

一般抽象実ヒルベルト空間での ニューラルネットのon-line学習

鈴木 昇一

On-Line Learnings of Neural Networks on Abstract Real Hilbert Space

Shoichi Suzuki

あらまし

適切な初期値から出発し、ニューラルネットワーク内の重みの更新（学習）を繰り返して理想出力を得る写像を求める方法は反復学習法（iterative learning method）と呼ばれる。反復法には2つの様式があり、それは、次のように述べられる：

（1#）オンライン学習（on-line learning）入出力事例毎に重みを更新する。

（2#）バッチ学習（batch learning）複数個の入出力事例の影響を1度に評価し、重みを更新する。

オンライン学習はシステムが稼動しているときでさえ、学習の効果が不足している場合、その時処理を中断して、重みを更新し、その後稼動を再開できる脳が採用している方法である。

従来の各種ニューラルネットワークでは、各ニューロンの入力、出力は共に数である。本論文では、この数を関数（可分な一般抽象ヒルベルト空間の元）に置き換えた各種ニューラルネットワークのon-line学習法が研究される。このように、本研究はこれまでの研究者が決して挑戦しなかった問題を解決しようとしている。

具体的には、最急降下法を適用して、

- （1）クラスタリングの機能を備えたパーセプトロン型ニューラルネット
- （2）入力パターンをカテゴリの代表パターンの近くに写像する連想ニューラルネット
- （3）1次ニューラルネット、並びに、パターンから抽出された特徴量の組を用いた1次ニューラルネット
- （4）再帰ニューラルネット
- （5）3層誤差逆伝播学習ニューラルネット、並びに、特徴量の組を用いた3層誤差逆伝播学習ニューラルネット
- （6）2次ニューラルネット

における稼動公式、並びに、on-line学習公式が研究される。いずれも、実用に耐える形式と構造を備えていることが特色である。（1）に関しては、その類似度関数 SM の構成への応用が論じられる。

本研究により、ニューラルネットワークによる情報処理に関する数理情報学（ヒルベルト空間での

多変数非線形関数解析学 [21]) は理論分野、実用分野双方において、新局面を迎えたことになった。実用に向けての新しい適用分野を変革し、飛躍的な進歩を遂げる数理的基礎が築かれたことになった。本論文により、脳による思考機能の数理的解明、並びに、身体運動機能の実現に向けて、ニューラルネットワークの理論がその本来の役割を果たせるようになったといえよう。

キーワード

- (1) ヒルベルト空間 (2) クラスタリング (3) 類似度関数 (4) パターン連想
- (5) 1 次ニューラルネット (6) パターン特徴量 (7) 再帰ニューラルネット
- (8) 3 層誤差逆伝播学習ニューラルネット (9) 最急降下法 (10) SS理論
- (11) オンライン学習

Abstract

The learning-updating method of asking for the mapping which leaves a suitable initial value, repeats renewal of the weights in a neural network, and obtains the specified desired outputs is called the repeating method. There are two styles in the repeating method and it is described as follows :

(1#)(on-line learning) The weights are updated for every example of an input and an output.

(2#)(batch learning) The influence of two or more input-and-output examples is evaluated at a time, and the weights are updated. □

On-line learning method is the learning method which interrupts processing on a momentary target and updates the weights and which was easy things, can resume operation , and the brain has adopted, when the effects of study run short, even while the system is working.

In the various conventional neural networks, both the input of each neuron and an output are numbers. In this paper, the on-line learning method of the various neural networks which transposed this number to the function which is an element of a separable general abstract Hilbert space \mathfrak{H} is studied. This research tends to solve the problem which many old researchers never challenged . The method of steepest descent is applied specifically. Formulae required to work and on-line study formulae required to on-line learn in the following six items(1)~(6) are studied:

(1)A perceptron-like neural network equipped with the function of clustering

(2) An associative neural network which maps an input pattern near the prototypical pattern of a category

(3) A first-order neural network and the first-order neural network using the group of the amount of the features extracted from the pattern

(4) A recurrent neural network

(5) A three-layer error-backpropagation-learning neural network and the three-layer error-back-propagation-learning neural network using the group of the amount of the features extracted from the pattern

(6)A second-order neural network □

It is the special feature to have form and structure that all are equal to practical use. About the item(1), the application to the composition of the similarity-measure function SM is discussed.

It means that the mathematical informatics about information processing by the neural network or the multi-variable nonlinear functional-analysis study in Hilbert space had greeted the new state in both the theoretical field and the practical use field by this research. It means that the new application field towards practical use is reformed, and the mathematical foundation which accomplishes fast progress was built. With this paper, it can say that a neural network's theory can play the original role now towards the mathematical elucidation of the thinking function by the brain, and realization of a body movement function.

Key words: (1) Hilbert space (2) clustering (3) similarity-measure function (4) association using patterns (5) first-order neural net (6) set of features extracted from pattern (7) recurrent neural net (8) three-layer backpropagation-learning neural-net (9) gradient descent procedure (10) SS theory suggested by S.Suzuki (11) on-line learning

第1章 まえがき

1. 1 本研究内容は従来のニューラルネットワーク理論、線形多変量解析学を変革させる

本論文は従来のニューラルネットワーク理論 [7], [8], [9], [17] を変革させる「一般抽象実ヒルベルト空間 \mathfrak{H} で動作可能な各種ニューラルネットの、事例からのon-line学習法」を研究したものである。

on-line学習とは、各学習時刻に全部の訓練事例を使って重みを修正させ、1度に学習を完結させようとする一括学習 (batch learning) と対比するものであり、各学習時刻に唯1つの訓練事例を使って、学習の効果が不足しているとき、稼動中でも重みの修正動作を可能ならしめる逐次学習 (sequential learning) のことである。

従来の線形の判別分析、主成分分析、回帰分析などの有限次元多変量解析手法を線形の場合を含めて非線形に一般化するには、従来の有限次元ニューラルネットワーク理論が役立つことはよく知られている [22]。非線形多変量解析学を無限次元化するには、本研究内容が役立つことは明白であることを強調しておかなければならない。

1. 2 S. Suzukiの、これまでのニューラルネット研究

パターン情報処理における可分な一般抽象ヒルベルト空間 \mathfrak{H} における量子力学的原理をS. Suzukiは提案し、この原理に基づいて、「入力パターン $\varphi \in \mathfrak{H}$ の代りとなり、認識システムの内部に確保され、ユニタリ座標変換に不変・共変的なパターンモデル」 $T\varphi \in \mathfrak{H}$ を提案している [1], [5], [6]。このパターンモデル $T\varphi$ はパターン φ から特徴抽出した後、確保されるパターンであり、SS理論 [3], [4] のaxiom 1を満たす典型的なものである。

また、S.Suzukiは、階層型ニューラルネットにおいて、平均値ゼロ・等分散の正規分布を適応誤差の確率分布とする誤差逆伝播学習が従来の最小自乗法に基づく誤差逆伝播学習であることを明らかにしたが [2]、本論文はこの文献 [2] で割愛された一般抽象ヒルベルト空間 \mathfrak{H} で動作可能な各種ニューラルネットを研究したものである。但し、実用にあたって不便が生じない程度に単純化するために複素空間ではなく、実空間で展開している。

1. 3 本研究内容は認識システムRECOGNITRONの構成に使える

パターン認識の働きは、入力パターンを「その帰属するであろうカテゴリ」の諸性質を典型的に代表するパターン（代表パターン）へ変換する写像（連想形認識写像）でなされる（prototype-based recognition）と考えてみよう。問題は、入力パターンが帰属するカテゴリの少数候補すら、当然ながら、事前に判明していないことである。S. Suzukiはパターンであると判明しているものの集合 Φ_B 、モデル構成作用素 T 、類似度関数 SM 、大分類関数に BSC 関する公理系（付録A）を打ちたて、パターン認識の数学的理論（SS理論）を提案し、多段階認識過程で入力パターン φ のモデル $T\varphi$ をある代表パターン ω_j のモデル $T\omega_j$ へと変換するこの種の連想形認識写像

$$\text{RECOGNITRON} =_{\Delta} \text{RECOGNITRON}(\mathfrak{C}(J) : \Omega(J)) = \langle \Phi_B, T, SM, BSC \rangle \quad (1.1)$$

を構築している。連想形認識写像RECOGNITRONの構造を予め用意しておいて、訓練事例の系列を使って構造内の助変数を学習で決定していくという考えである。処理の対象とする問題の入力パターン $\varphi \in \Phi$ を認識しようとする認識システムRECOGNITRONの認識動作は、連想形認識方程式（equation of associative recognition）

$$MK =_{\Delta} \langle T\varphi, J \rangle \sqcup TA(\mu)T \cdot MK \quad (1.2)$$

の解（カテゴリ帰属知識；categorical-membership knowledge） $MK =_{\Delta} \langle \phi, \lambda \rangle$ を求める多段階連想形認識過程である。正常な場合 $\phi = T\omega_j, \lambda = [j]$ が連想形認識方程式（1.2）の求解過程で得られる。式（1.1）の認識システムRECOGNITRONは、ありとあらゆる認識写像をシミュレーションできるという意味で万能である。この事実はニューラルネットワークの理論が万能であることが証明されている次の事実と相応している：

K.Funahashiは、3層ニューラルネットワークの出力が任意の連続関数を望むだけの精度で近似できるような構成に成功している [17]。更に、連続値出力ニューロンからなる再帰形ニューラルネットワークは、任意の決定性チューリング機械をシミュレーションできる [10]。ニューラルネットワークはこの意味で、情報処理に関し万能である。□

T , SM , BSC を構成するのに、本論文のニューラルネットを使うことができる。

1. 4 ニューラルネットワークをヒルベルト空間上の多変数非線形作用素と考え直したこと

これまでのニューラルネット理論では、ニューロンへの入力、ニューロンから出力、中間層のニューロンの入出力は1つの数であったが、本論文では、これらの数が関数（実は、一般抽象実ヒルベルト空間 \mathfrak{H} の元）で置き換えられた各種ニューラルネットワークが誤差逆伝播学習ニューラルネットワーク、再帰ニューラルネットワークを含めて、研究される。このような研究は本論文が初めてであり、ニューラルネットワーク理論は新局面を迎えたことになり、飛躍的に新しい応用が期待される。

ニューラルネットワークの稼動領域をヒルベルト空間化とすること（ニューラルネットワークをヒルベルト空間上の多変数非線形作用素と考え直すこと）には多くの困難を伴うが、本研究はこの種の多くの諸困難を克服してなされたことを強調しておきたい。

1. 5 SS理論での類似度関数の構成への1つの応用

簡単な応用の1つで、本研究内容の有益性を以下に示しておこう。

各 $w_\ell (\ell \in L)$ を実定数とし、各 $P_\ell : \mathfrak{H} \rightarrow \mathfrak{H}$ をからへの非線形であってもよい作用素とする。

列ベクトル

$$\vec{w} \equiv \text{col}(w_1 \quad w_2 \quad \cdots \quad w_n) \quad (1.3)$$

の代りに、パーセプトロン型作用素 (perceptron-like operator) [6]

$$P \equiv \sum_{\ell \in L} w_\ell \cdot P_\ell \quad (1.4)$$

を使うと、大抵の、数空間で動作するこれまでのニューラルネットワーク、例えば、誤差逆伝播学習ニューラルネットワークは関数空間で動作する形式に変換され、訓練事例からの学習が従来の学習法と同様に実行できることを示すのが、本論文の目的である。尚、閾値を考慮しないパーセプトロンは n 次元ユークリッド空間 R^n での、後述の式 (2.3) で表される 1 次写像であり、閾値 $-h_j$ を考慮した第 j 番目のパーセプトロンとは式 (4.2) で表される出力 y_j を持つ非線形写像である。

パターンの集合として、

$$\{\varphi_1, \varphi_2, \dots, \varphi_n\} (\supseteq \Omega) \quad (1.5)$$

を用意する。例えば、 $L = \{1, 2, 3, 4, 5, 6, 7\}$ であれば、各作用素 $P_\ell (\ell \in L)$ を次のように選ぶことができる：

$$(P_1\varphi)(x_1, x_2) = \varphi(x_1, x_2) \quad (\text{視野の中心点の抽出}) \quad (1.6)$$

$$(P_2\varphi)(x_1, x_2) = \varphi(x_1, x_2) + \frac{1}{2} \cdot \varphi(x_1, x_2 + 1) + \frac{1}{2} \cdot \varphi(x_1, x_2 - 1) \\ (\text{視野の中心点にある縦棒の抽出}) \quad (1.7)$$

$$(P_3\varphi)(x_1, x_2) = \varphi(x_1, x_2) + \frac{1}{2} \cdot \varphi(x_1 + 1, x_2) + \frac{1}{2} \cdot \varphi(x_1 - 1, x_2) \\ (\text{視野の中心点にある横棒の抽出}) \quad (1.8)$$

$$(P_4\varphi)(x_1, x_2) = \varphi(x_1, x_2) + \frac{1}{2} \cdot \varphi(x_1 + 1, x_2 + 1) + \frac{1}{2} \cdot \varphi(x_1 - 1, x_2 - 1) \\ (\text{視野の中心点にある傾き45度の棒の抽出}) \quad (1.9)$$

$$(P_5\varphi)(x_1, x_2) = \varphi(x_1, x_2) + \frac{1}{2} \cdot \varphi(x_1 - 1, x_2 + 1) + \frac{1}{2} \cdot \varphi(x_1 + 1, x_2 - 1) \\ (\text{視野の中心点にある傾き-45度の棒の抽出}) \quad (1.10)$$

$$(P_6\varphi)(x_1, x_2) = \frac{1}{9} \cdot [\varphi(x_1, x_2) - [\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}]\varphi(x_1, x_2)] \\ \doteq \frac{1}{9} \cdot [5 \cdot \varphi(x_1, x_2) + (-1) \cdot \varphi(x_1 + 1, x_2) + (-1) \cdot \varphi(x_1 - 1, x_2) \\ + (-1) \cdot \varphi(x_1, x_2 + 1) + (-1) \cdot \varphi(x_1, x_2 - 1)] \\ (\text{視野の中心点での、先鋭化形状成分の抽出}) \quad (1.11)$$

$$(P_7\varphi)(x_1, x_2) = \frac{1}{4} \cdot \int_{x_1}^{x_1+2} dy_1 \int_{x_2}^{x_2+2} dy_2 \varphi(y_1, y_2)$$

$$\begin{aligned} &\doteq \frac{1}{36} \cdot [\varphi(x_1, x_2) + 4 \cdot \varphi(x_1, x_2 + 1) + \varphi(x_1, x_2 + 2) \\ &\quad + 4 \cdot \varphi(x_1 + 1, x_2) + 16 \cdot \varphi(x_1 + 1, x_2 + 1) + 4 \cdot \varphi(x_1 + 1, x_2 + 2) \\ &\quad + \varphi(x_1 + 2, x_2) + 4 \cdot \varphi(x_1 + 2, x_2 + 1) + \varphi(x_1 + 2, x_2 + 2)] \end{aligned}$$

(視野の中心点での、平滑化形状成分の抽出)

∴ シンプソンの積分近似公式 (1.12)

パターン $\varphi \in \Phi$ が与えられた場合、そのモデル $T\varphi \in \Phi$ を求める。その後、 $T\varphi \in \Phi$ を 2. 2 節で求められた式 (2.33) の実数値重みの組 $\{w_\ell\}_{\ell \in L}$ を使ったパーセプトロン型作用素 P で写像し、パターン

$$P(T\varphi) = \sum_{\ell \in L} w_\ell \cdot (P_\ell T\varphi) \quad (1.13)$$

を求める。式 (2.27) の関数を最小にする実数値重みの組 $\{w_\ell\}_{\ell \in L}$ を求めることは、

An attempt to ensure that if $T\varphi_p$ and $T\varphi_q$ are close then $P(T\varphi_p)$ and $P(T\varphi_q)$ are close as well (1.14)

である。

非一致条件

$$i \neq j \in J \Rightarrow T\omega_i \neq T\omega_j \quad (1.15)$$

の下で、

$$SM(\varphi, \omega_j) = \frac{\left\| \sum_{\ell \in L} w_\ell \cdot P_\ell (T\varphi - T\omega_j) \right\|^{-2}}{\sum_{i \in J} \left\| \sum_{\ell \in L} w_\ell \cdot P_\ell (T\varphi - T\omega_i) \right\|^{-2}} \quad (1.16)$$

と定義される式 (A.19) の関数 SM は付録Aのaxiom 2を満たす。 \square

1. 6 その他の応用について

その他、多変数データから規則を発見する回帰問題・データマイニング、雑音に汚された画像の復元、画像理解 [18] への応用、音声認識システム [19] の構成への適用、独立成分分析 (混ぜ合わさった信号が観測されたときに各信号を分離する手法) [20] への応用、SS理論 [3]、[4] の大分類関数の構成への応用など、広大な適用分野が考えられる。

第2章 クラスタリングの機能を持つパーセプトロン型作用素 P のon-line学習

本章では、同一カテゴリに帰属するパターン対の2パターンがより近づくような (clustering), (出来れば、異なったカテゴリに帰属するパターン対の2パターンがより大きく分離できるような) パーセプトロン型作用素 P の重み $\{w_\ell\}_{\ell \in L}$ を求める。

2. 1 n 次元ユークリッド空間 R^n でのon-line学習

2. 1. 1 問題の定式化

Let column vector

$$\vec{x}_i = \text{col}(x_{1i} \ x_{2i} \ \cdots \ x_{ni}) \in R^n, \ i=1 \sim m \quad (2.1)$$

denote the set of pattern-samples in n-dimensional Euclidian space R^n .

実数値重みベクトル

$$\vec{w} = \text{col}(w_1 \ w_2 \ \cdots \ w_n) \in R^n \quad (2.2)$$

で決まる1次写像

$$\vec{y} = (w_1 \ w_2 \ \cdots \ w_n) \cdot \vec{x} = \sum_{i=1}^n w_i \cdot x_i \quad (2.3)$$

where

$$\vec{x} = \text{col}(x_1 \ x_2 \ \cdots \ x_n) \in R^n \quad (2.4)$$

を導入する。

The local structure of the image spaceを評価している2値数

$$S_{ij} =$$

$$\begin{cases} 1 & \text{if } \vec{x}_i \text{ is among the } k \text{ nearest neighbors of } \vec{x}_j \text{ or } \vec{x}_j \text{ is among the } k \text{ nearest} \\ & \text{neighbors of } \vec{x}_i \\ 0 & \text{otherwise} \end{cases} \quad (2.5)$$

を用意する。

The optimal projections preserving locality can be obtained by the following minimization problem :

$$\min_{\vec{w}} F \quad (2.6)$$

$$\text{where } F = \frac{1}{2} \cdot \sum_{i=1}^m \sum_{j=1}^m \left[\sum_{\ell=1}^n w_\ell \cdot x_{\ell i} - \sum_{\ell=1}^n w_\ell \cdot x_{\ell j} \right]^2 \cdot S_{ij} . \quad (2.7)$$

\vec{w} を変数とする式 (2.7) の F の最小化は、

An attempt to ensure that if \vec{x}_i and \vec{x}_j are close then $(w_1 \ w_2 \ \cdots \ w_n) \cdot \vec{x}_i$ and $(w_1 \ w_2 \ \cdots \ w_n) \cdot \vec{x}_j$ are close as well

(2.8)

を意味する。

文献[11]では、 F に最小をもたらす重み式 (2.2) の重み \vec{w} をある程度、解析的に求めている。

2. 1. 2 on-line学習

式 (2.7) の関数 F が最小となる式 (2.2) の重み \vec{w} をon-line学習で決定しよう。

適当な初期値 $w_\ell(t)|_{t=0}$ ($\ell \in L$) から出発して、関数

$$F(t) = \frac{1}{2} \cdot \sum_{i=1}^m \sum_{j=1}^m \left[\sum_{\ell=1}^n w_\ell(t) \cdot x_{\ell i} - \sum_{\ell=1}^n w_\ell(t) \cdot x_{\ell j} \right]^2 \cdot S_{ij} \quad (2.9)$$

を減少させるように、学習時刻 $t (= 0, 1, 2, \dots)$ での各 $w_\ell(t)$ ($\ell = 1 \sim n$) の更新式

$$w_\ell(t+1) = w_\ell(t) + \Delta w_\ell(t) \quad (2.10)$$

での各修正分 $(\Delta w_\ell)(t)$ ($\ell = 1 \sim n$) を求める。それには、最急降下法 (a gradient descent procedure) を適用し、

$$\forall \ell (= 1 \sim n), [\forall t (0 \leq t < \Delta), \varepsilon_\ell(t) > 0] \wedge \lim_{t \rightarrow \infty} \varepsilon_\ell(t) = 0 \quad (2.11)$$

を満たす各学習率 (learning rate) $\varepsilon_\ell(t)$ を十分小さく選び、各修正分 $(\Delta w_\ell)(t)$ ($\ell = 1 \sim n$) を

$$\Delta w_\ell(t) = -\varepsilon_\ell(t) \cdot \frac{\partial F(t)}{\partial w_\ell(t)} \quad (2.12)$$

