

# バックプロパゲーション・ニューラルネットワークに おける収束条件と学習の加速化

田口 功\*

## The Convergence Condition and Learning Acceleration in the Backpropagation Neural Network

Isao TAGUCHI

It is considered important to decrease the learning time for the backpropagation neural network. In this paper, the author has considered the convergence condition and the learning speed of a multilayered neural network by increasing  $k$  ( $k > 1$ ) for the Sigmoid Function of all neurons. As a result, a convergence condition is the input of all neurons to separate as far as possible from the origin for all input patterns in the multilayered neural network. Furthermore, the convergence speed is improved by increasing  $k$  ( $k > 1$ ), that the region satisfies the convergence condition for the input of all patterns. When a convergence condition is satisfied and  $k$  is increased, learning speed is very much improved in the backpropagation neural network.

---

\* たぐち・いさお：敬愛大学国際学部助教授 ニューロ・コンピューティング

Associate Professor of Computer Science, Faculty of International Studies, Keiai University; neurocomputing.

## 1. はじめに

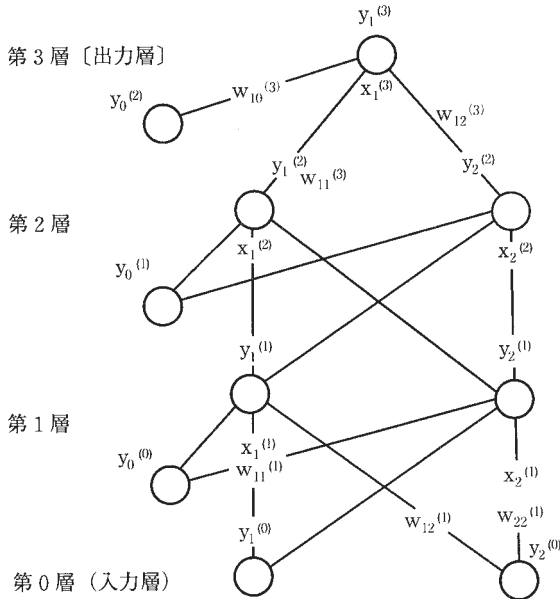
排他的論理と問題の学習課程で、時折正しい出力を与えない安定した解に収束することがある。ローカルミニマムにトラップされることを防ぐためにいくつかの提案がされている。重み係数を変化させる方法、内部ユニット数を変化させる方法、収束速度の改善としてモーメント項を追加する方法など、正しい出力を出し、更に収束速度を改善する方法などを含めて、いろいろ考えられている（ビール/ジャクソン 1993）。また、文献（Nielsen 1990）では、シグモイド関数に  $\sin$  関数のゆらぎを加えることによって、収束速度の改善が行われることを計算機でのシミュレーションで示した。なぜ  $\sin$  関数をゆらぎとして加えると収束速度の改善が行われるのかの式的な証明は、行われていない。本稿では、式的な関係は、シグモイド関数の  $k$  に対して  $k > 1$  の条件のもとで、誤差に対する重み係数偏微分式がすべて 0 になれば収束すると考え、その証明をグラフ（シグモイド関数の入力関係およびその微分形式）をもとに考察を行った。シミュレーションを行った結果、 $k$  の値を増加させればさせるほど収束速度が改善される理由は、すべてのニューロン入力が原点から離れることが必要条件であることが説明された。 $k$  の値が大きくなればなるほど、ニューロン入力は、比較的原点に近くとも収束速度が改善されることもわかった。更に、シミュレーションを行った結果、 $k$  の値にかかわらず、すべてのニューロンに対して、原点付近の入力  $x$  は、あたかも収束条件を満足する状態を探し、振動しているように観察された。更に、収束領域にはいると、各ニューロンの入力の大きさが原点から離れることが確かに必要な条件であることがグラフでもわかり、式的な考察と一致した。

後半では、 $k$  を切り替え学習を加速化するための注意点についての考察も行い、シミュレーションによって、その有効性を示した。

## 2. 3層ニューラルネットワークにおける 重み係数変化による誤差の変化について

図2.1に示すように、入力層（0層）の入力を $y_1^{(0)}$ 、 $y_2^{(0)}$ とし、1層の出力を $y_1^{(1)}$ 、 $y_2^{(1)}$ 、2層の出力を $y_1^{(2)}$ 、 $y_2^{(2)}$ 、3層の出力を $y_1^{(3)}$ とする。また、入力を持たない $y_0^{(0)}$ 、 $y_0^{(1)}$ 、 $y_0^{(2)}$ は、重み係数と共にしきい値を表現するための出力である。重み係数は、0層から1層の間を考えた場合、6個存在する。例えば、重み係数 $w_{21}^{(1)}$ とは、0層から1層の信号伝達を行い、0層の1番目の素子から、1層の2番目の素子の信号伝達の割合を決定する係数である。すなわち、 $w_{ij}^{(m)}$ と一般的に表現すれば、 $m-1$ 層の $j$ 番目の素子から、 $m$ 層の $i$ 番目に向かう信号伝達を行う重み係数を意味する。また、ここでは隠れ層は、一般的な重み係数の更新に対する性質を考察する

図2.1 3層ニューラルネットワーク



ために1層および2層の場合を考える（シミュレーションでは、1層のみで考察を行った）。

次に、0層に、入力信号が入力された場合の出力計算を行う。最初に、1層の1番目の素子への入力を計算すると、

$$\mathbf{x}_1^{(1)} = \mathbf{y}_1^{(0)} \mathbf{w}_{11}^{(1)} + \mathbf{y}_2^{(0)} \mathbf{w}_{12}^{(1)} - \mathbf{y}_0^{(0)} \mathbf{w}_{10}^{(1)} \quad (2-1)$$

となる。同様に、

$$\mathbf{x}_2^{(1)} = \mathbf{y}_1^{(0)} \mathbf{w}_{21}^{(1)} + \mathbf{y}_2^{(0)} \mathbf{w}_{22}^{(1)} - \mathbf{y}_0^{(0)} \mathbf{w}_{20}^{(1)} \quad (2-2)$$

となる。 $\mathbf{x}_1^{(1)}$  および  $\mathbf{x}_2^{(1)}$  に対して、出力は、一般的にシグモイド関数を用いて1層目の1番目の出力および2番目の出力は計算され、次に示す式で表される。

$$\mathbf{y}_1^{(1)} = \mathbf{F}(\mathbf{x}_1^{(1)}) \quad (2-3)$$

$$\mathbf{y}_2^{(1)} = \mathbf{F}(\mathbf{x}_2^{(1)}) \quad (2-4)$$

$$\mathbf{y}_0^{(1)} = 1 \quad (2-5)$$

ここで、(2-3)式における $\mathbf{y}_1^{(1)}$ は、隠れ層1層目の1番目出力、(2-4)式の $\mathbf{y}_2^{(1)}$ は、隠れ層1層目の2番目の出力を示す。隠れ層2層目の入力および出力は、

$$\mathbf{x}_1^{(2)} = \mathbf{y}_1^{(1)} \mathbf{w}_{11}^{(2)} + \mathbf{y}_2^{(1)} \mathbf{w}_{12}^{(2)} - \mathbf{y}_0^{(1)} \mathbf{w}_{10}^{(2)} \quad (2-6)$$

$$\mathbf{x}_2^{(2)} = \mathbf{y}_1^{(1)} \mathbf{w}_{21}^{(2)} + \mathbf{y}_2^{(1)} \mathbf{w}_{22}^{(2)} - \mathbf{y}_0^{(1)} \mathbf{w}_{20}^{(2)} \quad (2-7)$$

$$\mathbf{y}_1^{(2)} = \mathbf{F}(\mathbf{x}_1^{(2)}) \quad (2-8)$$

$$\mathbf{y}_2^{(2)} = \mathbf{F}(\mathbf{x}_2^{(2)}) \quad (2-9)$$

$$\mathbf{y}_0^{(2)} = 1 \quad (2-10)$$

となる。また3層目の入力および出力は、

$$\mathbf{x}_1^{(3)} = \mathbf{y}_1^{(2)} \mathbf{w}_{11}^{(3)} + \mathbf{y}_2^{(2)} \mathbf{w}_{12}^{(3)} - \mathbf{y}_0^{(2)} \mathbf{w}_{10}^{(3)} \quad (2-11)$$

$$\mathbf{y}_1^{(3)} = \mathbf{F}(\mathbf{x}_1^{(3)}) \quad (2-12)$$

となる。ここで期待出力を $d_1$ と置くと、その誤差は $\mathbf{y}_1^{(3)} - d_1$ となるので、2乗誤差関数は、

$$E = 1/2 (\mathbf{y}_1^{(3)} - d_1)^2 \quad (2-13)$$

となる。したがって、出力層に対する重み係数の偏微分関数 $\partial E / \partial \mathbf{w}_{10}^{(3)}$ 、 $\partial E / \partial \mathbf{w}_{11}^{(3)}$ 、 $\partial E / \partial \mathbf{w}_{12}^{(3)}$ は、偏微分に関する連鎖率とシグモイド関数の

微分に関する性質を利用すれば、

$$\begin{aligned}
 \partial E / \partial w_{10}^{(3)} &= \partial E / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial w_{10}^{(3)} \\
 &= \partial E / \partial y_1^{(3)} \cdot \partial y_1^{(3)} / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial w_{10}^{(3)} \\
 &= \partial / \partial y_1^{(3)} \cdot \{1/2 (y_1^{(3)} - d_1)^2\} \cdot \{ky_1^{(3)} (1 - y_1^{(3)})\} \cdot \\
 &\quad \{-y_0^{(2)}\} \\
 &= (y_1^{(3)} - d_1) ky_1^{(3)} (1 - y_1^{(3)}) (-y_0^{(2)}) \quad (2-14)
 \end{aligned}$$

$$\begin{aligned}
 \partial E / \partial w_{11}^{(3)} &= \partial E / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial w_{11}^{(3)} \\
 &= \partial E / \partial y_1^{(3)} \cdot \partial y_1^{(3)} / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial w_{11}^{(3)} \\
 &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) (y_1^{(2)}) \quad (2-15)
 \end{aligned}$$

$$\begin{aligned}
 \partial E / \partial w_{12}^{(3)} &= \partial E / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial w_{12}^{(3)} \\
 &= \partial E / \partial y_1^{(3)} \cdot \partial y_1^{(3)} / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial w_{12}^{(3)} \\
 &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) (y_2^{(2)}) \quad (2-16)
 \end{aligned}$$

