

階層型ニューラルネットワークと回帰分析の関係

渡辺 栄治, 清水 光

A Study on Relationships between Multi-layered Neural Networks and Regression Analysis

Eiji Watanabe, Hikaru Shimizu

ABSTRACT

This paper studies relationships between multi-layered neural networks and the regression analysis for function approximation problems in the presence of the observation noise. First, we compare NNs with regression models from the viewpoint of both model structures and parameter estimators. Based on the discussions, we propose a learning algorithm for avoidance of overfitting. Next, we propose a method for analysis of the internal representation of NN by using the principal component analysis. Finally, the modeling results of the two modeling methods are concretely shown by several numerical computations.

1. はじめに

各種信号処理の分野において、対象システムの正確な数学的モデルを構築することは重要な問題である。しかし、雑音等の影響により完全なモデルを構築することが一般に困難であるため、雑音が既知なる分布に従うという仮定を設定した各種モデリング手法が提案され、実システムのモデリングに適用されている。システムの静的な入出力関係に対しては回帰分析等が、一方、動的な入出力関係に対してはARモデリング等が、代表的なモデリング手法として挙げられる。回帰分析はモデリングの簡易さやモデル誤差の検定が可能である等の長所を有しているが、これらの長所はモデル誤差がガウス分布に従うという条件下で実現され、非ガウス分布の場合はモデリング精度が劣化することが知られている。

近年、入出力データ間における任意の非線形連続写像を近似的に実現できる手法としてニューラルネットワークモデル(以下ではNNと略す)が注目されている。NNが有する学習能力、並列処理能力、および汎化能力等に期待し、NNの各種分野への応用が試みられている。特に、汎化能力は工学的応用に際して重要な能力であり、適切な汎化能力を獲得するネットワーク構造や学習法に関する研究が急務な課題となっている[7] 相吉等 [6] は広い意味での汎化能力を、(i) 追加学習性能、(ii) 耐ノイズ性、(iii) 補間性能、(iv) 外延的類推性能、等に分類しており、各ケースに応じた汎化能力の検討が必要である。関数近似問題を対象とする場合、離散的に与えられた学習データを精度良く近似することが重要であり、未学習パターンに対する補間能力が汎化能力として定義される((iii)補間性能)。さらに、実データには観測雑音が含まれることを考慮する必要がある、そのようなデータに対する忠実な学習は“過剰学習”と呼ばれる困難な問題を生じる。

本稿では、観測雑音を含む関数近似問題を対象としてNNと回帰分析の関係について検討する。まず、モデル構造およびパラメータ推定手法の観点からNNと回帰分析について比較検討する。つぎに、雑音存在下においてNNによる関数近似手法が生じる問題点について検討し、特に過剰学習を抑制するための学習法を提案する。さらに、主成分分析の

適用によるNNの内部表現の一解析手法を提案する。最後に、数値計算によりNNおよび回帰モデルのモデリング精度の比較を行う。

2. 問題設定

本稿では、簡単化のために静的な単入出力システムを対象とする。入出力間の関係は未知なる連続な非線形関数 $f(x)$ により与えられるとし、雑音 ξ が付加された値 y のみが観測可能とする (Figure 1 参照)。

$$y_p = f(x_p) + \xi_p, (p=1, 2, \dots, P) \quad (1)$$

(1)式において、 p はパターン番号を表し、雑音 ξ_p は平均0および分散 σ^2 の統計量を有し、 x_p および y_p と独立な確率変数とする。雑音の分散 σ^2 および分布形状は未知とする。

[Figure 1 about here.]

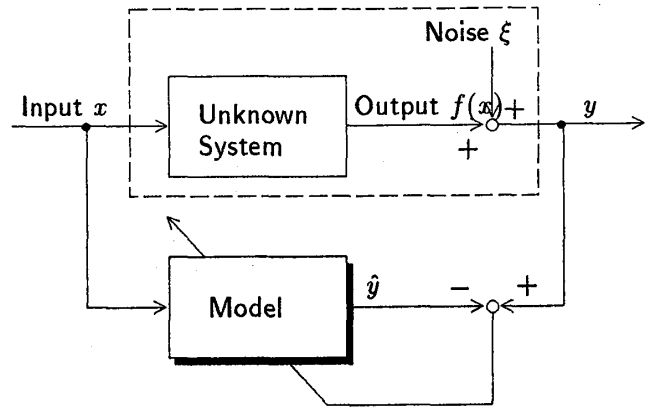
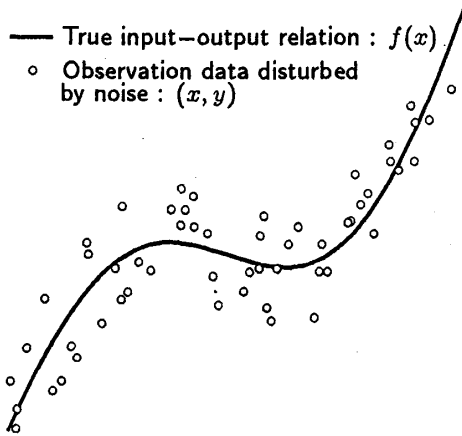


Figure 1: Observation data disturbed by noise

Figure 2: A block diagram for modeling

本稿で扱う問題は、入出力の観測値 (x_p, y_p) を用いて未知なる入出力関数 $f(x)$ を推定することである。すなわち、次式で定義される誤差関数 E を最小化することによりモデル $y = \hat{f}(x)$ を構築することである (Figure 2 参照)。

$$E = \frac{1}{P} \sum_{p=1}^P (y_p - \hat{y}_p)^2 \quad (2)$$

[Figure 2 about here.]