とすればよい。

そうすれば、不動点方程式

$$F(t+1) = F(t) \quad (2.13)$$

が成立する時刻 t が学習終了時刻である。このとき、求める各 $w_\ell(t)$ は

$$w_\ell = w_\ell(t), \ell \in L \quad (2.14)$$

と与えられる。

式 (2.12) の $\Delta w_\ell(t)$ は

$$\Delta w_\ell(t) = -\varepsilon_\ell(t) \cdot \frac{1}{2} \cdot \sum_{i=1}^m \sum_{j=1}^m 2 \cdot \left[\sum_{k=1}^n w_k(t) \cdot (x_{ki} - x_{kj}) \right] \cdot [x_{\ell i} - x_{\ell j}] \cdot S_{ij} \quad (2.15)$$

と計算されるが、ここで、 $q_{\ell k}$ を

$$q_{\ell k} \equiv \sum_{i=1}^m \sum_{j=1}^m (x_{\ell i} - x_{\ell j}) \cdot (x_{ki} - x_{kj}) \cdot S_{ij} \quad (2.16)$$

とおけば、

$$\Delta w_\ell(t)$$

$$= -\varepsilon_\ell(t) \cdot \sum_{k=1}^n q_{\ell k} \cdot w_k(t) \quad \because \text{各 } w_k(k=1 \sim n) \text{ は実定数} \quad (2.17)$$

と計算される。ここに、対称性

$$\forall \ell, \forall k(=1 \sim n), q_{\ell k} = q_{k\ell} \quad (2.18)$$

が成り立っている。尚、式 (2.9) の $F(t)$ は、

$$F(t) = \frac{1}{2} \cdot \sum_{i=1}^m \sum_{j=1}^m \left[\sum_{k=1}^n w_k(t) \cdot (x_{ki} - x_{kj}) \right] \left[\sum_{\ell=1}^n w_\ell(t) \cdot (x_{\ell i} - x_{\ell j}) \right] \cdot S_{ij} \quad (2.19)$$

を得、結局、

$$F(t) = \frac{1}{2} \cdot \sum_{k=1}^n \sum_{\ell=1}^n q_{k\ell} \cdot w_k(t) \cdot w_\ell(t) \quad (2.20)$$

と再表現される。

2. 2 無限次元一般抽象実ヒルベルト空間 \mathfrak{H} でのon-line学習

式 (1.4) のパーセプトロン型作用素 P を提案したのは、S.Suzukiである[6]。そこでは、 P 内の各作用素

$$P_\ell : \mathfrak{H} \rightarrow \mathfrak{H}, \ell \in L \quad (2.21)$$

は射影作用素である。以下では、各作用素 $P_\ell(\ell \in L)$ は7式 (1.6) ～ (1.12) に見られる如く、射影作用素とは限らない。

前節の、 R^n での内容が一般抽象実ヒルベルト空間 \mathfrak{H} で成り立つように、書き直す。

それには、式 (2.2) の実数値重みベクトルの代りに、式 (1.4) のパーセプトロン型作用素 P を用意しなければならない。

可分な一般抽象実ヒルベルト空間 \mathfrak{H} の元であるパターンの集合

$$\varphi_1, \varphi_2, \dots, \varphi_n \quad (2.22)$$

を用意する。但し、付録Aの式 (A.3) のモデル構成作用素 T を導入して、2条件

$$i \neq j \in J \Rightarrow p(i) \neq p(j) \in \{1, 2, \dots, n\} \quad (2.23)$$

$$\forall j \in J, \exists p(j) \in \{1, 2, \dots, n\}, T\omega_j = T\varphi_{p(j)} \quad (2.24)$$

を満たしているように、式 (2.22) のパターン集合を選ぶことが望ましい。

対 $[\Phi, T]$ がaxiom 1を満たすとしよう。パターン $\varphi \in \Phi$ のモデル $T\varphi \in \Phi$ が定義される。

次のように定義される各 $S_{pq}(p, q \in \{1, 2, \dots, n\})$ を決めておく。

$$S_{pq} = \begin{cases} 1 & \text{if } T\varphi_p \text{ is among the } k \text{ nearest neighbors of } T\varphi_q \text{ or } T\varphi_q \text{ is among the } k \\ & \text{nearest neighbors of } T\varphi_p \text{ (} T\varphi_p \text{ が } T\varphi_q \text{ に近い} \text{か、或いは、} T\varphi_q \text{ が } T\varphi_p \text{ に近い場合)} \\ 0 & \text{otherwise (} T\varphi_p \text{ が } T\varphi_q \text{ に近くなくて、かつ、} T\varphi_q \text{ が } T\varphi_p \text{ に近くない場合)} \end{cases} \quad (2.25)$$

□

$$S_{pq} \text{ can be thought of as a similarity measure between } T\varphi_p \text{ and } T\varphi_q. \quad (2.26)$$

関数

$$F \equiv \frac{1}{2} \cdot \sum_{p=1}^n \sum_{q=1}^n \left\| \sum_{\ell \in L} w_\ell \cdot P_\ell T\varphi_p - \sum_{\ell \in L} w_\ell \cdot P_\ell T\varphi_q \right\|^2 \cdot S_{pq} \quad (2.27)$$

を最小にする実数値重みの組 $\{w_\ell\}_{\ell \in L}$ を on-line 学習で求めよう。それは、次のように述べられる。

【式 (1.4) のパーセプトロン型作用素 P の on-line 学習法】

学習時刻 $t (= 0, 1, 2, \dots)$ での、 w_ℓ の値 $w_\ell(t)$ をと表す。

[initialization]

各重み変数 $w_\ell (\ell \in L)$ の初期値として、

$$w_\ell(t)|_{t=0} = \text{十分小さな正の定数、} \ell \in L \quad (2.28)$$

を採用する。

[induction]

各重み変数 $w_\ell (\ell \in L)$ の適応評価関数

$$F(t) \equiv \frac{1}{2} \cdot \sum_{p=1}^n \sum_{q=1}^n \left\| \sum_{\ell \in L} w_\ell(t) \cdot P_\ell T\varphi_p - \sum_{\ell \in L} w_\ell(t) \cdot P_\ell T\varphi_q \right\|^2 \cdot S_{pq} \quad (2.29)$$

を減少させるように、各重み $w_\ell(t) (\ell \in L)$ を更新する式として、

$$w_\ell(t+1) = w_\ell(t) + \Delta w_\ell(t) \quad (2.30)$$

を採用し、各 $w_\ell(t)$ の修正分 $(\Delta w_\ell)(t)$ を求める。最急降下法を適用して、

$$\Delta w_\ell(t) = -\varepsilon_\ell(t) \cdot \frac{\partial F(t)}{\partial w_\ell(t)} \quad (2.31)$$

と求める。ここに、 $\varepsilon_\ell(t) > 0$ は学習率である。

[termination]

不動点方程式

$$F(t+1) = F(t) \quad (2.32)$$

が成立する時刻が学習終了時刻である。このとき、求める各重み $w_\ell(t)$ は

$$w_\ell = w_\ell(t), \ell \in L \quad (2.33)$$

と与えられる。 \square

式 (2.31) の各更新分 $\Delta w_\ell(t) (\ell \in L)$ を計算しよう。

学習時刻 t での、式 (2.29) の関数 $F(t)$ は、

$$F(t) = \frac{1}{2} \cdot \sum_{p=1}^n \sum_{q=1}^n \left(\sum_{k \in L} w_k(t) \cdot P_k[T\varphi_p - T\varphi_q], \sum_{m \in L} w_m(t) \cdot P_m[T\varphi_p - T\varphi_q] \right) \cdot S_{pq} \quad (2.34)$$

と書き直されるから、

$$(\varphi, \eta) = (\eta, \varphi) \quad \because \quad \mathfrak{H} \text{ は実ヒルベルト空間} \quad (2.35)$$

を考慮し、式 (2.31) の $\Delta w_\ell(t)$ を計算すれば、

$$\begin{aligned} \Delta w_\ell(t) \\ = -\varepsilon_\ell(t) \cdot \frac{1}{2} \cdot \sum_{p=1}^n \sum_{q=1}^n 2 \cdot ([P_\ell(T\varphi_p - T\varphi_q], \sum_{m \in L} w_m(t) \cdot P_m[T\varphi_p - T\varphi_q]) \cdot S_{pq} \end{aligned} \quad (2.36)$$

である。ここで、 $g_{\ell m}$ を

$$g_{\ell m} \equiv \sum_{p=1}^n \sum_{q=1}^n (P_\ell[T\varphi_p - T\varphi_q], P_m[T\varphi_p - T\varphi_q]) \cdot S_{pq} \quad (2.37)$$

とおけば、

$$\begin{aligned} \Delta w_\ell(t) \\ = -\varepsilon_\ell(t) \cdot \sum_{m \in L} g_{\ell m} \cdot w_m(t) \quad \because \quad \text{各 } w_\ell (\ell \in L) \text{ は実定数} \end{aligned} \quad (2.38)$$

と計算される。ここに、対称性

$$\forall \ell, \forall m \in L, g_{\ell m} = g_{m\ell} \quad \because \quad \mathfrak{H} \text{ は実ヒルベルト空間} \quad (2.39)$$

が成り立っている。尚、式 (2.14) の関数 $F(t)$ は、

$$F(t) = \frac{1}{2} \cdot \sum_{k \in L} \sum_{\ell \in L} w_k(t) \cdot w_\ell(t) \sum_{p=1}^n \sum_{q=1}^n (P_k[T\varphi_p - T\varphi_q], P_\ell[T\varphi_p - T\varphi_q]) \cdot S_{pq} \quad (2.40)$$

であるから、結局、2 次形式の形に

$$F(t) = \frac{1}{2} \cdot \sum_{k \in L} \sum_{\ell \in L} g_{k\ell} \cdot w_k(t) \cdot w_\ell(t) \quad (2.41)$$

と再表現される。

本節で得られた式 (1.4) のパーセプトロン型作用素 P を使って、付録Aのaxiom 2を満たす式 (A.19) の類似度関数 SM を構成できることは、1. 5 節で解説されている。

第3章 パターンをカテゴリの代表パターンの近くに写像する連想ニューラルネット

本章では、文献 [12] の 3-3-4 項の、 n 次元ユークリッド空間 R^n でのパターン認識の働きの一括学習 (batch learning) (pp.128-131) を 3.1 節で説明し、その後、この内容を可分な一般抽象ヒルベルト空間 \mathfrak{H} で論じ、 \mathfrak{H} での on-line 学習法を研究する。

3.1 n 次元ユークリッド空間 R^n での一括学習

R^n での内積 (\vec{x}, \vec{y}) , ノルム $\|\vec{x}\|$ は

$$(\vec{x}, \vec{y}) = \sum_{q=1}^n x_q \cdot y_q, \|\vec{x}\| = \sqrt{(\vec{x}, \vec{x})} \quad (3.1)$$

where

$$\vec{x} = \text{col}(x_1 \ x_2 \ \cdots \ x_n), \vec{y} = \text{col}(y_1 \ y_2 \ \cdots \ y_n) \quad (3.2)$$

である。

入力パターン $\vec{x} \in R^n$ の類別、つまり、

$$\min_{i=1 \sim k} \|A\vec{x} - \nu_i\| = \|A\vec{x} - \nu_j\| \quad (3.3)$$

ならば、

$$\vec{x} \text{ belongs to the } j \text{ th category } \mathfrak{S}_j \quad (3.4)$$

が可能になるような行列 A を求めてみよう。

まず、the i th category \mathfrak{S}_i の代表パターン

$$\vec{\nu}_i = \text{col}(\nu_{1i} \ \nu_{2i} \ \cdots \ \nu_{ni}) \in R^n \quad (3.5)$$

を選ぶ ($i=1 \sim k$)。

同じカテゴリ \mathfrak{S}_j に帰属する複数個の訓練パターン

$$\vec{x}_{ij}, j=1, 2, \dots, N_i \quad (3.6)$$

を選定し、この訓練パターンをすべて同一の代表パターン $\vec{\nu}_i$ に変換できるような行列 A を求めればよい、つまり、すべてのカテゴリ番号 $i=1 \sim k$ について、等式

$$A\vec{x}_{ij} = \vec{\nu}_i, j=1, 2, \dots, N_i \quad (3.7)$$

が成立するような行列 A を求めればよいのであるが、このようなことは一般的には不可能である。それで、

$$A\vec{x}_{ij}, j=1 \sim N_i \quad (3.8)$$

をすべて、同一の代表パターン $\vec{\nu}_i$ の近傍に変換するような行列 A を求めることにしよう。

平均自乗誤差

$$\varepsilon_{i,j} = \|A\vec{x}_{ij} - \vec{v}_i\|^2 \quad (3.9)$$

の算術平均

$$\varepsilon_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \varepsilon_{i,j} \quad (3.10)$$

の総和

$$\begin{aligned} \varepsilon &= \sum_{i=1}^k \varepsilon_i = \sum_{i=1}^k \frac{1}{N_i} \sum_{j=1}^{N_i} \|A\vec{x}_{ij} - \vec{v}_i\|^2 \\ &= \sum_{i=1}^k \frac{1}{N_i} \sum_{j=1}^{N_i} (A\vec{x}_{ij} - \vec{v}_i, A\vec{x}_{ij} - \vec{v}_i) \end{aligned} \quad (3.11)$$

$$= \sum_{i=1}^k \frac{1}{N_i} \sum_{j=1}^{N_i} [(A\vec{x}_{ij}, A\vec{x}_{ij}) - (A\vec{x}_{ij}, \vec{v}_i) - (\vec{v}_i, A\vec{x}_{ij}) + (\vec{v}_i, \vec{v}_i)] \quad (3.12)$$

を最小にするような行列 A を求めることにしよう。

行列 A を未知とする行列方程式

$$\frac{\partial}{\partial A} \varepsilon = 0 \quad (3.13)$$

を解けば、

$$\frac{\partial}{\partial A} (A\vec{x}_{ij}, A\vec{x}_{ij}) = 2A\vec{x}_{ij} \cdot \vec{x}_{ij}^t \quad (3.14)$$

$$\frac{\partial}{\partial A} (A\vec{x}_{ij}, \vec{v}_i) = \vec{v}_i \cdot \vec{x}_{ij}^t \quad (3.15)$$

$$\frac{\partial}{\partial A} (\vec{v}_i, A\vec{x}_{ij}) = \vec{v}_i \cdot \vec{x}_{ij}^t \quad (3.16)$$

$$\frac{\partial}{\partial A} (\vec{v}_i, \vec{v}_i) = 0 \quad (3.17)$$

であるから、

$$A = \left[\sum_{i=1}^k \frac{1}{N_i} \sum_{j=1}^{N_i} \vec{v}_i \cdot \vec{x}_{ij}^t \right] \cdot \left[\sum_{i=1}^k \frac{1}{N_i} \sum_{j=1}^{N_i} \vec{x}_{ij} \cdot \vec{x}_{ij}^t \right]^{-1} \quad (3.18)$$

と求まる。ここに、 \vec{x}_{ij}^t は縦ベクトル \vec{x}_{ij} の転置であり、横ベクトルである。

3. 2 一般抽象実ヒルベルト空間 \mathfrak{H} でのon-line学習

$\bar{x} \in R^n$ から $A\bar{x} \in R^n$ を想起するという意味で想起行列と呼ばれてよい行列

$$A: R^n \rightarrow R^n \quad (3.19)$$

を式 (3.18) のごとく決定する前節のbatch learningを可分な一般抽象ヒルベルト空間 \mathfrak{H} でのon-line learningに直そう。

各 $w_\ell (\ell \in L)$ は実定数として、式 (2.21) の作用素 P_ℓ を導入し、式 (1.4) のパーセプトロン型作用素 P を考える。付録Aのaxiom 1を満たすパターン集合 $\Phi \subset \mathfrak{H}$ とモデル構成作用素 T との対 $[\Phi, T]$ をも導入する。 $T\varphi \in \Phi$ はパターン $\varphi \in \Phi$ の代りとなるパターンモデルである。

$$PT\varphi = \sum_{\ell \in L} w_\ell \cdot P_\ell(T\varphi) \quad (3.20)$$

へと変換する変換過程

$$\varphi \rightarrow P(T\varphi) \quad (3.21)$$

は、パターン $\varphi \in \Phi \subset \mathfrak{H}$ が与えられた時、モデル $T\varphi \in \Phi \subset \mathfrak{H}$ からパターン $PT\varphi$ を想起する想起過程と考えられる。

$\varphi_{i,k} \in \Phi \subset \mathfrak{H}$ は第 $i \in J$ 番目のカテゴリに帰属する第 $k (= 1 \sim n_i)$ 番目の訓練パターンであり、すべてのカテゴリ番号 $i \in J$ について、或るパターン番号 k_i が存在して、

$$\omega_i = \varphi_{i,k_i} \text{ が成り立つ} \quad (3.22)$$

としよう。但し、

$$\varphi_{i,1} = \omega_i, i \in J \quad (3.23)$$

としよう。

パーセプトロン型作用素 P に $\varphi_{i,k}$ を入力して得られるパターン

$$PT\varphi_{i,k} = \sum_{\ell \in L} w_\ell \cdot P_\ell T\varphi_{i,k} \quad (3.24)$$

がパターン $\varphi_{i,k}$ の帰属するカテゴリの代表パターン ω_i のモデル $T\omega_i$ と一致することが望ましい。この不一致差

$$PT\varphi_{i,k} - T\omega_i = \sum_{\ell \in L} w_\ell \cdot P_\ell T\varphi_{i,k} - T\omega_i \quad (3.25)$$

のノルム自乗誤差

$$\varepsilon_{i,k} \equiv \left\| \sum_{\ell \in L} w_\ell \cdot P_\ell T\varphi_{i,k} - T\omega_i \right\|^2 \quad (3.26)$$

の算術平均値

$$\varepsilon_i \equiv \frac{1}{n_i} \sum_{k=1}^{n_i} \varepsilon_{i,k} \quad (3.27)$$

の、すべてのカテゴリ $\mathcal{C}(J)$ にわたる総和の半分

$$\begin{aligned} F &\equiv \frac{1}{2} \cdot \sum_{i \in J} \varepsilon_i = \frac{1}{2} \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} \varepsilon_{i,k} \\ &= \frac{1}{2} \cdot \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} \left\| \sum_{\ell \in L} w_\ell \cdot P_\ell T \varphi_{i,k} - T \omega_i \right\|^2 \end{aligned} \quad (3.28)$$

が最小となる各重み $w_\ell (\ell \in L)$ を on-line 学習で決定しよう。

各重み $w_\ell (\ell \in L)$ の、学習時刻 $t(0, 1, 2, \dots)$ での重みを $w_\ell(t) (\ell \in L)$ と表す。

学習時刻 t での適応誤差関数

$$F(t) \equiv \frac{1}{2} \cdot \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} \left\| \sum_{\ell \in L} w_\ell(t) \cdot P_\ell T \varphi_{i,k} - T \omega_i \right\|^2 \quad (3.29)$$

が減少するように $t \rightarrow \infty$ としながら各重み $w_\ell(t) (\ell \in L)$ を求め、

$$w_\ell = \lim_{t \rightarrow \infty} w_\ell(t), \ell \in L \quad (3.30)$$

と決定しよう。その学習アルゴリズムは次のように述べられる。

【連想ニューラルネットの学習アルゴリズム】

[initialization]

初期値

$$w_\ell(t)|_{t=0} = \text{十分小さな正の定数}, \ell \in L \quad (3.31)$$

と選ぶ。

[induction]

式 (3.29) の関数 $F(t)$ を減少させるように、最急降下法を適用し、各重み $w_\ell(t) (\ell \in L)$ の更新式

$$w_\ell(t+1) = w_\ell(t) + \Delta w_\ell(t), \ell \in L \quad (3.32)$$

での各修正分 $(\Delta w_\ell)(t)$ を

$$\Delta w_\ell(t) = -\varepsilon_\ell(t) \cdot \frac{\partial F(t)}{\partial w_\ell(t)}, \varepsilon_\ell(t) > 0 \quad (3.33)$$

と計算する。

[termination]

不動点方程式

$$F(t+1) = F(t) \quad (3.34)$$

が成立する時刻 t が学習終了時刻である。このとき、求める各 $w_\ell(t)$ は

$$w_\ell = w_\ell(t), \ell \in L \quad (3.35)$$

と与えられる。 □

式 (3.33) の $\Delta w_\ell(t)$ を計算しよう。それは、

$$\begin{aligned} & \Delta w_\ell(t) \\ &= -\varepsilon_\ell(t) \cdot \frac{1}{2} \cdot \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{\partial}{\partial w_\ell} (\sum_{m \in L} w_m(t) \cdot P_m T \varphi_{i,k} - T \omega_i, \sum_{q \in L} w_q(t) \cdot P_q T \varphi_{i,k} - T \omega_i) \\ &= -\varepsilon_\ell(t) \cdot \frac{1}{2} \cdot \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} 2 \cdot \left(\frac{\partial}{\partial w_\ell} \sum_{m \in L} w_m(t) \cdot P_m T \varphi_{i,k} - T \omega_i, \sum_{q \in L} w_q(t) \cdot P_q T \varphi_{i,k} - T \omega_i \right) \\ &= -\varepsilon_\ell(t) \cdot \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} (P_\ell T \varphi_{i,k}, \sum_{q \in L} w_q(t) \cdot P_q T \varphi_{i,k} - T \omega_i) \\ &= -\varepsilon_\ell(t) \cdot \left[\sum_{q \in L} w_q(t) \cdot \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} (P_\ell T \varphi_{i,k}, P_q T \varphi_{i,k}) - \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} (P_\ell T \varphi_{i,k}, T \omega_i) \right] \\ &= -\varepsilon_\ell(t) \cdot \left[\sum_{q \in L} g_{\ell q} \cdot w_q(t) - b_\ell \right] \end{aligned} \quad (3.36)$$