となる。ここで (2-14) 式、(2-15) 式および (2-16) 式をみると、第 3 層の出力層への重み係数変化による微分式は、教師信号と実際の出力差 および シグモイド関数の微分値 ( $k$  の 1 乗を含む) および 第 2 層からの出力 (重み係数の位置によって異なる) の 3 つの要素の積によって決定されることがわかる。更に、第 1 層から第 2 層に結合している重み係数変化による誤差変化 ( $\partial E / \partial w_{21}^{(2)}$ ) を例として計算すると、

$$\begin{aligned}
 \partial E / \partial w_{21}^{(2)} &= \partial E / \partial x_2^{(2)} \cdot \partial x_2^{(2)} / \partial w_{21}^{(2)} \\
 &= \partial E / \partial y_2^{(2)} \cdot \partial y_2^{(2)} / \partial x_2^{(2)} \cdot \partial x_2^{(2)} / \partial w_{21}^{(2)} \\
 &= \partial E / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial y_2^{(2)} \cdot \partial y_2^{(2)} / \partial x_2^{(2)} \cdot \partial x_2^{(2)} / \partial w_{21}^{(2)} \\
 &= \partial E / \partial x_1^{(3)} \cdot w_{12}^{(3)} (ky_2^{(2)}) \cdot (1 - y_2^{(2)}) \cdot y_1^{(1)} \\
 &= \partial E / \partial y_1^{(3)} \cdot \partial y_1^{(3)} / \partial x_1^{(3)} \cdot w_{12}^{(3)} \cdot ky_2^{(2)} (1 - y_2^{(2)}) \cdot y_1^{(1)} \\
 &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) \cdot w_{12}^{(3)} \cdot ky_2^{(2)} (1 - y_2^{(2)}) \cdot \\
 &\quad y_1^{(1)} \quad (2-17)
 \end{aligned}$$

となる。(2-17) 式をみると、第 2 層から第 3 層の重み係数変化による誤差関数の偏微分式は、 $w_{12}^{(3)}$  の重み係数を通して、出力誤差  $(y_1^{(3)} - d_1)$  および第 3 層の出力微分  $(ky_1^{(3)} (1 - y_1^{(3)}))$  および  $w_{21}^{(2)}$  と結合している出力微分と  $w_{21}^{(2)}$  と結合している 1 層前の出力の積となることがわかる。

全く同様な考えで他の第2層における偏微分式は導くことができるので、その結果を以下に示す。

$$\begin{aligned} \partial E / \partial w_{11}^{(2)} &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) \cdot w_{11}^{(3)} \cdot ky_1^{(2)} \cdot \\ &\quad (1 - y_1^{(2)}) \cdot y_1^{(1)} \end{aligned} \quad (2-18)$$

$$\begin{aligned} \partial E / \partial w_{12}^{(2)} &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) \cdot w_{11}^{(3)} \cdot ky_1^{(2)} \cdot \\ &\quad (1 - y_1^{(2)}) \cdot y_2^{(1)} \end{aligned} \quad (2-19)$$

$$\begin{aligned} \partial E / \partial w_{10}^{(2)} &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) \cdot w_{11}^{(3)} \cdot ky_1^{(2)} \cdot \\ &\quad (1 - y_1^{(2)}) \cdot y_0^{(1)} \end{aligned} \quad (2-20)$$

$$\begin{aligned} \partial E / \partial w_{22}^{(2)} &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) \cdot w_{12}^{(3)} \cdot ky_2^{(2)} \cdot \\ &\quad (1 - y_2^{(2)}) \cdot y_2^{(1)} \end{aligned} \quad (2-21)$$

$$\begin{aligned} \partial E / \partial w_{20}^{(2)} &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) \cdot w_{12}^{(3)} \cdot ky_2^{(2)} \cdot \\ &\quad (1 - y_2^{(2)}) \cdot y_0^{(1)} \end{aligned} \quad (2-22)$$

次に第0層（入力層）と第1層の間の重み係数変化に対する偏微分式を考える。同様に考えれば、

$$\begin{aligned} \partial E / \partial w_{12}^{(1)} &= \partial E / \partial x_1^{(1)} \cdot \partial x_1^{(1)} / \partial w_{12}^{(1)} \\ &= \partial E / \partial x_1^{(1)} \cdot y_2^{(0)} \\ &= \partial E / \partial y_1^{(1)} \cdot \partial y_1^{(1)} / \partial x_1^{(1)} \cdot y_2^{(0)} \\ &= \partial E / \partial y_1^{(1)} \cdot ky_1^{(1)} (1 - y_1^{(1)}) \cdot y_2^{(0)} \\ &= \partial E / \partial y_1^{(1)} \cdot ky_1^{(1)} (1 - y_1^{(1)}) \cdot y_2^{(0)} \\ &= (\partial E / \partial x_1^{(2)} \cdot \partial x_1^{(2)} / \partial y_1^{(1)} + \partial E / \partial x_2^{(2)} \cdot \partial x_2^{(2)} / \partial y_1^{(1)}) \cdot \\ &\quad ky_1^{(1)} (1 - y_1^{(1)}) \cdot y_2^{(0)} \end{aligned} \quad (2-23)$$

となる。ここで、(2-17) 式を導く過程で使用した式を用い、

$$\begin{aligned} \partial E / \partial x_1^{(2)} &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) \cdot w_{11}^{(3)} \cdot ky_1^{(2)} (1 - y_1^{(2)}) \\ &= A \end{aligned} \quad (2-24)$$

$$\begin{aligned} \partial E / \partial x_2^{(2)} &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) \cdot w_{12}^{(3)} \cdot ky_2^{(2)} (1 - y_2^{(2)}) \\ &= B \end{aligned} \quad (2-25)$$

と置くと、

$$\begin{aligned} \partial E / \partial w_{12}^{(1)} &= (A \cdot w_{11}^{(2)} + B \cdot w_{21}^{(2)}) \cdot ky_1^{(1)} (1 - y_1^{(1)}) \cdot y_2^{(0)} \end{aligned} \quad (2-26)$$

となる。ここで  $(A \cdot w_{11}^{(2)} + B \cdot w_{21}^{(2)})$  は、重み係数を2回通しての出力差、 $ky_1^{(1)} \cdot (1 - y_1^{(1)})$  は、重み係数の先の出力微分、 $y_2^{(0)}$  は、重み係数の手前からの入力と考えることができる。他の重み係数に対する偏微分も同様に

$$\frac{\partial E}{\partial w_{11}^{(1)}} = (A \cdot w_{11}^{(2)} + B \cdot w_{21}^{(2)}) \cdot ky_1^{(1)} \cdot (1 - y_1^{(1)}) \cdot y_1^{(0)} \quad (2-27)$$

$$\frac{\partial E}{\partial w_{21}^{(1)}} = (A \cdot w_{12}^{(2)} + B \cdot w_{22}^{(2)}) \cdot ky_2^{(1)} \cdot (1 - y_2^{(1)}) \cdot y_1^{(0)} \quad (2-28)$$

$$\frac{\partial E}{\partial w_{22}^{(1)}} = (A \cdot w_{12}^{(2)} + B \cdot w_{22}^{(2)}) \cdot ky_2^{(1)} \cdot (1 - y_2^{(1)}) \cdot y_2^{(0)} \quad (2-29)$$

$$\frac{\partial E}{\partial w_{10}^{(1)}} = (A \cdot w_{12}^{(2)} + B \cdot w_{22}^{(2)}) \cdot ky_1^{(1)} \cdot (1 - y_1^{(1)}) \cdot y_0^{(0)} \quad (2-30)$$

$$\frac{\partial E}{\partial w_{20}^{(1)}} = (A \cdot w_{12}^{(2)} + B \cdot w_{22}^{(2)}) \cdot ky_2^{(1)} \cdot (1 - y_2^{(1)}) \cdot y_0^{(0)} \quad (2-31)$$

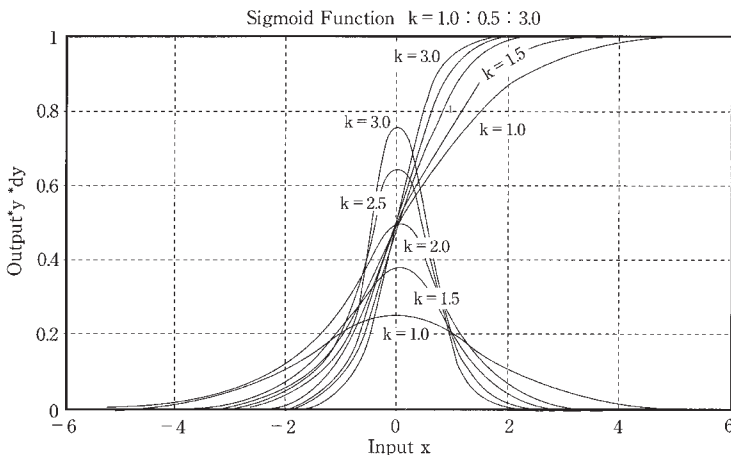
となる。以上をまとめて、各偏微分式を分解して考えると、次のことが言える。

(1) 出力計算終了後、重み係数更新を考えすべての偏微分式をみると、その計算式の中に共通の項が存在し、それらは、教師データ、各層出力、重み係数によって構成されている。更に、出力層の影響が隠れ層、入力層に次々に影響を与えている。

(2)  $k$ の次数をみると、出力層と結びつく重み係数偏微分の次数は1次、中間層と中間層を結びつける重み係数微分は2次、更に、入力層に向かうにつれて増加する。したがって、 $k$ によって重み係数微分の更新に変化をつけることができる。従来は、学習率によって、重み係数偏微分式に対して一様に ( $k=1$ として) 重み係数更新を行っていたことになる。

(3) 重み係数偏微分式をみると、シグモイド関数に対する微分項を必ず含む。したがって、 $y_1^{(1)}$ 、 $y_1^{(2)}$ 、 $y_1^{(3)}$  の値がすべて1または0に近づけば、重み係数に対する誤差の偏微分値は、すべて0に近づくことになる。これは、極小値(最小値)を意味する(美多・小郷 1979)。シグモイド関数の性

