換言するとこの射影ベクトル \mathbf{p} により定まる軸上では、応答変量の値の大きいデータ（クラス C ）は分散の小さい群をなし、応答変量の値の小さいデータ（クラス D ）はその群から離れて分布する。本提案法ではクラス C （応答変量の値の大きいファジィ集合）とクラス D （応答変量の値の小さいファジィ集合）の2クラスに分けるため、回帰曲線の関数形状は単峰形、単調増加、または単調減少であると仮定する。射影によってデータの構造をよく表現できる射影ベクトル \mathbf{p} を求めるための評価関数、すなわち射影指標を

$$Q(\mathbf{p}) = \frac{\mathbf{p}^T \sum_D \mathbf{p}}{\mathbf{p}^T \sum_C \mathbf{p}} \quad (10)$$

とする。射影ベクトル \mathbf{p} は、この $Q(\mathbf{p})$ を最大化することによって求めることができる。 $\partial Q(\mathbf{p}) / \partial \mathbf{p} = 0$ より

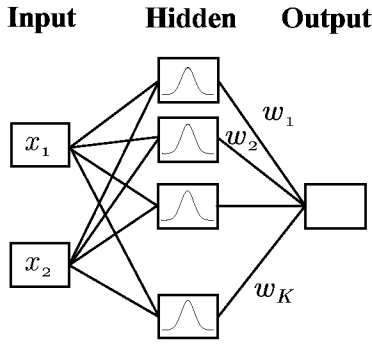


Fig. 1 RBF neural network.

$$\sum_D \mathbf{p}(\mathbf{p}^T \sum_C \mathbf{p}) - \sum_C \mathbf{p}(\mathbf{p}^T \sum_D \mathbf{p}) = 0 \quad (11)$$

となり、 $Q(\mathbf{p})$ の最大値を λ とすると

$$\sum_D \mathbf{p} = \lambda \sum_C \mathbf{p} \quad (12)$$

となる。 $Q(\mathbf{p})$ の最大化問題は(12)式で表される固有値問題に帰着される。

この最大固有値に対応する固有ベクトルが、求めるべき射影ベクトル \mathbf{p} である。2次元平面に射影する場合は第2固有値とそれに対応する固有ベクトルも求め、二つの固有ベクトルで定まる平面へデータを射影する。

3. RBFによる回帰

Fig. 1は一般的なRBF (Radial Basis Function) ネットワークの概念図を示している。RBF ネットワークは中間層にラジアル基底関数を用いた3層ニューラルネットワークの一種であり、各ニューロンからの出力の重み付き線形和が出力となる。シグモイド関数を用いたニューラルネットワークに比べ中間層の数が少なく構造が簡単であり、また、関数近似能力や学習効率に優れていると報告されている¹⁾。

RBF ネットワークにおいて説明変数(入力)の次元が高くなればネットワークを構成するパラメータの数が増加し、計算量が指数的に増加する。RBF ネットワークによる回帰はデータの平滑化による回帰手法の一種であり、決定すべきパラメータの数に比べ十分なデータが必要である。2.で述べた射影追跡回帰を用いて低次元空間の回帰を考えることによりRBF ネットワークのパラメータ数を小さくできる。

2.で求めた固有ベクトルを射影ベクトルとして、説明変数 \mathbf{x}_j の射影 y_j を求め、 y_j と基準化された応答変数 $\mu(z_j)$ との関係を、RBF ネットワークにより近似する。ガウス基底を用いるRBF ネットワークの出力を

$$s(y) = \sum_{k=1}^K A_k(y) \cdot w_k \quad (13)$$

$$A_k(y) = \exp\left(-\frac{(y - a_k)^2}{b_k}\right) \quad (14)$$

とする。ただし、 K はガウス基底の数である。 y を入力とするRBF ネットワークにより、応答変数を回帰しようとする場合の近似度合を表す評価関数を

$$E = \frac{1}{2} \sum_{j=1}^J (s(y_j) - \mu(z_j))^2 \quad (15)$$

とする。(15)式を最小とするようなRBF ネットワークの未知パラメータを最急降下法(学習)により求める。さらに、パラメータの更新時(学習時)のRBF ネットワークの出力をCRT上のコンピュータグラフィクスで確認することにより、適当な時点から射影ベクトルの要素とRBF ネットワークの未知パラメータを同時に更新し、近似精度をあげる。本研究の数値例においてはRBF ネットワークを構成するパラメータ \mathbf{w} を50回更新した後、 $\mathbf{a}, \mathbf{b}, \mathbf{p}$ を併せて更新するようにした。最急降下法に基づく学習則は

$$w_k^{NEW} = w_k^{OLD} - \eta \frac{\partial E}{\partial w_k}, \quad (k=1, 2, \dots, K) \quad (16)$$

$$a_k^{NEW} = a_k^{OLD} - \eta \frac{\partial E}{\partial a_k}, \quad (k=1, 2, \dots, K) \quad (17)$$

$$b_k^{NEW} = b_k^{OLD} - \eta \frac{\partial E}{\partial b_k}, \quad (k=1, 2, \dots, K) \quad (18)$$

$$p_i^{NEW} = p_i^{OLD} - \eta \frac{\partial E}{\partial p_i}, \quad (i=1, 2, \dots, n) \quad (19)$$

