

# ニューラルネットワークにおけるシグモイド関数の k (温度) を利用した学習の加速化

田口 功\*

## Learning Acceleration by Using the Sigmoid Function (Temperature: k) in the Neural Network

Isao TAGUCHI

It is considered important to decrease the learning time for the neural network. In this paper, the author has considered the learning speed of a multilayered neural network by increasing k ( $k > 1$ ) for the Sigmoid function of all neurons. Using many numerical approaches for changing k, the learning speed was improved. The performance of the multilayered neural network was investigated by a 4-bit parity check problem with a 2-2-1 neural network.

Improved results by increasing k and momentum version were investigated for many tables including four output data with four teaching signals. Results were calculated by simulation using the matlab.

---

\*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.

ニューラルネットワークにおいて、学習時間を減少させることが問題になっている。本稿では、すべてのニューロンモデルのシグモイド関数に対して、 $k > 1$  の条件で多層ニューラルネットワークの学習速度の改善をプログラムを作成し、考察した。 $k$  を変化させることによって、学習時間は非常に改善された。2-2-1 ニューラルネットワークにおいて、4 ビットパリティチェック問題を扱い、多層ニューラルネットワークの性能の向上を示した。 $k$  の増加とモーメント項による改善された結果は、4 個の出力と 4 個の教師信号として多くの表に示した。それらの表は、matlab を使用してシミュレーションを行った結果である。

## 1. はじめに

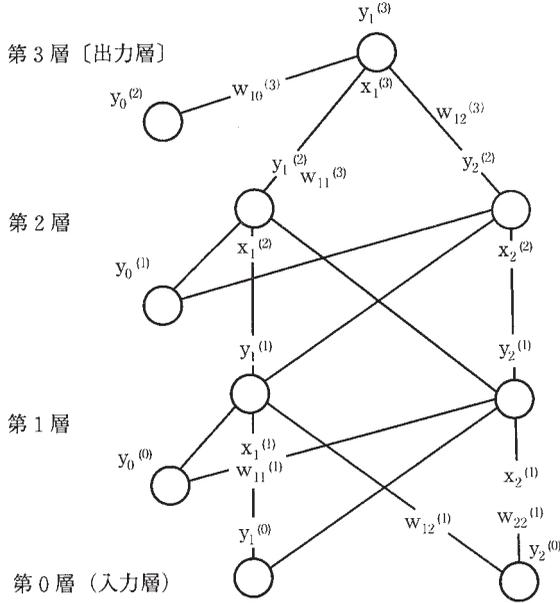
バックプロパゲーション・ニューラルネットワークは、ニューロコンピュータの発達の歴史を考えても最もよく知られている。ニューラルネットワークを構成するとは、すべての学習させたい入力に対して有効な出力を取り出すために誤差関数を定義し、その誤差を最小にする重み係数を決定することである。また、学習回数を減少させることも研究目標の 1 つである。従来、重み係数の更新は、最急降下法が用いられ、学習率は経験的に設定していた (八名・鈴木 1992a・1992b、八名 1993)。学習率を大きくすると、収束は速くなるが不安定になることはよく知られている。小さくすると、安定性は増加するが収束までに時間がかかることになる。文献 (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。本稿では、一般的に知られているシグモイド関数を利用してでもできることを示すと共に、その ( $y=1/(1+\exp(-kx))$ ) とする) 微分形式が ( $ky(1-y)$ ) となることを考慮し、学習率および  $k$  の両方を変化させることによる収束性、収束速度の改善方法および安定性を考察した。特に、 $k$  による収束速度の改善条件は、2 章のまとめ (3) で述べた。収束性や安定性の目安として、ヤコビ行列の固有値を取り入れた (美多・小郷 1979、本間・小島・矢

沢 1990)。更に、重み係数の初期値を乱数で決定後、 $k$  を変化させる時、期待出力に差があるように、実際の出力に対して、その大きさの順位が等しく、共に出力差があることが近似的に成立（以下比例的関係と呼ぶ）すれば（例えば、期待出力が 1 の時出力が 0.5 と仮定した時、期待出力が 0 の場合は 0.2 のように）、 $k$  の値をある程度大きくしても収束するが、比例的関係にない場合（例えば、期待出力 1 の時出力 0.5 と仮定した時、期待出力 0 の時出力 0.8）には、 $k$  の値を変えても期待される状態には収束しないことを示す。初期値の条件が好条件である時、 $k$  の値を増加させれば、反復回数は減少することも示した。また、従来学習率の値を大きくし、その値が従来方法では ( $k=1$ ) 収束しない場合、 $k$  の値を 1 以上に設定すると、収束する  $k$  が存在することも示した。また、重み係数の変化に対する誤差の偏微分係数の変化から定義した行列の固有値の関係（ヤコビ行列）は、すべての固有値が零になることが収束条件であることをシミュレーションを行って示した。更に、有効な出力を取り出すためには、乱数で重み係数を決定し、何回か学習が行われた後の出力結果のパターンが良好な場合には、 $k$  の値の調節で反復回数を改善できることを示し、反復計算ごとに学習率、シグモイド関数の  $k$  を変化させることによっても収束速度の改善が行われることを示した。最後に、モーメント項の付加によって、更なる改善が行われることを示した。ここでは、matlab ソフトを用い、さまざまなシミュレーションを行った。

## 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

第 1 図に示すように、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

第1図 3層ニューラルネットワーク



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ために1層および2層の場合を考える。

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、

$$x_1^{(1)} = y_1^{(0)} w_{11}^{(1)} + y_2^{(0)} w_{12}^{(1)} - y_0^{(0)} w_{10}^{(1)} \quad (2-1)$$

となる。同様に、

$$x_2^{(1)} = y_1^{(0)} w_{21}^{(1)} + y_2^{(0)} w_{22}^{(1)} - y_0^{(0)} w_{20}^{(1)} \quad (2-2)$$