図2.2 kの変化によるシグモイド関数および微分値の変化



質から、 $k$ の選択により、出力を1または0とする $x$ （入力）の領域を増加させることによってこれは可能となる。 $k$ の増加によって、出力が1または0の値を持つ領域が増加する様子を図2.2に示す。また、 $k$ の増加にしたがって、シグモイド関数の微分値が0となる領域が増加する様子も同時に示した。 $k$ は、1.0、1.5、2.0、2.5、3.0とした。

### 3. 排他的論理和問題の収束条件について

バックプロパゲーション・ニューラルネットワークは、ニューロコンピュータの発達の歴史を考えても最もよく知られている。ここでは、最も基本的な排他的論理和問題を取り上げ、学習に対する収束速度の改善に対する考察を行うこととする。ニューラルネットワークを構成するとは、すべての学習させたい入力に対して有効な出力を取り出すために誤差関数を定義し、その誤差を最小にする重み係数を決定することである。また、学習回数を減少させることも研究目標の1つである。従来、重み係数の更新は、最急降下法が用いられ、学習率は経験的に設定していた（八名・鈴木 1992a、同 1992b、ビール/ジャクソン 1993）。学習率を大きくすると、収束は速くなる



が不安定になることは、よく知られている。小さくすると、安定性は増加するが、収束までに時間がかかることになる。(松井・石見 1994) によれば、しきい値ゆらぎを持つ(正弦関数)ニューロンモデルを考え、ゆらぎによる学習回数の減少や収束率の改善が報告されている。また、ゆらぎ入力は、中間層のみに限定されている。本稿では、ゆらぎを与えるのではなくシグモイド関数のkを大きくすることによって、収束速度の改善が行われることを示すと共に、式的な検討も行った。本稿では、一般的に知られているシグモイド関数 ( $y=1/(1+\exp(-kx))$ ) をもとに、微分形式が出力をyとすると、 $(ky(1-y))$  となること、および、 $(ky(1-y))$  の値が、kを1より大きくすればするほど入力xに対して0に近づく領域が増加することを考慮した。更に、収束速度とkの関係をグラフと関連させながらkの値を1より大きくすると、収束速度が改善されることの式的な検討を行った。特に、収束時には、誤差に対する、重み係数の微分式を考えた場合、すべての偏微分値(すべてのニューロン)が0に近づくためには、すべてのニューロン入力xが原点からより離れることが必要であることを述べる。特に、平均二乗誤差の大きさをより小さくするためには、ニューロン入力xが原点からより離れることが必要な条件となる。しかし、k=1としたシグモイド関数では、この実現には、xの絶対値が非常に大きくなければならない。シグモイド関数 ( $y=1/(1+\exp(-kx))$ ) の微分形式が  $(ky(1-y))$  となり、出力層の  $(ky(1-y))$  の関係式は、重み係数の更新式にすべて影響を与えることになる。中間層および入力層のニューロンも同様に、誤差に対する重み係数の偏微分式に対して、 $(ky(1-y))$  の関係を含むことから出力層と同様に、kの値を1より大きくすることで、 $(ky(1-y))$  の絶対値が1より非常に小さくなる領域を増加させることが可能になる。特に、出力層のニューロンに対する入力xは重要となり、教師信号に近づくことと収束速度を考慮する必要があるために、中間層、並びに入力層と条件が異なることを述べる。

最初に、出力層の重み係数変化における、誤差の変化 ( $\partial E/\partial w_{10}^{(3)}$ ) を0とするための条件を考える。次の(2-15)式

$$\begin{aligned}
\partial E / \partial w_{11}^{(3)} &= \partial E / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial w_{11}^{(3)} \\
&= \partial E / \partial y_1^{(3)} \cdot \partial y_1^{(3)} / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial w_{11}^{(3)} \\
&= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) (y_1^{(2)}) \quad (2-15)
\end{aligned}$$

から次のことが言える。

(1) シグモイド関数の微分形式  $ky_1^{(3)} (1 - y_1^{(3)})$  は、グラフからもわかるように  $y_1^{(3)}$  が 0 に近づくかまたは 1 に近づくとき 0 に近づくことは明らかである。

(2)  $y_1^{(3)}$  は、シグモイド関数の出力であり、 $y_1^{(3)}$  が 0 に近づくとき、入力  $x_1^{(3)}$  は  $-\infty$  に近づく。 $y_1^{(3)}$  が 1 に近づくとき入力  $x_1^{(3)}$  は  $\infty$  に近づく。

(3) (1)および(2)から、シグモイド関数の微分形式  $ky_1^{(3)} (1 - y_1^{(3)})$  が 0 に近づくためには、入力  $x_1^{(3)}$  は  $-\infty$  に近づくか、または  $\infty$  に近づかねばならない。

(4) したがって、 $\partial E / \partial w_{10}^{(3)} = 0$  とするために、入力  $x_1^{(3)}$  は  $-\infty$  に近づくか、または  $\infty$  に近づかねばならない。

(5) 更に、(2-14) 式から、出力  $y_1^{(3)}$  が教師出力  $d_1$  に近づけば、他の項 ( $y_1^{(2)}$ ) の絶対値は、1 を超えないから、(2-14) 式は、0 となる。

(6) 排他的論理和問題を考えると、4 パターンの入力に対して 4 パターンの出力があるが、それぞれに対して、(1)から(5)までの条件を満足しなければならない。すなわち、入力が (1,1) のとき、出力  $y_1^{(3)}$  は、0 とならなくてはならない。したがって、(2)の条件は、入力  $x_1^{(3)}$  は  $-\infty$  に近づかねばならない。入力が (0,0) のとき、 $x_1^{(3)}$  は、同様に  $-\infty$  に近づかねばならない。逆に、(1,0) および (0,1) のときは、入力  $x_1^{(3)}$  は、 $\infty$  に近づかねばならない。

(7) (2-16) 式に対しても全く同様な条件が成立すれば、 $\partial E / \partial w_{12}^{(3)} = 0$  となる。ただし、後で述べるが、 $y_2^{(2)}$  は、中間層の収束条件を考えると、0 または 1 に近づけば、収束条件を満足することになることに注意する。また、しきい値を決定する重み係数に関する偏微分式  $\partial E / \partial w_{10}^{(3)}$  も他の出力層の偏微分式と全く同様な関係が成立すれば、0 となる。ただし、 $y_0^{(2)} = 1$  固定である。

次に、中間層の誤差に対する重み係数の偏微分式を

$$\begin{aligned}
 \partial E / \partial w_{21}^{(2)} &= \partial E / \partial x_2^{(2)} \cdot \partial x_2^{(2)} / \partial w_{21}^{(2)} \\
 &= \partial E / \partial y_2^{(2)} \cdot \partial y_2^{(2)} / \partial x_2^{(2)} \cdot \partial x_2^{(2)} / \partial w_{21}^{(2)} \\
 &= \partial E / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial y_2^{(2)} \cdot \partial y_2^{(2)} / \partial x_2^{(2)} \cdot \partial x_2^{(2)} / \partial w_{21}^{(2)} \\
 &= \partial E / \partial x_1^{(3)} \cdot w_{12}^{(3)} (ky_2^{(2)}) \cdot (1 - y_2^{(2)}) \cdot y_1^{(1)} \\
 &= \partial E / \partial y_1^{(3)} \cdot \partial y_1^{(3)} / \partial x_1^{(3)} \cdot w_{12}^{(3)} \cdot ky_2^{(2)} (1 - y_2^{(2)}) \cdot y_1^{(1)} \\
 &= (y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)}) \cdot w_{12}^{(3)} \cdot ky_2^{(2)} (1 - y_2^{(2)}) \cdot y_1^{(1)}
 \end{aligned} \tag{2-17}$$

を代表として考える。

(8) (2-17) 式をみると前半の、 $(y_1^{(3)} - d_1) \cdot ky_1^{(3)} (1 - y_1^{(3)})$  は、出力層の収束条件を満足すればよいということになるので、後半の  $w_{12}^{(3)} \cdot ky_2^{(2)} (1 - y_2^{(2)}) \cdot y_1^{(1)}$  の部分が、0 に収束すれば、 $\partial E / \partial w_{21}^{(2)} = 0$  となるので、そのための条件を述べる。

(9)  $w_{12}^{(3)}$  は、結合係数であるから、一般的には、0 とすることができない。

(10)  $ky_2^{(2)} (1 - y_2^{(2)})$  に対して、出力層の重み係数に対する偏微分の考えをそのまま適用すると、 $ky_2^{(2)} (1 - y_2^{(2)}) = 0$  にする条件は、 $y_2^{(2)} = F(x_2^{(2)})$  の関係からわかるように、 $y_2^{(2)} = 0$  または、 $y_2^{(2)} = 1$  に対する入力を考えればよいから、 $x_2^{(2)} = -\infty$  または、 $x_2^{(2)} = \infty$  の条件が必要となる。すなわち、中間層のニューロン入力値を原点から離すことによってこれは実現できる。

(11) (2-18) 式から (2-22) 式をみると、同様な関係が成立すれば、それらの関係式を 0 にできる。すなわち、出力層での入出力収束条件を満足し、かつ中間層 (2 層目) ニューロンの入力を原点から離すことが収束の条件となる。

(12) このことから、中間層 1 層目の条件は、同様に、中間層 (2 層目) ニューロンの入力を原点から離すこととを含むことになる。すなわち、入力層に近づくにしたがって前層の収束条件を満たすことと、ニューロン入力が原点から離れれば、離れるほど収束速度は、改善されることになる。

(13) すなわち、重み係数更新に対する、バックプロパゲーション則の適用順序にそのまま収束性をあてはめると、すべてのニューロン入力 that 原点から離れれば離れるほど、重み係数に対する偏微分値が0に近づき誤差の精度も上昇すると考えられる。