である。ただし、 η は正の学習係数である。また、学習則における未知パラメータに対する勾配ベクトルは、それぞれ

$$\frac{\partial E}{\partial w_k} = \sum_{j=1}^J (s(y_j) - \mu(z_j)) \exp\left(-\frac{(y_j - a_k)^2}{b_k}\right) \quad (k=1, 2, \dots, K) \quad (20)$$

$$\frac{\partial E}{\partial a_k} = \sum_{j=1}^J (s(y_j) - \mu(z_j)) s(y_j) \frac{2(y_j - a_k)}{b_k} \quad (k=1, 2, \dots, K) \quad (21)$$

$$\frac{\partial E}{\partial b_k} = \sum_{j=1}^J (s(y_j) - \mu(z_j)) s(y_j) \left(\frac{y_j - a_k}{b_k} \right)^2 \quad (k=1, 2, \dots, K) \quad (22)$$

$$\frac{\partial E}{\partial p_i} = \left(\sum_{j=1}^J (s(y_j) - \mu(z_j)) \sum_{k=1}^K \left(s(y_j) \frac{-2(y_j - a_k)}{b_k} x_{ij} \right) \right) \quad (i=1, 2, \dots, n) \quad (23)$$

となる．固有値法において射影指標の定義からデータの下にある曲線の形状として単調増加，単調減少，単峰形を仮定しているが，この仮定を厳密に満たしていない場合も提案手法を適用することができる．

RBF ネットワークを用いて非線形回帰を行う場合，評価関数が局所最小値に陥ることがあり，ガウス基底の数や未知パラメータの初期値の決定は重要な問題である．RBF ネットワークを構成するパラメータ w_k, a_k, b_k はそれぞれ一つのガウス基底の高さ，位置，幅という意味を持っている．そのため，応答変量の値が大きくなった隆起部分にガウス基底を配置すると学習が局所最小値に陥らずパラメータの最適値を求めやすい．本提案法では，固有値法により低次元空間での射影 y と応答変量 z とを散布図として視覚化し，その散布図からおおまかな回帰曲線の形状を推定することによりガウス基底の数や未知パラメータの初期値を適切に決定することができる．

$s(\mathbf{p}_1 \mathbf{x}, \mathbf{p}_2 \mathbf{x})$ と表される2変数のRBFを用いる場合も， \mathbf{p}_1 と \mathbf{p}_2 の初期値を適切に定めないと局所最適な射影ベクトルが求まることになる．データの下にある関数形状に関する仮定をある程度満たしていれば，提案の固有値法により適切な射影ベクトルの初期値を求めることができる．2. で求めた最大固有値に対応する固有ベクトルを \mathbf{p}_1 ，第2固有値に対応する固有ベクトルを \mathbf{p}_2 として，説明変量 \mathbf{x} の射影 y_1, y_2 を求め，それらと基準化された応答変量 $\mu(z_j)$ の散布図を3次元グラフィックスとして示し，2次元RBFネットワークの未知パラメータの初期値を決定する．そしてパラメータ $w_k, a_k, b_k, \mathbf{p}_1, \mathbf{p}_2$ を最急降下法により求める．2次元RBFネットワークは

$$s(y_1, y_2) = \sum_{k=1}^K A_k(y_1, y_2) \cdot w_k \quad (24)$$

$$A_k(y_1, y_2) = \exp \left(-\frac{(y_1 - a_{1k})^2}{b_{1k}} - \frac{(y_2 - a_{2k})^2}{b_{2k}} \right) \quad (25)$$

とする．

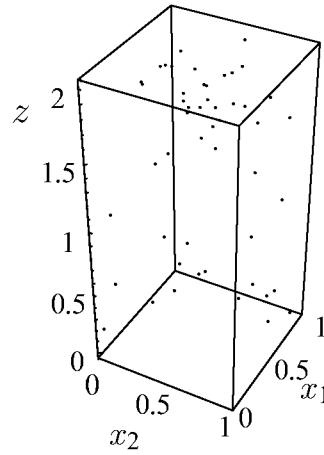


Fig. 2 3-D scatterplot of the data for the first example. z is plotted on the vertical axis, x_1 and x_2 on the horizontal axes.