となる。 $x_1^{(1)}$ および  $x_2^{(1)}$ 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。

$$y_1^{(1)} = F(x_1^{(1)}) \quad (2-3)$$

$$y_2^{(1)} = F(x_2^{(1)}) \quad (2-4)$$

$$y_0^{(1)} = 1 \quad (2-5)$$

ここで、(2-3) 式における  $y^{(1)}$  は、隠れ層 1 層目の 1 番目出力、(2-4) 式の  $y_2^{(1)}$  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、

$$x_1^{(2)} = y_1^{(1)} w_{11}^{(2)} + y_2^{(1)} w_{12}^{(2)} - y_0^{(1)} w_{10}^{(2)} \quad (2-6)$$

$$x_2^{(2)} = y_1^{(1)} w_{21}^{(2)} + y_2^{(1)} w_{22}^{(2)} - y_0^{(1)} w_{20}^{(2)} \quad (2-7)$$

$$y_1^{(2)} = F(x_1^{(2)}) \quad (2-8)$$

$$y_2^{(2)} = F(x_2^{(2)}) \quad (2-9)$$

$$y_0^{(2)} = 1 \quad (2-10)$$

となる。また 3 層目の入力および出力は、

$$x_1^{(3)} = y_1^{(2)} w_{11}^{(3)} + y_2^{(2)} w_{12}^{(3)} - y_0^{(2)} w_{10}^{(3)} \quad (2-11)$$

$$y_1^{(3)} = F(x_1^{(3)}) \quad (2-12)$$

となる。ここで期待出力を  $d_1$  と置くと、その誤差は  $y_1^{(3)} - d_1$  となるので、2 乗誤差関数は、

$$E = 1/2 (y_1^{(3)} - d_1)^2 \quad (2-13)$$

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  $\partial E / \partial w_{10}^{(3)}$ 、 $\partial E / \partial w_{11}^{(3)}$ 、 $\partial E / \partial w_{12}^{(3)}$  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、

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

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

$$\partial E / \partial w_{12}^{(3)} = \partial E / \partial x_1^{(3)} \cdot \partial x_1^{(3)} / \partial w_{12}^{(3)}$$

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

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、

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

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 ( $k y_1^{(3)} (1 - y_1^{(3)})$ ) 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。

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

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

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

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

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

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、

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

となる。ここで、(2-17) 式を導く過程で使用した式を用い、

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

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

と置くと、

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

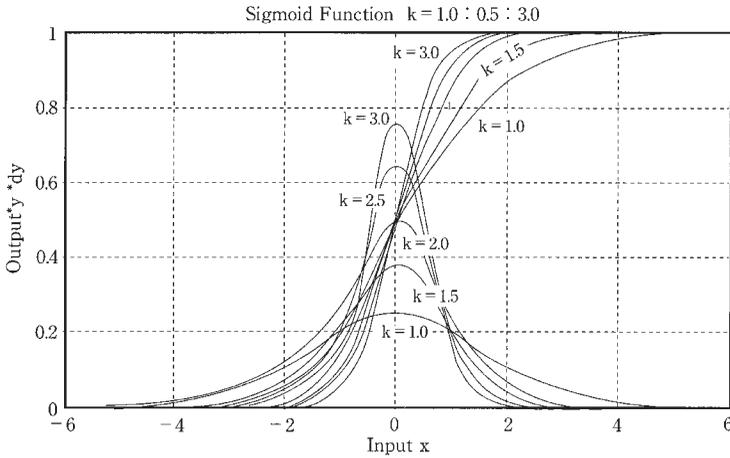
となる。ここで (2-26) 式の  $(A \cdot w_{11}^{(2)} + B \cdot w_{21}^{(2)})$  は、重み係数を2回通しての出力差、 $k y_1^{(1)} (1 - y_1^{(1)})$  は、重み係数の先の出力微分、 $y_2^{(0)}$  は、重み係数の手前からの入力と考えることができる。他の重み係数に対する偏微分も同様に

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

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

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

第2図 k の変化によるシグモイド関数および微分値の変化



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第2図に示す。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. 重み係数の更新について

#### 1. 従来の重み係数の更新について

排他的論理和問題を学習する場合の例は、非常によく知られている。ニューラルネットワークに、入力データ(AおよびB)を4組与え、期待出力(C)を4組与える問題である(第1表)。

第1表において、AとBの排他的論理和をとると、その結果はCになる。すなわち、入力のどちらか一方が1の時出力は1となる。ニューラルネットワークに入力ABを与えた場合には、0に近い値または1に近い値が出力されれば、学習が成功したことになるわけである(実際の実行結果は後で示す)。

次に、重み係数の更新を考える前に、実際の動作について述べる。

- (1) 重み係数の初期値を乱数を使用して与える(例えば、-0.1から0.1)。
- (2) 入力パターンをネットに与える。
- (3) (1)および(2)によって与えられたデータによって出力計算を行う。
- (4) 誤差評価を行うために2乗誤差関数Eの重み係数 $w_{ij}^{(m)}$ に関する偏微分の計算を行う。
- (5) 重み係数の更新を次式にしたがって行う。

第1表 排他的論理和問題

A	B	C
1.0	1.0	0.0
0.0	1.0	1.0
1.0	0.0	1.0
0.0	0.0	0.0

$$w_{ij}^{(m)}[n+1] = w_{ij}^{(m)}[n] - \eta \left( \frac{\partial E}{\partial w_{ij}^{(m)}} \right) \quad (3-1)$$

ここで  $\eta$  は、学習率を表す。

(6)  $E$  が十分小さくなるまで(2)から(5)までを繰り返す。

(7) 入力パターンが 1 個の場合には、これで終了となる。

また、学習データが複数存在する場合には、期待出力データも同数存在し、複数となる。更新方法には 2 通りの考え方が存在する。

〔方法 1〕 1 組の学習データに対する 2 乗誤差式を考え、順次学習データをネットに提示し、上で述べた(2)から(6)にしたがって重み係数を更新する。各学習データに対する  $E$  がすべて十分小さくなるまで繰り返す (逐次更新)。

〔方法 2〕 学習データが 4 組存在し、期待出力も 4 組存在する場合を考える。期待出力とネットの実際の出力の 2 乗誤差の総和は、

$$E = (1/2) (E_1 + E_2 + E_3 + E_4) \quad (3-2)$$

とし、 $E$  の各層の重み係数に対する偏微分  $\partial E / \partial w_{ij}^{(m)}$  は、各学習データに対する和として、

$$\frac{\partial E}{\partial w_{ij}^{(m)}} = \frac{\partial E_1}{\partial w_{ij}^{(m)}} + \frac{\partial E_2}{\partial w_{ij}^{(m)}} + \frac{\partial E_3}{\partial w_{ij}^{(m)}} + \frac{\partial E_4}{\partial w_{ij}^{(m)}} \quad (3-3)$$

とする。(3-3) 式の右辺の各項は、方法 1 で述べたようにして計算可能であるから、各学習データにわたって総和をとってから、(3-1) 式を用いて重み係数の更新を行う (一括更新)。

ここで問題となるのが、どちらの更新方法を用いても (3-1) 式における学習率  $\eta$  は定数であり、 $\eta$  を大きくすると、収束速度は勿論速くなるが、大き過ぎると発散してしまうということである。従来、この  $\eta$  に関しては、経験的に値を決定していた。