と計算される。ここに、 $g_{\ell q}, b_\ell$ は

$$g_{\ell q} \equiv \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} (P_\ell T \varphi_{i,k}, P_q T \varphi_{i,k}) \quad (3.39)$$

$$b_\ell \equiv \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} (P_\ell T \varphi_{i,k}, T \omega_i) \quad (3.40)$$

と定義されている。

式 (3.36) から判明することは、連立 1 次方程式

$$\sum_{q \in L} g_{\ell q} \cdot w_q(t) = b_\ell, \ell \in L \quad (3.41)$$

の解 $w_\ell(t), \ell \in L$ が求まった段階で、

$$\Delta w_\ell(t) = 0, \ell \in L \quad (3.42)$$

を得、学習終了時刻 t が求まることがわかる。

$$F(t) = \frac{1}{2} \cdot \sum_{i \in J} \frac{1}{n_i} \sum_{k=1}^{n_i} (\sum_{m \in L} w_m(t) \cdot P_m T \varphi_{i,k} - T \omega_i, \sum_{q \in L} w_q(t) \cdot P_q T \varphi_{i,k} - T \omega_i) \quad (3.43)$$

であるから、結局、式 (3.29) の関数 $F(t)$ は 2 次形式の形に

$$F(t) = \frac{1}{2} \cdot \sum_{k \in L} \sum_{\ell \in L} g_{k\ell} \cdot w_k(t) \cdot w_\ell(t) \quad (3.44)$$

と再表現される。

第4章 1次ニューラルネットのon-line学習

本章では、 R^n, Φ で稼動する1次ニューラルネットのon-line学習を研究する。 R^n で稼動する1次ニューラルネットで、ニューロン発火関数（activation function）として符号関数に選び、その出力を2値に変換すれば、このニューラルネットはパーセプトロン（perceptron）と呼ばれるものになる。

4.1 n 次元ユークリッド空間 R^n での1次ニューラルネット

4.1.1 システム方程式

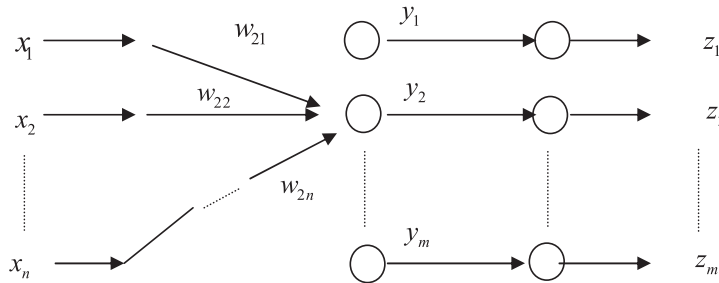


Fig.4-1 The first order neural network

以下では、文献 [13] で説明されている1次ニューラルネットのon-line学習を少し、詳細に論じる。
1次ニューラルネットの模式はFig.4-1で表現されている。

まず、感覚層（sensory layer）からの出力

$$\vec{x} = \text{col}[x_1 \ x_2 \ \cdots \ x_n] \in R^n \quad (4.1)$$

を連合層（associative layer）に入力すれば、システム方程式（稼動方程式）

$$y_j = \text{sgn}\left(\sum_{i=1}^n w_{ji} \cdot x_i + h_j\right), j=1 \sim m \quad (4.2)$$

に従って決まる出力

$$\vec{y} = \text{col}[y_1 \ y_2 \ \cdots \ y_m] \in R^m \quad (4.3)$$

が得られる。ここに、ニューロン発火関数として、符号関数（sign function）

$$\text{sgn}(u) = \begin{cases} -1 & \cdots u < 0 \text{ のとき} \\ +1 & \cdots u \geq 0 \text{ のとき} \end{cases} \quad (4.4)$$

が選ばれている。最後に、出力層（response layer）から、

$$\vec{z} = \text{col}[z_1 \quad z_2 \quad \cdots \quad z_m] \in R^m \quad (4.5)$$

$$\text{ここに, } z_j = \frac{1+y_j}{2} =$$

$$\begin{cases} 1 \cdots y_j = +1 \text{ のとき} \\ 0 \cdots y_j = -1 \text{ のとき} \end{cases} \quad (4.6)$$

が得られる。

4. 1. 2 学習の目的

第 $q(=1, 2, \dots, \ell)$ 番目の訓練パターン $\langle \vec{a}(q), \vec{b}(q) \rangle$ を用意する。ここに、

$$\vec{a}(q) = \text{col}(a_1(q) \quad a_2(q) \quad \cdots \quad a_n(q)) \in R^n \quad (4.7)$$

$$\vec{b}(q) = \text{col}(b_1(q) \quad b_2(q) \quad \cdots \quad b_m(q)) \in R^m \quad (4.8)$$

である。任意の q について、 $\vec{x} = \vec{a}(q)$ を入力すれば、 $\vec{y} = \vec{b}(q)$ が出力されるように、

$$\text{各重み } w_{ji} (i=1 \sim n, j=1 \sim m) \text{ と各閾値 } h_j (j=1 \sim m) \quad (4.9)$$

を学習する。

4. 1. 3 on-line学習

求めようとする各重み $w_{ji} (i=1 \sim n, j=1 \sim m)$ 、各閾値 $h_j (j=1 \sim m)$ の学習時刻のときの各重み $w_{ji} (i=1 \sim n, j=1 \sim m)$ 、各閾値 $h_j (j=1 \sim m)$ を各々、

$$w_{ji}(t) (i=1 \sim n, j=1 \sim m), h_j(t) (j=1 \sim m) \quad (4.10)$$

と表す。学習が終了後、求めようとする各重み $w_{ji} (i=1 \sim n, j=1 \sim m)$ 、各閾値 $h_j (j=1 \sim m)$ は、

$$w_{ji} = \lim_{t \rightarrow \infty} w_{ji}(t), i=1 \sim n, j=1 \sim m \quad (4.11)$$

$$h_j = \lim_{t \rightarrow \infty} h_j(t), j=1 \sim m \quad (4.12)$$

と得られる。

各重み $w_{ji} (i=1 \sim n, j=1 \sim m)$ 、各閾値 $h_j (j=1 \sim m)$ を学習するアルゴリズムは、次のように述べられる。

【1次ニューラルネットの学習アルゴリズム】

①初期化 (initialization)

学習開始時の $t=0$ のとき、初期値

$$w_{ji}(t)|_{t=0}, h_j(t)|_{t=0} \quad (4.13)$$

を十分小さい正数に選ぶ。

②帰納 (recursion)

$t = 0, 1, 2, \dots$ として、修正分

$$\begin{aligned} \Delta w_{ji}(t) &= \varepsilon_{ji}(t) \cdot [b_j(q) - \{\sum_{k=1}^n w_{jk}(t) \cdot a_k(q) + h_j(t)\}] \cdot a_i(q) \\ \Delta h_j(t) &= \varepsilon_j(t) \cdot [b_j(q) - \{\sum_{k=1}^n w_{jk}(t) \cdot a_k(q) + h_j(t)\}] \end{aligned} \quad (4.14)$$

を決め、現在の学習時刻 t での $w_{ji}(t), h_j(t) (i=1 \sim n, j=1 \sim m)$ を次の時刻 $t+1$ での

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji}(t), i=1 \sim n, j=1 \sim m \quad (4.15)$$

$$h_j(t+1) = h_j(t) + \Delta h_j(t), j=1 \sim m \quad (4.16)$$

へと修正する。但し、学習率 $\varepsilon_{ji}(t), \varepsilon_j(t) (j=1 \sim m, i=1 \sim n)$ は正値関数であり、極限性質

$$\lim_{t \rightarrow \infty} \varepsilon_{ji}(t) = \lim_{t \rightarrow \infty} \varepsilon_j(t) = 0 \quad (4.17)$$

を満たすとしなければならない。

訓練パターン対の系列

$$\langle \vec{a}(q), \vec{b}(q) \rangle, q=1, 2, \dots, \ell \quad (4.18)$$

に如何ほど適応しているかの指標としての、時刻 t の適応誤差関数

$$F(t) \equiv \sum_{q=1}^{\ell} f_q(t) \quad (4.19)$$

$$\text{ここに, } f_q(t) \equiv \sum_{s=1}^m 2^{-1} [b_s(q) - \{\sum_{k=1}^n w_{sk}(t) \cdot a_k(q) + h_s(t)\}]^2 \quad (4.20)$$

を導入すれば、適応誤差指標 $F(t)$ は $t \rightarrow \infty$ のときに 0 収束することが望ましい。実は、上述の $\Delta w_{ji}(t), \Delta h_j(t)$ は、最急降下法 (method of steepest descent) を適用して、

$$\Delta w_{ji}(t) = -\varepsilon_{ji}(t) \cdot \frac{\partial f_q(t)}{\partial w_{ji}(t)} \quad (4.21)$$

$$\Delta h_j(t) = -\varepsilon_j(t) \cdot \frac{\partial f_q(t)}{\partial h_j(t)} \quad (4.22)$$

を計算したものである。

③終了 (termination)

不動点方程式

$$F(t+1) = F(t) \quad (4.23)$$

が成立する学習終了時刻 t を求め、

$$w_{ji} = w_{ji}(t), i = 1 \sim n, j = 1 \sim m \quad (4.24)$$

$$h_j = h_j(t), j = 1 \sim m \quad (4.25)$$

と決定する。 \square

尚、現在の学習時刻 t での $w_{ji}(t), h_j(t) (i = 1 \sim n, j = 1 \sim m)$ を次の時刻 $t+1$ での $w_{ji}(t+1), h_j(t+1) (i = 1 \sim n, j = 1 \sim m)$ へ修正することが $b_j(q)$ へ近づけるようになされていることは、

\bar{x} として、 $\bar{x} = \bar{a}(q)$ を入力したとき、任意の $j (= 1 \sim m)$ について、不等式

$$g_j(t+1) \Big|_{\bar{x}=\bar{a}(q)} \begin{cases} < g_j(t) \Big|_{\bar{x}=\bar{a}(q)} \cdots b_j(q) < g_j(t) \Big|_{\bar{x}=\bar{a}(q)} \text{ のとき} \\ = g_j(t) \Big|_{\bar{x}=\bar{a}(q)} \cdots b_j(q) = g_j(t) \Big|_{\bar{x}=\bar{a}(q)} \text{ のとき} \\ > g_j(t) \Big|_{\bar{x}=\bar{a}(q)} \cdots b_j(q) > g_j(t) \Big|_{\bar{x}=\bar{a}(q)} \text{ のとき} \end{cases} \quad (4.26)$$

が成立していることからわかる。ここに、

$$g_j(t) \equiv \sum_{k=1}^n w_{jk}(t) \cdot a_k(q) + h_j(t) \quad (4.27)$$

を導入すれば、この不等式の成立は、等式

$$\begin{aligned} g_j(t+1) \Big|_{\bar{x}=\bar{a}(q)} &= \sum_{k=1}^n w_{jk}(t+1) \cdot a_k(q) + h_j(t+1) \\ &= g_j(t) \Big|_{\bar{x}=\bar{a}(q)} + \sum_{k=1}^n \Delta w_{jk}(t) \cdot a_k(q) + \Delta h_j(t) \\ &= g_j(t) \Big|_{\bar{x}=\bar{a}(q)} + \sum_{k=1}^n \varepsilon_{jk}(t) \cdot [b_j(q) - \{\sum_{i=1}^n w_{ji}(t) \cdot a_i(q) + h_j(t)\}] \cdot a_k(q)^2 \\ &\quad + \varepsilon_j(t) \cdot [b_j(q) - \{\sum_{i=1}^n w_{ji}(t) \cdot a_i(q) + h_j(t)\}] \end{aligned} \quad (4.28)$$

から明らかである。

4. 2 一般抽象実ヒルベルト空間 \mathfrak{H} での1次ニューラルネット

$$w_{ji}(j=1\sim m, i=1\sim n), w_j(j=1\sim m) \quad (4.29)$$

を実定数として、また、 φ_0 を \mathfrak{H} の元として、更に

$$h_i:\mathfrak{H} \rightarrow \mathfrak{H}, i=1\sim n \quad (4.30)$$

$$f_j:R \rightarrow R, j=1\sim m, \text{ここに, } R \text{ は実数全体の集合} \quad (4.31)$$

$$g_j:\mathfrak{H} \rightarrow \mathfrak{H}, j=1\sim m \quad (4.32)$$

$$P_i:\mathfrak{H} \rightarrow \mathfrak{H}, i=1\sim n \quad (4.33)$$

として、システム方程式（稼動方程式）として、

$$\varphi_i = h_i(\varphi'_i), i=1\sim n \quad (4.34)$$

$$\eta_j = f_j\left(\sum_{i=1}^m w_{ji} \cdot P_i \varphi_i + w_j \cdot \varphi_0\right), j=1\sim m \quad (4.35)$$

$$\phi_j = g_j(\eta_j), j=1\sim m \quad (4.36)$$

を提案する。

$$Q_j \varphi_i = \sum_{i=1}^m w_{ji} \cdot P_i \varphi_i \quad (4.37)$$

は、第 $i(=1\sim m)$ 番目のパターンが入力されたときの、第 $i(=1\sim m)$ 番目のパーセプトロン型作用素

$$Q_j = \sum_{i=1}^m w_{ji} \cdot P_i \quad (4.38)$$

からの出力であることに注意しておく。

上述のシステム方程式は座標値 $x \in M$ を明示すれば、次のようになる：

$$\varphi_i(x) = h_i(\varphi'_i)(x), i=1\sim n \quad (4.39)$$

$$\eta_j(x) = f_j\left(\sum_{i=1}^m w_{ji} \cdot (P_i \varphi_i)(x) + w_j \cdot \varphi_0(x)\right), j=1\sim m \quad (4.40)$$

$$\phi_j(x) = g_j(\eta_j)(x), j=1\sim m \quad (4.41)$$

□

学習時刻 $t(=0, 1, 2, \dots)$ での $w_{ji}(j=1\sim m, i=1\sim n), w_j(j=1\sim m)$ を

$$w_{ji}(t)(j=1 \sim m, i=1 \sim n), w_j(t)(j=1 \sim m) \quad (4.42)$$

と表すと、学習時刻 $t(=0, 1, 2, \dots)$ で得られる $\eta_j(j=1 \sim m)$ は

$$\eta_{jt} \equiv f_j\left(\sum_{i=1}^m w_{ji}(t) \cdot P_i \varphi_i + w_j(t) \cdot \varphi_0\right), j=1 \sim m \quad (4.43)$$

と表される。

学習時刻 $t(=0, 1, 2, \dots)$ に $\varphi_i(i=1 \sim n), \varphi_0$ を入力したとき、得られなければならない理想出力が $\eta'_j(j=1 \sim m)$ であるとして、最急降下法を適用して、適応誤差

$$E(t) \equiv \sum_{j=1}^m \frac{1}{2} \cdot \|\eta_{jt} - \eta'_j\|^2 = \sum_{j=1}^m \frac{1}{2} \cdot (\eta_{jt} - \eta'_j, \eta_{jt} - \eta'_j) \quad (4.44)$$

を最小とするように、更新式

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji}(t) \quad (4.45)$$

$$w_j(t+1) = w_j(t) + \Delta w_j(t) \quad (4.46)$$

の修正分 $\Delta w_{ji}(t), \Delta w_j(t)$ を決定すれば、次のようになる。

$$\Delta w_{ji}(t) = -\varepsilon_{ji}(t) \cdot \left(\frac{df_j(u)}{du} \Big|_{u=\sum_{k=1}^m w_{jk}(t) \cdot P_k \varphi_k + w_j(t) \cdot \varphi_0} \cdot P_i \varphi_i, \eta_{jt} - \eta'_j \right) \quad (4.47)$$

$$\Delta w_j(t) = -\varepsilon_j(t) \cdot \left(\frac{df_j(u)}{du} \Big|_{u=\sum_{k=1}^m w_{jk}(t) \cdot P_k \varphi_k + w_j(t) \cdot \varphi_0} \cdot \varphi_0, \eta_{jt} - \eta'_j \right) \quad (4.48)$$

となる。その理由は次のようになる。最急降下法によれば、

$$\Delta w_{ji}(t) = -\varepsilon_{ji}(t) \cdot \frac{\partial E(t)}{\partial w_{ji}(t)} \quad (4.49)$$

$$\Delta w_j(t) = -\varepsilon_j(t) \cdot \frac{\partial E(t)}{\partial w_j(t)} \quad (4.50)$$

であり、具体的に計算すれば、

$$\begin{aligned} \Delta w_{ji}(t) &= -\varepsilon_{ji}(t) \cdot \frac{\partial}{\partial w_{ji}(t)} \sum_{s=1}^m \frac{1}{2} \cdot (\eta_{st} - \eta'_s, \eta_{st} - \eta'_s) \\ &= -\varepsilon_{ji}(t) \cdot \left(\frac{\partial}{\partial w_{ji}(t)} \eta_{jt}, \eta_{jt} - \eta'_j \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間} \end{aligned} \quad (4.51)$$

であり、

$$\frac{\partial}{\partial w_{ji}(t)} \eta_{jt} = \frac{df_j(u)}{du} \Big|_{u=\sum_{k=1}^m w_{jk}(t) \cdot P_k \varphi_k + w_j(t) \cdot \varphi_0} \cdot P_i \varphi_i \quad (4.52)$$

である。また、

$$\begin{aligned} \Delta w_j(t) &= -\varepsilon_j(t) \cdot \frac{\partial}{\partial w_j(t)} \sum_{s=1}^m \frac{1}{2} \cdot (\eta_{st} - \eta'_s, \eta_{st} - \eta'_s) \\ &= -\varepsilon_j(t) \cdot \left(\frac{\partial}{\partial w_j(t)} \eta_{jt}, \eta_{jt} - \eta'_j \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間} \end{aligned} \quad (4.53)$$

であり、

$$\frac{\partial}{\partial w_j(t)} \eta_{jt} = \frac{df_j(u)}{du} \Big|_{u=\sum_{k=1}^m w_{jk}(t) \cdot P_k \varphi_k + w_j(t) \cdot \varphi_0} \cdot \varphi_0 \quad (4.54)$$

である。

4. 3 抽出された実数値特徴量を用いた、一般抽象実ヒルベルト空間 \mathfrak{H} での1次ニューラルネット
パターンから抽出される実数値特徴量の組を用いて、前節の内容を書き直したのが本節の内容である。

4. 3. 1 システム方程式

付録Aのaxiom 1を満たすモデル構成作用素 (model construction operator) T の構造形式

$$T \bullet = \sum_{\ell \in L} u(\bullet, \ell) \cdot \phi_\ell \quad (4.55)$$

を採用することで可能となった特徴量の組を用いたニューラルネットの稼動方程式(システム方程式)、on-line学習法が研究される。式 (4.55) に登場している系 $\{\phi_\ell\}_{\ell \in L}$ は \mathfrak{H} での1次独立な系であり、 $u(\varphi, \ell)$ はパターン φ から抽出された第 $\ell \in L$ 番目の実数値特徴量である。このニューラルネットは格別に、本論文の独創である。このニューラルネットは容易に実現可能という意味で、格別に、実用価値が高いと予想しており、特徴量の組を入力する従来の有限次元のユークリッド空間で動作するニューラルネットと馴染みが深い。

$\{\nu_\ell\}_{\ell \in L}, \{\sigma_p\}_{p \in P}$ は \mathfrak{H} での1次独立な系とする。

パターン φ から抽出される2種類の第 $\ell \in L$ 番目、第 $q \in Q$ 番目の実数値特徴量を $u_1(\varphi, \ell) \in R_{real}$ (実数全体の集合)、 $u_2(\varphi, q) \in R_{real}$ と表そう。

$$u_1(\varphi, \ell) \in R_{real}, \ell \in L \quad (4.56)$$

$$u_2(\varphi, q) \in R_{real}, q \in Q \quad (4.57)$$

はパターンから抽出される2種類の特徴量の組である。

$$Q \subseteq P, J \subseteq I \quad (4.58)$$

とする。

$$w_{pq}(p \in P, q \in Q \subseteq P), w_p(p \in P) \quad (4.59)$$

を実定数として、更に

$$f_p : R_{real} \rightarrow R_{real}, p \in P, \text{ここに, } R_{real} \text{ は実数全体の集合} \quad (4.60)$$

$$g_k : \mathfrak{H} \rightarrow \mathfrak{H}, k \in K \quad (4.61)$$

として、システム方程式（稼動方程式）は次のように記述される：

$$\varphi_i = \sum_{\ell \in L} u_1(\varphi'_i, \ell) \cdot v_\ell, i \in I \quad (4.62)$$

$$\eta_j = \sum_{p \in P} f_p \left(\sum_{q \in Q} w_{pq} \cdot u_2(\varphi_j, q) - w_p \right) \cdot \sigma_p, j \in J \subseteq I \quad (4.63)$$

$$\phi_k = g_k(\eta_k), k \in K \subseteq J \quad (4.64)$$

□

上述のシステム方程式は座標値 $x \in M$ を明示すれば、次のようになる：

$$\varphi_i(x) = \sum_{\ell \in L} u_1(\varphi'_i, \ell) \cdot v_\ell(x), i \in I \quad (4.65)$$

$$\eta_j(x) = \sum_{p \in P} f_p \left(\sum_{q \in Q} w_{pq} \cdot u_2(\varphi_j, q) - w_p \right) \cdot \sigma_p(x), j \in J \subseteq I \quad (4.66)$$

$$\phi_k(x) = g_k(\eta_k)(x), k \in K \subseteq J \quad (4.67)$$

□

4. 3. 2 on-line学習

学習時刻 $t(=0, 1, 2, \dots)$ に

$$\varphi_i = \sum_{\ell \in L} u_1(\varphi'_i, \ell) \cdot v_\ell \in \mathfrak{H}, i \in I \quad (4.68)$$