4. 数値例

4.1 直線に射影する場合

提案の固有値法ではデータの下にある回帰曲線の形状を単峰形，単調増加，単調減少を仮定したが，4. では仮定を満たす場合と完全には満たさない場合の二つの場合の数値例を示す．仮定を満たす場合として一様乱数によりデータ $\mathbf{x}_j \in [0, 1]$ ($n=4, j=1, \dots, 50$) を作成し，その応答変量を $z_j^* = \sin(\pi \cdot (0.75x_{1j} + 0.75x_{2j} + 0.75x_{3j} + 0.75x_{4j}) - \pi)$ とする50個のデータに提案手法を適用する．Fig. 2に説明変量 (x_1, x_2) と応答変量 z_j の散布図を示す． (x_1, x_2) 以外のどの説明変量をとっても同様に，説明変量と応答変量の特徴的な関係は見いだせない．固有値法により求めた固有値および固有ベクトルは $\lambda=5.56$ および $\mathbf{p}=(0.4315, 0.5573, 0.5565, 0.4397)$ である．Fig. 3に固有値法により求めた固有ベクトルによる射影 y_j と応答変量 z_j の散布図を示し，Fig. 4に学習後の射影ベクトルを用いたデータ点 (y_j, z_j) の散布図を示す．またFig. 5に学習後のRBFネットワークの出力を示す．Fig. 5は本数値例に用いた関数とほぼ一致している．この場合のRBFネットワークを構成するガウス基底の数は $K=10$ ，未知パラメータの初期値は $w_k=0$ ， a_k は定義域内を均等に分割した値， b は0.01とした．Fig. 3の散布図から単峰形であることがわかるので，ガウス基底は均等に配置するだけで良いと判断した．学習係数 $\tau=0.0001$ とし，学習回数は10000回とした．用いた計算機はPC9801BX (CPU i486DX2, 66MHz) で計算時間は8分である．学習後の誤差は $2E/J=4.0 \times 10^{-4}$ である．

次に回帰曲線の形状についての仮定を完全に満たさな

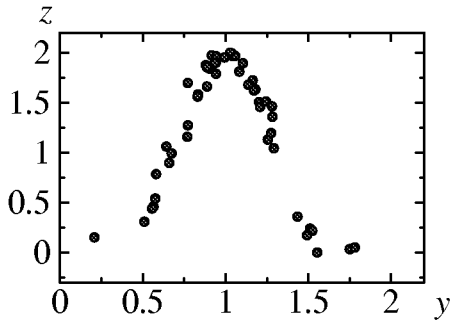


Fig. 3 2-D scatterplot of data obtained by the eigenvalue method for the first example. z is plotted against projection y .

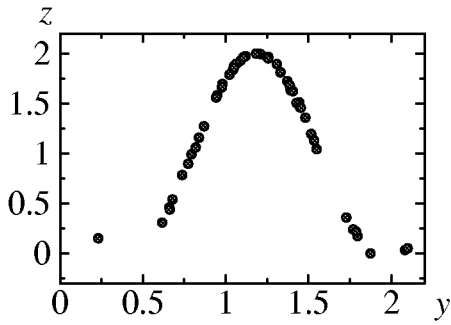


Fig. 4 2-D scatterplot of data after learning for the first example. z is plotted against projection y obtained by the vector p after learning.

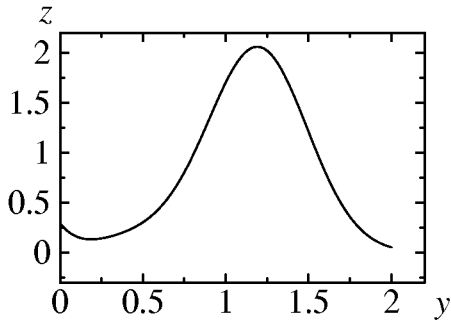


Fig. 5 Output of the RBF neural network after learning.

い場合でも本提案手法が適用できることを示すための数値例を示す。説明変量の次元 $n=4$ ，データ数 $J=150$ とする。一つ目の数値例と同様に説明変量 (x_1, x_2) と応答変量 z_j との散布図をFig. 6に，固有値法により求めた固有ベク

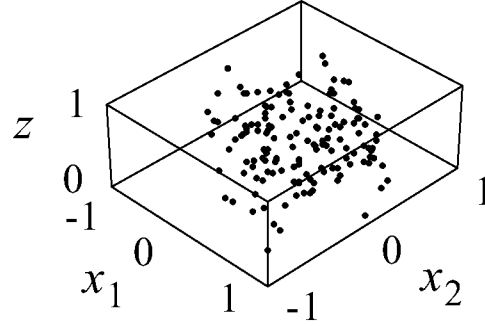


Fig. 6 3-D scatterplot of the data for the 2nd example.