## 2. シグモイド関数の $K$ を考慮した重み係数更新法

ここでは、問題を簡単化するために第 1 図に対して、隠れ層の第 1 層がない構成を考える。すなわち、重み係数は、 $w_{11}^{(2)} w_{12}^{(2)} w_{10}^{(2)} w_{11}^{(1)} w_{12}^{(1)} w_{10}^{(1)} w_{21}^{(1)} w_{22}^{(1)} w_{20}^{(1)}$  の 9 個から構成される。したがって、

$$\frac{\partial E}{\partial w_{11}^{(2)}} = (y_1^{(2)} - d_1) \cdot k y_1^{(2)} (1 - y_1^{(2)}) (y_1^{(1)}) \quad (3-4)$$

$$\partial E / \partial w_{12}^{(2)} = (y_1^{(2)} - d_1) \cdot k y_1^{(2)} (1 - y_1^{(2)}) (y_2^{(1)}) \quad (3-5)$$

$$\partial E / \partial w_{10}^{(2)} = (y_1^{(2)} - d_1) \cdot k y_1^{(2)} (1 - y_1^{(2)}) (-1) \quad (3-6)$$

$$\begin{aligned} \partial E / \partial w_{11}^{(1)} &= (y_1^{(2)} - d_1) \cdot k y_1^{(2)} (1 - y_1^{(2)}) w_{11}^{(2)} \\ &\quad k y_1^{(1)} (1 - y_1^{(1)}) (y_1^{(0)}) \end{aligned} \quad (3-7)$$

$$\begin{aligned} \partial E / \partial w_{12}^{(1)} &= (y_1^{(2)} - d_1) \cdot k y_1^{(2)} (1 - y_1^{(2)}) w_{11}^{(2)} \\ &\quad k y_1^{(1)} (1 - y_1^{(1)}) (y_2^{(0)}) \end{aligned} \quad (3-8)$$

$$\begin{aligned} \partial E / \partial w_{10}^{(1)} &= (y_1^{(2)} - d_1) \cdot k y_1^{(2)} (1 - y_1^{(2)}) w_{11}^{(2)} \\ &\quad k y_1^{(1)} (1 - y_1^{(1)}) (y_0^{(0)}) \end{aligned} \quad (3-9)$$

$$\begin{aligned} \partial E / \partial w_{21}^{(1)} &= (y_1^{(2)} - d_1) \cdot k y_1^{(2)} (1 - y_1^{(2)}) w_{12}^{(2)} \\ &\quad k y_2^{(1)} (1 - y_2^{(1)}) (y_1^{(0)}) \end{aligned} \quad (3-10)$$

$$\begin{aligned} \partial E / \partial w_{22}^{(1)} &= (y_1^{(2)} - d_1) \cdot k y_1^{(2)} (1 - y_1^{(2)}) w_{12}^{(2)} \\ &\quad k y_2^{(1)} (1 - y_2^{(1)}) (y_2^{(0)}) \end{aligned} \quad (3-11)$$

$$\begin{aligned} \partial E / \partial w_{20}^{(1)} &= (y_1^{(2)} - d_1) \cdot k y_1^{(2)} (1 - y_1^{(2)}) w_{12}^{(2)} \\ &\quad k y_2^{(1)} (1 - y_2^{(1)}) (y_0^{(0)}) \end{aligned} \quad (3-12)$$

となる。ここで、 $\partial E / \partial w_{11}^{(2)}$ 、 $\partial E / \partial w_{12}^{(2)}$ 、 $\partial E / \partial w_{10}^{(2)}$ 、 $\partial E / \partial w_{11}^{(1)}$ 、 $\partial E / \partial w_{12}^{(1)}$ 、 $\partial E / \partial w_{10}^{(1)}$ 、 $\partial E / \partial w_{21}^{(1)}$ 、 $\partial E / \partial w_{22}^{(1)}$ 、 $\partial E / \partial w_{20}^{(1)}$ の要素をまとめて縦ベクトル  $\mathbf{GE}$  と置くと、

$$\begin{aligned} \mathbf{GE} &= [\partial E / \partial w_{11}^{(2)} \partial E / \partial w_{12}^{(2)} \partial E / \partial w_{10}^{(2)} \partial E / \partial w_{11}^{(1)} \\ &\quad \partial E / \partial w_{12}^{(1)} \partial E / \partial w_{10}^{(1)} \partial E / \partial w_{21}^{(1)} \partial E / \partial w_{22}^{(1)} \\ &\quad \partial E / \partial w_{20}^{(1)}]^T \end{aligned} \quad (3-13)$$

となる。最急降下法を用いた従来の更新方法は、単位行列を  $\mathbf{I}$ （ここでは、9次）とし、学習率を  $\eta$  とし、更に、重み係数を

$$\mathbf{W} = [w_{11}^{(2)} w_{12}^{(2)} w_{10}^{(2)} w_{11}^{(1)} w_{12}^{(1)} w_{10}^{(1)} w_{21}^{(1)} w_{22}^{(1)} w_{20}^{(1)}]^T \quad (3-14)$$

とすると、一般的には、

$$\mathbf{W} [n+1] = \mathbf{W} [n] - \eta \cdot \mathbf{I} \cdot \mathbf{GE} \quad (3-15)$$

となる。ここで、 $n$  は、学習サイクルを表すパラメータである。(3-15)式に対して、(3-13)式を用い、次の操作を考え、9行9列の行列  $\mathbf{H}$  を作成する。

(1) ベクトル  $\mathbf{GE}$  を、ベクトル  $\eta \cdot \mathbf{I} \cdot \mathbf{GE}$  の最初の要素（1行目）で割っ

た縦ベクトルを行列  $H$  の 9 行 9 列の 1 列目の要素とする。

(2) ベクトル  $\eta$ ,  $I$ ,  $GE$  の 2 行目の要素で割った縦ベクトルを行列  $H$  の 2 列目の要素とする。

(3) 同様な操作を続けて 9 行 9 列の行列  $H$  を求める。この行列要素は、学習サイクルで、各結合係数の変化に対する  $E$  の重み係数に対する偏微分 ( $\partial E / \partial w_{ij}^{(m)}$ ) 行列となる。行列  $H$  の要素は、結合係数の変化に対する  $\partial E / \partial w_{ij}^{(m)}$  の変化を表すことになる。