3. 回帰モデルとNNモデル

3. 1 回帰モデル

真の入出力関数 $f(x)$ が非線形関数である場合、観測値 (x, y) を曲線に当てはめることが考えられ、以下に示すような回帰モデルが用いられる [8]。一つは、基底関数に入力 x に関する多項式を用いた多項式回帰モデルであり、

$$\hat{y}_p = a_0 + \sum_{i=1}^n a_i x_p^i$$

他方は、多項式回帰モデルにおける、(i)高次項を含む場合の多重共線性、(ii)次数 n の増加に伴う回帰係数 a_i の再計算等の欠点を補うために、基底関数に直交多項式を用いた直交多項式回帰モデルである。

$$\hat{y}_p = a_0 P_0(x_p) + \sum_{i=1}^n a_i P_i(x_p)$$

ただし、 $P_i(x_p)$ は次式を満足する。

$$\sum_{p=1}^P P_i(x_p) P_j(x_p) = 0, \quad i \neq j, \quad i, j = 1, 2, \dots, n$$

上述の2種類の回帰モデルでは、雑音 ϵ がガウス分布に従うという制約条件が必要であるが、回帰係数 a_i の解釈の容易さ等の長所を有している。回帰係数 a_i は次式の誤差関数 E を最小にするように求められる。

$$E = \sum_{p=1}^P (y_p - \hat{y}_p)^2$$

$$= \sum_{p=1}^P (y_p - (a_0 P_0(x_p) + \sum_{i=1}^n a_i P_i(x_p)))^2$$

最小2乗法により回帰係数ベクトル $\mathbf{a} = (a_0, a_1, \dots, a_n)^T$ は以下のように与えられる。

$$\mathbf{a} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

ただし、

$$\mathbf{X} = \begin{pmatrix} P_0(x_1) & P_1(x_1) & \dots & P_n(x_1) \\ P_0(x_2) & P_1(x_2) & \dots & P_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(x_P) & P_1(x_P) & \dots & P_n(x_P) \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_P \end{pmatrix}$$

3. 2NNモデル

NNは多数のユニットから構成される M 層構造を有している。ここでは、同一層内のユニット間結合やフィードバック結合を持たないフィードフォワード型構造を対象とする。P番目の学習パターンに対して、第 m 層第 i ユニットにおける入力 $i_{p_i}^m$ と出力 $o_{p_i}^m$ 間の関係は以下のように表される。

$$i_{p_i}^m = \sum_{j=1}^{n_{m-1}} w_{ij}^{m-1} o_{p_j}^{m-1} + \theta_i^m \quad (9)$$

$$o_{p_i}^m = f_u(i_{p_i}^m) \quad (10)$$

ここで、 $n_m (m=1, 2, \dots, M)$ は第 m 層内のユニット数を示す。また、 w_{ij}^{m-1} 及び θ_i^m は、第 m 層第 i ユニットと第 $m-1$ 層第 j ユニット間の荷重係数及び第 m 層第 i ユニットのしきい値を表す。さらに、入出力層ユニット特性関数 f_u には線形関数を用いられ、中間層ユニット特性関数 f_u には次式で与えられるシグモイド関数を用いられている。

$$f_u(x) = \frac{1}{1 + e^{-x}} \quad (11)$$

入力 \mathbf{x} およびNN出力 $\hat{\mathbf{y}}$ との関係は次式で与えられる。

$$\hat{y}_p = \theta_1^2 + \sum_{i=1}^{n_2} w_{i1}^2 f_u(w_{i1}^1 x_p + \theta_1^1) \quad (12)$$

入出力層ユニット数は対象システムの入出力次元に対応して設定されるが、適切な中間層数及び各中間層ユニット数を事前に決定することは困難である。しかし、船橋 [10] により十分な中間層ユニットを有する3層NNが任意の精度で非線形連続関数を近似できることが示されており、本稿では3層NNを対象とする。

NNの学習法として最急降下法を基本とした誤差逆伝搬学習法(BP法) [2] が広く用いられている。BP法では、出力層学習パターン y_p に対する誤差関数 E を最小にするように荷重係数が修正される。

$$E = \frac{1}{2} \sum_{p=1}^P (y_p - o_{p_1}^3)^2 \quad (13)$$

ここで、 $o_{p_1}^3$ はモデル出力値 y_p を表す。

$$w_{ij}^{m-1} = w_{ij}^{m-1} + \alpha \delta_i^m o_{p_j}^{m-1} \quad (14)$$

$$\theta_i^m = \theta_i^m + \alpha \delta_i^m \quad (15)$$

ただし、

$$\delta_i^m = \begin{cases} (y_p - o_{pi}^m) f'(i_{pi}^m) & \text{for } m=3, \\ \sum_j (\delta_j^{m+1} w_{ji}^{m+1}) f'(i_{pi}^m) & \text{for } m=2. \end{cases}$$

ここで、 α は学習係数を表す。

学習誤差 e_p および中間-出力層間荷重係数 w_{ij}^3 の符号にしたがい、中間層ユニット出力値 o_{pj}^2 は Table 1 に示すように更新される。Table 1 において、学習誤差 e_p および中間-出力層間荷重係数 w_{ij}^3 の符号が異なる (w_{ij}^3 の符号が適切でない) 場合、(11) 式より中間層ユニット出力 o_{pj}^2 がその学習パターンに対して影響を及ぼさないように ($o_{pj}^2 \rightarrow 0$) 入力-中間層間荷重係数 w_{ij}^2 を更新することがわかる。

[Table 1 about here.]