が入力されたとき、現実出力

$$\eta_{jt} = \sum_{p \in P} f_p \left(\sum_{q \in Q} w_{pq}(t) \cdot u_2(\varphi_j, q) - w_p(t) \right) \cdot \sigma_p, j \in J \quad (4.69)$$

が得られたとしよう。ここに、学習時刻 $t(=0, 1, 2, \dots)$ での $w_{pq}(p \in P, q \in Q \subseteq P), w_p(p \in P)$ を、

$$w_{pq}(t)(p \in P, q \in Q \subseteq P), w_p(t)(p \in P) \quad (4.70)$$

と表していることに注意する。

学習時刻 $t(=0,1,2,\dots)$ に各 η_{jt} に得られなければならない理想出力が η'_j になることを期待した適応誤差

$$E(t) \equiv \sum_{j=1}^m \frac{1}{2} \cdot \|\eta_{jt} - \eta'_j\|^2 = \sum_{s=1}^m \frac{1}{2} \cdot (\eta_{st} - \eta'_s, \eta_{st} - \eta'_s) \quad (4.71)$$

を最小とするように、最急降下法に従い、更新式

$$w_{pq}(t+1) = w_{pq}(t) + \Delta w_{pq}(t) \quad (4.72)$$

$$w_p(t+1) = w_p(t) + \Delta w_p(t) \quad (4.73)$$

の修正分

$$\Delta w_{pq}(t) = -\varepsilon_{pq}(t) \cdot \frac{\partial E(t)}{\partial w_{pq}(t)} \quad (4.74)$$

$$\Delta w_p(t) = -\varepsilon_p(t) \cdot \frac{\partial E(t)}{\partial w_p(t)} \quad (4.75)$$

を決定すれば、次のようになる。 $\varepsilon_{pq}(t) > 0$, $\varepsilon_p(t) > 0$ は学習率である。

$$\Delta w_{pq}(t) = -\varepsilon_{pq}(t) \cdot \sum_{j \in J} u_2(\varphi_j, q) \cdot \left(\frac{df_p(u)}{du} \right) \Big|_{u=\sum_{r \in Q} w_{pr}(t) \cdot u_2(\varphi_j, r) - w_p(t)} \cdot \sigma_p, \eta_{jt} - \eta'_j \quad (4.76)$$

$$\Delta w_p(t) = -\varepsilon_p(t) \cdot \sum_{j \in J} (-1) \cdot \left(\frac{df_p(u)}{du} \right) \Big|_{u=\sum_{r \in Q} w_{pr}(t) \cdot u_2(\varphi_j, r) - w_p(t)} \cdot \sigma_p, \eta_{jt} - \eta'_j \quad (4.77)$$

となる。その理由は次のようになる。具体的に計算すれば、

$$\begin{aligned} \Delta w_{pq}(t) &= -\varepsilon_{pq}(t) \cdot \sum_{j \in J} \frac{1}{2} \cdot \frac{\partial}{\partial w_{pq}(t)} (\eta_{jt} - \eta'_j, \eta_{jt} - \eta'_j) \\ &= -\varepsilon_{pq}(t) \cdot \sum_{j \in J} \frac{1}{2} \cdot 2 \left(\frac{\partial}{\partial w_{pq}(t)} \eta_{jt}, \eta_{jt} - \eta'_j \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間} \\ &= -\varepsilon_{pq}(t) \cdot \sum_{j \in J} \left(\frac{\partial}{\partial w_{pq}(t)} \eta_{jt}, \eta_{jt} - \eta'_j \right) \end{aligned} \quad (4.78)$$

であり、

$$\frac{\partial}{\partial w_{pq}(t)} \eta_{jt} = \sigma_p \cdot \left(\frac{df_p(u)}{du} \right) \Big|_{u=\sum_{r \in Q} w_{pr}(t) \cdot u_2(\varphi_j, r) - w_p(t)} \cdot u_2(\varphi_j, q) \quad (4.79)$$

である。また、

$$\begin{aligned}
 \Delta w_p(t) &= -\varepsilon_p(t) \cdot \sum_{j \in J} \frac{1}{2} \cdot \frac{\partial}{\partial w_p(t)} (\eta_{jt} - \eta'_j, \eta_{jt} - \eta'_j) \\
 &= -\varepsilon_p(t) \cdot \sum_{j \in J} \frac{1}{2} \cdot 2 \left(\frac{\partial}{\partial w_p(t)} \eta_{jt}, \eta_{jt} - \eta'_j \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間} \\
 &= -\varepsilon_p(t) \cdot \sum_{j \in J} \left(\frac{\partial}{\partial w_p(t)} \eta_{jt}, \eta_{jt} - \eta'_j \right) \tag{4.80}
 \end{aligned}$$

であり、

$$\frac{\partial}{\partial w_p(t)} \eta_{jt} = \sigma_p \cdot \frac{df_p(u)}{du} \Big|_{u=\sum_{r \in Q} w_{pr}(t) \cdot u_2(\varphi_j, r) - w_p(t)} \cdot (-1) \tag{4.81}$$

である。

第5章 再帰ニューラルネットワークのon-line学習

再帰ニューラルネットとは、過去に出力された自らの活動を現在の活動に利用できる機能を備えたニューラルネットである。本章では、

- (1) 外部からの入力を受け入れる入力ニューロン群
- (2) 理想出力が得られることを想定した制約下での出力ニューロン群
- (3) 制約が課せられていない出力ニューロン群

の3群からなるように分割されたFig.5-1の再帰ニューラルネットが、お互いに情報をやり取りしながら、 n 次元ユークリッド空間 R^n 、或いは、可分な一般抽象実ヒルベルト空間 \mathfrak{H} で活動を行う場合、訓練事例の系列を使ったon-line学習法が研究される。重みの修正分 $\Delta w_{k\ell}$ の計算は、再帰性のため、注意を必要とし、計算結果は複雑である。

5.1 有限次元ユークリッド空間 R^n での学習

文献 [14] の内容を少し、詳しく説明すれば次のようになる。

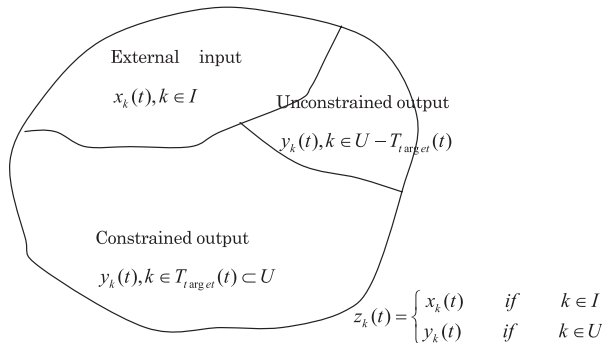


Fig.5-1 Fully recurrent neural networks

第 k 番目のニューロンの出力 $z_k(t)$ は、

$$z_k(t) = \begin{cases} x_k(t) & \text{if } k \in I \\ y_k(t) & \text{if } k \in U \end{cases} \quad (5.1)$$

であるとしよう。 $x_k(t)$ は時刻 t にニューラルネットに与えられる外部からの入力である。 $y_k(t)$ は時刻 t でニューラルネットから得られる出力である。

再帰ニューラルネットのシステム方程式（稼動方程式）として、

$$s_k(t) = \sum_{\ell \in U \cup I} w_{k\ell} \cdot z_\ell(t), k \in U \cup I \quad (5.2)$$

$$y_k(t+1) = f_k(s_k(t)), k \in U \quad (5.3)$$

where

$$w_{kk} = 0 \quad \text{if } k \in I \quad (5.4)$$

$$w_{k\ell} = 1 \quad \text{if } k \in U \wedge \ell \in I \quad (5.5)$$

$$f_k: R \rightarrow R \text{ is the squashing function of the } k \text{ th neural unit} \quad (5.6)$$

が採用されるとしよう。

時刻 $t(=0,1,2,\dots)$ に

$$z_k(t) = x_k(t), k \in I \quad (5.7)$$

が入力されたとき、時刻 $t+1$ に理想出力（target output） $d_k(t+1)$ が

$$y_k(t+1), k \in T_{\text{target}}(t+1) \subset U \quad (5.8)$$

に得られなければならないとしよう。the k th time-varying error $e_k(t+1)$ は、

$$e_k(t+1) = \begin{cases} d_k(t+1) - y_k(t+1) & \text{if } k \in T_{\text{target}}(t+1) \subset U \\ 0 & \text{if } k \in U - T_{\text{target}}(t+1) \end{cases} \quad (5.9)$$

である。適応に伴う総自乗誤差 E は

$$E = \frac{1}{2} \cdot \sum_{q \in U} e_q(t+1)^2 = \frac{1}{2} \cdot \sum_{q \in T_{\text{target}}(t+1)} e_q(t+1)^2 \quad (5.10)$$

である。

このとき、時刻 t の各重み $w_{k\ell}$ から、時刻 $t+1$ の各重み $w'_{k\ell}$ へ

$$w'_{k\ell} = w_{k\ell} + \Delta w_{k\ell}, k \in U, \ell \in U \quad (5.11)$$

という形で変化させよう。最急降下法に従い、適応総誤差 E を最小とするように、修正分 $\Delta w_{k\ell}$ を、

$$\Delta w_{k\ell} = -\varepsilon_{k\ell}(t) \cdot \frac{\partial E}{\partial w_{k\ell}}, \text{ where } \varepsilon_{k\ell}(t) > 0 \wedge \lim_{t \rightarrow \infty} \varepsilon_{k\ell}(t) = 0 \quad (5.12)$$

と求めよう。 $\varepsilon_{k\ell}(t)$ は学習率である。

計算結果は次の通りである：

$$p_{k\ell}^q(t+1) \equiv \frac{\partial y_q(t+1)}{\partial w_{k\ell}}, q \in T_{\text{target}}(t+1), k \in U, \ell \in U \quad (5.13)$$

とおくと、

$$\Delta w_{k\ell}(t) = \varepsilon_{k\ell}(t) \cdot \sum_{q \in T_{\text{target}}(t+1)} e_q(t+1) \cdot p_{k\ell}^q(t+1), k \in U, \ell \in U \quad (5.14)$$

と求められる。但し、 $p_{k\ell}^q(t)$ は、

$$p_{k\ell}^q(t+1) = \frac{df_q(s)}{ds} \Big|_{s=s_q(t)} \cdot [\sum_{r \in U} w_{qr} \cdot p_{k\ell}^r(t) + \delta_{kq} \cdot x_\ell(t)] \quad (5.15)$$

where

$$\delta_{kq} = \begin{cases} 1 & \text{if } k = q \\ 0 & \text{if } k \neq q \end{cases} \quad (5.16)$$

と、計算される。 □

その理由は次の通りである。

まず、

$$\begin{aligned} -\frac{\partial E}{\partial w_{k\ell}} &= \sum_{q \in T_{\text{target}}(t+1)} e_q(t+1) \cdot \frac{\partial y_q(t+1)}{\partial w_{k\ell}} \\ &= \sum_{q \in T_{\text{target}}(t+1)} e_q(t+1) \cdot p_{k\ell}^q(t+1) \end{aligned} \quad (5.17)$$

であり、

$$p_{k\ell}^q(t+1) = \frac{\partial y_q(t+1)}{\partial w_{k\ell}} = \frac{df_q(s)}{ds} \Big|_{s=s_q(t)} \cdot \frac{\partial s_q(t)}{\partial w_{k\ell}} \quad (5.18)$$

である。ここで、

$$\frac{\partial s_q(t)}{\partial w_{k\ell}} = \sum_{r \in U} \frac{\partial w_{qr} \cdot z_r(t)}{\partial w_{k\ell}} + \sum_{r \in I} \frac{\partial w_{qr} \cdot z_r(t)}{\partial w_{k\ell}} \quad (5.19)$$

$$\begin{aligned}
 &= \sum_{r \in U} \frac{\partial w_{qr} \cdot y_r(t)}{\partial y_r(t)} \cdot \frac{\partial y_r(t)}{\partial w_{k\ell}} + \delta_{kq} \cdot x_\ell(t) \\
 &= \sum_{r \in U} w_{qr} \cdot \frac{\partial y_r(t)}{\partial w_{k\ell}} + \delta_{kq} \cdot x_\ell(t) \\
 &= \sum_{r \in U} w_{qr}(t) \cdot p_{k\ell}^r(t) + \delta_{kq} \cdot x_\ell(t)
 \end{aligned} \tag{5.20}$$

である。

5. 2 一般抽象実ヒルベルト空間 \mathfrak{H} でのon-line学習

パターン $\varphi_k \langle t \rangle$ は、

$$\begin{aligned}
 \varphi_\ell \langle t \rangle = & \\
 & \begin{cases} \varphi'_\ell \langle t \rangle \cdots \ell \in I \text{ のとき} \\ \phi_\ell \langle t \rangle \cdots \ell \in U \text{ のとき} \end{cases}
 \end{aligned} \tag{5.21}$$

であるとしよう。 $\varphi'_\ell \langle t \rangle$ は時刻 t にニューラルネットに与えられる外部からの入力である。 $\phi_\ell \langle t \rangle$ は時刻 t でニューラルネットから得られる出力である。

$$w_{k\ell}, w_k(k, \ell \in U \cup I): \text{実定数} \tag{5.22}$$

$$P_\ell(\ell \in U \cup I): \mathfrak{H} \rightarrow \mathfrak{H} \tag{5.23}$$

$$\varphi_\ell \langle t \rangle (\ell \in U \cup I), \varphi_0 \langle t \rangle \in \mathfrak{H} \tag{5.24}$$

として、システム方程式（稼動方程式）

$$\eta_k \langle t \rangle = \sum_{\ell \in U \cup I} w_{k\ell} \cdot P_\ell \varphi_\ell \langle t \rangle - w_k \cdot \varphi_0 \langle t \rangle, k \in U \cup I \tag{5.25}$$

$$\phi_k(t+1) = f_k(\eta_k \langle t \rangle), k \in U \tag{5.26}$$

where

$$w_{kk} = 0 \quad \text{if } k \in I \tag{5.27}$$

$$w_{k\ell} = 1 \quad \text{if } k \in U \wedge \ell \in I \tag{5.28}$$

$$f_k: R \rightarrow R \text{ is the squashing function of the } k \text{th neural unit} \tag{5.29}$$

を提案する。

時刻 $t(=0,1,2,\dots)$ に

$$\varphi_k(t) = \varphi'_k < t >, k \in I \quad (5.30)$$

が入力されたとき、時刻 $t+1$ に理想出力 (target output) $\phi'_k(t+1)$ が $\phi_k(t+1), k \in T_{target}(t+1) \subset U$ に得られなければならないとしよう。the k th time-varying error $e_k(t+1)$ は、

$$e_k(t+1) = \begin{cases} \phi'_k(t+1) - \phi_k(t+1) & \text{if } k \in T_{target}(t+1) \subset U \\ 0 & \text{if } k \in U - T(t+1) \end{cases} \quad (5.31)$$

である。適応に伴う総ノルム自乗誤差 E は

$$E = \frac{1}{2} \cdot \sum_{q \in U} \|e_q(t+1)\|^2 = \frac{1}{2} \cdot \sum_{q \in T_{target}(t+1)} (e_q(t+1), e_q(t+1)) \quad (5.32)$$

である。

このとき、時刻 t の各重み $w_{k\ell}, w_k$ から、時刻 $t+1$ の各重み $w'_{k\ell}, w'_k$ へ

$$w'_{k\ell} = w_{k\ell} + \Delta w_{k\ell}, k \in U, \ell \in U \quad (5.33)$$

$$w'_k = w_k + \Delta w_k, k \in U, \ell \in U \quad (5.34)$$

という形で変化させよう。最急降下法に従い、適応総誤差 E を最小とするように、修正分 $\Delta w_{k\ell}$ を、

$$\Delta w_{k\ell} = -\varepsilon_{k\ell}(t) \cdot \frac{\partial E}{\partial w_{k\ell}}, \text{ where } \varepsilon_{k\ell}(t) > 0 \wedge \lim_{t \rightarrow \infty} \varepsilon_{k\ell}(t) = 0 \quad (5.35)$$

$$\Delta w_k = -\varepsilon_k(t) \cdot \frac{\partial E}{\partial w_k}, \text{ where } \varepsilon_k(t) > 0 \wedge \lim_{t \rightarrow \infty} \varepsilon_k(t) = 0 \quad (5.36)$$

と求めよう。

計算結果は次の通りである：

$$p_{k\ell}^q(t+1) \equiv \frac{\partial \phi_q < t+1 >}{\partial w_{k\ell}}, q \in T_{target}(t+1), k \in U, \ell \in U \quad (5.37)$$

とおくと、

$$\Delta w_{k\ell}(t) = \varepsilon_{k\ell}(t) \cdot \sum_{q \in T_{target}(t)} (p_{k\ell}^q(t+1), e_q(t+1)), k \in U, \ell \in U \quad (5.38)$$

と求められる。但し、 $p_{k\ell}^q(t+1)$ は、

$$p_{k\ell}^q(t+1) = \frac{df_q(s)}{ds} \Big|_{s=\eta_q < t >} \cdot \left[\sum_{r \in U} w_{qr} \cdot \frac{\partial P_r \phi_r < t >}{\partial w_{k\ell}} + \delta_{kq} \cdot P_\ell \phi'_\ell < t > \right] \quad (5.39)$$

where δ_{kq} denotes the Kronecker delta. □

その理由は次の通りである。

先ず、

$$\begin{aligned}
 -\frac{\partial E}{\partial w_{k\ell}} &= \sum_{q \in T_{\text{target}}(t)} \left(\frac{\partial}{\partial w_{k\ell}} \phi_q \langle t+1 \rangle, e_q(t+1) \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間} \\
 &= \sum_{q \in T_{\text{target}}(t)} (p_{k\ell}^q(t+1), e_q(t+1))
 \end{aligned} \tag{5.40}$$

であり、

$$p_{k\ell}^q(t+1) = \frac{\partial \phi_q \langle t+1 \rangle}{\partial w_{k\ell}} = \frac{df_q(s)}{ds} \Big|_{s=\eta_q \langle t \rangle} \cdot \frac{\partial \eta_q \langle t \rangle}{\partial w_{k\ell}} \tag{5.41}$$

である。ここで、

$$\begin{aligned}
 \frac{\partial \eta_q \langle t \rangle}{\partial w_{k\ell}} &= \sum_{r \in U} \frac{\partial w_{qr} \cdot P_r \phi_r \langle t \rangle}{\partial w_{k\ell}} + \sum_{r \in I} \frac{\partial w_{qr} \cdot P_r \phi_r \langle t \rangle}{\partial w_{k\ell}} \\
 &= \sum_{r \in U} \frac{\partial w_{qr} \cdot P_r \phi_r \langle t \rangle}{\partial P_r \phi_r \langle t \rangle} \cdot \frac{\partial P_r \phi_r \langle t \rangle}{\partial w_{k\ell}} + \delta_{kq} \cdot P_\ell \phi'_\ell \langle t \rangle \\
 &= \sum_{r \in U} w_{qr} \cdot \frac{\partial P_r \phi_r \langle t \rangle}{\partial w_{k\ell}} + \delta_{kq} \cdot P_\ell \phi'_\ell \langle t \rangle
 \end{aligned} \tag{5.42}$$

である。上記の学習が可能となるためには、少なくとも、

$$\frac{\partial P_r \phi_r \langle t \rangle}{\partial w_{k\ell}}, r \in U, k \in U, \ell \in U \cup I \tag{5.43}$$

を近似的に計算出来る必要がある。

$\Delta w_k(t)$ を求めよう。

$$p_k^q(t+1) \equiv \frac{\partial \phi_q \langle t+1 \rangle}{\partial w_k}, q \in T_{\text{target}}(t+1), k \in U, \ell \in U \tag{5.44}$$

とおくと、

$$\Delta w_k(t) = \varepsilon_k(t) \cdot \sum_{q \in T_{\text{target}}(t)} (p_k^q(t+1), e_q(t+1)), k \in U, \ell \in U \tag{5.45}$$

と求められる。但し、 $p_k^q(t+1)$ は、

$$p_k^q(t+1) = \frac{\partial \phi_q \langle t+1 \rangle}{\partial w_k} = \frac{df_q(s)}{ds} \Big|_{s=\eta_q \langle t \rangle} \cdot \delta_{kq} \cdot (-\phi_0 \langle t \rangle) \tag{5.46}$$

である。

その理由は次の通りである。

先ず、

$$\begin{aligned}
 -\frac{\partial E}{\partial w_{k\ell}} &= \sum_{q \in T_{\text{target}}(t)} \left(\frac{\partial}{\partial w_k} \phi_q \langle t+1 \rangle, e_q(t+1) \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間} \\
 &= \sum_{q \in T_{\text{target}}(t)} (p_k^q(t+1), e_q(t+1))
 \end{aligned} \tag{5.47}$$