トルによる射影 y_j と応答変量 z_j の散布図をFig. 7に，学習後のデータ点 (y_j, z_j) の散布図をFig. 8に示す。また，学習後のRBFネットワークの出力をFig. 9に示す。本数値例で用いたデータはFig. 9で得られたように多峰形を示す関数ではあるが，グラフ中央部は応答変量が比較的大きな値（0.5以上）であり，グラフ両端部は比較的小きな値（0.5以下）となっている。すなわち単峰形の仮定を完全には満たさないデータであるが，回帰曲線（Fig. 9）を示唆する散布図（Fig. 7）を固有値法により得ることができる。

本数値例において，RBFネットワークを構成するガウス基底の数は $K=18$ とした。RBFネットワークを構成するパラメータ w_k, a_k, b_k は，各ガウス基底の高さ，位置，幅を意味している。固有値法により求めたFig. 7で応答変量の大きな部分（隆起部分）に，ガウス基底を配置するように a_k の初期値を設定した。また $w_k=0$ ， $b_k=0.001$ とした。学習係数 $\tau=0.0005$ とし，10000回の学習を行った。計算時間はPC9801BX（CPU i486DX2，66MHz）で12分である。学習後の誤差は $2E/J=1.91 \times 10^{-6}$ である。Fig. 10に示すようにガウス基底の初期位置を定義域内に均等に配置した場合は局所最小値に陥ってうまく近似できない。このことから提案の固有値法と勾配法の併用法では，求めようとする回帰曲線の形状を事前に推測できるので，適切に未知パラメータの初期値を設定できる。

4.2 平面に射影する場合

直線への射影のみでは十分に近似を行えないデータに対しては2次元平面への射影追跡回帰を行う。一様乱数によりデータ $\mathbf{x}_j \in [0,1]$ ($n=4, j=1, \dots, 200$) を作成し，応答変量を

$$z^* = \exp(-5(y_1^* - 0.3)^2) + \exp(-8y_1^{*2}) + \exp(-5y_2^{*2}) \quad (26)$$

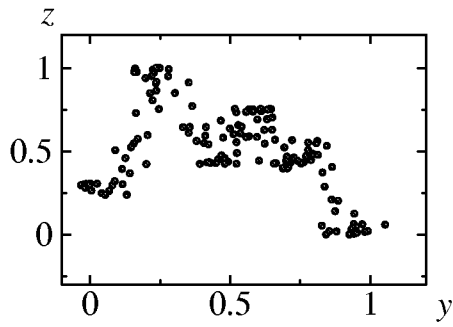


Fig. 7 2-D scatterplot of data obtained by the eigenvalue method for the 2nd example.

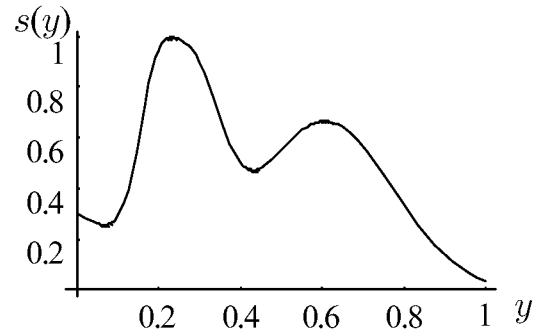


Fig. 10 Output of the RBF neural network for the 2nd example when the initial values of parameters a_k were uniformly spaced.

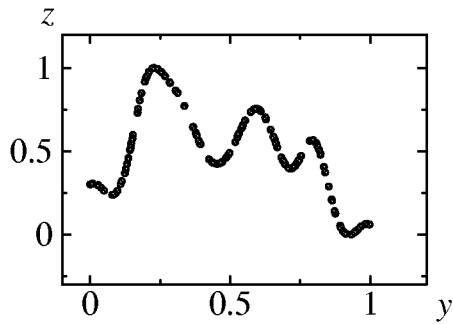


Fig. 8 2-D scatterplot of data after learning for the 2nd example.

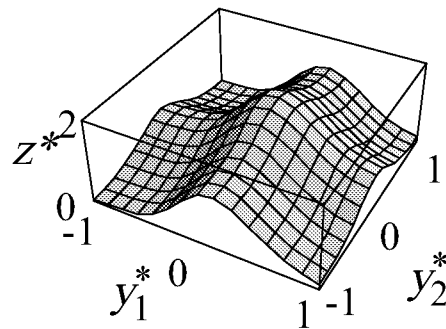


Fig. 11 3-D graphics of the function in eq.(26).

$$y_1^* = 0.5x_{1j} + 0.5x_{2j} - 0.5x_{3j} - 0.5x_{4j} \quad (27)$$

$$y_2^* = 0.5x_{1j} - 0.5x_{2j} + 0.5x_{3j} - 0.5x_{4j} \quad (28)$$

とした. (26)式で表される関数の形状をFig.11に示す. Fig.12に説明変量 (x_1, x_2) と応答変量 z との散布図を示す. 他の説明変量を用いた場合も同様に説明変量と応答変量の特徴的関係は見いだせない. 固有値法により求めた最大固有値に対する固有ベクトル \mathbf{p}_1 による射影 y_1 と第2固有値に対する固有ベクトル \mathbf{p}_2 による射影 y_2 に対する応答変量 z の散布図を Fig.13に示す. また, Fig.14に学習後の2次元RBFネットワークの出力を示す. Fig.14は, データを作成するために用いた上記の関数 (Fig.11) を良く近似している.