Table 1: Behaviour of hidden units o_{pj}^2 for learning error e_p

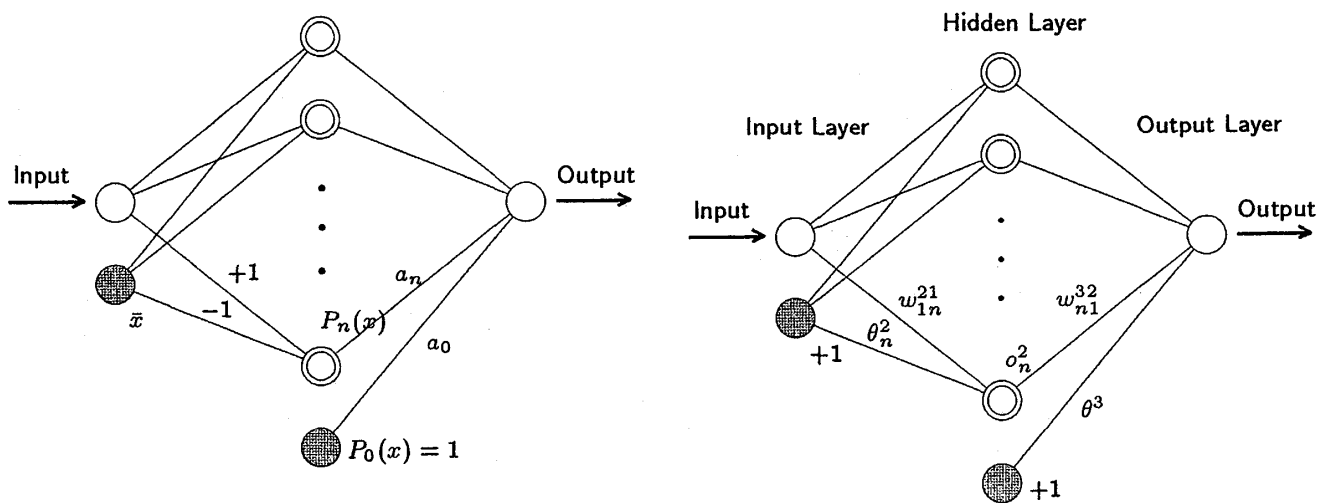
Learning error e_p	Hidden-output weights w_{1j}^{32}	Generalized error of hidden units δ_j^2	Hidden outputs o_{pj}^2
+	+	+	↗
+	-	-	↘
-	+	-	↘
-	-	+	↗

3. 3モデル構造の比較

¹出力層ユニット特性関数は、対象問題毎にシグモイド関数および線形関数を使い分けられているが、NN非線形写像能力は中間層ユニット特性関数の非線形性に依存するため、質的な相違はない。

Figure 3に回帰モデル(直交多項式回帰モデル)およびNNモデル構造を示す。

[Figure 3 about here.]



(a) Structure of orthogonal polynomial regression model

(b) Structure of 3-layered NN model

Figure 3: Structures of both models

NNモデルにおける中間層-出力層間荷重係数 w_{ij} と直交多項式回帰モデルにおける回帰係数 a_i が対応しているが、NNモデルにおける入力-中間層間荷重係数 w_{ij} に対して直交多項式回帰モデルでは1として固定されていることがわかる。すなわち、回帰モデルでは対象システムを明示的な基底関数による線形結合で表現するのに対して、NNモデルでは入力-中間層間荷重係数 w_{ij} で決定される中間層ユニットの特性関数による線形結合で表現している。回帰モデルでは、基底関数は外部から与えられるものであり、基底関数の選択がモデリング性能に多大な影響を与える。一方、NNモデルにおいて中間層ユニットの特性は w_{ij} により適応的に変更可能であるが、 w_{ij} がモデリング精度に多大な影響を与えることがわかる。

4. 過剰学習回避を目的とした誤差逆伝搬学習法

4. 1 雑音存在下におけるBP法の問題点

BP法における誤差関数 E を以下のように書き換える。

$$\begin{aligned} E &= \frac{1}{P} \sum_{p=1}^P (f(x_p) + \xi_p - \hat{y}_p)^2 \\ &= \frac{1}{P} \sum_{p=1}^P (f(x_p) - \hat{y}_p)^2 + \frac{2}{P} \sum_{p=1}^P (f(x_p) - \hat{y}_p) \xi_p + \frac{1}{P} \sum_{p=1}^P \xi_p^2 \\ &= E^* + \sigma^2 \end{aligned} \quad (17)$$