であり、

$$p_k^q(t+1) = \frac{\partial \phi_q \langle t+1 \rangle}{\partial w_k} = \frac{df_q(s)}{ds} \Big|_{s=\eta_q \langle t \rangle} \cdot \frac{\partial \eta_q \langle t \rangle}{\partial w_k} \tag{5.48}$$

である。ここで、

$$\begin{aligned}
 \frac{\partial \eta_q \langle t \rangle}{\partial w_k} &= \frac{\partial -w_q \cdot \varphi_0 \langle t \rangle}{\partial w_k} \\
 &= \delta_{kq} \cdot (-\varphi_0 \langle t \rangle)
 \end{aligned} \tag{5.49}$$

である。

第6章 3層の誤差逆伝播学習ニューラルネットの稼働動作と、誤差逆伝播学習の動作

誤差逆伝播学習ニューラルネットは複数の層からなり、入力層、複数の中間層、出力層間にはこの方向にニューロン同士の結合はあるが、各々の層ではニューロン同士の結合がないような多層ニューラルネット (multi-layer neural network) である。このニューラルネットでは、学習動作では、情報の流れが通常の稼働動作での流れと逆に、出力層から複数の中間層へ、更に、複数の中間層から入力層へ向かっていることから、この名称が付けられている。

本章では、入力層、1 個の中間層 (隠れ層)、出力層からなる 3 層誤差逆伝播学習ニューラルネット (three-layer backpropagation learning network) のon-line学習法を研究するが、4 層以上の多層ニューラルネットの設計法については、適応誤差確率分布を想定した状況下で最尤法を適用して、文献 [B2] で研究されている。

6.1 n 次元ユークリッド空間でのon-line学習

文献 [15] の付録CにThree-layer backpropagation learning networkのon-line学習法が説明されている。このon-line学習法を本節では説明するが、このon-line学習法は、ニューロン発火関数として、非負シモイド関数(sigmoid function)を使っている場合 [16] と異なり、一般の発火関数 f_j, g_k を採用している。

6. 1. 1 3層ニューラルネットの稼働動作

本節では、3層ニューラルネットが稼働しているときの動作を方程式で記述する。

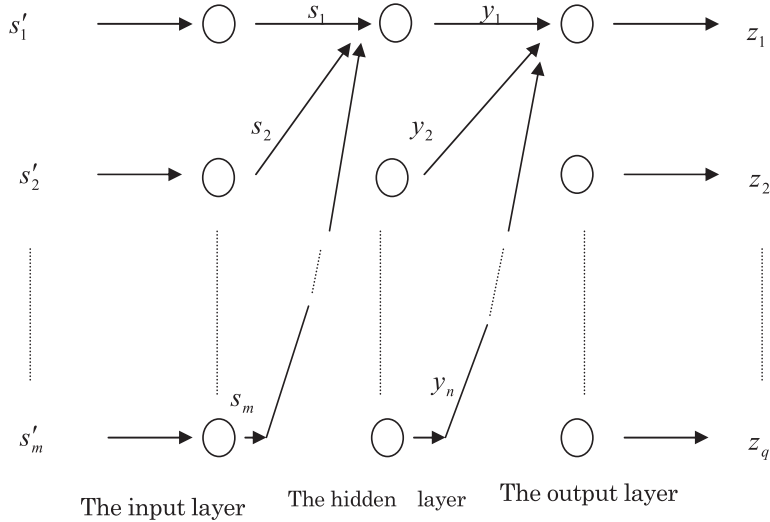


Fig.6.1 The three-layer backpropagation learning network, where the input layer consists of m neurons, the hidden layer n neurons, and the output Layer q neurons.

第1層が m 個のニューロンからなり、第2層が n 個のニューロンからなり、第3層が q 個のニューロンからなる3層ニューラルネットを考えよう。そのネットワーク稼働方程式は、

$$s_i = h_i(s'_i) : \text{第1層（入力層）内の第}i\text{番目のニューロンからの出力 } (i=1\sim m) \quad (6.1)$$

$$y_j = f_j\left(\sum_{i=1}^m w_{ji} \cdot s_i - a_j\right) : \text{第2層（隠れ層、或いは、中間層）内の第}j\text{番目ニューロンからの出力 } (j=1\sim n) \quad (6.2)$$

$$z_k = g_k\left(\sum_{j=1}^n v_{kj} \cdot y_j - b_k\right) : \text{第3層（出力層）内の第}k\text{番目のニューロンからの出力 } (k=1\sim q) \quad (6.3)$$

である。ここに、

$$h_i, f_j, g_k : \text{第1、2、3層のニューロン発火関数（activation function）} \quad (6.4)$$

$$s' = \text{col}(s'_1, s'_2, \dots, s'_m) \text{ (}m\text{次元縦ベクトル)} : \text{第1層への入力} \quad (6.5)$$

$$s = \text{col}(s_1, s_2, \dots, s_m) \text{ (}m\text{次元縦ベクトル)} : \text{第1層からの出力であり、同時に第2層への入力} \quad (6.6)$$

$y = col(y_1, y_2, \dots, y_n)$ (n次元縦ベクトル)：第2層からの出力であり、同時に第3層への入力 (6.7)

$z = col(z_1, z_2, \dots, z_q)$ (q次元縦ベクトル)：第3層からの出力 (6.8)

6. 1. 2 on-line学習の目標

前章の3層ニューラルネットをbackpropagation learning network (BPLN) とみなす。

時刻 $t (=1, 2, \dots)$ に、＜訓練入力, その訓練入力から得られなければならない理想出力＞という対としての訓練事例 (training example)

$$\langle \bar{s}, d\bar{o} \rangle \quad (6.9)$$

を与える。ここに、

$$\bar{s} = col(s_1, s_2, \dots, s_m) \quad (m\text{次元縦ベクトル}) \quad (6.10)$$

は訓練入力 (training input) であり、

$$d\bar{o} = col(do_1, do_2, \dots, do_m) \quad (q\text{次元縦ベクトル}) \quad (6.11)$$

はこの訓練入力を入力したとき、得られなければならない理想出力(desired output)である。

目標は、時刻 t での現実出力 (actual output) $\bar{z}(t)$ の各成分 $z_k(t)$ が出来るだけ理想出力 $d\bar{o}$ の各成分 do_k と一致するように、つまり、時刻 t に $s_i = h_i(s'_i), i=1 \sim m$ が入力されたとき、 t での適応誤差 $z_k(t) - do_k$ の自乗 $[z_k(t) - do_k]^2$ を第3層内のすべての出力ニューロンにわたり最小とするように、つまり、

$$E(t) \equiv E(t; \langle \bar{s}, d\bar{o} \rangle) \equiv \sum_{k=1}^q 2^{-1} \cdot [z_k(t) - do_k]^2 \rightarrow \min \quad (6.12)$$

が成立するように、各重み、各閾値

$$w_{ji}, a_j, v_{kj}, b_k \quad (i=1 \sim m, j=1 \sim n, k=1 \sim q) \quad (6.13)$$

を、最急降下法 (method of steepest descent) を適用した誤差逆伝播学習の逐次法 ($t=1, 2, \dots$) で求めることである。6. 1. 4 項に、その結果を示し、6. 1. 6 項でその誘導が示される。

For the moment, assume that the network is run starting at t_0 up to some final time t_1 . We take as the objective the minimization of the total error

$$E_{total}(t_0, t_1) \equiv \sum_{t=t_0}^{t_1} E(t) = \sum_{t=t_0}^{t_1} E(t; \langle \bar{s}, d\bar{o} \rangle) \quad (6.14)$$

Over this trajectory.

6. 1. 3 誤差逆伝播逐次学習における重み・閾値の更新式

学習を開始するにあたって、学習開始前の $t=0$ のときの初期値については、

$$w_{ji}(0) = \text{変数} \frac{1}{m+1} \text{の十分小さい関数値} \quad (6.15)$$

$$a_j(0) = \text{変数} \frac{m}{m+1} \left(= \frac{1}{\frac{1}{m} + 1} \right) \text{の十分小さい関数値} \quad (6.16)$$

$$v_{kj}(0) = \text{変数} \frac{1}{n+1} \text{の十分小さい関数値} \quad (6.17)$$

$$b_k(0) = \text{変数} \frac{n}{n+1} \left(= \frac{1}{\frac{1}{n} + 1} \right) \text{の十分小さい関数値} \quad (6.18)$$

と設定すればよい。

次に、4つの正值関数（学習率） $\varepsilon_{kj}(3;t), \varepsilon_k(4;t), \varepsilon_{ji}(1;t), \varepsilon_j(2;t)$ を選ぶ。極限性質

$$\lim_{t \rightarrow \infty} \varepsilon_{kj}(t) = 0 \quad (6.19)$$

が満たされていることが望ましい。残りの3正值関数 $\varepsilon_k(4;t), \varepsilon_{ji}(1;t), \varepsilon_j(2;t)$ についても同様である。

時刻 $t(=1, 2, 3, \dots)$ での訓練事例 $s', do >$ と、現実出力 z とを用い、逐次更新式

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji}(t) \quad (\text{第2層、つまり、隠れ層或いは、中間層の重みの更新式}) \quad (6.20)$$

$$a_j(t+1) = a_j(t) + \Delta a_j(t) \quad (\text{第2層、つまり、隠れ層或いは、中間層の閾値の更新式}) \quad (6.21)$$

$$v_{kj}(t+1) = v_{kj}(t) + \Delta v_{kj}(t) \quad (\text{第3層、つまり、出力層の重みの更新式}) \quad (6.22)$$

$$b_k(t+1) = b_k(t) + \Delta b_k(t) \quad (\text{第3層、つまり、出力層の閾値の更新式}) \quad (6.23)$$

の形式で、現在の時刻 t での各重み、各閾値

$$w_{ji}(t), a_j(t), v_{kj}(t), b_k(t) (i=1 \sim m, j=1 \sim n, k=1 \sim q) \quad (6.24)$$

を、1つだけ未来の各重み、各閾値

$$w_{ji}(t+1), a_j(t+1), v_{kj}(t+1), b_k(t+1) (i=1 \sim m, j=1 \sim n, k=1 \sim q) \quad (6.25)$$

へと、更新する。勿論、現在の時刻 t での現実出力 z をもたらす第1層からの現在の時刻 t での現実出力 s , 第2層からの現在の時刻 t での現実出力 y をも使うことになる。

6. 1. 4 3層ニューラルネットの稼働動作（誤差逆伝播によるon-line学習法）

本項では、3層ニューラルネットが学習しているときの動作を方程式で記述する。つまり、本節では、具体的に、訓練事例（学習に利用される入力とその理想出力との対）が入力される毎にその都度逐次的に w_{ji}, a_j, v_{kj}, b_k を更新するのに必要な諸公式が示される。

得られた4更新分

$$\Delta w_{ji}(t), \Delta a_j(t), \Delta v_{kj}(t), \Delta b_k(t) (i=1 \sim m, j=1 \sim n, k=1 \sim q) \quad (6.26)$$

の計算結果は、6. 1. 6項によれば、次の通りである。

$$e_k \equiv [do_k - z_k] \cdot \frac{dg_k(x)}{dx} \Big|_{x=\sum_{j=1}^n v_{kj}(t) \cdot y_j - b_k(t)} \quad (6.27)$$

$$e'_j \equiv \left[\sum_{k=1}^q v_{kj}(t) \cdot e_k \right] \cdot \frac{df_j(x)}{dx} \Big|_{x=\sum_{i=1}^m w_{ji}(t) \cdot s_i - a_j(t)} \quad (6.28)$$

として、

$$\Delta v_{kj} = \varepsilon_{kj}(3:t) \cdot e_k \cdot y_j \quad (6.29)$$

$$\Delta b_k = \varepsilon_k(4:t) \cdot e_k \cdot (-1) \quad (6.30)$$

$$\Delta w_{ji}(t) = \varepsilon_{ji}(1:t) \cdot e'_j \cdot s_i \quad (6.31)$$

$$\Delta a_j(t) = \varepsilon_j(2:t) \cdot e'_j \cdot (-1) \quad (6.32)$$

第3層の一般化された式(6.27)の各適応誤差 e_k が第2層へ、更に第1層へと伝播しながら、 w_{ji}, a_j, v_{kj}, b_k が更新されることが更新分 $\Delta w_{ji}, \Delta a_j, \Delta v_{kj}, \Delta b_k$ の上記の4表現式(6.29)～(6.32)からわかる。

6. 1. 5 学習終了時刻 t (収束判定)の方法

訓練事例の集合

$$\langle \vec{s}'(\ell), d\vec{o}(\ell) \rangle, \ell \in L \quad (6.33)$$

与えられた場合、各事例 $\langle \vec{s}'_\ell, d\vec{o}_\ell \rangle$ を各時刻 t にその生起頻度に比例して繰り返し与えなければならない。

求めようとする重み、閾値の、式(6.13)の組が式(6.24)であるように可能ならしめる学習終了時刻 t では、

$$\text{すべての } \ell \in L \text{ にわたり、 } \vec{s}(\ell) = \vec{h}(\vec{s}'(\ell)) \text{ が入力された} \quad (6.34)$$

とき、得られる出力 \vec{z} を

$$\vec{z}(t; \vec{s}(\ell)) \quad (6.35)$$

と表すと、等式

$$\vec{z}(t; \vec{s}(\ell)) = d\vec{o}(\ell) \quad (6.36)$$

が成立しなければならない。実際には、学習時刻 $t(=0, 1, 2, \dots)$ の関数

$$F(t) \equiv \sum_{\ell \in L} \sum_{k=1}^q |z_k(t; \vec{s}(\ell)) - do_k(\ell)| \quad (6.37)$$

を使って、次の①、②のいずれかで学習終了時刻 t を決めればよい:

①十分小さい正数 ε を予め決めておいて、

$$F(t) < \varepsilon \quad (6.38)$$

②不動点方程式

$$F(t+1) = F(t) \quad (6.39)$$

の成立. □

6. 1. 6 on-line誤差逆伝播学習式の誘導

本項では、4式 (6.29) ~ (6.32) を導く。最急降下法を適用して、下記の①,②,③,④のように計算される。

$$\textcircled{1} \Delta v_{kj}(t) = -\varepsilon_{kj}(3;t) \cdot \partial E(t) / \partial v_{kj}(t) \quad (6.40)$$

$$= -\varepsilon_{kj}(3;t) \cdot [z_k - do_k] \cdot \partial z_k / \partial v_{kj}(t)$$

$$= -\varepsilon_{kj}(3;t) \cdot [z_k - do_k] \cdot \frac{dg_k(x)}{dx} \Big|_{x=\sum_{j=1}^n v_{kj}(t) \cdot y_j - b_k(t)} \cdot y_j \quad (6.41)$$

$$= \varepsilon_{kj}(3;t) \cdot [do_k - z_k] \cdot \frac{dg_k(x)}{dx} \Big|_{x=\sum_{j=1}^n v_{kj}(t) \cdot y_j - b_k(t)} \cdot y_j$$

$$= \varepsilon_{kj}(3;t) \cdot e_k \cdot y_j$$

$$\textcircled{2} \Delta b_k(t) = -\varepsilon_k(4;t) \cdot \partial E(t) / \partial b_k(t) \quad (6.42)$$

$$= -\varepsilon_k(4;t) \cdot [z_k - do_k] \cdot \partial z_k / \partial b_k(t)$$

$$= -\varepsilon_k(4;t) \cdot [z_k - do_k] \cdot \frac{dg_k(x)}{dx} \Big|_{x=\sum_{j=1}^n v_{kj}(t) \cdot y_j - b_k(t)} \cdot (-1)$$

$$= \varepsilon_k(4;t) \cdot [do_k - z_k] \cdot \frac{dg_k(x)}{dx} \Big|_{x=\sum_{j=1}^n v_{kj}(t) \cdot y_j - b_k(t)} \cdot (-1) \quad (6.43)$$

$$= \varepsilon_k(4;t) \cdot e_k \cdot (-1)$$

$$\textcircled{3} \Delta w_{ji}(t) = -\varepsilon_{ji}(1;t) \cdot \partial E(t) / \partial w_{ji}(t) \quad (6.44)$$

$$\begin{aligned} &= -\varepsilon_{ji}(1;t) \cdot \left[\sum_{k=1}^q \frac{\partial E(t)}{\partial z_k} \cdot \frac{\partial z_k}{\partial y_j} \right] \cdot \frac{\partial y_j}{\partial w_{ji}(t)} \\ &= -\varepsilon_{ji}(1;t) \cdot \left[\sum_{k=1}^q [z_k - do_k] \cdot \frac{dg_k(x)}{dx} \Big|_{x=\sum_{j=1}^n v_{kj}(t) \cdot y_j - b_k(t)} \cdot v_{kj}(t) \right] \\ &\quad \cdot \frac{df_j(x)}{dx} \Big|_{x=\sum_{i=1}^m w_{ji}(t) \cdot s_i - a_j(t)} \cdot s_i \end{aligned} \quad (6.45)$$

$$\begin{aligned} &= \varepsilon_{ji}(1;t) \cdot \left[\sum_{k=1}^q v_{kj}(t) \cdot [do_k - z_k] \cdot \frac{dg_k(x)}{dx} \Big|_{x=\sum_{j=1}^n v_{kj}(t) \cdot y_j - b_k(t)} \right] \\ &\quad \cdot \frac{df_j(x)}{dx} \Big|_{x=\sum_{i=1}^m w_{ji}(t) \cdot s_i - a_j(t)} \cdot s_i \\ &= \varepsilon_{ji}(1;t) \cdot e'_j \cdot s_i \end{aligned}$$

$$\textcircled{4} \Delta a_j(t) = -\varepsilon_j(2;t) \cdot \partial E(t) / \partial a_j(t) \quad (6.46)$$

$$\begin{aligned} &= -\varepsilon_j(2;t) \cdot \left[\sum_{k=1}^q \frac{\partial E(t)}{\partial z_k} \cdot \frac{\partial z_k}{\partial y_j} \right] \cdot \frac{\partial y_j}{\partial a_j(t)} \\ &= -\varepsilon_j(2;t) \cdot \left[\sum_{k=1}^q [z_k - do_k] \cdot \frac{dg_k(x)}{dx} \Big|_{x=\sum_{j=1}^n v_{kj}(t) \cdot y_j - b_k(t)} \cdot v_{kj}(t) \right] \cdot \\ &\quad \frac{df_j(x)}{dx} \Big|_{x=\sum_{i=1}^m w_{ji}(t) \cdot s_i - a_j(t)} \cdot (-1) \end{aligned} \quad (6.48)$$

$$= \varepsilon_j(2;t) \cdot \left[\sum_{k=1}^q v_{kj}(t) \cdot [do_k - z_k] \cdot \frac{dg_k(x)}{dx} \Big|_{x=\sum_{j=1}^n v_{kj}(t) \cdot y_j - b_k(t)} \right] \cdot$$

$$\left. \frac{df_j(x)}{dx} \right|_{x=\sum_{i=1}^m w_{ji}(t) \cdot s_i - a_j(t)} \cdot (-1)$$

$$= \varepsilon_j(2:t) \cdot e'_j \cdot (-1)$$

6. 1. 7 ニューラルネットの発火関数 $f(x)$ の選定

各ニューロンの発火関数 h_i, f_j, g_k としては、各正実数パラメータをニューロン毎に変えて得られる次の4種類の関数から選べばよい。

①非負シグモイド関数 $f(x) = \frac{1}{1 + e^{-ax}}, a > 0$

$$\lim_{x \rightarrow -\infty} f(x) = 0, f(0) = \frac{1}{2}, \lim_{x \rightarrow +\infty} f(x) = 1$$

$$\frac{df(x)}{dx} = a \cdot f(x)[1 - f(x)]$$

②上述①の近似関数（非負区分的1次関数） $f(x) =$

$$\begin{cases} 0 & \text{if } x \leq -c_1 \\ \frac{d_1}{c_1} \cdot x + d_1 & \text{if } -c_1 < x \leq 0 \\ \frac{d_2 - d_1}{c_2} \cdot x + d_1 & \text{if } 0 < x < +c_2 \\ +d_2 & \text{if } +c_2 \leq x \end{cases}$$

$$c_1, c_2 > 0, d_2 > d_1 > 0$$

③正負シグモイド関数 $f(x) = 2 \cdot \frac{1}{1 + e^{-ax}} - 1, a > 0$

$$\lim_{x \rightarrow -\infty} f(x) = -1, f(0) = 0, \lim_{x \rightarrow +\infty} f(x) = +1$$

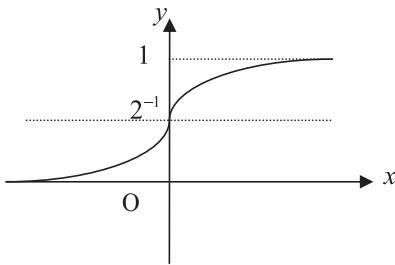
$$\frac{df(x)}{dx} = \frac{a}{2} \cdot [1 + f(x)][1 - f(x)]$$