4.3 判別分析への応用

本項では, 平面への射影による回帰手法を判別分析に応用する. 用いるデータは, R.A.Fisherによる線形判別関数の研究に用いられたあやめの3群判別データ¹³⁾である. 入力データとして四つの特徴量 (がく片の長さ, がく片の幅, 花卉の長さ, 花卉の幅) と, クラスとしてあや

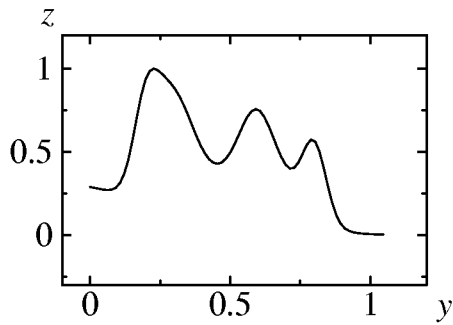


Fig. 9 Output of the RBF neural network for the 2nd example when the initial values of parameters a_k were decided by utilizing human ability of pattern discovery.

として計算した 200 個 のデータに提案手法を適用した. ただし

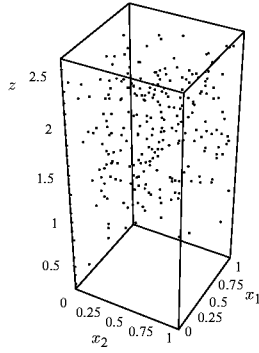


Fig. 12 3-D scatterplot of the data used in the two-dimensional PPR.

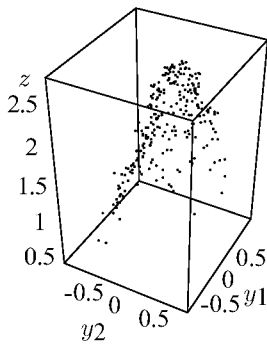


Fig. 13 The 3-D scatterplot of projections by using eigenvectors in the two-dimensional PPR.

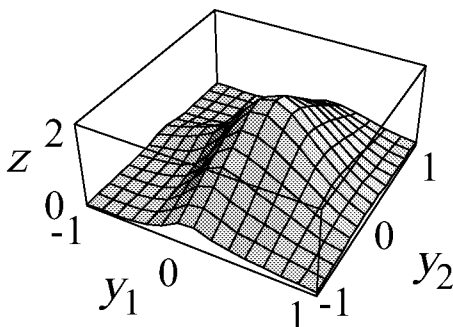


Fig. 14 3-D graphics of the output of obtained RBF neural network.

めの品種 (Iris setosa, Iris versicolor, Iris virginica) が与えられている。データ総数は150個で Iris setosa, Iris versicolor, Iris virginica のそれぞれが50個ずつ含まれている。データの出力値はカテゴリーを表しているの、提案の手法に用いるため以下のように数値データに置き換

える。学習用データの出力値を Iris setosa のみを1として他を0とする場合、Iris versicolor のみを1として他を0とする場合、Iris virginica のみを1として他を0とする場合の各々について提案手法を適用する。

判別分析において所与のデータすべてを用いてモデル同定した場合、得られるモデルはデータに含まれるノイズを内包するため真のモデルを得ることができない¹⁷⁾といわれている。そのため k -fold Cross Validation 法¹⁴⁾ や leaving-one-out 法¹⁵⁾ のような交叉確認法が用いられている。交叉確認法はモデル同定に所与のデータの一部を用い、他の残りでモデルの性能を評価する手法である。 k -fold CV 法ではデータを k 分割し全データの $(k-1)/k$ でモデル同定を行い、残りの $1/k$ で評価を行う過程を k 回繰り返す、その k 回の誤判別率の平均をとる手法である。leaving-one-out 法はこの k をデータ数として分割数 k を最大にした k -fold CV 法の特別な場合である。