本稿では、行列  $H$  の固有値を求めることによって収束の判定を試みた。そのシミュレーションの結果は、後章で述べる。

## 4. 排他的論理和を例としたバックプロパゲーション学習の実行

### 1. 学習率及び $k$ の変化による反復回数の変化と安定性

#### (1) $k = 1$ の場合の学習率変化による反復回数の変化

同じ重み係数の初期値を用い、学習率の増加に対する反復計算回数の変化を matlab ソフトを用い、プログラム化し、排他的論理和計算を行った。この場合の初期値としては、文献 (八名・鈴木 1992a) で使用されている値を参考にした。重み係数ベクトル  $W$  は、 $W = [w_{11}^{(2)} w_{12}^{(2)} w_{10}^{(2)} w_{11}^{(1)} w_{12}^{(1)} w_{10}^{(1)} w_{21}^{(1)} w_{22}^{(1)} w_{20}^{(1)}]^T = [1.1747 \ 1.6147 \ 0.8833 \ 0.9701 \ -1.2679 \ 0.7173 \ -1.7471 \ 1.5631 \ 1.0705]^T$  とした。ここでは、2 乗誤差の総和の平均が 0.01 以下になった時に反復計算を打ち切った。結果を第 2 表に示す。

第 2 表をみると、学習率が大きくなるにしたがって一般的に言われているように反復回数は減少することがわかる。しかし、学習率が 2.7 を超えると、発散してしまい、収束しない。

#### (2) 学習率 $\eta = 0.9$ 、 $k$ を変化させた場合の収束回数の変化

4・1 の (1) と同様な重み係数の初期値を設定し、 $k$  による反復回数の変化を確かめた。その結果を第 3 表に示す。 $k$  を 6.0 10.0、20.0、25.0、30.0、50.0 と大きくするにしたがって、反復回数は減少することがわかる。

第2表 学習率の変化による反復回数

学習率 $\eta$	反復回数	学習率 $\eta$	反復回数	学習率 $\eta$	反復回数
0.5	576	1.3	235	2.1	161
0.6	483	1.4	220	2.2	281
0.7	417	1.5	207	2.3	154
0.8	381	1.6	196	2.4	152
0.9	328	1.7	187	2.5	151
1.0	298	1.8	179	2.6	154
1.1	273	1.9	172	2.7	発散
1.2	271	2.0	166		

第3表 k による反復計算結果

<u>k=6</u>		期待出力=[0 1 1 0]			
反復回数		出力 (最大2乗誤差0.01)			
1	0.0055	0.9803	0.6178	0.0051	
2	0.9846	0.9995	1.0000	0.9311	
3	0.3151	0.9936	0.9990	0.2391	
4	0.0108	0.7395	0.9564	0.0084	
5	0.0847	0.9930	0.9925	0.0436	
6	0.0622	0.9907	0.9900	0.0337	
7	0.0526	0.9891	0.9884	0.0292	
<u>k=10</u>		期待出力=[0 1 1 0]			
反復回数		出力			
1	0.0001	0.9992	0.8855	0.0001	
2	0.1601	0.9999	1.0000	0.1483	
3	0.0050	0.9969	0.9988	0.0046	
<u>k=20</u>		期待出力=[0 1 1 0]			
反復回数		出力			
1	0.0000	1.0000	0.9966	0.0000	
<u>k=25</u>		期待出力=[0 1 1 0]			
反復回数		出力			
1	0.0000	1.0000	0.9993	0.0000	
<u>k=30</u>		期待出力=[0 1 1 0]			
反復回数		出力			
1	0.0000	1.0000	0.9998	0.0000	
<u>k=50</u>		期待出力=[0 1 1 0]			
反復回数		出力			
1	0.0000	1.0000	1.0000	0.0000	

## 2. $k > 1$ の場合の収束性と安定性

4・1 で使用した重み係数を初期値として使用し、 $k=2$  の場合、 $k=3$  の場合のシミュレーションを行い、収束性と安定性を考察する。ここでは、学習率を0.9とし、 $k=2.0$ の場合（第4表）と $k=3.0$ （第5表）の場合を検討した。2乗誤差の平均は、0.01から更に条件を強め0.001まで計算を行った。その結果、 $k=2$  の場合には、誤差が0.01以下になるまでに73回の反復を要した。これは、4・1 で述べた従来の学習率を2.6とした場合よりはるかに良い結果となっている。4・1 における学習率を2.6にしても50%以下の反復回数となっている。誤差を0.001以下にするためには、425回の反復回数を必要とした。その結果を第4表に示す。

第4表  $k=2$ ,  $\eta=0.9$  の場合の出力

反復回数	期待出力=[0 1 1 0] 出力				行列の固有値
1	0.2739	0.6520	0.4281	0.2225	53.1561他0
5	0.5316	0.6847	0.5194	0.3145	3.3443
10	0.4931	0.7318	0.5426	0.2766	0.9452
15	0.4387	0.7629	0.5989	0.2412	-0.4160
20	0.3711	0.7888	0.6683	0.2096	-1.7459
25	0.3086	0.8119	0.7297	0.1833	-2.4109
30	0.2600	0.8319	0.7753	0.1625	-2.4745
35	0.2241	0.8486	0.8078	0.1463	-2.3016
40	0.1973	0.8623	0.8314	0.1335	-2.0809
45	0.1768	0.8734	0.8492	0.1231	-1.8719
50	0.1608	0.8827	0.8630	0.1146	-1.6884
55	0.1478	0.8904	0.8741	0.1074	-1.5308
60	0.1372	0.8970	0.8831	0.1013	-1.3960
65	0.1284	0.9027	0.8907	0.0960	-1.2804
70	0.1208	0.9076	0.8971	0.0914	-1.1807
73	0.1168	0.9102	0.9005	0.0889 (平均誤差0.01)	-0.7509
150	0.0658	0.9443	0.9440	0.0562	-0.4879
200	0.0550	0.9533	0.9531	0.0474	-0.3589
250	0.0480	0.9591	0.9590	0.0416	-0.2830
300	0.0431	0.9633	0.9632	0.0375	-0.2332
350	0.0394	0.9664	0.9663	0.0343	-0.1982
400	0.0364	0.9689	0.9688	0.0318	-0.1722
425	0.0352	0.9700	0.9699	0.0308 (誤差0.001)	-0.1615

第5表 k=3,  $\eta=0.9$ の場合の出力

反復回数	期待出力=[0 1 1 0] 出力				行列の固有値
1	0.1096	0.7862	0.4381	0.0847	-40.8710他0
5	0.3465	0.8447	0.8598	0.1832	27.6260
10	0.1927	0.8436	0.8516	0.1066	-38.1967
15	0.1608	0.8796	0.8834	0.0926	-3.4658
20	0.1362	0.8975	0.8994	0.0814	-2.8235
25	0.1193	0.9097	0.9107	0.0735	-2.3539
30	0.1070	0.9186	0.9191	0.0676	-2.0106
35	0.0976	0.9255	0.9256	0.0629	-1.7512
40	0.0900	0.9309	0.9309	0.0591	-1.5494
45	0.0839	0.9354	0.9352	0.0559	-1.3883
50	0.0785	0.9392	0.9389	0.0531	-1.2570
60	0.0706	0.9451	0.9448	0.0487	-1.0560
70	0.0644	0.9497	0.9493	0.0452	-0.9098
80	0.0596	0.9533	0.9529	0.0424	-0.7988
90	0.0556	0.9563	0.9558	0.0400	-0.7117
100	0.0522	0.9588	0.9583	0.0380	-0.6415
150	0.0413	0.9670	0.9666	0.0312	-0.4290
181	0.0371	0.9702	0.9699	0.0284 (誤差0.001)	-0.3576