④上述③の近似関数（正負区分的1次関数） $f(x)$

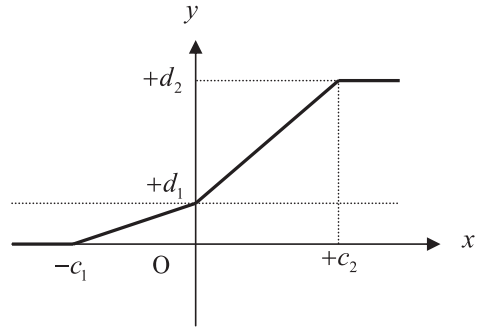
$$\begin{cases} -d_1 & \text{if } x \leq -c_1 \\ \frac{d_1}{c_1} \cdot x & \text{if } -c_1 < x \leq 0 \end{cases}$$

$$\begin{cases} \frac{d_2}{c_2} \cdot x & \text{if } 0 < x < +c_2 \\ +d_2 & \text{if } +c_2 \leq x \end{cases}$$

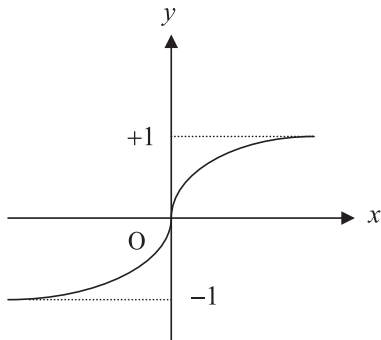
$$c_1, c_2, d_1, d_2 > 0$$



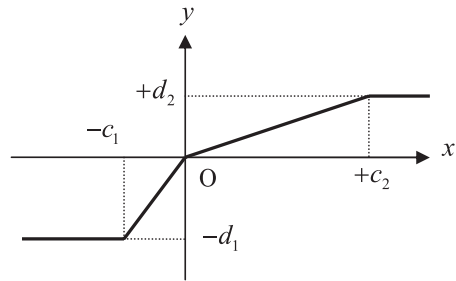
$$f(x) = \frac{1}{1 + e^{-ax}}, a > 0$$



非負区分的 1 次関数 $f(x)$



正負シグモイド関数 $f(x)$



正負区分的 1 次関数 $f(x)$

6. 2 一般抽象実ヒルベルト空間 \mathfrak{H} での誤差逆伝播学習

本節では、前節のユークリッド関数で稼動する 3 層誤差逆伝播ニューラルネットが一般抽象実ヒルベルト空間 \mathfrak{H} で稼動するニューラルネットに転換される。

6. 2. 1 システム方程式

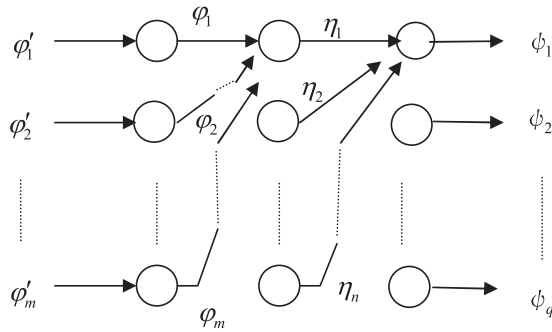


Fig.6-2 Three-layer backpropagation learning network

システム方程式として、

$$\phi_i = h_i(\phi'_i), i = 1 \sim m \quad (6.49)$$

$$\eta_j = f_j\left(\sum_{i=1}^m w_{ji} \cdot P_i \phi_i - w_j \cdot \phi_0\right), j = 1 \sim n \quad (6.50)$$

$$\phi_k = g_k\left(\sum_{j=1}^n v_{kj} \cdot Q_j \eta_j - v_k \cdot \eta_0\right), k = 1 \sim q \quad (6.51)$$

where

$$w_{ji}, w_j, v_{kj}, v_k (j = 1 \sim n, i = 1 \sim m, k = 1 \sim q) : \text{実定数} \quad (6.52)$$

$$h_i, f_j, g_k (j = 1 \sim n, i = 1 \sim m, k = 1 \sim q) : \text{実数値関数} \quad (6.53)$$

$$P_i : \mathfrak{S} \rightarrow \mathfrak{S} (i = 1 \sim m) \quad (6.54)$$

$$Q_j : \mathfrak{S} \rightarrow \mathfrak{S} (j = 1 \sim n) \quad (6.55)$$

on condition that there are

$$\frac{\partial Q_j \eta_j}{\partial \eta_j} (j = 1 \sim n) \quad (6.56)$$

を採用する (Fig.6-2)。

例えば、

$$Q_j \eta_j = q_j(\eta_j) \quad (6.57)$$

$$q_j(u) = 2 \cdot \frac{1}{1 + \exp[-a_j \cdot u]} - 1, a_j > 0 \quad (6.58)$$

ならば、

$$\frac{\partial Q_j \eta_j}{\partial \eta_j} = \frac{a_j}{2} \cdot [1 + q_j(\eta_j)] \cdot [1 - q_j(\eta_j)] \quad (6.59)$$

6. 2. 2 一般抽象実ヒルベルト空間 \mathfrak{H} での、on-lineの誤差逆伝播学習法

時刻 $t (= 0, 1, 2, \dots)$ に $\varphi_i = h_i(\varphi'_i) (i = 1 \sim m)$ が入力されたとき、現時出力 $\phi_k (k = 1 \sim q)$ が得られたとする。各現実出力 $\phi_k (k = 1 \sim q)$ が理想出力 $\phi'_k (k = 1 \sim q)$ になることを期待した適応誤差

$$E(t) \equiv \sum_{\ell=1}^q \frac{1}{2} \cdot \|\phi_\ell - \phi'_\ell\|^2 = \sum_{\ell=1}^q \frac{1}{2} \cdot (\phi_\ell - \phi'_\ell, \phi_\ell - \phi'_\ell) \quad (6.60)$$

を考える。

次に、4つの正值関数（学習率） $\varepsilon_{kj}(3;t), \varepsilon_k(4;t), \varepsilon_{ji}(1;t), \varepsilon_j(2;t)$ を選ぶ。極限性質

$$\lim_{t \rightarrow \infty} \varepsilon_{kj}(t) = 0 \quad (6.61)$$

が満たされていることが望ましい。残りの3正值関数 $\varepsilon_k(4;t), \varepsilon_{ji}(1;t), \varepsilon_j(2;t)$ についても同様である。4つの正值関数 $\varepsilon_{kj}(3;t), \varepsilon_k(4;t), \varepsilon_{ji}(1;t), \varepsilon_j(2;t)$ を例えば、共に 0.1 か、それより小と選べばよい。

時刻 $t (= 0, 1, 2, \dots)$ での訓練事例 $\langle \varphi_i, \phi_\ell \rangle$ と、現実出力 $\phi_k (k = 1 \sim q)$ とを用い、逐次更新式

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji}(t) \quad (\text{第2層、つまり、隠れ層或いは、中間層の重みの更新式}) \quad (6.62)$$

$$w_j(t+1) = w_j(t) + \Delta w_j(t) \quad (\text{第2層、つまり、隠れ層或いは、中間層の閾値の更新式}) \quad (6.63)$$

$$v_{kj}(t+1) = v_{kj}(t) + \Delta v_{kj}(t) \quad (\text{第3層、つまり、出力層の重みの更新式}) \quad (6.64)$$

$$v_k(t+1) = v_k(t) + \Delta v_k(t) \quad (\text{第3層、つまり、出力層の閾値の更新式}) \quad (6.65)$$

の形式で、現在の時刻 t での各重み、各閾値

$$w_{ji}(t), w_j(t), v_{kj}(t), v_k(t) (i = 1 \sim m, j = 1 \sim n, k = 1 \sim q) \quad (6.66)$$

を、1つだけ未来の各重み、各閾値

$$w_{ji}(t+1), w_j(t+1), v_{kj}(t+1), v_k(t+1) (i = 1 \sim m, j = 1 \sim n, k = 1 \sim q) \quad (6.67)$$

へと、更新する。勿論、現在の時刻 t での現実出力 $\phi_k (k = 1 \sim q)$ をもたらす第1層からの現在の時刻 t での現実出力 $\varphi_i (i = 1 \sim m)$ と第2層からの現在の時刻 t での現実出力 $\eta_j (j = 1 \sim n)$ をも使うことになる。

修正分

$$\Delta v_{kj}, \Delta v_k, \Delta w_{ji}, \Delta w_j \quad (j = 1 \sim n, i = 1 \sim m, k = 1 \sim q) \quad (6.68)$$

を計算した結果は次の通りである。

先ず、

$$e_k \equiv \frac{dg_k(u)}{du} \Big|_{u=\sum_{r=1}^n v_{kr}(t) \cdot Q_r \eta_r - v_k(t) \cdot \eta_0} \quad (6.69)$$

$$e_{kj} \equiv e_k \cdot \frac{\partial Q_j \eta_j}{\partial \eta_j} \cdot \frac{df_j(u)}{du} \Big|_{u=\sum_{r=1}^m w_{jr}(t) \cdot P_r \varphi_r - w_j(t) \cdot \varphi_0} \quad (6.70)$$

を導入しておく。

$$\begin{aligned} \textcircled{1} \Delta v_{kj}(t) &= -\varepsilon_{kj}(3:t) \cdot \frac{\partial E(t)}{\partial v_{kj}(t)} \\ &= \varepsilon_{kj}(3:t) \cdot (e_k \cdot Q_j \eta_j, \phi'_k - \phi_k) \end{aligned} \quad (6.71)$$

$$\begin{aligned} \textcircled{2} \Delta v_k(t) &= -\varepsilon_{kj}(4:t) \cdot \frac{\partial E(t)}{\partial v_k(t)} \\ &= \varepsilon_{kj}(4:t) \cdot (e_k \cdot (-\eta_0), \phi'_k - \phi_k) \end{aligned} \quad (6.72)$$

$$\begin{aligned} \textcircled{3} \Delta w_{ji}(t) &= -\varepsilon_{ji}(1:t) \cdot \frac{\partial E(t)}{\partial w_{ji}(t)} \\ &= \varepsilon_{ji}(1:t) \cdot \sum_{\ell=1}^q (e_{\ell j} \cdot P_i \varphi_i, \phi'_\ell - \phi_\ell) \end{aligned} \quad (6.73)$$

$$\begin{aligned} \textcircled{4} \Delta w_j(t) &= -\varepsilon_j(2:t) \cdot \frac{\partial E(t)}{\partial w_j(t)} \\ &= \varepsilon_j(2:t) \cdot \sum_{\ell=1}^q (e_{\ell j} \cdot (-\varphi_0), \phi'_\ell - \phi_\ell) \end{aligned} \quad (6.74)$$

6. 2. 3 $\Delta v_{kj}(t), \Delta v_k(t), \Delta w_{ji}(t), \Delta w_j(t)$ の計算

前項の 4 式 (6.71) ～ (6.74) を導こう。それは以下のように示される。

$$\begin{aligned} \textcircled{1} \Delta v_{kj}(t) &= \\ &= -\varepsilon_{kj}(3:t) \cdot \frac{\partial}{\partial v_{kj}(t)} 2^{-1}(\phi_k - \phi'_k, \phi_k - \phi'_k) \\ &= -\varepsilon_{kj}(3:t) \cdot \left(\frac{\partial}{\partial v_{kj}(t)} \phi_k, \phi_k - \phi'_k \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間} \end{aligned} \quad (6.75)$$

であるが、

$$\frac{\partial}{\partial v_{kj}(t)} \phi_k = e_k \cdot Q_j \eta_j \quad (6.76)$$

であるから、結局、

$$\Delta v_{kj}(t) = \varepsilon_{kj}(3:t) \cdot (e_k \cdot Q_j \eta_j, \phi'_k - \phi_k) \quad (6.77)$$

である。同様に、

$$\begin{aligned} \textcircled{2} \Delta v_k(t) &= \\ &= -\varepsilon_k(4:t) \cdot \frac{\partial}{\partial v_k(t)} 2^{-1}(\phi_k - \phi'_k, \phi_k - \phi'_k) \\ &= -\varepsilon_k(4:t) \cdot \left(\frac{\partial}{\partial v_k(t)} \phi_k, \phi_k - \phi'_k \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間} \end{aligned} \quad (6.78)$$

であるが、

$$\frac{\partial}{\partial v_k(t)} \phi_k = -\eta_0 \quad (6.79)$$

であるから、結局、

$$\Delta v_k(t) = \varepsilon_k(4:t) \cdot (e_k \cdot (-\eta_0), \phi'_k - \phi_k) \quad (6.80)$$

が得られる。

$$\begin{aligned} \textcircled{3} \Delta w_{ji}(t) &= \\ &= -\varepsilon_{ji}(1:t) \cdot \sum_{\ell=1}^q \frac{\partial}{\partial w_{ji}(t)} 2^{-1}(\phi_\ell - \phi'_\ell, \phi_\ell - \phi'_\ell) \\ &= -\varepsilon_{ji}(1:t) \cdot \sum_{\ell=1}^q \left(\frac{\partial}{\partial w_{ji}(t)} \phi_\ell, \phi_\ell - \phi'_\ell \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間} \end{aligned} \quad (6.81)$$

であるが、

$$\begin{aligned} &\frac{\partial}{\partial w_{ji}(t)} \phi_\ell \\ &= \left[\sum_{k=1}^q \frac{\partial \phi_\ell}{\partial \phi_k} \cdot \frac{\partial \phi_k}{\partial \eta_j} \right] \cdot \frac{\partial \eta_j}{\partial w_{ji}(t)} \\ &= \left[\sum_{k=1}^q \delta_{\ell k} \cdot \frac{\partial \phi_k}{\partial \eta_j} \right] \cdot \frac{\partial \eta_j}{\partial w_{ji}(t)} \end{aligned}$$

$$\begin{aligned}
 &= \frac{\partial \phi_\ell}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial w_{ji}(t)} \\
 &= \frac{\partial \phi_\ell}{\partial \eta_j} \cdot \frac{df_j(u)}{du} \Big|_{u=\sum_{r=1}^m w_{jr}(t) \cdot P_r \varphi_r - w_j(t) \cdot \varphi_0} \cdot P_i \varphi_i \\
 &= \frac{dg_\ell(u)}{du} \Big|_{u=g_\ell(\sum_{r=1}^n v_{ir}(t) \cdot Q_r \eta_r - v_i(t) \cdot \eta_0)} \cdot \frac{\partial Q_j \eta_j}{\partial \eta_j} \cdot \left[\frac{df_j(u)}{du} \Big|_{u=\sum_{r=1}^m w_{jr}(t) \cdot P_r \varphi_r - w_j(t) \cdot \varphi_0} \cdot P_i \varphi_i \right] \\
 &= e_\ell \cdot \frac{\partial Q_j \eta_j}{\partial \eta_j} \cdot \frac{df_j(u)}{du} \Big|_{u=\sum_{r=1}^m w_{jr}(t) \cdot P_r \varphi_r - w_j(t) \cdot \varphi_0} \cdot P_i \varphi_i \\
 &= e_{\ell j} \cdot P_i \varphi_i
 \end{aligned} \tag{6.82}$$

であるから、

$$\begin{aligned}
 &\Delta w_{ji}(t) \\
 &= -\varepsilon_{ji}(1:t) \cdot \sum_{\ell=1}^q (e_{\ell j} \cdot P_i \varphi_i, \phi_\ell - \phi'_\ell) \\
 &= \varepsilon_{ji}(1:t) \cdot \sum_{\ell=1}^q (e_{\ell j} \cdot P_i \varphi_i, \phi'_\ell - \phi_\ell)
 \end{aligned} \tag{6.83}$$

である。同様に、

$$\begin{aligned}
 &\textcircled{4} \Delta w_j(t) = \\
 &= -\varepsilon_j(2:t) \cdot \sum_{\ell=1}^q \frac{\partial}{\partial w_j(t)} 2^{-1}(\phi_\ell - \phi'_\ell, \phi_\ell - \phi'_\ell) \\
 &= -\varepsilon_j(2:t) \cdot \sum_{\ell=1}^q \left(\frac{\partial}{\partial w_j(t)} \phi_\ell, \phi_\ell - \phi'_\ell \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間}
 \end{aligned} \tag{6.84}$$

であるが、

$$\begin{aligned}
 &\frac{\partial}{\partial w_j(t)} \phi_\ell \\
 &= \left[\sum_{k=1}^q \frac{\partial \phi_\ell}{\partial \phi_k} \cdot \frac{\partial \phi_k}{\partial \eta_j} \right] \cdot \frac{\partial \eta_j}{\partial w_j(t)} \\
 &= \left[\sum_{k=1}^q \delta_{\ell k} \cdot \frac{\partial \phi_k}{\partial \eta_j} \right] \cdot \frac{\partial \eta_j}{\partial w_j(t)}
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{\partial \phi_\ell}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial w_j(t)} \\
 &= \frac{\partial \phi_\ell}{\partial \eta_j} \cdot \frac{df_j(u)}{du} \Big|_{u=\sum_{r=1}^m w_{jr}(t) \cdot P_r \phi_r - w_j(t) \cdot \phi_0} \cdot (-\phi_0) \\
 &= \frac{dg_\ell(u)}{du} \Big|_{u=g_\ell(\sum_{r=1}^n v_{jr}(t) \cdot Q_r \eta_r - v_\ell(t) \cdot \eta_0)} \cdot \frac{\partial Q_j \eta_j}{\partial \eta_j} \cdot \left[\frac{df_j(u)}{du} \Big|_{u=\sum_{r=1}^m w_{jr}(t) \cdot P_r \phi_r - w_j(t) \cdot \phi_0} \cdot (-\phi_0) \right] \\
 &= e_\ell \cdot \frac{\partial Q_j \eta_j}{\partial \eta_j} \cdot \frac{df_j(u)}{du} \Big|_{u=\sum_{r=1}^m w_{jr}(t) \cdot P_r \phi_r - w_j(t) \cdot \phi_0} \cdot (-\phi_0) \\
 &= e_{\ell j} \cdot (-\phi_0)
 \end{aligned} \tag{6.85}$$

であるから、

$$\begin{aligned}
 &\Delta w_{ji}(t) \\
 &= -\varepsilon_j(2:t) \cdot \sum_{\ell=1}^q (e_{\ell j} \cdot (-\phi_0), \phi_\ell - \phi'_\ell) \\
 &= \varepsilon_j(2:t) \cdot \sum_{\ell=1}^q (e_{\ell j} \cdot (-\phi_0), \phi'_\ell - \phi_\ell)
 \end{aligned} \tag{6.86}$$

が得られる。

6. 3 一般抽象実ヒルベルト空間 \mathfrak{H} での、特徴量の組を用いた誤差逆伝播学習

パターンから抽出された特徴量の組を使った形式に、一般抽象実ヒルベルト空間で稼動する誤差逆伝播学習ニューラルネットを直そう。それには、4. 3 節と同様に、モデル構成作用素 T の構造形式である式 (4.55) を採用しなければならない。

6. 3. 1 システム方程式

$\{v_\ell\}_{\ell \in L}, \{\sigma_p\}_{p \in P}, \{\theta_r\}_{r \in R}$ は \mathfrak{H} での 1 次独立な系とする。

パターン φ から抽出される 3 種類の第 $\ell \in L$ 番目、第 $q \in Q$ 番目、第 $s \in S$ 番目の実数値特徴量を $u_1(\varphi, \ell) \in R_{real}$, $u_2(\varphi, q) \in R_{real}$, $u_3(\varphi, s) \in R_{real}$ と表そう。

$$u_1(\varphi, \ell), \ell \in L \tag{6.87}$$

$$u_2(\varphi, q), q \in Q \tag{6.88}$$

$$u_3(\varphi, s), s \in S \tag{6.89}$$

はパターンから抽出される 3 種類の特徴量の組である。ここに R_{real} は実数全体の集合である。

$$\text{第1層の特徴抽出写像 } u_1 : \Phi \times L \rightarrow \{u \mid -1 \leq u \leq +1\} \quad (6.90)$$

$$\text{第2層の特徴抽出写像 } u_2 : \Phi \times Q \rightarrow \{u \mid -1 \leq u \leq +1\} \quad (6.91)$$

$$\text{第3層の特徴抽出写像 } u_3 : \Phi \times S \rightarrow \{u \mid -1 \leq u \leq +1\} \quad (6.92)$$

が導入される。

$$w_{pq}(p \in P, q \in Q), w_p(p \in P), v_{rs}(r \in R), v_r(s \in S) \quad (6.93)$$

を実定数とし、更に

$$\text{第 } p \in P \text{ 番目のニューロンの発火関数 } f_p : R_{real} \rightarrow \{u \mid -1 \leq u \leq +1\}, p \in P \quad (6.94)$$

$$\text{第 } r \in R \text{ 番目のニューロンの発火関数 } g_r : R_{real} \rightarrow \{u \mid -1 \leq u \leq +1\}, r \in R \quad (6.95)$$

とし、特徴量を用いた、一般抽象実ヒルベルト空間 \mathfrak{H} での誤差逆伝播学習 3 層ニューラルネットのシステム方程式として、

$$\text{第1層 } \varphi_i = \sum_{\ell \in L} u_1(\varphi'_i, \ell) \cdot v_\ell, i \in I \quad (6.96)$$

$$\text{第2層 } \eta_j = \sum_{p \in P} f_p \left(\sum_{q \in Q} w_{pq} \cdot u_2(\varphi_j, q) - w_p \right) \cdot \sigma_p, j \in J \quad (6.97)$$

$$\text{第3層 } \phi_k = \sum_{r \in R} g_r \left(\sum_{s \in S} v_{rs} \cdot u_3(\eta_k, s) - v_r \right) \cdot \theta_r, k \in K \quad (6.98)$$

$$\text{ここに、 } I \supseteq J \supseteq K \quad (6.99)$$

を採用しよう。但し、2 条件

$$\text{条件 1 } \quad \forall p \in P, \exists \ell(q) \in L (q = 1, 2, \dots : q \geq 2), (\sigma_p, v_{\ell(q)}) \neq 0 \quad (6.100)$$

$$\text{条件 2 } \quad \forall r \in R, \exists p(q) \in P (q = 1, 2, \dots : q \geq 2), (\theta_r, \sigma_{p(q)}) \neq 0 \quad (6.101)$$