ただし、

$$E^* = \frac{1}{P} \sum_{p=1}^P (f(x_p) - \hat{y}_p)^2 \quad (18)$$

雑音の影響により E の最小化と E^* の最小化が必ずしも一致しないために、 E の最小化だけでは真の入出力関数 $f(x)$ とかけ離れた近似関数が獲得される場合がある。誤差関数 E^* と未知分散 σ^2 との関係に着目すると両者の関係とその原因は以下のように列挙できる²⁾。

ただし、 E^* は雑音が存在しない場合にBP法によって得られる微小な関数近似誤差を表す。

$$1. E \gg E^* + \sigma^2$$

Figure 4(a)に示すように、中間層ユニット数の不足や局所解の影響による特定の学習パターンの過剰学習が考えられる。

$$2. E \ll E^* + \sigma^2$$

Figure 4(b)に示すように、逆に中間層ユニット数の過剰による全学習パターンの過剰学習が考えられる。

$$3. E \approx E^* + \sigma^2$$

雑音の影響を除去した望ましい近似関数が獲得できていると考えられる。

[Figure 4 about here.]

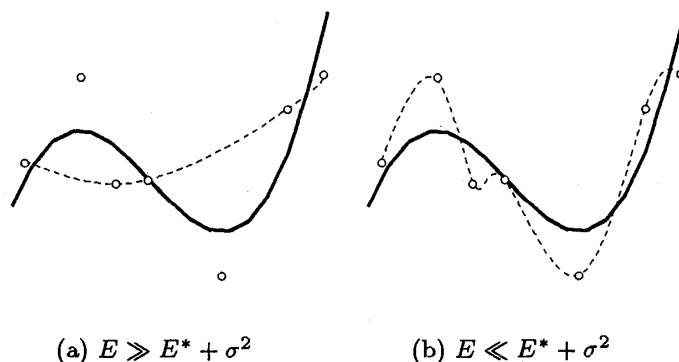


Figure 4: Relationships between E and σ^2

したがって、雑音の影響を除去することにより $E \approx E^* + \sigma^2$ の関係を満足するような学習を行うことが必要である。しかし、(i) 雑音の分散 σ^2 が未知であるため上述の関係を区別できない、(ii) $E \approx E^* + \sigma^2$ 関係を満足しても、学習誤差分布を誤差関数 E から容易に把握できない、等の問題点が生じる³。

4. 2 過剰学習回避を目的とした誤差逆伝搬学習法

学習誤差 e_p は(19)式から明らかなように雑音 ξ_p による影響と荷重係数値の不適切さ $(f(x_p) - y_p)$ の両者の要因から構成される。

$$\begin{aligned} e_p &= y_p - \hat{y}_p \\ &= (f(x_p) - \hat{y}_p) + \xi_p \end{aligned} \quad (19)$$

¹ $E \gg E^* + \sigma^2$ の関係を特定の学習パターンに対する過剰学習、² $E \ll E^* + \sigma^2$ の関係を全ての学習パターンに対する過剰学習として分類できる。

³ Chen等 [1] は、ロバスト推定の観点からoutlier(突発的な大きな誤差)の影響を無視する新たな時変型評価関数を提案している。outlierを残差(学習誤差)の分布状況から推定することにより、cutoff領域を学習の進行に従って更新している。雑音に関する統計情報を必要としない優れた学習法であるが、(i) 残差の分布形状を陰にGauss分布と仮定している、(ii) outlierの存在確率が必要である、(iii) outlier以外の一様分布等に従う雑音に対する有効性、(iv) 多くの設定パラメータが必要である、等の問題点が考えられる。

一般に、両者の分離は困難であり、各学習誤差間の偏りも両者の分離をより困難にしている。いま、この困難さを軽減するために以下の様な仮定を設定する。

仮 定

真の入出力関数 $f(x)$ は連続関数であり、Sigmoid関数を基底関数としたNNにより任意の精度で近似することが可能である。一方、NNによる雑音 ξ の近似精度は $f(x)$ に対する近似精度よりもはるかに悪い。

このような仮定の下で、雑音存在下において過剰学習を低減するための方策として、学習誤差の均等化、過小および過大な学習誤差の軽視が考えられる [11]。

まず、新たな誤差関数 $N(\varepsilon_{\beta}^+, \varepsilon_{\beta}^-)$ を次式のように定義する。

$$N(\varepsilon_{\beta}^+, \varepsilon_{\beta}^-) = \frac{1}{2} \sum_{p=1}^P I_p(\varepsilon_{\beta}^+, \varepsilon_{\beta}^-) \{e_p\}^2 \quad (20)$$

ただし、

$$e_p = y_p - o_{p1}^M \quad (21)$$

$$I_p(\varepsilon_{\beta}^+, \varepsilon_{\beta}^-) = \begin{cases} 1 & \text{for } e_p \geq \varepsilon_{\beta}^+ \text{ or } e_p \leq \varepsilon_{\beta}^-, \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

ただし、 $\varepsilon_{\beta}^+ (> 0)$ および $\varepsilon_{\beta}^- (< 0)$ は収束条件を表す。 $N(\varepsilon_{\beta}^+, \varepsilon_{\beta}^-) = 0$ が成立する場合、全ての学習誤差 e_p に対して次式が成立することを意味する。

$$\varepsilon_{\beta}^- < e_p < \varepsilon_{\beta}^+ \quad (23)$$

$\sum_{p=1}^P I_p(\varepsilon_{\beta}^+, \varepsilon_{\beta}^-) I_p \geq \beta$ が成立する場合に、 ε_{β}^+ および ε_{β}^- が修正される。

$$\begin{cases} \varepsilon_{\beta}^+ = \varepsilon_{\beta}^+ \varepsilon_{rate}^q \\ \varepsilon_{\beta}^- = \varepsilon_{\beta}^- \varepsilon_{rate}^q \end{cases} \quad (24)$$