第4表および第5表から言えることは、安定収束する場合には、固有値がすべて0にしたいに近づくことが条件と言える。 $k$ の値が大きくなれば同じ初期値に対して反復回数は減少する。従来の学習率 $\eta$ と似たような結果となる（学習率が増加すればするほど反復回数は減少する。しかし、大きくしすぎると収束しない）が、式的には、シグモイド関数の $k$ のほうが誤差に関する偏微分式をみると層ごとに、 $k$ 倍、 $k^2$ 倍、……、と変化するために、すべての偏微分要素に対して等倍ではない。 $k=1$ の場合は等しい。また、従来は $\eta$ を変化させるだけであった。しかし、 $k$ の利用も収束速度改善に対して有効と考えられる。

4・1においては、学習率 $\eta$ が2.6の時に、反復回数は、ほぼ150回（平均2乗誤差最大値0.01に対して）で最高であった。 $k=2$ 、 $k=3$ の場合は、それに対して、更に反復回数は改善され減少していることがわかる。学習率 $\eta$ と $k$ の組み合わせで（ $k$ の値を1以上にすることによって）反復回数は、改善されると考えることができる。収束するにしたがって、固有値減少の様子が、

第4表および第5表からわかる。

### 3. 学習率 $\eta = 2.7$ の場合の $k$ による反復回数の改善法

ここでは、 $k$  の値が1以下の場合も検討した。重み係数の初期値は、 $4 \cdot 1$  の時と同様な値を用いた。ここでは、 $k$  の値を0.1から3.6まで変化させることによって、その収束回数の変化を調べた。従来  $k$  の値は、1として通常用いられていたが66以下の場合には、すべて収束しなかった。結論としては、1以上の  $k$  を使用して反復回数の改善を行うことが可能であることが第6表からわかる。また、 $k$  の値には、1以上であっても限界があり、大きくしすぎると、反復回数の改悪となる。その結果を第6表に示す。学習率2.7で  $k = 2.1$  の場合には反復回数が29回となり、非常に収束状態が良好であった。

第6表  $k$  の変化による反復回数の変化  
(平均2乗誤差0.01の場合)

k	収束回数	k	収束回数
0.02	648回発散	2.4	31回収束
0.1	594回発散	2.7	51回収束
0.2	発散	3.0	133回収束
0.4	472回発散	3.3	348回収束
0.6	314回発散	3.6	34049回発散
0.8	発散		
1.2	102回収束		
1.5	59回収束		
2.1	29回収束		

### 4. $k$ および $\eta$ を利用した初期値の探索 (重み係数)

#### (1) 期待出力に収束しない初期値

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、 $k = 1$ 、学習率2.7とし63回の反復計算後  $k = 4.0$  に切り替えた。63回目の出力は  $Y = [0.0712 \ 0.9055 \ 0.4957 \ 0.5109]$  であった。ここで、期待出力 =  $[0 \ 1 \ 1 \ 0]$  である。切り替え

第7表 期待出力に収束しない初期値

	回数	出力			
期待出力	63	0.0712	0.9055	0.4957	0.5109
[0 1 1 0]	64	0.0007	0.9969	0.5032	0.5035

第8表 期待出力に収束する初期値

	回数	出力			
期待出力	50	0.7919	0.9843	0.9310	0.5692
[0 1 1 0]	51	0.0048	0.6975	0.0105	0.0032
	52	0.0140	0.8733	0.0303	0.0092
(k=2)	53	0.0324	0.9183	0.0858	0.0201
	54	0.1801	0.9639	0.5234	0.0968
	55	0.6982	0.9856	0.9592	0.4603
	56	0.0030	0.7381	0.0137	0.0022
	57	0.0079	0.8782	0.0418	0.0058
	58	0.0235	0.9291	0.1956	0.0157
	59	0.4354	0.9828	0.9571	0.2804
	60	0.0110	0.8906	0.2198	0.0078
(k=4)	61	0.3054	0.9774	0.9730	0.2212
	62	0.0002	0.9972	0.9391	0.0002

後の64回目の出力は、第7表となった。期待出力に対して3番目、4番目のデータの誤差が大きく勿論平均2乗誤差は、非常に大きいことがわかる。 $k=1$ で63回目の出力も3番目、4番目のデータで誤差が非常に大きかった。したがって、63回目の出力は、 $k=4$ に切り替えても期待出力に近い出力とはならないと考えられる。

## (2) 期待出力に収束する場合の初期値(乱数で重み係数を変える場合)

次に、 $4 \cdot 4$ の(1)に対してもう一度乱数を発生させ、 $W = [w_{11}^{(2)} w_{12}^{(2)} w_{10}^{(2)} w_{11}^{(1)} w_{12}^{(1)} w_{10}^{(1)} w_{21}^{(1)} w_{22}^{(1)} w_{20}^{(1)}]^T = [0.6885 \ 0.8682 \ 0.6295 \ 0.7362 \ 0.7254 \ 0.9995 \ 0.8886 \ 0.2332 \ 0.3063]^T$ 、 $k=2$ とし、同様に60回の反復計算を行った場合には、出力は、50回から示すと、第8表となる。50回の反復計算から60回の反復計算による出力をみると、期待出力=[0 1 1 0]の大きさの順に出力が出ている。更に、出力差も存在することが必要と思われる。(1)の場合には、期待出力に対して出力差が少なすぎるために $k$ の値を切り替えても期待出力付近に収束しないものと考えられる。4個の出

力が期待出力に向かう出力を見いだすことが必要と考えられる。すなわち、出力の変動を確認しながらkの値を切り替え反復回数を減少させればよい結果をもたらすと考えられる。

### (3) 最初に(1)と等しい重み係数を使用した場合の収束する重み係数の探索

4・4の(1)は、乱数によって決定した重み係数  $W = [w_{11}^{(2)} w_{12}^{(2)} w_{10}^{(2)} w_{11}^{(1)} w_{12}^{(1)} w_{10}^{(1)} w_{21}^{(1)} w_{22}^{(1)} w_{20}^{(1)}]^T = [0.2190 \ 0.0470 \ 0.6789 \ 0.6793 \ 0.9347 \ 0.3835 \ 0.5194 \ 0.8310 \ 0.0346]^T$  を用い、 $k = 1$ 、学習率=2.7で学習を行ったが、期待値に近い出力は得られなかった。そこで、ここでは同じ重み係数の初期値を用い、学習率=0.9一定とし、 $k = 3.0$ から反復計算ごとにk