2-fold CV 法により、150個のデータから Iris setosa, Iris versicolor, Iris virginica の3種が均等に含まれるように75個を選んで学習用データとし、残り75個のデータを評価用データとして用いる場合について述べる。まず、上記の三つの場合各々に対して、固有値法による最大固有値に対応する固有ベクトルと第2固有値に対応する固有ベクトルを用いて、説明変量 (Iris の四つの特徴量) を平面 (2次元) へ射影する。Versicolor を1として他を0とした場合の固有値法による平面への射影を Fig. 15 に示す。三つの場合それぞれについての平面への射影と応答変量 $\{0,1\}$ との回帰曲面を求める。3群の判別のためのネットワーク図を Fig. 16 に示す。求めた三つの場合の回帰曲線を Figs. 17~19 に示す。さらに、RBF ネットワークの出力の大きさが三つの中で最大である領域を3種類の Iris 各々の判別域として求めた。その結果を Fig. 20 に示す。この Iris の3品種の判別域を用いて75個の評価用データに対して判別を行った。同様に学習用データと評価用データを入れ換え判別を行った結果、それらの誤判別率の平均は0.027となった。

150個のデータすべてを学習用データとして提案の固有値法を適用し、平面へ射影した結果を Fig. 21 に示す。Fig. 21 で用いた射影ベクトルには Versicolor の応答変量を1とした場合に求めた最大固有値に対する固有ベクトルと第2固有値に対応する固有ベクトルである。2-fold CV 法の場合と同様にそれぞれの場合に対し回帰曲面を求め、それぞれの回帰曲面の値が最大となる領域を求めたものが Fig. 22 であり、全データを用いた場合の判別域である。150個のデータすべてを学習用データとした場合には、誤判別率は0.00であった。

S.M.Weiss and I.Kapouleas は Iris データを用いて種々

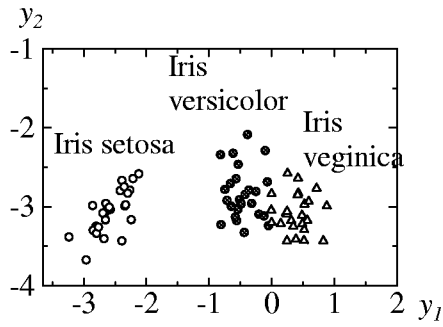


Fig. 15 Projection of the Iris data(Cv).

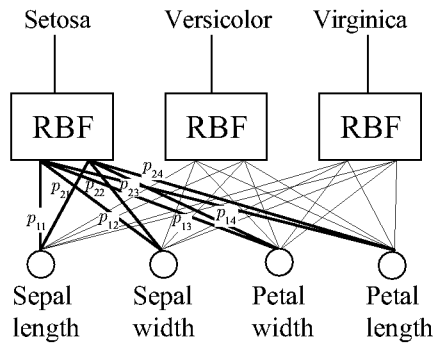


Fig. 16 RBF network for classification of Iris data.

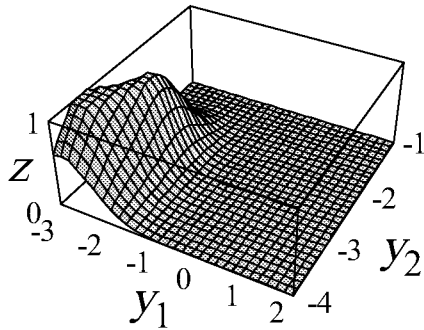
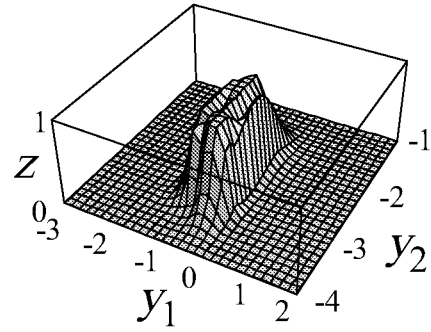
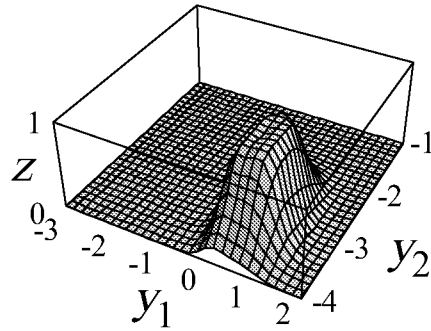
Fig. 17 3-D graphics of the RBF neural network when the value of response variables for *Iris setosa* class is set to 1.Fig. 18 3-D graphics of the RBF neural network when the value of response variables for *Iris versicolor* class is set to 1.Fig. 19 3-D graphics of the RBF neural network when the value of response variables for *Iris virginica* class is set to 1.

Fig. 20 Decision regions created by RBF neural network(Cv).

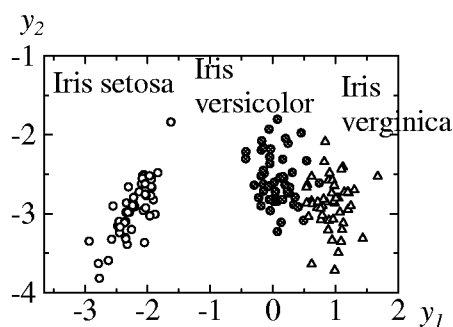


Fig. 21 Projection of all 150 cases by the eigenvalue method(App).