## 4. 計算機による収束性についてのシミュレーション

### 1. 平均 2 乗誤差0.01、 $k=2$ の場合

ここでは、中間層は1つでニューロン数を2とした。重み係数初期ベクトル $W$ は乱数を発生させ、 $W=[w_{11}^{(2)} w_{12}^{(2)} w_{10}^{(2)} w_{11}^{(1)} w_{12}^{(1)} w_{10}^{(1)} w_{21}^{(1)} w_{22}^{(1)} w_{20}^{(1)}]^T=[0.2190 \ 0.0470 \ 0.6789 \ 0.6793 \ 0.9347 \ 0.3835 \ 0.5194 \ 0.8310 \ 0.0346]^T$ とした。更に、ここで、入力データ (AおよびB) は $[1 \ 0 \ 1 \ 0]$   $[1 \ 1 \ 0 \ 0]$ とした。したがって、期待出力 $= [0 \ 1 \ 1 \ 0]$ である。学習係数は、1.53一定とした。 $k=1$ の場合は、4万3,202回の学習回数を要し非常に収束速度が悪かった。すべてのニューロンの $k$ を2とした場合の学習回数は、108回となった。4個の出力に対して、2乗誤差 ( $E=1/2 (y_i^{(3)} - d_i)^2$   $i=1,4$ ) を図4.1—図4.4に示した。ここで4個の誤差の平均が、0.01以下となったときに学習を終わりにした。更に、重み係数の更新に対して、使用される各層の  $\partial E / \partial w_{ij}^{(m)}$  \* 学習係数の変化の様子を図4.5—図4.13に示した。更に、4個の入力パターンに対して出力層に対してのニューロン入力の様子を図4.14—図4.17に示した。図4.14—図4.17をみると、ともに収束時には、振動状態が0に近い値となっていることがわかる。また、学習回数が95回を超えると出力層のニューロン入力が4パターンに対して、それぞれ $[-1 \ 1 \ 1 \ -1.2]$ に近い値になっていることがわかる。したがって、出力層への4パターンの教師信号に対応して、ニューロンの誤差に対する出力は、 $[0 \ 1 \ 1 \ 0]$ になっていることがわかる。したがって3で述べた収束条件、すなわち、原点からニューロン入力が離れることの結果と一致すると考えられる。平均2乗誤差を0.01から0.001にした結果を4・2に示す。

## 2. 平均 2 乗誤差0.001、 $k=2.5$ の場合

ここでは、4・1と比較して、平均 2 乗誤差が0.001以下となった場合に学習を終了するという設定にした。学習率、重み係数は全く等しい値を用いた。図4.18—図4.21までは、4パターンに対する 2 乗誤差を示し、図4.22—図4.25は、出力層ニューロンの 4パターンに対する入力 $x$ 、図4.26—図4.29は、中間層の 1 番目のニューロンの 4パターンに対する入力 $x$ 、図4.30—図4.33は、中間層の 2 番目のニューロンの 4パターンに対する入力 $x$ をグラフで示したものである。出力層の入力値は、平均 2 乗誤差を0.001と設定したために、[-1.5 -1.4 -1.4 -1.4]となった。中間層のニューロンへの入力値は、最終的には、[0.95 -1.7 -2.0 -4.5 2.75 0.85 0.85 -1.0]となった。これらの結果から、少なくとも出力層へのニューロンへの入力は、精度を上げるにしたがって原点からより離れなければならないことがわかる。更に、中間層についても原点から離れることが収束条件となることがわかる。すなわち、入力が原点から離れることと、すべての重み係数に対して、 $\partial E / \partial w_{ij}^{(m)} \approx 0$ の条件が成立することが収束条件になることから、このような結果になったものと考えられる。