を設けておくことが望ましいことがある。

上述のシステム方程式は座標値 $x \in M$ を明示すれば、次のようになる。

$$\varphi_i(x) = \sum_{\ell \in L} u_1(\varphi'_i, \ell) \cdot v_\ell(x), i \in I \quad (6.102)$$

$$\eta_j(x) = \sum_{p \in P} f_p \left(\sum_{q \in Q} w_{pq} \cdot u_2(\varphi_j, q) - w_p \right) \cdot \sigma_p(x), j \in J \quad (6.103)$$

$$\phi_k(x) = \sum_{r \in R} g_r \left(\sum_{s \in S} v_{rs} \cdot u_3(\eta_k, s) - v_r \right) \cdot \theta_r(x), k \in K \quad (6.104)$$

となることに注意しておく。

6. 3. 2 on-line学習

学習時刻 $t (= 0, 1, 2, \dots)$ に

$$\varphi_i = \sum_{\ell \in L} u_1(\varphi'_i, \ell) \cdot v_\ell, i \in I \quad (6.105)$$

が入力されたとき、得られた各現実出力

$$\phi_k = \sum_{r \in R} g_r \left(\sum_{s \in S} v_{rs}(t) \cdot u_3(\eta_k, s) - v_r(t) \right) \cdot \theta_r, k \in K \quad (6.106)$$

が理想出力 $\phi'_k (k \in K)$ になることを期待した適応誤差

$$E(t) \equiv \sum_{k \in K} \frac{1}{2} \cdot \|\phi_k - \phi'_k\|^2 = \sum_{k \in K} \frac{1}{2} \cdot (\phi_k - \phi'_k, \phi_k - \phi'_k) \quad (6.107)$$

を最小とするように、

$$w_{pq}(t+1) = w_{pq}(t) + \Delta w_{pq}(t), p \in P, q \in Q \quad (6.108)$$

$$w_p(t+1) = w_p(t) + \Delta w_p(t), p \in P \quad (6.109)$$

$$v_{rs}(t+1) = v_{rs}(t) + \Delta v_{rs}(t), r \in R, s \in S \quad (6.110)$$

$$v_r(t+1) = v_r(t) + \Delta v_r(t), r \in R \quad (6.111)$$

の修正分

$$\Delta w_{pq}(t) = -\varepsilon_{pq}(1:t) \cdot \frac{\partial E(t)}{\partial w_{pq}(t)} \quad (6.112)$$

$$\Delta w_p(t) = -\varepsilon_p(2:t) \cdot \frac{\partial E(t)}{\partial w_p(t)} \quad (6.113)$$

$$\Delta v_{rs}(t) = -\varepsilon_{rs}(3:t) \cdot \frac{\partial E(t)}{\partial v_{rs}(t)} \quad (6.114)$$

$$\Delta v_r(t) = -\varepsilon_r(4:t) \cdot \frac{\partial E(t)}{\partial v_r(t)} \quad (6.115)$$

を決定すれば、次のようになる。ここに、

$$\eta_j = \sum_{p \in P} f_p \left(\sum_{q \in Q} w_{pq}(t) \cdot u_2(\varphi_j, q) - w_p(t) \right) \cdot \sigma_p \in \mathfrak{S}, j \in J \quad (6.116)$$

であり、学習時刻 $t (= 0, 1, 2, \dots)$ での式 (6.93) を

を、

$$w_{pq}(t)(p \in P, q \in Q), w_p(t)(p \in P), v_{rs}(t)(r \in R, s \in S), v_r(t)(r \in R) \quad (6.117)$$

と表していることに注意する。

現在の時刻 t での各現実出力 $\phi_k (k \in K)$ をもたらす第1層からの現在の時刻 t での各現実出力 $\varphi_i (i \in I)$ 、第2層からの現在の時刻 t での現実出力 $\eta_j (j \in J)$ をも使うことになる。

4 修正分 $\Delta w_{pq}, \Delta w_p, \Delta v_{rs}, \Delta v_r$ の計算結果は以下の通りである。

先ず、

$$e_{rk} \equiv \frac{dg_r(u)}{du} \Big|_{u=\sum_{s \in S} v_{rs}(t) \cdot u_3(\eta_k, s) - v_r(t)} \cdot \theta_r \quad (6.118)$$

$$e_{rpk} \equiv e_{rk} \cdot \frac{df_p(u)}{du} \Big|_{u=\sum_{q \in Q} w_{pq}(t) \cdot u_2(\varphi_k, q) - w_p(t)} \cdot \sigma_p \quad (6.119)$$

を導入しておく。例えば、

$$u_3(\eta_j, s) = 2 \cdot \frac{1}{1 + \exp[-a_s \cdot (\eta_j - \eta'_j)]} - 1 \quad (6.120)$$

と与えることができる。

$$\begin{aligned} \textcircled{1} \Delta v_{rs}(t) &= -\varepsilon_{rs}(3:t) \cdot \frac{\partial E(t)}{\partial v_{rs}(t)} \\ &= \varepsilon_{rs}(3:t) \cdot \sum_{k \in K} u_3(\eta_k, s) \cdot (e_{rk}, \phi'_k - \phi_k) \end{aligned} \quad (6.121)$$

$$\begin{aligned} \textcircled{2} \Delta v_r(t) &= -\varepsilon_r(4:t) \cdot \frac{\partial E(t)}{\partial v_r(t)} \\ &= \varepsilon_r(4:t) \cdot \sum_{k \in K} (-1) \cdot (e_{rk}, \phi'_k - \phi_k) \end{aligned} \quad (6.122)$$

$$\begin{aligned}
 \textcircled{3} \Delta w_{pq}(t) &= -\varepsilon_{pq}(1:t) \cdot \frac{\partial E(t)}{\partial w_{pq}(t)} \\
 &= \varepsilon_{pq}(1:t) \cdot \sum_{k \in K} \left(\frac{\partial \phi_k}{\partial w_{pq}(t)}, \phi'_k - \phi_k \right)
 \end{aligned} \tag{6.123}$$

ここに、

$$\frac{\partial \phi_k}{\partial w_{pq}(t)} = \frac{\partial u_3(\eta_j, s)}{\partial \eta_j} \cdot u_2(\varphi_j, q) \cdot \sum_{r \in R} e_{rpk} \tag{6.124}$$

$$\begin{aligned}
 \textcircled{4} \Delta w_p(t) &= -\varepsilon_p(2:t) \cdot \frac{\partial E(t)}{\partial w_p(t)} \\
 &= \varepsilon_p(2:t) \cdot \sum_{k \in K} \left(\frac{\partial \phi_k}{\partial w_p(t)}, \phi'_k - \phi_k \right)
 \end{aligned} \tag{6.125}$$

ここに、

$$\frac{\partial \phi_k}{\partial w_p(t)} = \frac{\partial u_3(\eta_j, s)}{\partial \eta_j} \cdot (-1) \cdot \sum_{r \in R} e_{rpk} \tag{6.126}$$

6. 3. 3 $\Delta v_{rs}(t), \Delta v_r(t), \Delta w_{pq}(t), \Delta w_p(t)$ の計算

4 式 (6.112) ～ (6.115) を具体的に計算しよう。それは以下のようになる。

$$\begin{aligned}
 \textcircled{1} \Delta v_{rs}(t) &= \\
 &= -\varepsilon_{rs}(3:t) \cdot \sum_{k \in K} 2^{-1} \cdot 2 \left(\frac{\partial}{\partial v_{rs}(t)} \{ \phi_k - \phi'_k \}, \phi_k - \phi'_k \right) \quad \because \mathfrak{H} \text{ は実ヒルベルト空間} \\
 &= -\varepsilon_{rs}(3:t) \cdot \sum_{k \in K} \left(\frac{\partial}{\partial v_{rs}(t)} \phi_k, \phi_k - \phi'_k \right)
 \end{aligned} \tag{6.127}$$

$$\begin{aligned}
 &= -\varepsilon_{rs}(3:t) \cdot \sum_{k \in K} \left(\frac{\partial}{\partial v_{rs}(t)} \{ g_r \left(\sum_{s \in S} v_{rs}(t) \cdot u_3(\eta_k, s) - v_r(t) \right) \cdot \theta_r \}, \phi_k - \phi'_k \right) \\
 &= -\varepsilon_{rs}(3:t) \cdot \sum_{k \in K} \left(\theta_r \cdot \frac{dg_r(u)}{du} \Big|_{u=\sum_{s \in S} v_{rs}(t) \cdot u_3(\eta_k, s) - v_r(t)} \cdot u_3(\eta_k, s), \phi_k - \phi'_k \right) \\
 &= -\varepsilon_{rs}(3:t) \cdot \sum_{k \in K} (e_{rk} \cdot u_3(\eta_k, s), \phi_k - \phi'_k) \\
 &= \varepsilon_{rs}(3:t) \cdot \sum_{k \in K} u_3(\eta_k, s) \cdot (e_{rk}, \phi'_k - \phi_k)
 \end{aligned} \tag{6.128}$$

である。同様に、

$$\begin{aligned}
 & \textcircled{2} \Delta v_r(t) = \\
 & = -\varepsilon_r(4:t) \cdot \sum_{k \in K} 2^{-1} \cdot 2 \left(\frac{\partial}{\partial v_r(t)} \{ \phi_k - \phi'_k \}, \phi_k - \phi'_k \right) \quad \because \quad \mathfrak{H} \text{ は実ヒルベルト空間} \\
 & = -\varepsilon_r(4:t) \cdot \sum_{k \in K} \left(\frac{\partial}{\partial v_r(t)} \phi_k, \phi_k - \phi'_k \right) \tag{6.129} \\
 & = -\varepsilon_r(4:t) \cdot \sum_{k \in K} \left(\frac{\partial}{\partial v_r(t)} \{ g_r(\sum_{s \in S} v_{rs}(t) \cdot u_3(\eta_k, s) - v_r(t)) \cdot \theta_r \}, \phi_k - \phi'_k \right) \\
 & = -\varepsilon_r(4:t) \cdot \sum_{k \in K} \left(\theta_r \cdot \frac{dg_r(u)}{du} \Big|_{u=\sum_{s \in S} v_{rs}(t) \cdot u_3(\eta_k, s) - v_r(t)} \cdot (-1), \phi_k - \phi'_k \right) \\
 & = -\varepsilon_r(4:t) \cdot \sum_{k \in K} (e_{rk} \cdot (-1), \phi_k - \phi'_k) \\
 & = \varepsilon_r(4:t) \cdot \sum_{k \in K} (-1) \cdot (e_{rk}, \phi'_k - \phi_k) \tag{6.130}
 \end{aligned}$$

である。

$$\begin{aligned}
 & \textcircled{3} \\
 & w_{pq}(t) \text{ が変わる} \Rightarrow f_p \left(\sum_{q \in Q} w_{pq}(t) \cdot u_2(\varphi_j, q) - w_p(t) \right) \text{ が変わる} \\
 & \text{すべての } \eta_j \text{ が変わる} (j \in J) \Rightarrow \text{すべての } \phi_k \text{ が変わる} (k \in K) \tag{6.131}
 \end{aligned}$$

ことに注意して、計算すれば、

$$\begin{aligned}
 & \Delta w_{pq}(t) \\
 & = -\varepsilon_{pq}(1:t) \cdot \sum_{k \in K} 2^{-1} \cdot 2 \left(\frac{\partial}{\partial w_{pq}(t)} \{ \phi_k - \phi'_k \}, \phi_k - \phi'_k \right) \quad \because \quad \mathfrak{H} \text{ は実ヒルベルト空間} \\
 & = -\varepsilon_{pq}(1:t) \cdot \sum_{k \in K} \left(\frac{\partial}{\partial w_{pq}(t)} \phi_k, \phi_k - \phi'_k \right) \tag{6.132} \\
 & = \varepsilon_{pq}(1:t) \cdot \sum_{k \in K} \left(\frac{\partial \phi_k}{\partial w_{pq}(t)}, \phi'_k - \phi_k \right) \tag{6.133}
 \end{aligned}$$

ここに、

$$\begin{aligned}
 & \frac{\partial \phi_k}{\partial w_{pq}(t)} \\
 &= \sum_{j \in J} \frac{\partial \phi_k}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial w_{pq}(t)} \\
 &= \frac{\partial \phi_k}{\partial \eta_k} \cdot \frac{\partial \eta_k}{\partial w_{pq}(t)} \\
 &= \frac{\partial \sum_{r \in R} g_r \left(\sum_{s \in S} v_{rs}(t) \cdot u_3(\eta_k, s) - v_r(t) \right) \cdot \theta_r}{\partial \eta_k} \\
 & \quad \cdot \frac{\partial \sum_{p \in P} f_p \left(\sum_{q \in Q} w_{pq}(t) \cdot u_2(\varphi_k, q) - w_p(t) \right) \cdot \sigma_p}{\partial w_{pq}(t)} \\
 &= \sum_{r \in R} \theta_r \cdot \frac{dg_r(u)}{du} \Big|_{u = \sum_{s \in S} v_{rs}(t) \cdot u_3(\eta_k, s) - v_r(t)} \cdot \frac{\partial u_3(\eta_k, s)}{\partial \eta_k} \\
 & \quad \sigma_p \cdot \frac{df_p(u)}{du} \Big|_{u = \sum_{q \in Q} w_{pq}(t) \cdot u_2(\varphi_k, q) - w_p(t)} \cdot u_2(\varphi_k, q) \\
 &= \sum_{r \in R} e_{rk} \cdot \frac{\partial u_3(\eta_j, s)}{\partial \eta_j} \\
 & \quad \sigma_p \cdot \frac{df_p(u)}{du} \Big|_{u = \sum_{q \in Q} w_{pq}(t) \cdot u_2(\varphi_k, q) - w_p(t)} \cdot u_2(\varphi_j, q) \\
 &= \frac{\partial u_3(\eta_j, s)}{\partial \eta_j} \cdot u_2(\varphi_j, q) \cdot \sum_{r \in R} e_{rpk} \tag{6.134}
 \end{aligned}$$

同様に、

$$\textcircled{4} \quad w_p(t) \text{ が変る} \Rightarrow f_p \left(\sum_{q \in Q} w_{pq}(t) \cdot u_2(\varphi_j, q) - w_p(t) \right) \text{ が変る}$$

$$\text{すべての} \eta_j \text{ が変る} (j \in J) \Rightarrow \text{すべての} \phi_k \text{ が変る} (k \in K)$$

ことに注意して、計算すれば、

(6.135)

$$\Delta w_p(t)$$

$$= -\varepsilon_p(2:t) \cdot \sum_{k \in K} 2^{-1} \cdot 2 \left(\frac{\partial}{\partial w_p(t)} \{ \phi_k - \phi'_k \}, \phi_k - \phi'_k \right) \quad \because \quad \mathfrak{H} \text{ は実ヒルベルト空間}$$

$$= -\varepsilon_p(2:t) \cdot \sum_{k \in K} \left(\frac{\partial}{\partial w_p(t)} \phi_k, \phi_k - \phi'_k \right) \quad (6.136)$$

$$= \varepsilon_p(2:t) \cdot \sum_{k \in K} \left(\frac{\partial \phi_k}{\partial w_p(t)}, \phi'_k - \phi_k \right) \quad (6.137)$$

ここに、

$$\begin{aligned} & \frac{\partial \phi_k}{\partial w_p(t)} \\ &= \sum_{j \in J} \frac{\partial \phi_k}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial w_p(t)} \\ &= \frac{\partial \phi_k}{\partial \eta_k} \cdot \frac{\partial \eta_k}{\partial w_p(t)} \\ &= \frac{\partial \sum_{r \in R} g_r \left(\sum_{s \in S} v_{rs}(t) \cdot u_3(\eta_k, s) - v_r(t) \right) \cdot \theta_r}{\partial \eta_k} \\ & \quad \cdot \frac{\partial \sum_{p \in P} f_p \left(\sum_{q \in Q} w_{pq}(t) \cdot u_2(\phi_k, q) - w_p(t) \right) \cdot \sigma_p}{\partial w_p(t)} \\ &= \sum_{r \in R} \theta_r \cdot \frac{dg_r(u)}{du} \Big|_{u = \sum_{s \in S} v_{rs}(t) \cdot u_3(\eta_k, s) - v_r(t)} \cdot \frac{\partial u_3(\eta_k, s)}{\partial \eta_k} \\ & \quad \sigma_p \cdot \frac{df_p(u)}{du} \Big|_{u = \sum_{q \in Q} w_{pq}(t) \cdot u_2(\phi_k, q) - w_p(t)} \cdot (-1) \\ &= \sum_{r \in R} e_{rk} \cdot \frac{\partial u_3(\eta_j, s)}{\partial \eta_j} \end{aligned}$$

$$\begin{aligned}
 & \sigma_p \cdot \frac{df_p(u)}{du} \Big|_{u=\sum_{q \in Q} w_{pq}(t) \cdot u_2(\varphi_k, q) - w_p(t)} \cdot (-1) \\
 &= \frac{\partial u_3(\eta_j, s)}{\partial \eta_j} \cdot (-1) \cdot \sum_{r \in R} e_{rpk}
 \end{aligned} \tag{6.138}$$

第7章 むすび

我々は明示的知識を記号列（言語；シンボルの列）で表現することに慣れている。男顔と女顔の違いを認識するのに非明示的知識を用いているのは確かである。非明示的知識を表現するには、パターン（非言語表現）が都合がよい。我々の脳は明示的知識、非明示的知識をどのように記憶して使っているのだろうか？意識的処理には記号列モデル（symbol model）を使い、無意識的処理にはパターン（pattern）を使っているのだろうか？

知識を表現するのに、

- (1) 記号列モデルを使うか、
- (2) 分散的表現モデル（パターン;parallel-distributed model）、ニューラルネットモデル（コネクションISTモデル;connectionist model）を使うか

は、議論が分かれる。意識下推論の実現には記号列モデルが適しており、無意識下連想の実現にはニューラルネットモデル（コネクションISTモデル）が適している。

本論文は、全く未知の問題に挑戦した。それは、S.Suzukiの、付録Aのaxiom 1を満たすモデル構成作用素 T の構造形式を採用することで可能となった特徴量の組を用いたニューラルネットを除いて、可分な一般抽象ヒルベルト空間 \mathfrak{H} が有限次元のユークリッド空間に一致させ、登場している作用素を恒等作用素にする特別な場合、従来の結果に一致するように、稼動方程式、on-line学習法が研究されたことを指している。本研究内容はこれ迄のニューラルネットワーク理論、線形多変量解析学を大きく変革させるだろう。特徴量の組を用いたニューラルネットの稼動方程式、on-line学習法は特に、本論文の、全くの独創である。このニューラルネットは容易に実現可能という意味で、格別に、実用価値が高いと予想しており、特徴量の組を入力する従来の有限次元のユークリッド空間で動作するニューラルネットと馴染みが深い。

可分な一般抽象ヒルベルト空間 \mathfrak{H} で稼動する各種ニューラルネットモデルのon-line学習がこれまで他の研究者により提案されなかったのは、学習が可能ないように稼動方程式（システム方程式）を提案できなかったからである。本論文は、学習が可能ないように \mathfrak{H} での稼動方程式を提案したことにも触れておかねばならない。on-line学習法のみならず、この事実（稼動方程式の提案）も本論文の功績である。

再帰ニューラルネット、誤差逆伝播ニューラルネットのシステム方程式の性質を数理的に解明することは、非線形関数解析学を飛躍的に向上させることに結びつくと予想している。

連想ニューラルネット、1次ニューラルネット、再帰ニューラルネット、誤差逆伝播ニューラルネットの諸稼動方程式の構造形式は、パーセプトロン型ニューラルネットは無論のこと、S.Suzukiの提

案した式 (1.4) のパーセプトロン型作用素 [6] が基調となって、S.Suzukiが熟慮の上、本論文で提案したものである。他の構造形式をとれば、on-line学習が非実用的になると感触を得ている。付録Bに2次ニューラルネットを研究しておいた。