第9表 k=3.0から減少させた場合の初期値の探索 (学習率0.9一定)

学習率	回数	k	出力			
0.90 一定	1	3.00	0.1404	0.2049	0.1889	0.2214
	10	2.89	0.1891	0.1818	0.1864	0.1838
	20	2.79	0.6192	0.6399	0.6486	0.6591
	30	2.69	0.3357	0.4441	0.4398	0.4642
	40	2.59	0.2715	0.5062	0.4883	0.4985
	50	2.49	0.1550	0.5038	0.4839	0.3848
	60	2.39	0.1057	0.5364	0.5300	0.2584
	70	2.29	0.1010	0.6135	0.6128	0.1611
	80	2.19	0.1372	0.8067	0.8069	0.2068
	90	2.09	0.1309	0.8307	0.8308	0.1792
	100	1.99	0.1277	0.8408	0.8409	0.1649
	110	1.89	0.1278	0.8444	0.8445	0.1589
	120	1.79	0.1305	0.8437	0.8438	0.1581
	127	1.72	0.1337	0.8413	0.8414	0.1597
0.90 一定	1	1.80	0.1209	0.8598	0.8599	0.1451
	3	一定	0.0477	0.7746	0.8149	0.1073
	5		0.0582	0.8543	0.8710	0.1530
	7		0.0616	0.8684	0.8797	0.1554
	9		0.0635	0.8744	0.8828	0.1497
	11		0.0649	0.8792	0.8857	0.1443
	13		0.0659	0.8832	0.8882	0.1395
	15		0.0665	0.8868	0.8911	0.1351
	17		0.0669	0.8899	0.8936	0.1310
	19		0.0670	0.8927	0.8960	0.1273
	21		0.0670	0.8950	0.8982	0.1239
	23		0.0668	0.8975	0.9003	0.1207

の値を0.01間隔で減少させて127回の計算を行った。その結果を第9表に示す。反復回数が40回前後で出力が徐々に期待出力に向かっていることがわかる。更に80回を超えるといっそう期待出力に向かっていることがわかる。73回目の反復計算により得られた重み係数を  $W = [-3.1805 \quad -3.2698 \quad -1.5778 \quad 1.8379 \quad 1.8367 \quad 2.8856 \quad -2.5040 \quad -2.4999 \quad -0.8213]^T$  を用い、 $k = 1.8$ 一定として反復学習を行った結果を第9表の後半に示す。23回の反復学習で誤差は、0.01以下となった。また、同じ重み係数の初期値で、 $k = 1.6$ では、49回の反復学習を必要とした。更に、 $k = 1.4$ では96回、 $k = 1.2$ では162回でそれぞれ誤差0.01以下となった。 $k$ の値を反復計算ごとに変化させることによって、学習率一定として探索することも可能となることがわかった。また、期待出力に収束させる初期値が見つからない場合もあるが、反復計算に対して、 $k$ を変化させることも初期値探索には有効であると考えられる。

次に、同様に学習率を1.35、1.80、2.25、2.70一定とし、 $k = 3$ から0.01間隔で減少させた場合、1.80の時70回の計算でよい初期値を見つけることができた。その結果を第10表に示す。なお、70回目における重み係数値は  $W = [-2.8179 \quad 2.5581 \quad -1.3864 \quad 2.8878 \quad -2.8456 \quad -1.6865 \quad 2.0987 \quad -2.4304 \quad 0.8175]^T$  となった。また、学習率1.35における結果を第11表に示す。この場合は、1番目および2番目の期待出力に対して、出力は好結果となるが、3、4番目の期待出力に対しては学習が全くできていない状態と考えられる。第9表および第10表を比較すると当然学習率が大きいほうが $k$ の値が等しい

第10表  $k = 3.0$ から減少させた場合の初期値の探索(学習率1.8一定)

学習率	回数	k	出力			
1.8 一定	1	2.99	0.1404	0.2049	0.1889	0.2214
	10	2.89	0.9875	0.9927	0.9796	0.9885
	20	2.79	0.7560	0.9188	0.6843	0.6824
	30	2.69	0.3247	0.9333	0.2630	0.2679
	40	2.59	0.1334	0.9203	0.1088	0.0763
	50	2.49	0.7271	0.9756	0.7291	0.3208
	60	2.39	0.4148	0.9610	0.6565	0.1150
	70	2.29	0.0821	0.9403	0.9028	0.0562

第11表 k=3.0から減少させた場合の初期値の探索(学習率1.35一定)

学習率	回数	k	出力			
1.35 一定	1	2.99	0.1404	0.2049	0.1889	0.2214
	10	2.89	0.0240	0.0452	0.0150	0.0207
	20	2.79	0.4013	0.7139	0.2699	0.3858
	30	2.69	0.4931	0.9210	0.4118	0.4908
	40	2.59	0.2841	0.9257	0.3768	0.4247
	50	2.49	0.3142	0.9451	0.6047	0.6345
	60	2.39	0.2062	0.9443	0.6080	0.6333
	70	2.29	0.1510	0.9349	0.5770	0.5977
	80	2.19	0.1167	0.9201	0.5191	0.5374
	90	2.09	0.1027	0.9093	0.4870	0.5042
	100	1.99	0.1021	0.9049	0.4842	0.5013

場合には速く好条件の初期値を見いだせることがわかる。kを変化させても学習率の値によって好条件の初期値を見いだせる場合とそうでない場合があると考えられる。

## 5. 学習率およびkを変化させることによる加速化法

多層ニューラルネットワークの学習に対して、1次偏微分に基づく加速化法や2次偏微分に基づく加速化法は、多数考えられている。最急降下法に対して、シグモイド関数の温度および学習率を反復計算ごとに変化させながら学習を行い加速を考えた例はあまり知られていないようである。ここでは、4章で述べた方法で、重み係数の収束する初期値が選択された場合のkおよび学習率の変動による、収束性の加速化についてシミュレーションを行い、kの効果について報告する。

第12表には、k=1一定および学習率0.9で、重み係数の初期値は、 $W = [-3.1805 \ -3.2698 \ -1.5778 \ 1.8379 \ 1.8367 \ 2.8856 \ -2.5040 \ -2.4999 \ -0.8213]^T$ とした(4章の4で収束する初期値として求められた重み係数値)場合の結果を示す。

次に、k=1.0をkの初期値とし、1回の反復学習ごとに、0.0001増加させ、学習率は、1回の反復計算で、0.09増加させて計算を行った結果を第

第12表 k, 学習率一定の場合(90回の反復では収束しない)