ただし、収束条件修正率 ε_{rate} は $0 < \varepsilon_{rate} < 1$ を満足し、 q は収束条件の修正回数を表す。また、初期収束条件 ε_{β} および ε_{β} は学習初期における学習誤差の最大値を設定する。

$$\begin{cases} \varepsilon_{ini}^+ = \max e_p \\ \varepsilon_{ini}^- = \min e_p \end{cases} \quad (25)$$

BP法と同様に、 $N(\varepsilon)$ を最小にするための誤差逆伝搬量 δ_{pi}^m は次式で表される。

$$\delta_{pi}^m = \begin{cases} I_p(\varepsilon_{\beta}^+, \varepsilon_{\beta}^-) e_p f'(i_{pi}^m) & \text{for } m=3, \\ \sum_{j=1}^{m+1} (\delta_{pj}^{m+1} w_{ij}^{m+1}) f'(i_{pi}^m) & \text{for } m=2. \end{cases} \quad (26)$$

提案学習法による計算量は誤差逆伝搬量の更新式における誤差指標 $I_p(\varepsilon_{\beta}^+, \varepsilon_{\beta}^-)$ の乗算分だけ増加しているが、 $I_p(\varepsilon_{\beta}^+, \varepsilon_{\beta}^-)$ が0に等しい場合は誤差逆伝搬量の更新を省略することが可能であり、実質的な計算量はBP法に較べて低減化できる。まず、学習誤差 e_p の分布に着目する。Figure 5に学習誤差 e_p と収束条件 $\varepsilon_{\beta}^+, \varepsilon_{\beta}^-$ の関係を示す。

[Figure 5 about here.]

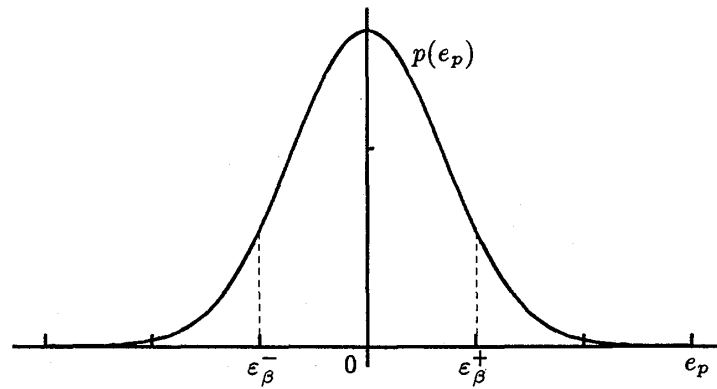


Figure 5: Relationship between the probability density function of learning errors and the convergence condition

5. NNモデルにおける内部表現の解析

両者のモデル構造について検討したように、直交多項式等を基底関数として用いた回帰モデルでは対象システムの複雑さに対応して次数が増加する。また、次数自体の数理的な意味も明確であり、モデリング後における説明変数と目的変数間を把握し易い等の長所を有している。一方、Sigmoid関数を基底関数としたNNでは、回帰モデルの次数に相当する中間層ユニット数の増加により対象システムの複雑さに対応している。しかしながら、各中間層ユニットが有する数理的な意味については明確では無く、モデリング後の入出力関係を解析することは困難である。さらに、NNにおける荷重係数値の非一意性も内部表現の解析を困難にしている。NNによる入出力表現の内部表現として中間層が重要な役割を果たしていることが指摘されており、Sekita等 [3] は主成分分析により、またTamura等 [4] は特異値分解により、冗長な中間層ユニットを削除するための手法を提案している。本節では、Sekita等と同様に主成分分析 [9] を中間層ユニット出力に適用することにより、NN内部表現の適切な特性について検討する。

いま、学習誤差 e_p はその定義から次式のように書き換えられる。

$$\begin{aligned} e_p &= y_p - \hat{y}_p \\ &= y_p - (\sum_i w_i o_{pi}) + \theta \end{aligned} \quad (27)$$

ここで、中間層ユニット出力 o_{pi} と学習誤差 e_p とは線形関係にある。次に、主成分分析の適用について簡単に触れる。 o_{pi} の共分散行列 $A(n \times n)$ の要素 a_{ij} は次式で与えられる。

$$a_{ij} = \frac{1}{P} \sum_{p=1}^P (o_{pi} - \bar{o}_i)(o_{pj} - \bar{o}_j), \quad \bar{o}_i = \frac{1}{P} \sum_{p=1}^P o_{pi} \quad (28)$$