Table 1 Comparative performance on Fisher's Iris Data.

Method	Apparent Error	Leaving-one-out Error
Linear	0.020	0.020
Quadratic	0.020	0.027
Nearest neighbor	0.000	0.040
Bayes independence	0.047	0.067
Bayes 2nd order	0.040	0.160
Neural net(BP)	0.017	0.033
PVM rule	0.027	0.040
Optimal rule size 2	0.020	0.020
CART tree	0.040	0.047
proposed PPR	0.000	0.027*

* 2-fold CV の結果

Fig. 22 Decision region created by RBF neural network(App).

の判別分析手法の性能比較を行い、その結果を報告している¹⁶⁾。その比較結果に提案手法の結果を付加してTable 1に転載する。用いられている判別手法は、五つの統計的判別手法¹⁷⁾、ニューラルネットワークによる判別手法、三つの機械学習による判別手法^{18)–20)}である。Table 1第1列は用いた判別手法を表し、第2列(Apparent error)は150個全てのデータを学習用データとした場合の誤判別率を表し、第3列はLeaving-one-out法を用いた場合の誤判別率である。ただし、提案手法の欄の第3列は2-fold CV法を用いた結果である。Leaving-one-out法は、 J 個のデータに対して、 $J-1$ 個を学習用データとし、残りの1個のデータで判別を行う過程を J 回行い、 J 回の平均を誤判別率とする手法である。そのため、Leaving-one-out法による誤判別率は2-fold CV法より小さくなる傾向がある。Grabisch and Dispot²¹⁾はIrisデータを用いて判別分析における2-fold CV法とLeaving-one-out法との比較結果を示していて、その結果でもLeaving-one-out法の誤差判別率が1%~3%小さくなっている。提案法はRBFのパラメータの初期値を人が決めるという過程があるため、Leaving-one-out法のようにデータ件数と同じ回数の判別結果を求めることが時間的に困難である。そのため2回の計算で済む2-fold CV法を用いて誤差判別率を評価した。Leaving-one-out法を用いた場合には誤差判別率が小さくなることを考慮すると、提案した射影追跡回帰による判別分析手法は、他の手法と比べて良い判別結果を示している。

本提案法は、固有値法により高次元空間における特徴的なデータ構造を表す射影方向を見つけ、さらに勾配法によりその方向の微調整を行う固有値法と勾配法の併用法である。線形射影を用いる提案法はFisherの線形判別

手法に非線形性を加味した手法となっているが、高次元空間のままで判別域を求めるニューラルネットワークによる手法とは異なり、判別域を可視化できる。そのため、低次元空間での散布図により人間のパターン認識能力を活用して、基底関数の初期配置などのパラメータの初期値を適切に決定できる。したがって、勾配法による学習がローカルミニマムに陥る欠点を改善していると考えられる。

5. おわりに

本研究では、射影追跡によって説明変量の次元を1次元または2次元に低減し、回帰を行う射影追跡回帰手法を提案した。提案手法の特徴は、射影ベクトルを求めるにあたってファジィ集合のメンバシップ関数の考え方を導入し、固有値法を用いることである。固有値法により説明変量の射影と応答変量との散布図が視覚化でき、回帰を行う前に説明変量の変域や回帰曲線のおおまかな形状を知ることができる。そのために基底関数の初期配置を適切に行うことができる。また Fisher の Iris データを用いた判別分析に射影追跡回帰を適用し、その有用性を示した。提案法では固有値法の結果に基づいて人間が基底関数を配置するので、自動化レベルは低いが従来法よりも良い結果を得ることができた。

固有値法で提案した射影指標はどのようなデータにも適用できる万能のものではない。例えば線形モデル (AR モデル) からの周期的な時系列データに提案法を適用すると射影後の応答変量の散布図は楕円形になり、次のステップでの学習により射影ベクトルが大きく変化する。したがって対象とするデータに則した指標の改善を考えていくことが必要である。また、学習に要する計算時間が長いので逐次射影法や代数的再構成法などにより学習の高速化を図ることが今後の課題である。

参考文献

- 1) J. Moody and C. J. Darken: Fast learning networks of locally-tuned processing units; *Neural Computation*, Vol. 1, No. 2, pp. 281-294 (1989)
- 2) D. S. Broomhead and D. Lowe: Multivariate functional interpolation and adaptive networks; *Complex Systems*, Vol. 2, No. 3, pp. 321-355(1988)
- 3) T. Poggio and F. Girosi: Regularization algorithms for learning that are equivalent to multilayer networks; *Sciences*, Vol. 247, No. 4945, pp. 978-982(1990)
- 4) 市橋, 佐藤, 三好, 長坂: ニューロ・ファジィ有限要素法と誤差解析; システム制御情報学会論文誌, Vol. 8, No. 3, pp. 122-129(1995)
- 5) H. Ichihashi and I. B. Turksen: A neuro-fuzzy approach to data analysis of pairwise comparisons; *Int. J. of Approximate Reasoning*, Vol. 9, No. 3, pp. 227-248 (1993)
- 6) M. Brown and C. Harris: *Neurofuzzy Adaptive Mod-*

elling and Control, Prentice Hall, (1994)