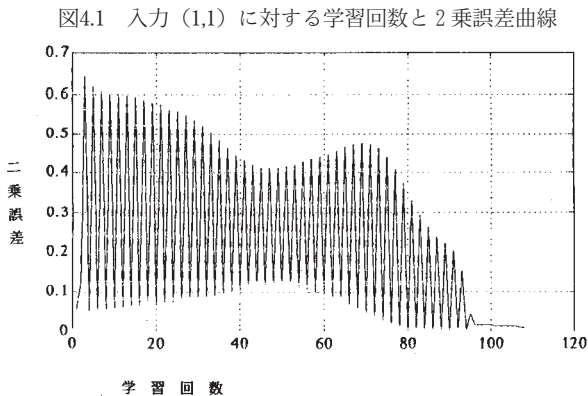


図4.2 入力 (0,1) に対する学習回数と2乗誤差曲線

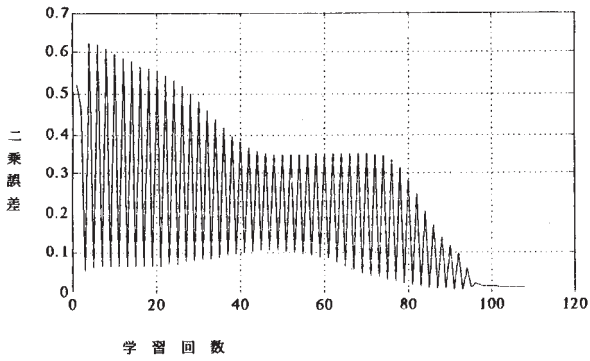


図4.3 入力 (1,0) に対する学習回数と2乗誤差曲線

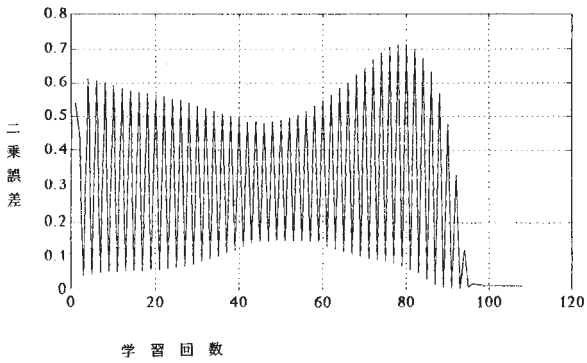


図4.4 入力 (0,0) に対する学習回数と2乗誤差曲線

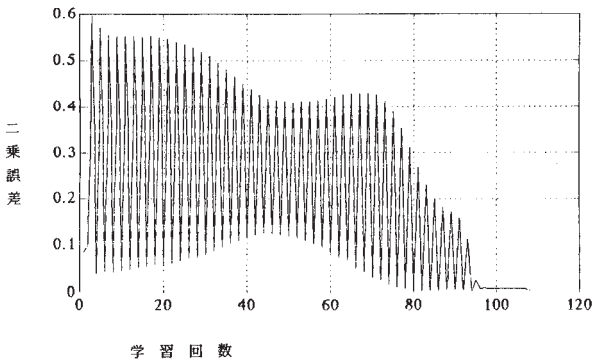


図4.5 学習回数と $1.53 * \partial E / \partial w_{11}^{(2)}$ 曲線

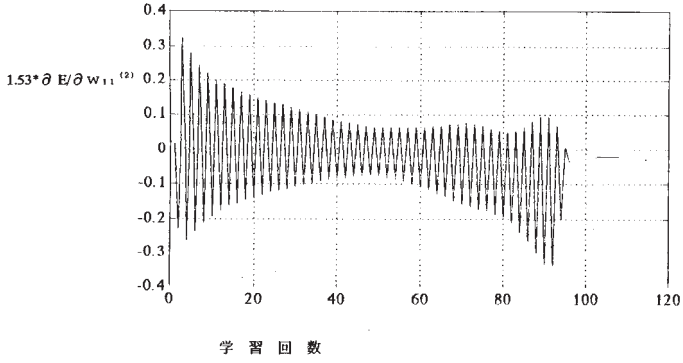


図4.6 学習回数と $1.53 * \partial E / \partial w_{12}^{(2)}$ 曲線

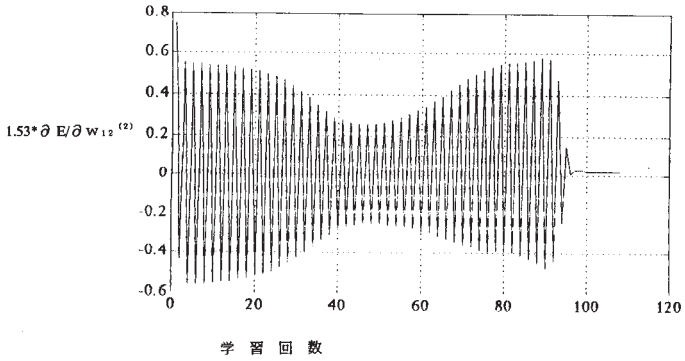


図4.7 学習回数と $1.53 * \partial E / \partial w_{10}^{(2)}$ 曲線

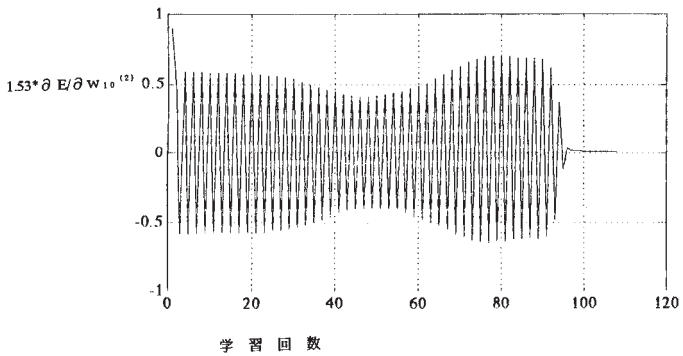


図4.8 学習回数と $1.53 * \partial E / \partial w_{11}^{(1)}$ 曲線

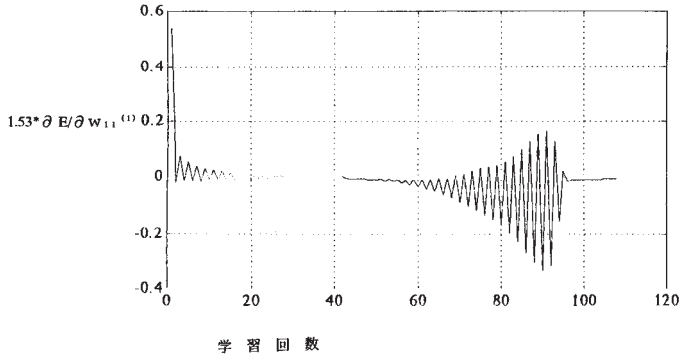


図4.9 学習回数と $1.53 * \partial E / \partial w_{21}^{(1)}$ 曲線

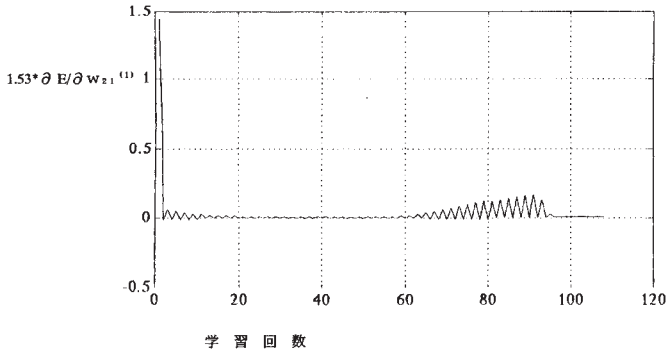


図4.10 学習回数と $1.53 * \partial E / \partial w_{10}^{(1)}$ 曲線

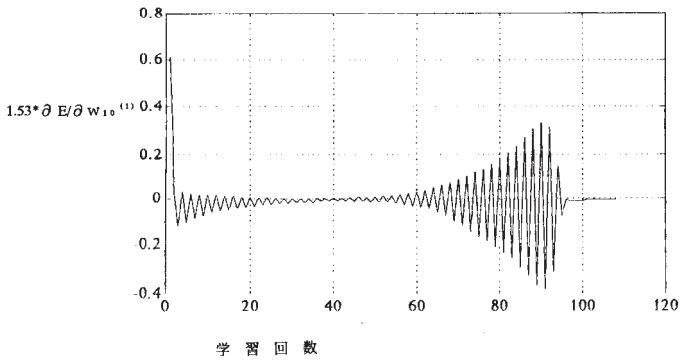




図4.11 学習回数と  $1.53 * \partial E / \partial w_{12}^{(1)}$  曲線

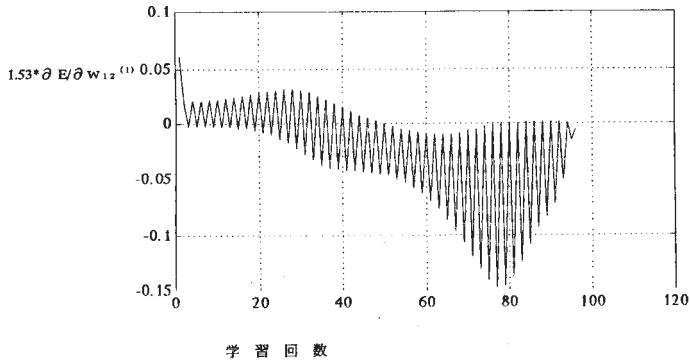


図4.12 学習回数と  $1.53 * \partial E / \partial w_{22}^{(1)}$  曲線

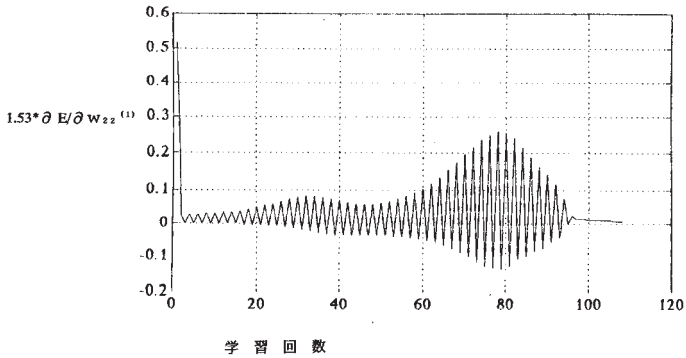


図4.13 学習回数と  $1.53 * \partial E / \partial w_{20}^{(1)}$  曲線

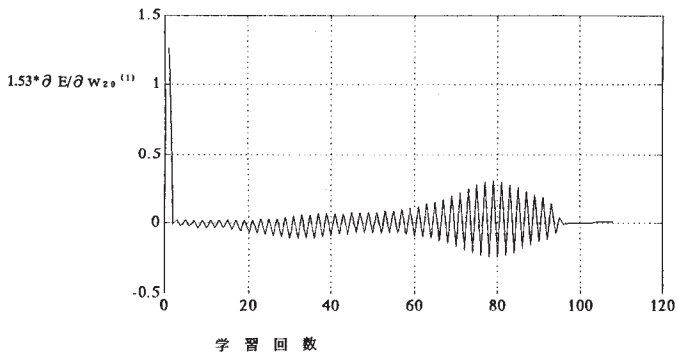


図4.14 学習回数と出力層のシグモイド関数入力 $x_1(1)$

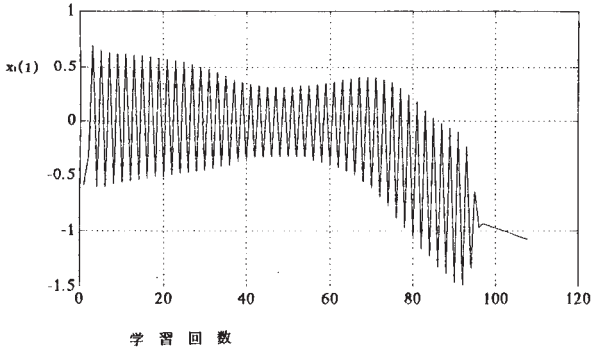


図4.15 学習回数と出力層のシグモイド関数入力 $x_1(2)$

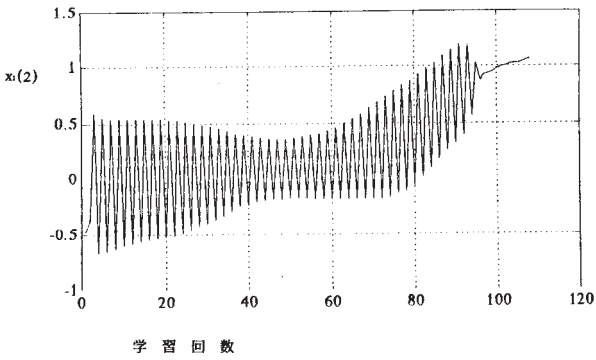


図4.16 学習回数と出力層のシグモイド関数入力 $x_1(3)$

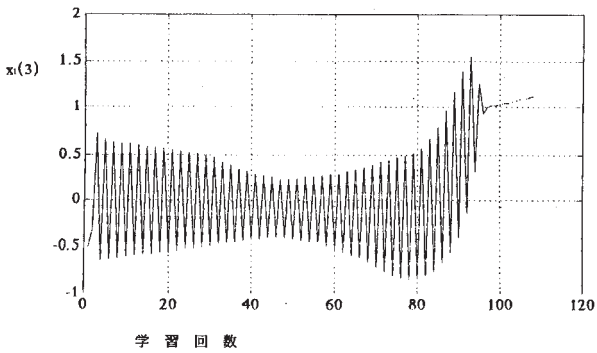


図4.17 学習回数と出力層のシグモイド関数入力 $x_i(4)$

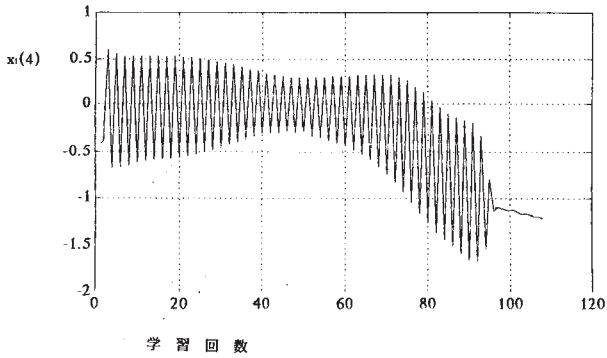