直交行列 $Q(P \times n)$ を用いて共分散行列 A を対角化することにより、積和行列 A の固有値 $\lambda_i (i=1, 2, \dots, n)$ が求められる。

$$Q^T A Q = \text{diag} (\lambda_1, \lambda_2, \dots, \lambda_n) \quad (29)$$

$\tilde{n} (< n)$ 個の変数が主成分である場合、変数 \bar{o}_{p_i} を合成変数 $\hat{o}_{p_i} (i=1, 2, \dots, \tilde{n})$ に次元圧縮できる。その場合、変数 \bar{o}_{p_i} が有する情報を固有値の総和 $\sum_{i=1}^{\tilde{n}} \lambda_i$ とみなすと、(29)式および主成分抽出基準より次元圧縮による情報損失を次式で評価することができる。

$$I_{Loss} = \sum_{i=1}^n \lambda_i - \sum_{i=1}^{\tilde{n}} \frac{1}{P} \sum_{p=1}^P \{ \bar{o}_{p_i} \}^2 \quad (30)$$

ここで、右辺第2項は合成変数 \hat{o}_{p_i} の分散和を表す。情報損失 I_{Loss} を次式の様に書き換える。

$$\begin{aligned} I_{Loss} &= \frac{1}{P} \sum_{p=1}^P I_{Loss}^p \\ &= \frac{1}{P} \sum_{p=1}^P \left\{ \sum_{i=1}^n \lambda_i - \sum_{i=1}^{\tilde{n}} \{ \bar{o}_{p_i} \}^2 \right\} \end{aligned} \quad (31)$$

上式において、 I_{Loss}^p は学習パターン毎の情報損失を評価している。したがって、各学習パターン間における I_{Loss}^p のバラツキが少ない程、中間層ユニットが各学習パターンに対して均等に“知識”を獲得したものと解釈できる。ここでは、各学習パターンに対する情報損失 I_{Loss}^p の分布を表現するための一指標として以下のようなエントロピー H を提案する。

$$H = - \sum_{p=1}^P r_p \log r_p \quad (32)$$

ただし、

$$r_p = \frac{\{ I_{Loss}^p \}^2}{\sum_{q=1}^P \{ I_{Loss}^q \}^2} \quad (33)$$

6. 数値計算

6. 1 モデリング精度の比較

真のシステムは次式の実線形関数 $f(x)$ で表現されるものとする。

$$f(x) = \frac{1}{5} (x^3 + x^2 + x + 1) \quad (34)$$

入力パターン x_p は $[-2, 2]$ の範囲においてランダムに50個選び、平均0、分散 σ^2 のガウスおよび一様乱数を雑音 ξ_p として用いた。3層NN(線形-Sigmoid-線形)を用い、各層のユニット数は1-3-1と設定した。また、学習係数 α を0.01、収束条件修正率 ϵ_{mse} を0.99と設定した。モデリング精度は次式で表される補間誤差 E_{Inter} として評価した。

$$E_{Inter} = \frac{1}{200} \sum_{p=1}^{200} (f(x_p) - \hat{y}_p)^2 \quad (35)$$

ガウス分布に従う雑音を付加した場合の各モデルによるモデリング精度の比較をFigure 6に示す。ただし、多項式回帰モデルの次数は3と設定した。ガウス分布に従う雑音を付加した場合は、回帰モデルによるモデリング精度が良好であることが示されている。

[Figure 6 about here.]

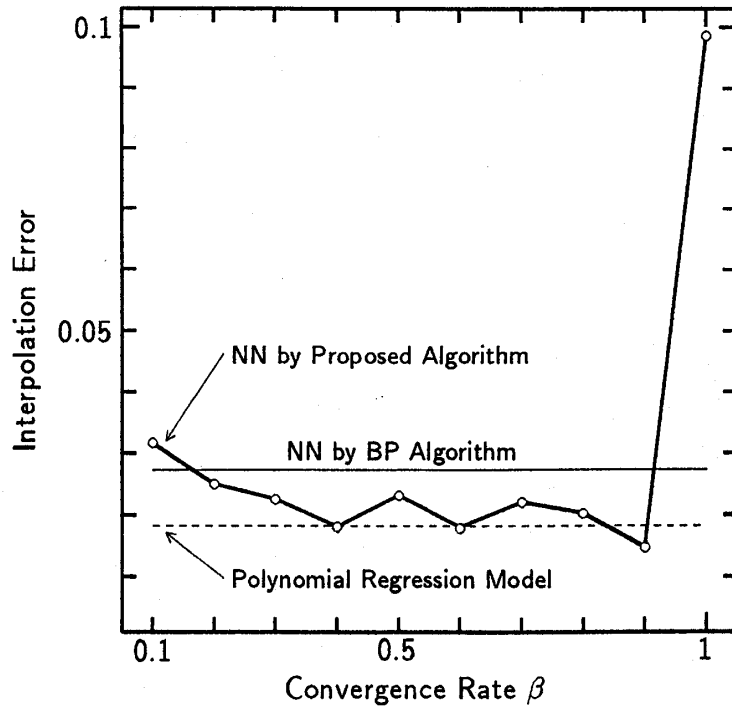


Figure 6: Comparison of interpolation error in the presence of gaussian noise

ガウス分布に従う雑音を付加した場合の各モデルによるモデリング結果をFigure 7に示す。ただし、提案学習法における収束率 β は0.8と設定した。

[Figure 7 about here.]