学習率	回数	k	出力	誤差評価<0.01		
0.90 一定	1	1.0	0.2971	0.5593	0.5596	0.3411
	10	一定	0.2592	0.6938	0.6500	0.3935
	20		0.2501	0.7049	0.6827	0.3493
	30		0.2340	0.7192	0.7074	0.3191
	40		0.2181	0.7349	0.7283	0.2941
	50		0.2037	0.7505	0.7467	0.2722
	60		0.1909	0.7654	0.7631	0.2530
	70		0.1795	0.7792	0.7777	0.2359
	80		0.1693	0.7918	0.7908	0.2209
	90		0.1602	0.8032	0.8025	0.2075

第13表  $k = 1.0 + N * 0.0001$ , 学習率 =  $0.99 + 0.01 * N$  とした場合

学習率	回数	k	出力	誤差評価<0.01		
0.99	1	1.0001	0.2971	0.5593	0.5598	0.3411
1.44	5	1.0006	0.2579	0.6920	0.6378	0.4142
1.89	10	1.0011	0.2581	0.6975	0.6639	0.3736
2.34	15	1.0016	0.2424	0.7108	0.6945	0.3350
2.79	20	1.0021	0.2217	0.7309	0.7238	0.2996
3.24	25	1.0026	0.1989	0.7556	0.7527	0.2653
3.69	30	1.0031	0.1772	0.7816	0.7804	0.2330
4.14	35	1.0036	0.1575	0.8063	0.8058	0.2039
4.59	40	1.0041	0.1402	0.8285	0.8282	0.1787
5.04	45	1.0046	0.1252	0.8476	0.8474	0.1573
5.49	50	1.0051	0.1125	0.8638	0.8636	0.1395
5.94	55	1.0056	0.1018	0.8774	0.8773	0.1247
6.39	60	1.0061	0.0926	0.8889	0.8889	0.1123
6.66	64	1.0065	0.0863	0.8969	0.8968	0.1038

13表に示した。第12表および第13表を比較すると、明らかに、重み係数の良好な値が設定されている場合には、k および学習率をわずかに増加させることによって加速性が増加している。

次に、従来から加速法として知られている学習率の増加（反復計算中は一定値）による加速性を調べることにする。ただし、ここでは、反復計算ごとに徐々に0.09増加させた結果を第14表に示す。反復回数は、第13表、第14表を比較すると1回だけkを増加したほうが少なくなっている。この結果、kの値を0.0001反復計算ごとに増加させてもあまり効果はないこ

第14表 k=1.0一定, 学習率=0.99+0.01 \* N とした場合

学習率	回数	k	出力 誤差評価<0.01			
0.99	1	1.00	0.2971	0.5593	0.5596	0.3411
1.44	5	(一定)	0.2574	0.6933	0.6336	0.4302
1.89	10		0.2584	0.6972	0.6636	0.3740
2.34	15		0.2439	0.7103	0.6939	0.3356
2.79	20		0.2223	0.7301	0.7230	0.3005
3.24	25		0.1998	0.7546	0.7517	0.2664
3.69	30		0.1782	0.7804	0.7792	0.2343
4.14	35		0.1586	0.8051	0.8045	0.2052
4.59	40		0.1413	0.8271	0.8268	0.1801
5.04	45		0.1264	0.8462	0.8460	0.1588
5.49	50		0.1137	0.8624	0.8622	0.1410
5.94	55		0.1029	0.8760	0.8759	0.1261
6.39	60		0.0938	0.8876	0.8875	0.1137
6.84	65		0.0860	0.8974	0.8973	0.1032

とがわかる。

そこで、学習率は0.9一定とし、k = 1 を初期値とし反復増分を0.01とした場合、0.03とした場合および0.05, 0.07, 0.1と増加させた場合の計算結果を第15表にまとめて示す。第15表をみると、k の値を反復計算ごとに増加させた場合、その増加の幅が大きくなればなるほど、収束率は増加することがわかる。

## 6. モーメント項を追加することによる 重みの修正速度の改善

5章においては、重み係数の更新は、次の (3-15) 式によって行っていた。

$$W [n+1] = W [n] - \eta \cdot I \cdot GE \quad (3-15)$$

一般的によく知られているように、学習に対しては、モーメント項を追加し加速を行う。ここでは、前章で行ったシグモイド関数の k をわずかに反復計算ごとに増加させながら、更に、モーメント項を追加し、重み係数の更新を行った。すなわち、重み更新式は、(3-15) 式に対してモーメ

第15表 k=1.0に対して反復計算ごとにkを増加させた場合、  
学習率=0.9一定とした場合

学習率	回数	k	出力 誤差評価<0.01			
0.9 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	5	(1.0)	0.2459	0.7126	0.6494	0.4286
	10	増加分 (0.01)	0.2301	0.7279	0.6835	0.3579
	15		0.2099	0.7514	0.7217	0.3108
	20		0.1867	0.7759	0.7562	0.2703
	25		0.1637	0.8004	0.7874	0.2345
	30		0.1424	0.8240	0.8153	0.2028
	35		0.1234	0.8459	0.8399	0.1750
	40		0.1067	0.8658	0.8615	0.1508
	45		0.0922	0.8834	0.8802	0.1300
50	0.0796	0.8988	0.8965	0.1120		
0.9 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	5	(1.0)	0.2202	0.7423	0.6767	0.3981
	10	増加分 (0.03)	0.1745	0.7918	0.7490	0.2862
	15		0.1298	0.8426	0.8177	0.2064
	20		0.0924	0.8855	0.8709	0.1463
	23		0.0746	0.9066	0.8959	0.1183
0.9 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	5	(1.0)	0.1952	0.7707	0.7036	0.3656
	10	増加分 (0.05)	0.1268	0.8470	0.8081	0.2213
	15		0.0746	0.9080	0.8881	0.1302
	16		0.0667	0.9175	0.9001	0.1167
0.9 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	5	(1.0)	0.1716	0.7975	0.7301	0.3325
	10	増加分 (0.07)	0.0892	0.8914	0.8575	0.1673
	12		0.0661	0.9194	0.8945	0.1253
0.9 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	3	(1.0)	0.2110	0.8020	0.7345	0.4562
	5	増加分 (0.1)	0.1395	0.8346	0.7688	0.2844
	7		0.0956	0.8847	0.8373	0.1946
9	0.0632		0.9236	0.8918	0.1314	

ント項 $\alpha (W [n] - W [n-1])$ を加え、

$$W [n+1] = W [n] - \eta \cdot I \cdot GE + \alpha (W [n] - W [n-1]) \quad (6-1)$$