図4.18 入力 (1,1) に対する学習回数と 2 乗誤差曲線

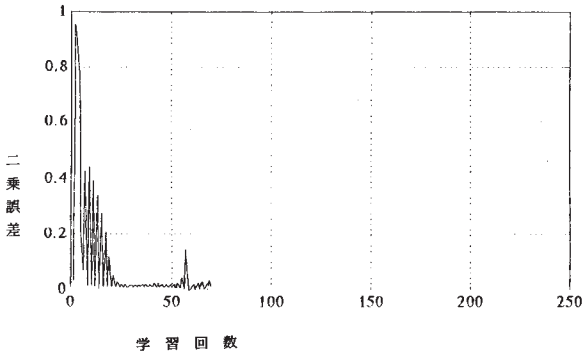


図4.19 入力 (0,1) に対する学習回数と 2 乗誤差曲線

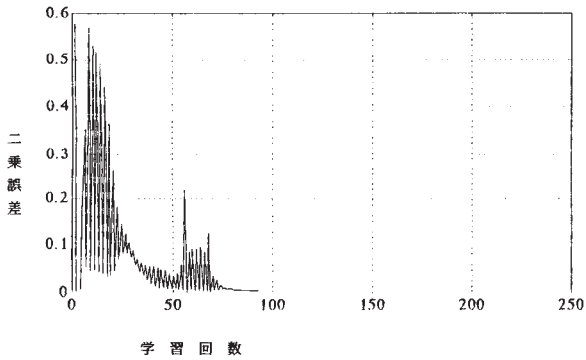


図4.20 入力 (1,0) に対する学習回数と 2 乗誤差曲線

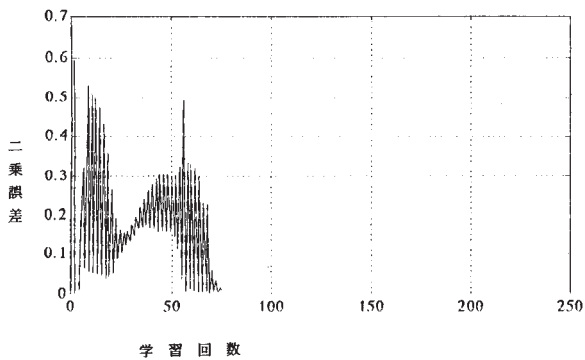


図4.21 入力 (0,0) に対する学習回数と 2 乗誤差曲線

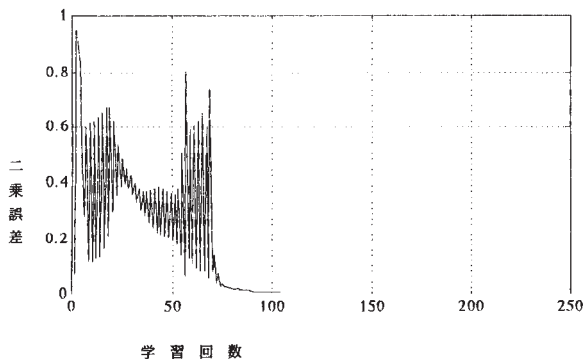


図4.22 学習回数と出力層のシグモイド関数入力  $x_1(1)$

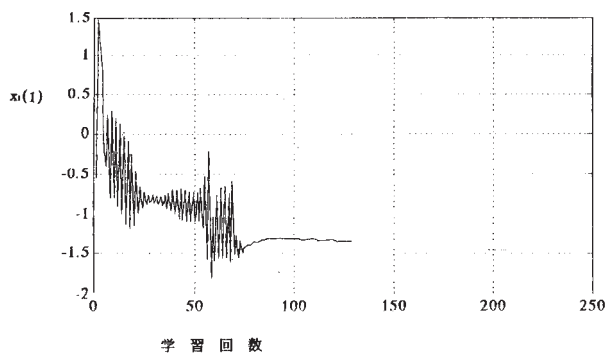


図4.23 学習回数と出力層のシグモイド関数入力 $x_i(2)$

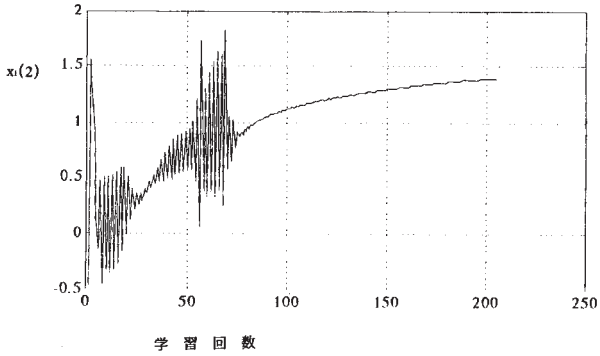


図4.24 学習回数と出力層のシグモイド関数入力 $x_i(3)$

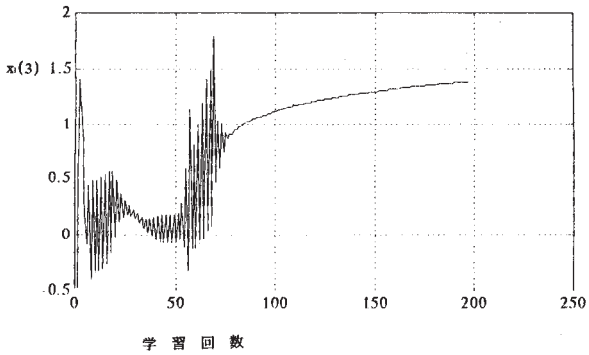


図4.25 学習回数と出力層のシグモイド関数入力 $x_i(4)$

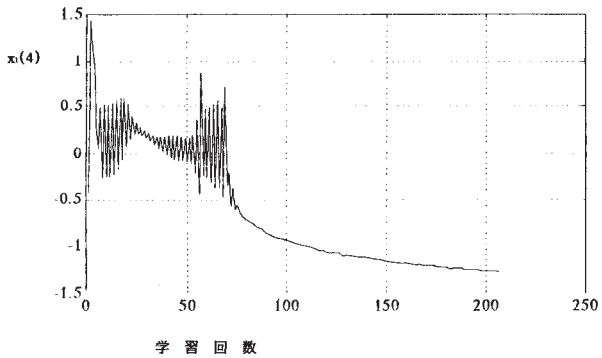


図4.26 学習回数と中間層のシグモイド関数入力 $x_2(1)$

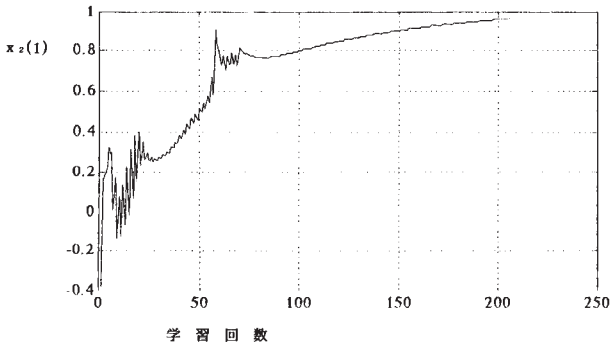


図4.27 学習回数と中間層のシグモイド関数入力 $x_2(2)$

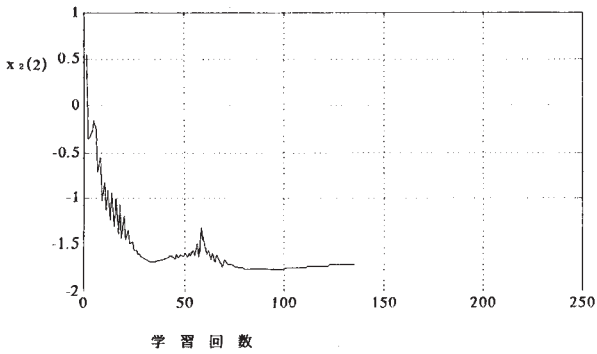


図4.28 学習回数と中間層のシグモイド関数入力 $x_2(3)$

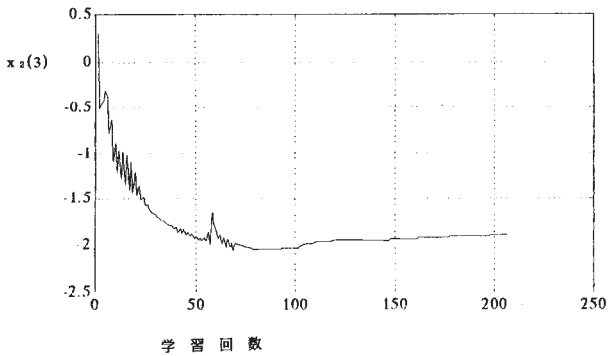


図4.29 学習回数と中間層のシグモイド関数入力 $x_2(4)$

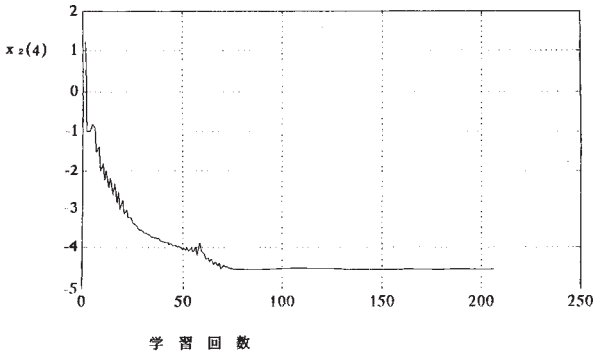


図4.30 学習回数と中間層のシグモイド関数入力 $x_3(1)$

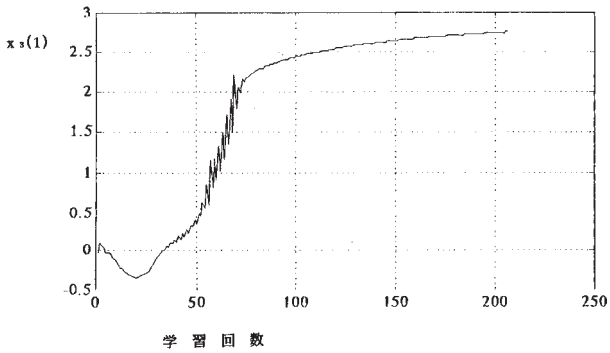


図4.31 学習回数と中間層のシグモイド関数入力 $x_3(2)$

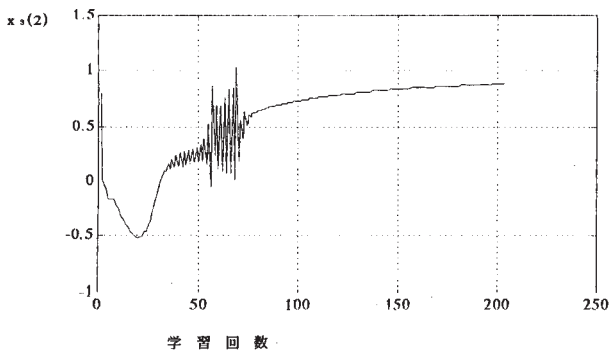


図4.32 学習回数と中間層のシグモイド関数入力 $x_3(3)$

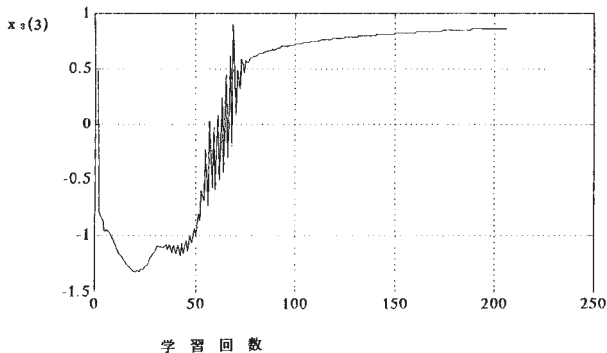
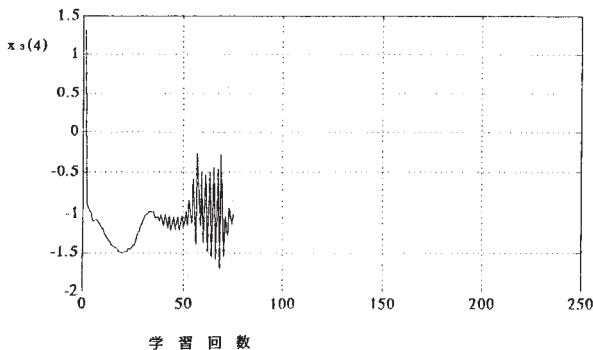


図4.33 学習回数と中間層のシグモイド関数入力 $x_3(4)$



## 5. 排他的論理和問題の収束速度の改善について

4で述べた収束条件のもとに、シグモイド関数の $k$ を切り替え、その値を大きくすることによって、学習回数を非常に少なくすることができることを示す。すなわち、すべての入力パターンに対して、すべての層へのすべての入力 $x$ が原点から離れることと、3の(6)の条件を満足する領域において、 $k$ の値を大きくすることによってより早く少ない誤差条件に合う学習を可能にすることを示す（特に、出力層の4個のパターン入力）。