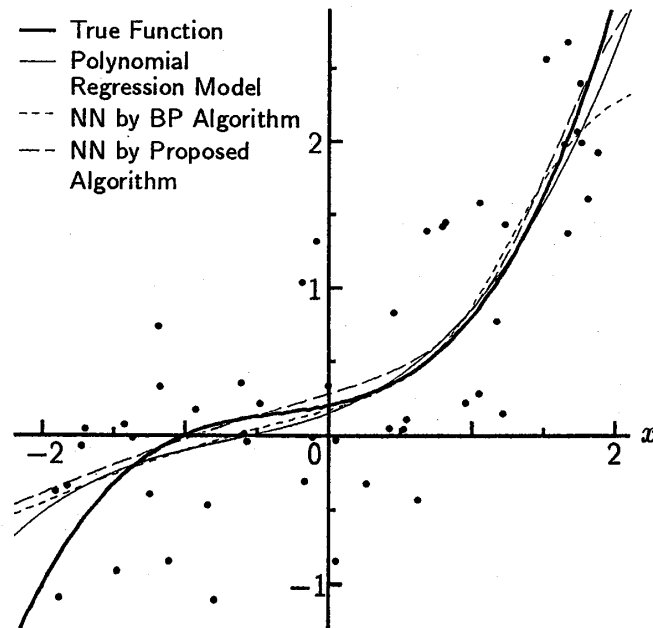


Figure 7: Modeling results in the presence of gaussian noise

一様分布に従う雑音を付加した場合の各モデルによるモデリング精度の比較をFigure 8に示す。提案学習法において、適切な収束率 β を設定することにより、良好なモデリング精度を獲得できることが示されている。

[Figure 8 about here.]

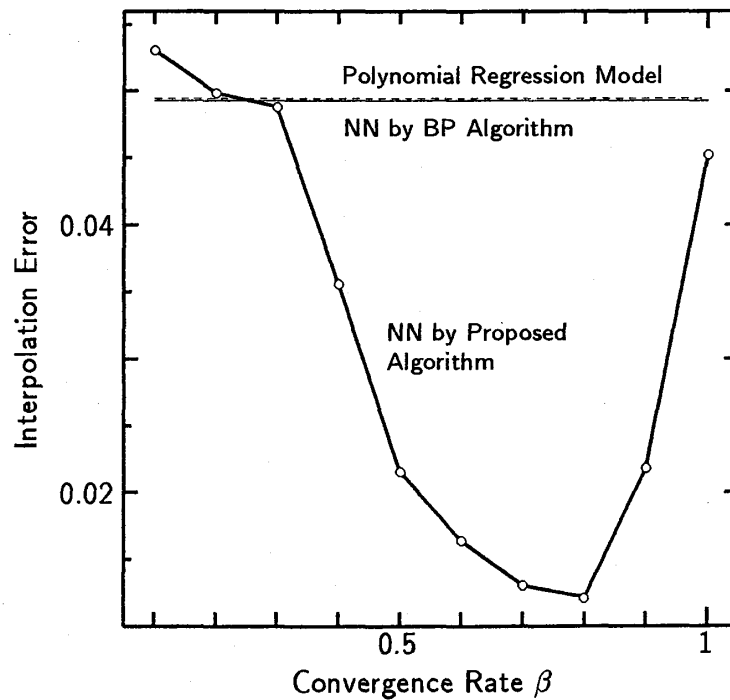


Figure 8: Comparison of interpolation error in the presence of uniformly distributed noise

一様分布に従う雑音を付加した場合の各モデルによるモデリング結果をFigure 9に示す。ただし、提案学習法における収束率 β は0.8と設定した。

[Figure 9 about here.]

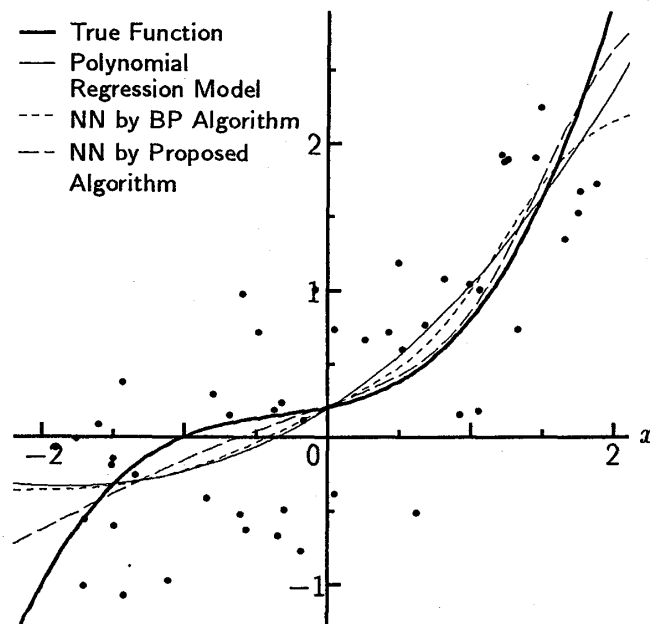


Figure 9: Modeling results in the presence of uniformly distributed noise

6. 2 内部表現の比較

主成分分析の適用により、多項式回帰モデル、NNモデル(学習法:BP法)における内部表現の比較を行う。ここで、多項式回帰モデルでは x, x^2, x^3 の各項を、NNモデルでは o_{p1}, o_{p2}, o_{p3} の各項を内部表現として用いた。また、NNモデルにおいて荷重係数初期値の相違から異なる補間能力を獲得し、補間能力が最良および最悪の各ケースの内部表現を用いた。

第1および第2主成分分布をFigure 10に示す。Figure 10から、補間能力が最良であるNNの内部表現が多項式回帰モデルの内部表現と類似しており、このNNが適切な内部表現を獲得していることがわかる。基底関数形状の相違から、多項式回帰モデルおよびNNモデルの内部表現を直接比較することは困難であるが、主成分分析の適用によりその比較が容易になることが示されている。

[Figure 10 about here.]