とした。ここで、 $0 < \alpha < 1$ であり、重み係数の初期値は、5章で使用された $W = [-3.1805 \ -3.2698 \ -1.5778 \ 1.8379 \ 1.8367 \ 2.8856 \ -2.5040 \ -2.4999 \ -0.8213]^T$ を用いた。ここでは、

第16表 モーメント項を加えた場合の反復学習回数の変化

学習率	回数	k	出力	誤差評価<0.01		
0.9 一定 モーメント項 なし	1	初期値	0.2971	0.5593	0.5596	0.3411
	10	(1.0)	0.2290	0.6781	0.6966	0.3783
	20	増加分	0.1930	0.7479	0.7517	0.2938
	30	(0.01)	0.1472	0.8071	0.8081	0.2220
	40		0.1094	0.8557	0.8561	0.1641
	50		0.0808	0.8931	0.8932	0.1206
	52		0.0761	0.8994	0.8994	0.1133
0.9 一定 $\alpha=0.05$ 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	3	(1.0)	0.2330	0.7769	0.7161	0.4978
	5	増加分	0.1751	0.7961	0.7336	0.3624
	7	(0.01)	0.1327	0.8394	0.7852	0.2741
	9		0.0955	0.8847	0.8419	0.2032
	11		0.0643	0.9238	0.8926	0.1441
	12		0.0515	0.9400	0.9142	0.1191
0.9 一定 $\alpha=0.1$ 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	3	(1.0)	0.1636	0.7214	0.7576	0.4431
	5	増加分	0.1028	0.8043	0.8196	0.2748
	7	(0.01)	0.0541	0.8982	0.8985	0.1520
	8		0.0356	0.9347	0.9318	0.1045
0.9 一定 $\alpha=0.3$ 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	2	(1.0)	0.1775	0.8264	0.8041	0.5131
	3	増加分	0.0855	0.9403	0.9243	0.4257
	4	(0.01)	0.0287	0.9816	0.9760	0.1750
0.18 一定 $\alpha=0.025$ 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	5	(1.0)	0.2640	0.6731	0.6614	0.3588
	10	増加分	0.2117	0.7648	0.7544	0.2983
	15	(0.01)	0.1489	0.8460	0.8378	0.2175
	20		0.0912	0.9115	0.9062	0.1393
	21		0.0813	0.9221	0.9174	0.1256
0.18 一定 $\alpha=0.025$ 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	3	(1.0)	0.2588	0.6712	0.6590	0.3548
	5	増加分	0.2189	0.7466	0.7351	0.3132
	7	(0.05)	0.1734	0.8136	0.8037	0.2585
	9		0.1282	0.8689	0.8619	0.1989
	11		0.0888	0.9139	0.9083	0.1436
	12		0.0722	0.9315	0.9269	0.1194
0.18 一定 $\alpha=0.025$ 一定	1	初期値	0.2971	0.5593	0.5596	0.3411
	5	(1.0)	0.2748	0.6539	0.6424	0.3691
	10	増加分	0.2418	0.7220	0.7112	0.3318
	15	(0.0)	0.1986	0.7852	0.7756	0.2777
	20		0.1517	0.8436	0.8357	0.2170

	25		0.1074	0.8945	0.8886	0.1583
	28		0.0840	0.9200	0.9153	0.1267
0.18	1	初期値	0.1836	0.7652	0.7654	0.2150
増加分	3	(1.5)	0.1491	0.8300	0.8197	0.2322
(0.09)	5	増加分	0.1265	0.8513	0.8421	0.1902
$\alpha = 0.025$	7	(0.005)	0.1058	0.8734	0.8656	0.1545
一定	9		0.0872	0.8951	0.8887	0.1250
	10		0.0787	0.9053	0.8997	0.1121

①最初に、学習率を0.9一定、モーメント項に対する係数 $\alpha$ を $\alpha = 0$ 、0.05、0.1、0.3と変化させた場合のデータを第16表の初めに示した。ただし、 $k$ の値は反復計算ごとに0.01増加させた。その結果、モーメント項がない場合には、52回の反復計算を必要としたのに対して、 $\alpha = 0.05$ の時12回、 $\alpha = 0.1$ の時8回、 $\alpha = 0.3$ の時には4回の反復計算で平均2乗誤差 $< 0.01$ を満足する結果となった。この結果、 $\alpha$ の値が大きくなればなるほど反復回数が減少することがわかる。

②学習率0.18一定、 $\alpha = 0.025$ 一定とし、 $k$ の値を初期値1.0とし、反復ごとに0.01、0.05、0.0と変えた場合の結果を第16表の中間に示した。その結果、増加分が0.05の時12回、0.01の時21回、0.0の時28回という結果となった。したがって、モーメント項を付加した場合でも反復学習ごとに $k$ の値を大きくすればするほど、学習回数は減少することとなった。

③最後に、②に対して、 $k$ の初期値を1.5とし、 $k$ の増加分を0.005として計算を行った結果、学習回数は10回となった。学習率の値を小さくした割には、 $k$ の初期値を大きくすることで、学習回数は減少させることができた。

## 7. おわりに

本研究では、ニューラルネットワークの重み係数の更新方法に対して、学習率、シグモイド関数の $k$ を取り入れ、更新に対する収束性や初期値の設定法および収束速度の改善を考察した。実際に計算機シミュレーション

ンを行って、行列  $H$  の固有値がすべて 0 になれば収束することもシミュレーションを行って示した。従来は、 $\eta$  を変化させるのみであったが、本稿では、シグモイド関数の  $k$  および学習率の両方を考慮（反復計算ごとに増加など）すれば、収束性、初期値探索、収束速度の改善に利用できることを示した。最後に、モーメント項を付加することによって、更なる改善が行われることを示した。

〔謝辞〕 本研究に対して、ご指導いただいた、東京工業大学の美多勉教授、同大学特別研究員の東南大学忻欣副教授、劉延年副教授、並びに UNIX 関係でいろいろお世話になった種子田昭彦院生や他の院生に感謝いたします。

(参考文献)

八名和夫・鈴木義武共著『ニューロコンピューティング』、海文堂、1992年、44-49ページ (1992a)。

Robert Hecht Nielsen, *Neurocomputing*, 1990, pp. 124-137, Addison Wesley.

松井伸之・石見憲一「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、1994年、1208-1213ページ。

八名和夫・鈴木義武共著『ニューロ情報処理技術』、海文堂、1992年、73-85ページ (1992b)。

美多勉・小郷寛共著『システム制御理論入門』、実教出版株式会社、1979年、55-57ページ。

本間光一・小島紀男・矢沢志雄作共著『マトリクスとシステム』、東海大学出版会、1990年、8ページ。

R・ビール/T・ジャクソン著（八名和夫監訳）『ニューラルコンピューティング入門』、海文堂、1993年、68-80ページ。