- 7) J. H. Friedman and W. Stuetzle: Projection pursuit regression; *J. of American Statistics Association*, Vol. 76, No. 376, pp. 817-823 (1981)
- 8) J. H. Friedman and J. W. Tukey: A projection pursuit algorithm for exploratory data analysis; *IEEE Trans. on Computers*, Vol. C-23, No. 9, pp. 881-890 (1974)
- 9) P. J. Huber: Projection pursuit; *The Annals of Statistics*, Vol. 13, No. 2, pp. 425-475 (1985)
- 10) 岩崎: 射影追跡: その考え方と実際; 計算機統計学, Vol. 4, No. 2, pp. 41-56 (1991)
- 11) 今井, 佐藤: 射影追跡法における感度分析; 計算機統計学, Vol. 5, No. 2, pp. 101-106 (1992)
- 12) J. N. Hwang et al.: Reregression modeling in back-propagation and projection pursuit; *IEEE Trans. on Neural Networks*, Vol. 5, No. 3, pp. 342-353 (1994)
- 13) R. A. Fisher: The use of multiple measurements in taxonomic problems; *Annals of Eugenics*, Vol. 7, No. 2, pp. 179-188 (1936)
- 14) M. Stone: Cross validity choice and assessment of statistical predictions; *J. of the Royal Statistical Society*, Vol. 36, No. 2, pp. 111-147 (1974)
- 15) K. Fukunaga: *Introduction to Statistical Pattern Recognition*, Academic Press, (1972)
- 16) S. M. Weiss and I. Kapouleas: An empirical comparison of pattern recognition, neural nets, and machine learning classification methods; *Proc. of the 10th Int. Joint Conference on Artificial Intelligence*, pp. 781-787 (1989)
- 17) R. Duda and P. Hart: *Pattern Classification and Scene Analysis*, Wiley, (1973)
- 18) S. Weiss, R. Galen and P. Tadepalli: Optimizing the predictive value of diagnostic decision rules; *Proc. of the 6th Annual National Conference on Artificial Intelligence*, pp. 521-526 (1987)
- 19) J. Quinlan: Induction of decision trees; *Machine Learning*, Vol. 1, No. 1, pp. 81-106 (1986)
- 20) L. Breiman, J. Friedman, R. Olshen and C. Stone: *Classification and Regression Trees*, Monterey, Ca., Wadsworth (1984)
- 21) M. Grabisch and F. Dispot: A comparison of some methods of fuzzy classification on real data; *Proc. of 2nd International Conference on Fuzzy Logic, Neural Nets and Soft Computing IIZUKA '92*, pp. 659-662 (1992)

著者略歴

みよ し てつ や
三 好 哲 也 (正会員)



1963年 4月28日生。1989年3月大阪府立大学大学院工学研究科経営工学専攻博士前期課程修了。同年4月大阪府庁入庁、1994年4月大阪府立大学工学部助手となり現在に至る。ファジィ制御、ニューロ・ファジィなどのファジィ応用研究に従事。日本経営工学会、システム制御情報学会などの会員。

いち はし ひで とも
市 橋 秀 友 (正会員)



1948年12月25日生。1971年大阪府立大学工学部経営工学科卒業。同年松下電器産業(株)入社。1981年大阪府立大学工学部経営工学科助手。1987年同講師。1989年同助教授。1993年同教授。現在に至る。工学博士。ファジィ多目的計画法、ファジィ制御、ニューロ・ファジィなどのファジィ理論の応用研究に従事。国際ファジィシステム学会、日本ファジィ学会、システム制御情報学会、日本経営工学会などの会員。

なか お けい すけ
中 尾 圭 介 (非会員)



1973年2月24日生。1996年3月大阪府立大学工学部経営工学科卒業。同年4月阪神高速道路公団入社。現在に至る。在学中はニューロ・ファジィとその応用に関する研究に従事。

なが さか かず のり
長 坂 一 徳 (非会員)



1941年10月21日生。1965年大阪府立大学大学院工学研究科機械工学専攻博士前期課程修了。同年同大学工学部経営工学科助手。1988年同講師。現在に至る。工学博士。多入力変数を含むシステムの同定、予測に関する研究に従事。日本ファジィ学会、日本経営工学会、精密工学会などの会員。