ここでは、学習回数を $N$ とし、入力データ（AおよびBは $[1\ 0\ 1\ 0]$   $[1\ 1\ 0$



0] とした場合) に対し、出力層への4入力 $x(1)$ 、 $x(2)$ 、 $x(3)$ 、 $x(4)$ に対する変化を考察する。最初に、入力が(1,1)のときを考える。当然、出力は0とならなくてはならないので、 $x(1)=x_1^{(3)}$  は、 $-\infty$ に近づかねばならない。したがって、学習に対して、

$$\partial x(1) / \partial N \leq 0 \quad (5-1)$$

の関係が満足され、学習回数の増加に対して、原点からますます負の値に離れるか、または原点に近づく傾向にあるが、現時点では負の状態にあり、変動は少ないという条件が満足されていなければならない。入力が(0,0)のときも全く同様に、

$$\partial x(4) / \partial N \leq 0 \quad (5-2)$$

の関係が成立しなければならない。

逆に、入力(0 1) および(1 0) に対しては、学習回数を増やすにつれて、原点からますます正の方向に離れる条件、

$$\partial x(2) / \partial N \geq 0 \quad (5-3)$$

および、

$$\partial x(3) / \partial N \geq 0 \quad (5-4)$$

が、成立しなければならない。

原点から出力層への入力がより離れることが良い学習の条件(学習回数の減少)が整ったと考えられるので、(5-1)式から(5-4)式の関係が成立する時点で、3で述べた収束条件、すなわち、 $k$ の値を学習計算途中で切り替え増加させることによって、 $\partial E / \partial w_{ij}^{(m)} = 0$  に対する関係を満足させることができるものと考えられる( $k$ が増加することによって、すべての $\partial E / \partial w_{ij}^{(m)}$ はより0に近づく)。次の章で、切り替えた結果、(5-1)式から(5-4)式の関係が成立した場合に $k$ の値を増加した場合と、(5-1)式から(5-4)式の関係が成立しない場合に $k$ の値を増加した場合の結果を示す。

## 6. 収束速度の改善に対するシミュレーション

### 1. 切り替え条件が満足されている場合

平均2乗誤差0.001以下になったとき、学習を終了させる設定とした。k=2.5とし、中間層は1つでニューロン数を2とした。重み係数初期ベクトルWは乱数を発生させ、 $W = [w_{11}^{(2)} \ w_{12}^{(2)} \ w_{10}^{(2)} \ w_{11}^{(1)} \ w_{12}^{(1)} \ w_{10}^{(1)} \ w_{21}^{(1)} \ w_{22}^{(1)} \ w_{20}^{(1)}]^T = [0.2190 \ 0.0470 \ 0.6789 \ 0.6793 \ 0.9347 \ 0.3835 \ 0.5194 \ 0.8310 \ 0.0346]^T$ とした。更に、ここで、入力データ(AおよびB)は $[1 \ 0 \ 1 \ 0] \ [1 \ 1 \ 0 \ 0]$ とした。したがって、期待出力= $[0 \ 1 \ 1 \ 0]$ である。学習係数は、1.53一定とした。70回の学習終了後、k=4.0とした場合の学習回数に対する出力を図6.1から図6.4に示す。図6.5から図6.8に出力層ニューロンの4パターンに対する入力xの様子を示す。70回まで、k=2.5で学習を行い、71回目で、k=4.0に切り替えた結果である。図6.5から図6.8をみると、5で述べたように、55回から70回のグラフの様子で(直線)、(5-1)式から(5-4)式の条件を満足している状態であると考えられる。これらの条件を満足した状態で、k=4に切り替えたとき出力層への入力x(1)およびx(4)は、負の傾きを持ち、入力値は、原点から離れる結果となった。また、x(2)およびx(3)は逆に、正の傾きを持った。同時に、4個の入力パターンに対して、出力も期待出力に更に近づいた。これは、図6.1から図6.4に示す結果から明らかである。70回を超えた時点で急激に、4出力とも期待出力に近づくことがわかる。結局、切り替え後、2回の学習で、平均2乗誤差が0.001以下となった。図6.1から図6.8を通して言えることは、54回以降、ともに振動状態がなくなっていることである。

### 2. 切り替え条件が満足されていない場合

6・1では、(5-1)式から(5-4)式の条件を満足している状態でのkの増加による切り替えをグラフに示した。ここでは6・1と同様な条件で学習

計算を行った。ただし  $k=4.0$  に切り替える学習回数は30回とした。この結果、切り替え後72回の学習を行う結果となった。切り替え時点で、6・1の例題のように、(5-1)式から(5-4)式の条件を満足していないとき（振動状態）、 $k$ の値を切り替えても収束速度の改善は、行われなかったことを示した。

学習回数に対する出力を図6.9から図6.12に示し、図6.13から図6.16に出力層ニューロンの4パターンに対する入力 $x$ の様子を示す。6・1で示した結果と大きく異なる点は、切り替え時点から収束時点の様子が全く異なっている。単純に学習回数に対して、減少傾向、増加傾向の曲線とはならず、振動を繰り返し、安定性に欠け、学習回数を増加させる結果となった。