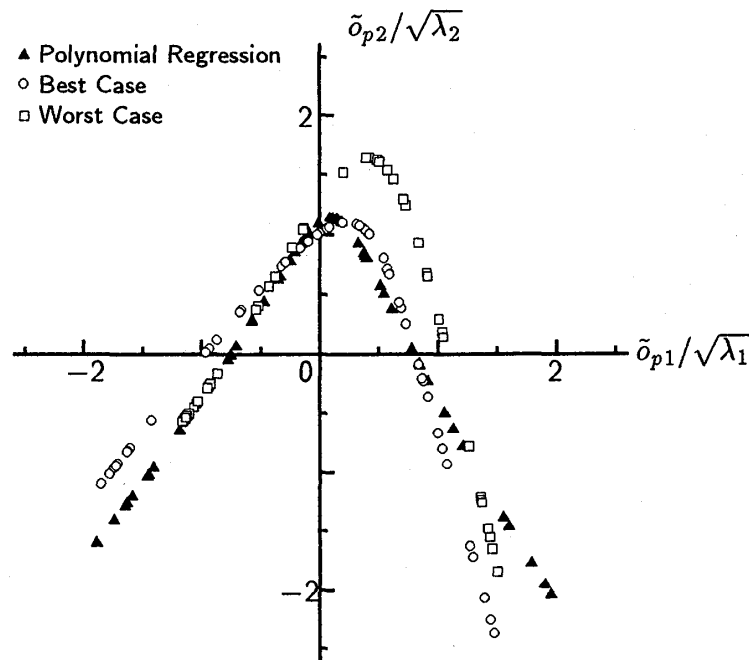


Figure 10: Distribution of principal components

7. おわりに

本稿では、観測雑音を含む関数近似問題を対象としてNNと回帰分析の関係について検討した。まず、モデル構造およびパラメータ推定手法の観点からNNと回帰分析について比較検討し、NNモデルにおける内部表現(中間層ユニット出力)が重要な役割を果たすこと示した。つぎに、雑音存在下においてNNによる関数近似手法が生じる問題点について検討し、特に過剰学習を抑制するための学習法を提案した。さらに、主成分分析の適用によるNNの内部表現の一解析手法を提案した。最後に、数値計算によりNNおよび回帰モデルのモデリング精度の比較を行った。数値計算結果より、雑音分布に依存するものの過剰学習を防止することを目的とした提案学習法は適切な収束率 β を設定することにより、BP法及び回帰モデルに較べて良好なモデリング能力を獲得できることを示した。しかし、提案学習法の有効性を詳細に検討するために、各種雑音への対応、事前情報への対応や実データへの適用について検討する必要がある。また、他の手法 [1], [5] 等との比較についても今後検討したい。

【参考文献】

- [1] D. S. Chen and R. C. Jain. A robust back propagation learning algorithm for function approximation. IEEE Trans. on Networks, Vol. 5, No. 3, pp. 467-479, 1994.

- [2] D. E. Rumelhart, J. L. McClelland, and the PDP Research Group. Parallel Distributed Processing, Vol. I. MIT Press, Cambridge, MA, 1986.
- [3] I. Sekita, T. Kureta, D. K. Y. Chiu, and H. Asoh. A method to reduce redundant hidden nodes. IEICE Trans. on Inf. and Syst., Vol. E77-D, No. 4, pp. 443-449, 1994.
- [4] S. Tamura, M. Tateishi, M. Matsumoto, and S. Akita. Determination of the number of redundant hidden units in a three-layered feed-forward neural networks. In Proc. of IEEE/INNS Joint Conference on Neural Networks, Vol. I, pp. 335-338, Nagoya, JAPAN, 1993.
- [5] A. R. Webb. Functional approximation by feedforward networks: A least-squares approach to generalization. IEEE Trans. on Neural Networks, Vol. 5, No. 3, pp. 363-371, 1994.
- [6] 相吉英太郎, 根岸直樹, 羽深嘉宣. 最適化手法を用いたニューラルネットワークの汎化能力獲得学習法. 第3回 FANシンポジウム講演論文集, pp. 371-322, 1993.
- [7] 喜多一. ニューラルネットワークの汎化能力. システム制御情報学会誌, Vol.36, No.10, pp. 625-633, 1992.
- [8] 早川毅. 回帰分析の基礎. 朝倉書店, 1986.
- [9] 田中豊, 脇本和昌. 多変量解析法. 現代数学社, 1987.
- [10] 船橋賢一. ニューラルネットワークのcapabilityについて. 電子情報通信学会技術研究報告MBE88-52, 1988.
- [11] 渡辺栄治, 清水光. 関数近似問題に対する階層型ニューラルネットワークの汎化能力に関する検討. 第4回 FANシンポジウム講演論文集, 1994.