図6.1 入力 (1,1) に対する学習回数と出力 $y_1^{(3)}(1)$

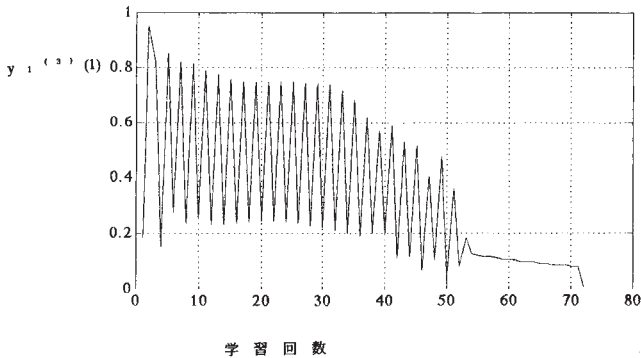


図6.2 入力 (0,1) に対する学習回数と出力 $y_1^{(3)}(2)$

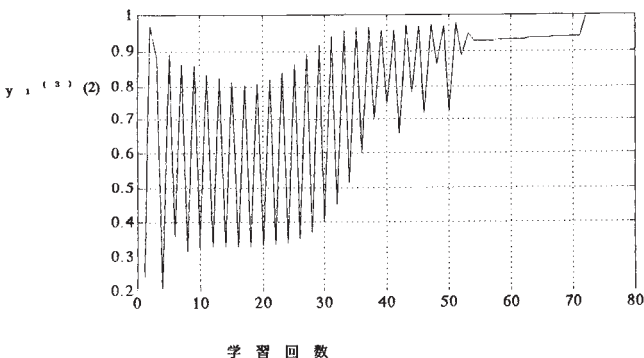


図6.3 入力 (1,0) に対する学習回数と出力 $y_1^{(3)}(3)$

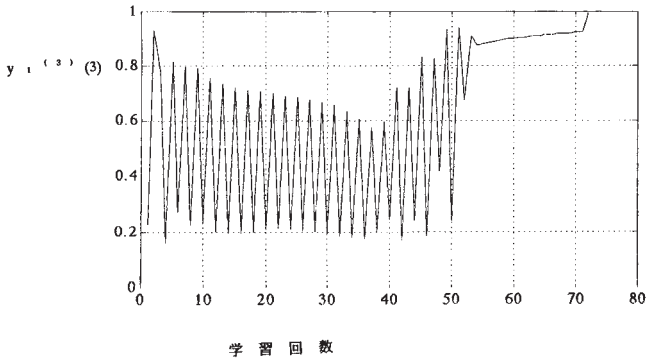


図6.4 入力 (0,0) に対する学習回数と出力 $y_1^{(3)}(4)$

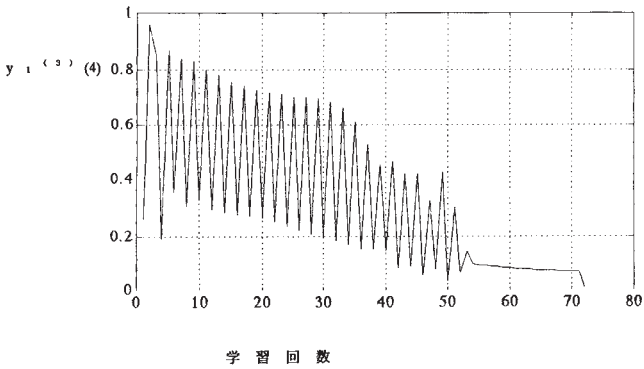


図6.5 学習回数と出力層のシグモイド関数入力 $x_1(1)$

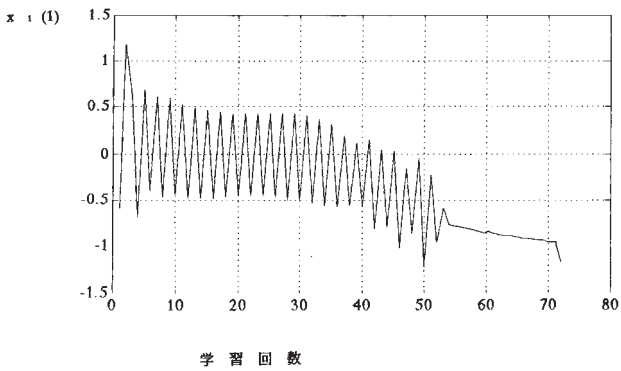


図6.6 学習回数と出力層のシグモイド関数入力 $x_1(2)$

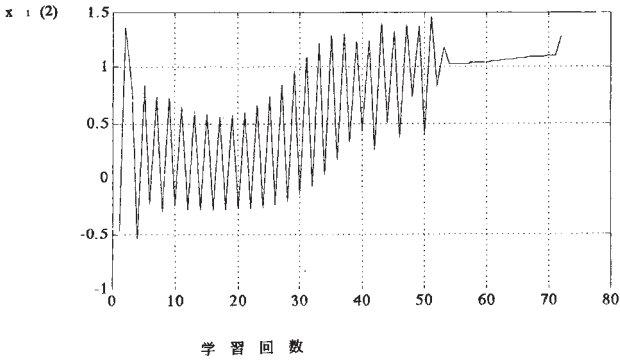


図6.7 学習回数と出力層のシグモイド関数入力 $x_1(3)$

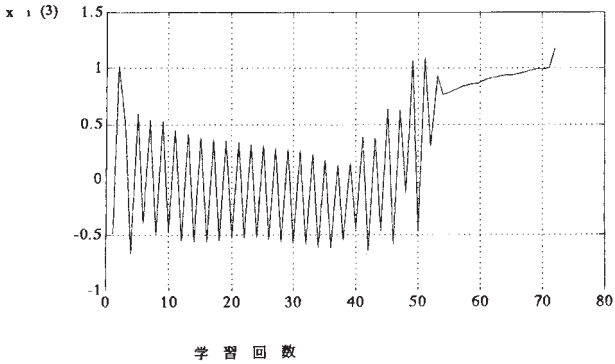


図6.8 学習回数と出力層のシグモイド関数入力 $x_1(4)$

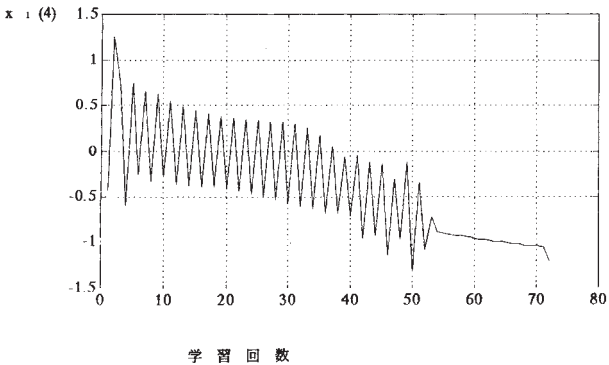


図6.9 入力 (1,1) に対する学習回数と出力 $y_1^{(3)}(1)$

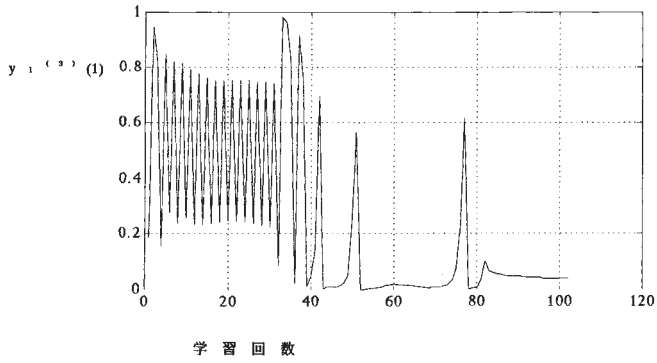


図6.10 入力 (0,1) に対する学習回数と出力 $y_1^{(3)}(2)$

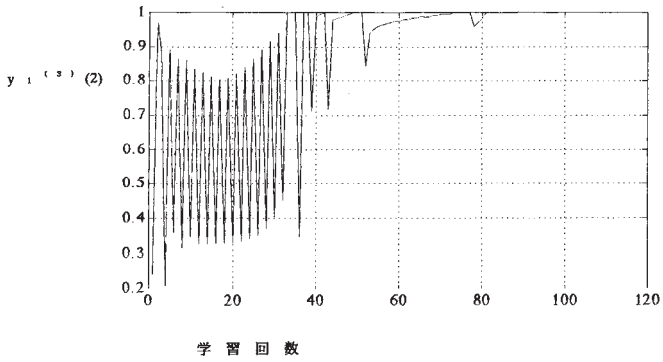


図6.11 入力 (1,0) に対する学習回数と出力 $y_1^{(3)}(3)$

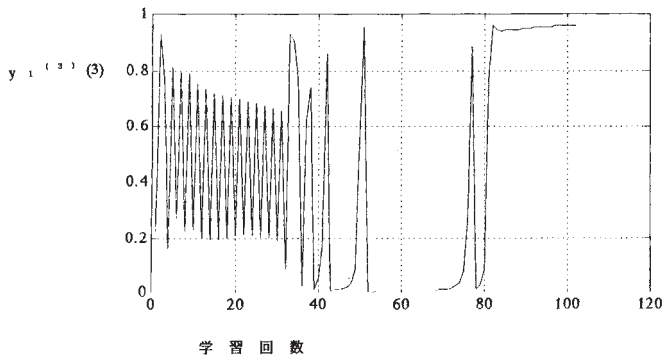


図6.12 入力 (0,0) に対する学習回数と出力 $y_1^{(3)}(4)$

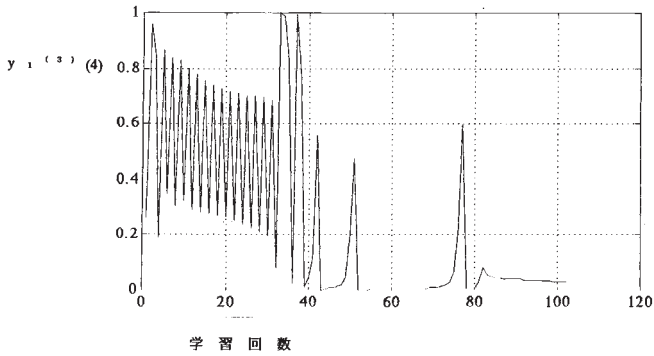


図6.13 学習回数と出力層のシグモイド関数入力 $x_1(1)$

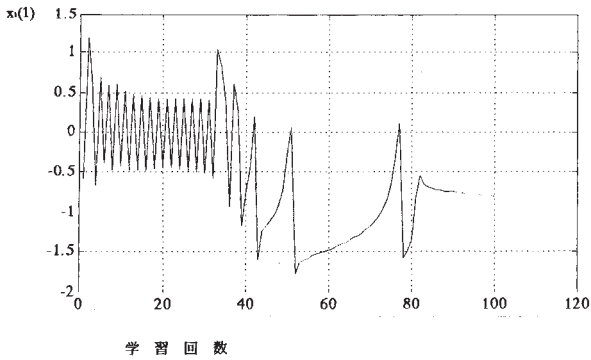


図6.14 学習回数と出力層のシグモイド関数入力 $x_1(2)$

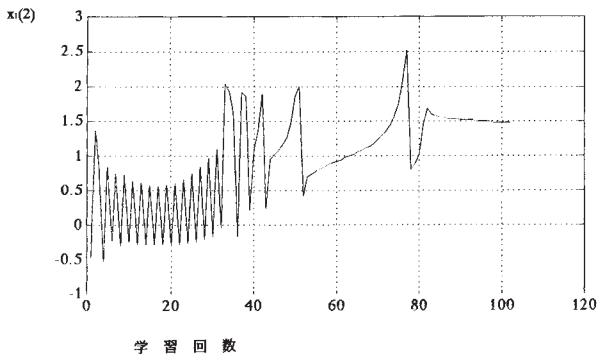


図6.15 学習回数と出力層のシグモイド関数入力 $x_i(3)$

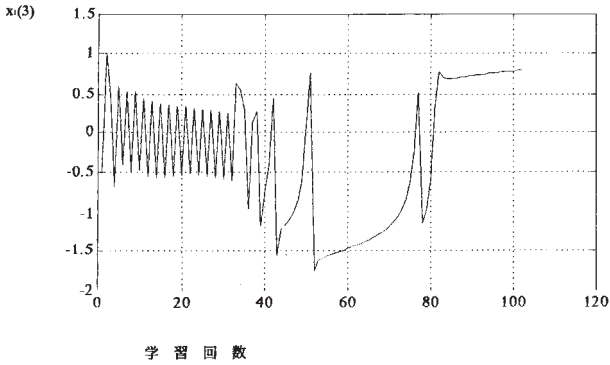
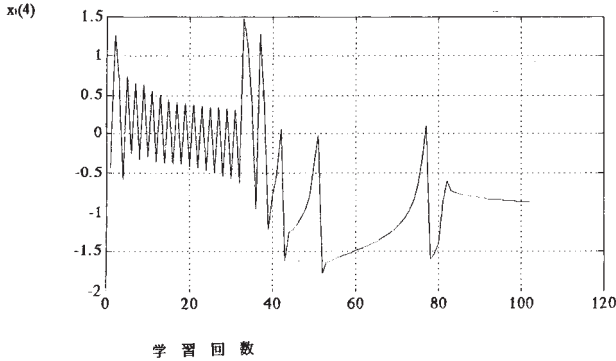


図6.16 学習回数と出力層のシグモイド関数入力 $x_i(4)$



## 7. おわりに

従来から、シグモイド関数を用いる場合、 $k=1$ として学習がされてきた。しかし、学習に時間がかかることが知られていた。また、期待しない出力に収束した場合の収束点からの抜け道として $k$ を時折変化させてみることは、知られていた。しかし、本稿では、すべての期待出力に収束するための収束条件を、シグモイド関数の微分形式を取り上げ、偏微分形式を  $0 (\partial E / \partial w_{ij}^{(m)} \doteq 0)$  の条件に適用して、期待される出力を出すための条



件を考察した。更に、 $k$ の値を増加させることによって、収束領域が広がるために、学習速度が改善されることも示した。後半では、 $k$ の切り替えに対する考察も行った。

〔謝辞〕 本研究に対して、ご指導いただいた、東京工業大学美多勉教授、特別研究員析副教授、劉延年副教授に感謝いたします。

(参考文献)

- 八名和夫・鈴木義武、1992a、『ニューロコンピューティング』、海文堂、44-49ページ。
- Robert Hecht Nielsen, 1990, *Neurocomputing*, Addison Wesley, pp. 124-137.
- 松井伸之・石見憲一、1994、「しきい値ゆらぎをもつニューロンモデルを用いた階層型ニューラルネットワーク」『電気学会論文誌C』114巻11号、1208-1213ページ。
- 八名和夫・鈴木義武、1992b、『ニューロ情報処理技術』、海文堂、73-85ページ。
- 美多勉・小郷寛、1979、『システム制御理論入門』、実教出版、55-57ページ。
- 本間光一・小島紀男・矢沢志雄作、1990、『マトリクスとシステム』、東海大学出版会、8ページ。
- R. ビール/T. ジャクソン著（八名和夫監訳）、1993、『ニューラルコンピューティング入門』、海文堂、68-80ページ。
- 田口 功、1998a、「ニューラルネットワークにおけるシグモイド関数の  $k$  を利用した学習の加速化」『敬愛大学国際研究』1号、229-254ページ。
- 田口 功、1998b、「シグモイド関数を利用した階層型ニューラルネットワークの学習に対する収束条件について」、電気学会電子・情報・システム部門大会、589-592ページ。