学習率の選定はさておいて、実用に耐えるon-line学習諸公式が提出された。

シミュレーションを行い、その効用を確認し、本研究の価値を高める必要がある。

文 献

- [1] 鈴木昇一：認識工学, 柏書房, Feb.1975
- [2] 鈴木昇一：ニューラルネットの新数理, 近代文芸社, Sept.1996
- [3] 鈴木昇一：パターン認識問題の数理的一般解決, 近代文芸社, June 1997
- [4] 鈴木昇一：認識知能情報論の新展開, 近代文芸社, Aug.1998
- [5] 鈴木昇一：測度的不変量検出形認識系の構成理論、電子通信学会論文誌 (D)、vol.55-D、No.8、pp.513-538, Aug.1972
- [6] 鈴木昇一：パターンのエントロピーモデル, 電子情報通信学会論文誌 (D-II)、vol.J77-D-II、no.10, pp.2220-2238, Nov.1994
- [7] 都築誉史、浅川伸一：コネクショニストモデルと認知科学—特集編集にあたって— (特集—高次認知機能の創発とコネクショニストモデル)、Cognitive Studies, vol.10, no.1, pp.3-9, Mar.2003
- [8] 守一雄、都築誉史、楠見孝：コネクショニストモデルと心理学(脳のシミュレーションによる心の理解)、北大路書房、Jun.2001
- [9] M.シュピッツァー (Manfred Spitzer)：脳 回路網の中の精神 (ニューラルネットが描く地図)、村井俊哉・山岸洋訳、新曜社、Nov.2001
- [10] 田中賢：チューリング等価なニューラルネットワークの簡略化、電子情報通信学会論文誌D-II、vol.J88-D-II、no.5, pp.916-923, May 2005
- [11] Ke Lu, Xiaofei He: Image retrieval based on incremental subspace learning, Pattern Recognition, vol.38, pp.2047-2054, 2005
- [12] 濱裕光、柳重堪、阮牧：シグナルシステムの数学、森北出版、Dec.1997
- [13] 森田昌彦、村田和彦、諸上茂光、末光厚夫：選択的不感化法を適用した層状ニューラルネットの情報統合能力、電子情報通信学会論文誌D-II、vol.J87-D-II、no.12, pp.2242-2252, Dec.2004
- [14] Ronald J. Williams, David Zipser: A learning algorithm for continually running fully recurrent neural networks, Neural Computation, vol.1, pp.270-280, 1989
- [15] 鈴木昇一：パターン系列 (動画像、会話音声) の、dynamical systemによる連想理論と、連想器SPATEMTRON、情報研究 (文教大学・情報学部)、no.30, pp.139-186, Jan.2004
- [16] U.Rajendra Acharya, P.Subbanna Bhat, S.S.Iyengar, Ashok Rao, Sumeet Dua: "Classification of heart rate data using artificial neural network and fuzzy equivalence relation", Pattern Recognition, vol.36, pp.61-68, 2003
- [17] 中野良平：ニューラル情報処理の数理的基礎 (情報システム工学 = MKC-A3)、数理工学社、Nov.2005
- [18] 白井良明、谷内田正彦：パターン情報処理 (新コンピュータサイエンス講座)、オーム社、

Aug.1998

- [19] 鹿野清宏、伊藤克亘、河原達也、武田一哉、山本幹雄:音声認識システム (IT Text 情報処理学会編集)、オーム社、May 2001
- [20] Aapo Hyvärinen (アーポ・ヒバリネン) ,Juha Karhunen (ユハ・カルーネン) ,Erkki Oja (エルキ・オヤ) :詳解 独立成分分析 (信号解析の新しい世界)、根本幾・川勝真喜訳、東京電機大学出版局、Feb.2005
- [21] 高橋渉：非線形関数解析学 (現代数学ゼミナール7)、近代化学社、Sept.1989
- [22] 豊田秀樹：非線形多変量解析 (統計ライブラリー ニューラルネットによるアプローチ)、朝倉書店、Feb.1996

付録A.SS公理系と認識システムRECOGNITRON

S.Suzukiが提案したパターン情報処理に関する数学的理論 [3], [4] は4つの公理axiom 1~4からなるSS公理系で組み立てられている。本付録Aでは、その内、4つの公理axiom 1~4を説明する。SS公理系を適用して構築された連想形認識システム

$$\text{RECOGNITRON} =_{\Delta} \text{RECOGNITRON}(\mathcal{G}(J) : \Omega(J)) \equiv \langle \Phi_B, T, SM, BSC \rangle \quad (\text{A.1})$$

の多段階連想形認識の動作、カテゴリ帰属知識空間、構造受精作用素と構造受精変換、同値関係、半順序関係、カテゴリ帰属知識 MK を未知とする連想形認識方程式

$$MK =_{\Delta} \langle T\varphi, J \rangle \sqcup TA(\mu)T \cdot MK \quad (\text{A.2})$$

と、その解 MK の諸性質、人工知能学における探索理論から眺めた「処理の対象とする問題の入力パターン φ についての不動点探索形多段階連想形認識過程、つまり、連想形認識方程式 (A.2) の求解過程」などについては、2文献 [3], [4] で説明されている。

A1.パターン集合 Φ と、モデル構成作用素 T の対 $[\Phi, T]$ が満たさなければならない公理 1

認識システムRECOGNITRONがモデル $T\varphi \in \Phi$ を見たり聞いたりしたならば、原パターン $\varphi \in \Phi$ と同じに見えたり聞こえたりする同一知覚原理が成立するならば、処理の対象とする問題のパターン φ の集合 Φ とモデル構成作用素と呼ばれる写像

$$T : \Phi \rightarrow \Phi \quad (\text{A.3})$$

のなす対 $[\Phi, T]$ が満たさなければならないaxiom 1について説明しよう。

処理の対象とするパターン φ の集合 Φ はある可分なヒルベルト空間 \mathcal{H} の、零元 0 を含むある部分集合であり、この Φ 、並びに式 (A.3) の写像 T の対 $[\Phi, T]$ は4性質

- ① (零パターンの不動点性) $\varphi = 0 \Rightarrow T\varphi = \varphi$
- ② (正定数倍に対するモデルの不変性) 任意のパターン φ について、任意の正定数 a に対し、 $T(a \cdot \varphi) = T\varphi$
- ③ (モデルのモデルは元のモデルである性質) 任意のパターン φ について、 $T(T\varphi) = T\varphi$
- ④ (非零モデルの存在) 或るパターン φ が存在して、 $T\varphi \neq 0$

を含む形で、次のaxiom 1を満たさなければならない。4性質①～④は次のaxiom 1の (i), (ii), (iii) の3後半、並びに (iv) である。このとき、写像 T はモデル構成作用素 (model-construction operator) と呼ばれ、 $T\varphi \in \Phi$ は $\varphi \in \Phi$ の代りとなり得るという意味で、パターン $\varphi \in \Phi$ のパターンモデル、或いは簡単にモデル (model) と呼ばれる。

Axiom 1 (パターン集合 Φ とモデル構成作用素 T との対 $[\Phi, T]$ の満たすべき公理) [3], [4]

(i) (零元 0 の Φ への埋込性、零元 0 の T - 不動点性)

$$0 \in \Phi \wedge T0 = 0.$$

(ii) (Φ の錐性、 T の正定数倍吸収性)

$$\forall \varphi \in \Phi, a \cdot \varphi \in \Phi \wedge T(a \cdot \varphi) = T\varphi \quad \text{for any positive real number } a.$$

(iii) (Φ への埋込性、 T のベキ等性)

$$\forall \varphi \in \Phi, T\varphi \in \Phi \wedge T(T\varphi) = T\varphi.$$

(iv) (T の非零写像性)

$$\exists \varphi \in \Phi, T\varphi \neq 0.$$

□

パターンモデルの $T\varphi$ 集合

$$T \cdot \Phi \equiv \{T\varphi \mid \varphi \in \Phi\} \tag{A.4}$$

を考える。上述のaxiom 1からわかるように、パターン集合 Φ は、埋込性

$$T \cdot \Phi \subset \Phi \tag{A.5}$$

を満たし、原点 (= 0) を始点とし、 Φ の任意の点を通る半直線を含むような集合、つまり、錐 (cone) であらねばならない。

パターンと判明している φ の集合 (基本領域; basic domain) $\Phi_B (\ni 0)$ と、すべての正実定数の集合 R^{++} とを用意する。 Φ_B は零元 0 を含まなければならない。

パターンと判明している元の集合 (基本領域; basic domain) (axiom 1の (i) の前半から、 $0 \in \Phi_B$ を導入する。上述の4性質①～④から成立しなければならない集合論的再帰領域方程式 (set-theoretic reflective domain equation)

$$\Phi = \Phi_B \cup T \cdot \Phi \cup R^{++} \cdot \Phi \tag{A.6}$$

の解 Φ は以下の式 (A.9) のようにと表示される。ここに、 $R^{++}, R^{++}\Phi$ は

$$R^{++} : \text{正実数全体の集合} \tag{A.7}$$

$$R^{++} \cdot \Phi \equiv \{r^{++} \cdot \varphi \mid r^{++} \in R^{++}, \varphi \in \Phi\} \tag{A.8}$$

である。

次の定理A.1は、axiom 1を満たす対 $[\Phi, T]$ を決定している。

[定理A.1] (パターン集合 Φ とモデル構成作用素との対 $[\Phi, T]$ の基本構成定理)

パターンと判明している φ の集合（基本領域） $\Phi_B (\ni 0)$ と、すべての正実定数の集合 R^{++} とを用意する。

式 (A.3) の写像 T が axiom 1 の (i), (ii), (iii) の 3 後半、並びに、(iv) を満たすとしよう。このとき、次の (1#), (2#) が成り立つ:

(1#) 処理の対象とする問題のパターン φ の集合 Φ を、

$$\begin{aligned} \Phi &= R^{++} \cdot (\Phi_B \cup T \cdot \Phi_B) \\ &\equiv \{r^{++} \cdot \varphi_B \mid r^{++} \in R^{++}, \varphi_B \in \Phi_B\} \cup \{r^{++} \cdot T\varphi_B \mid r^{++} \in R^{++}, \varphi_B \in \Phi_B\} \end{aligned} \quad (\text{A.9})$$

の如く設定すれば、

$$(1\#-a) \quad T \cdot \Phi = T \cdot \Phi_B \subset \Phi$$

\therefore axiom 1 の (ii), (iii) の 2 後半

$$(1\#-b) \quad R^{++} \cdot \Phi = \Phi$$

\therefore axiom 1 の (ii) の後半

が成立し、axiom の (i), (ii), (iii) の 3 前半を Φ は満たし、結局、対 $[\Phi, T]$ は axiom 1 を満たす。

(2#) 逆に、 $(0 \in) \Phi_B$ を部分集合に持つ Φ が axiom 1 の (i), (ii), (iii) の 3 前半を満たすとするれば、

$$\Phi \supseteq \Phi_B \cup R^{++} \cdot \Phi \cup T \cdot \Phi \quad (\text{A.10})$$

が成立するが、ここで、特に、包含式 (A.10) において等号が成立するような最小の Φ を採用すれば、つまり、領域方程式 (A.6) の成立を仮定すれば、axiom 1 を満たす対 $[\Phi, T]$ の Φ は式 (A.9) のように表され、(1#-a)、(1#-b) も成立する。

(証明) (1#) (イ) は文献[4]、付録1の定理 A 1. 1 である。(2#) (ロ) は文献 [3]、pp.64-66 (2.4節) で証明されている。 \square

A2. 類似度関数 SM の満たさなければならない公理2

任意のパターン $\varphi \in \Phi$ が、記憶されている有限個の元（代表的なパターン）からなる集合 Ω 内の任意の代表パターン ω_j とどの程度似ているか、違っているかを計量する手段を設定することが、認識の働きを確保するために必要とされる。この類似性計量のための手段が類似度関数 SM である。

“正常なパターン” (well-formed pattern) は、ある 1 つのカテゴリ (category) \mathfrak{C}_j (第 $j \in J$ 番目の類概念) のみに帰属しているものとし、このような \mathfrak{C}_j の集まり (有限集合)

$$\mathfrak{C} \equiv \mathfrak{C}(J) \equiv \{\mathfrak{C}_j \mid j \in J\} \quad (\text{A.11})$$

を想定する。 \mathfrak{C}_j の備えている性質を典型的に持っている (第 $j \in J$ 番目の) 代表パターン (prototypical pattern) $\omega_j (\neq 0)$ を 1 つ選定する。 \mathfrak{C}_j は、典型 (prototype) としての代表パターン ω_j を中心とした緩やかな (第 $j \in J$ 番目の) カテゴリであることを仮定したことに注意しておく。ここに、

$$\Omega \equiv \Omega(J) \equiv \{\omega_j \mid j \in J\} \subset \Phi \quad (\text{A.12})$$

が式 (A.11) の全カテゴリ集合 $\mathfrak{C}(J)$ に 1 対 1 に対応する代表パターンの集合である。式 (A.12) の系 Ω は、

$$\text{複素定数 } a_j \text{ の組 } \{a_j \mid j \in J\} \text{ について } \sum_{j \in J} a_j \cdot \omega_j = 0 \Rightarrow \forall j \in J, a_j = 0 \quad (\text{A.13})$$

が成立しているという意味で、1次独立 (linearly independent) でなければならない。 Ω を視察で決定できる場合があるが、訓練パターン系列から Ω を適応的に決定する方法については、文献 [B3] の付録 I で説明されている。

Axiom 1 を満たす式 (A.3) のモデル構成作用素 T によって、式 (A.12) の代表パターン集合 Ω が変換されて得られる系

$$T \cdot \Omega \equiv \{T\omega \mid \omega \in \Omega\} = \{T\omega_j \mid j \in J\} \quad (\text{A.14})$$

も 1 次独立であると要請する。このとき、類似度関数 (similarity-measure function)

$$SM : \Phi \times \Omega \rightarrow \{s \mid 0 \leq s \leq 1\} \quad (\text{A.15})$$

を導入し、

$$SM(\varphi, \omega_j) = 1, 0 \text{ に従って、パターン } \varphi \in \Phi \text{ は各々、} \omega_j \text{ と確定的な類似度関係、相違関係にあり、}$$

$$\text{また、} 0 < SM(\varphi, \omega_j) < 1 \text{ の場合は、あいまいな類似・相違関係にある} \quad (\text{A.16})$$

と、 SM を解釈しよう。

式 (A.15) の関数 SM は次の axiom 2 を満たすように構成されねばならない。Axiom 2 の (i) では、クロネッカー (Kronecker) の δ 記号

$$\delta_{ij} = 1 \quad \text{if} \quad i = j, = 0 \quad \text{if} \quad i \neq j \quad (\text{A.17})$$

が導入されているが、特に axiom 2 の (i) なるこの正規直交性は、候補カテゴリの分離・抽出が効果的に行われ、

$$\text{候補カテゴリの鋭利な削減 (a sharp reduction)} \quad (\text{A.18})$$

をもたらすために要請されている。

1 より大きくない非負実数 $SM(\varphi, \omega_j)$ は、パターン $\varphi \in \Phi$ が、第 $j \in J$ 番目のカテゴリ \mathfrak{C}_j の諸性質を備えている代表パターン ω_j と似ている程度を表している、と解釈できる。

類似度関数 (similarity-measure function) と呼ばれる関数

$$SM : \Phi \times \Omega \rightarrow \{s \mid 0 \leq s \leq 1\} \quad (\text{A.19})$$

は、次の axiom 2 を満たすように構成されねばならない。

Axiom 2 (類似度関数 SM の満たすべき公理)

(i) (正規直交性; orthonormality)

$$\forall i, \forall j, SM(\omega_i, \omega_j) = \delta_{ij} (= 1 \quad \text{if} \quad i = j, = 0 \quad \text{if} \quad i \neq j)$$

(ii) (規格化条件、確率性、正規性; probability condition, normalization)

$$\forall \varphi \in \Phi, \sum_{j \in J} SM(\varphi, \omega_j) = 1.$$

(iii) (写像 T の下での不変性; invariance under mapping T)

$$\forall \varphi \in \Phi, \forall j \in J, SM(T\varphi, \omega_j) = SM(\varphi, \omega_j).$$

□

上述のaxiom 2の (i) ~ (iii) について簡単に説明しておこう。

SM の解釈式 (A.16) の下で、(i) は、相異なるカテゴリの代表パターン同士は確定的な相違関係にあり、同一カテゴリの代表パターン同士は確定的な類似度関係にあることを要請している。(ii) は、任意のパターン φ について、すべてのカテゴリについての類似度の総和は 1 であることを要請している。つまり、パターン φ は少なくとも 1 つのカテゴリ \mathfrak{C}_j に帰属していることを要請している。(iii) は、パターンモデル $T\varphi$ は原パターン φ と任意のカテゴリについて同一類似度を持つことを要請している。ということは、パターンモデル $T\varphi$ を見たり、聞いたりするならば、原パターン φ と同じように見えたり、聞こえたりすること (同一知覚原理; A2章を参照) を要請していることになる。

尚、不等式、等式

$$[\forall j \in J, 0 < p(\mathfrak{C}_j) < 1] \wedge \sum_{j \in J} p(\mathfrak{C}_j) = 1 \quad (\text{A.20})$$

を満たす第 $j \in J$ 番目のカテゴリ \mathfrak{C}_j の出現確率 $p(\mathfrak{C}_j)$ を導入しておく。

A3. 大分類関数 BSC の満たさなければならない公理3

次のaxiom 3を満たすものとして大分類関数 (rough classifier, binary-state classifier) と呼ばれる2値関数

$$BSC : \Phi \times J \rightarrow \{0, 1\} \quad (\text{A.21})$$

を導入し、解釈

パターン $\varphi \in \Phi$ の帰属する候補カテゴリの 1 つが第 $j \in J$ 番目の \mathfrak{C}_j であるならば、

$$BSC(\varphi, j) = 1 \text{ であることが望ましい} \quad (\text{A.22})$$

を採用しよう。

Axiom 3 (大分類関数 BSC の満たすべき公理)

(i) (カテゴリ抽出能力; category separability)

$$\forall j \in J, BSC(\omega_j, j) = 1.$$

(ii) (写像 T の下での不変性; invariance under mapping T)

$$\forall \varphi \in \Phi, \forall j \in J, BSC(T\varphi, j) = BSC(\varphi, j).$$

□

この際、注意すべきは、

$BSC(\varphi, j) = 0$ であっても、パターン $\varphi \in \Phi$ の帰属する候補カテゴリの 1 つは、第 $j \in J$ 番目の \mathfrak{C}_j でないとは限らない

(A.23)

としていることである。 $BSC(\varphi, j)=1$ であれば、第 $j \in J$ 番目のカテゴリ \mathfrak{C}_j は、パターン $\varphi \in \Phi$ が帰属する候補カテゴリの1つであると、解釈できる。

また、axiom 3の (i) からわかるように、カテゴリ間の相互排除性 (the mutual exclusion of the one category from the other categories)

$$\forall j \in J, \forall i \in J - \{j\}, BSC(\omega_i, j) = 0 \quad (\text{A.24})$$

を公理として要請していない事実に注意しておこう。この事実を補うのが実は、式 (A.15) の類似度関数 SM が満たさなければならないとしている axiom 2の (i) (正規直交性) である。

式 (A.15) の類似度関数が式 (A.18) でいう“候補カテゴリの鋭利な削減”を持つためには、axiom 2, (i) の正規直交性を満たす必要があることがA3節で指摘されたが、 $SM(\varphi, \omega_j)$ の代りに $SM(\varphi, \omega_j) \cdot BSC(\varphi, j)$ を用いれば、パターン φ が帰属するかも知れない候補カテゴリを益々、鋭利に削減できると期待される。

A4. パターン集合 Φ の意味領域としてのカテゴリ帰属知識空間 $\langle \Phi, 2^J \rangle$

カテゴリ番号 j の全集合 J の部分集合

$$\gamma = \{i, j, \dots k\} (\subseteq J) \quad (\text{A.25})$$

は、重複要素があってもよいリスト

$$\gamma = [i \ j \ \dots \ k] \in 2^J \equiv \{\lambda \mid \lambda \subseteq J\} \quad (\text{A.26})$$

として、表現されることがある。ここに、 2^J は集合 J のすべての部分集合のなす集合、つまり、“すべてのカテゴリ番号の集合 J のべき集合 (power set)” をで表わしている。このカテゴリ番号リスト $\gamma \in 2^J$ から定まるような $\mathfrak{C}(J)$, $\Omega(J)$ の部分集合

$$\mathfrak{C}(\gamma) \equiv \{\mathfrak{C}_i \mid i \in \gamma\} \subseteq \mathfrak{C}(J), \gamma \subseteq J \quad (\text{A.27})$$

$$\Omega(\gamma) \equiv \{\omega_i \mid i \in \gamma\} \subseteq \Omega(J), \gamma \subseteq J \quad (\text{A.28})$$

も定義しておく。

式 (A.1) の認識システム RECOGNITRON がパターン $\varphi \in \Phi$ に対し、

「パターン $\varphi \in \Phi$ が、式 (A.3.1) の全カテゴリ集合 \mathfrak{C}_j の部分集合 $\mathfrak{C}(\gamma)$ 内の何れか1つのカテゴリ \mathfrak{C}_j に帰属する可能性がある」

(A.29)

というような“パターン $\varphi \in \Phi$ のカテゴリ帰属知識 (categorical membership-knowledge)”を持っているとする。この知識を、

$$\langle \varphi, \gamma \rangle \in \langle \Phi, 2^J \rangle \quad (\text{A.30})$$

と表す。登場している集合

$$\langle \Phi, 2^J \rangle \equiv \{\langle \varphi, \gamma \rangle \mid \varphi \in \Phi, \gamma \in 2^J\} \quad (\text{A.31})$$

は、カテゴリ帰属知識空間 (categorical membership-knowledge space) と呼ばれ、すべてのパターン

$\varphi \in \Phi$ と、すべてのカテゴリ $\gamma \in 2^J$ 番号のリストとのなす順序の付いた対リスト (an ordered pair list) $\langle \varphi, \gamma \rangle$ の集合である。

パターン集合 Φ が構文領域 (syntactical domain) であるとすれば、カテゴリ帰属知識空間 $\langle \Phi, 2^J \rangle$ は構文領域 Φ の意味領域 (semantic domain) である。

A5. 公理 4 と、カテゴリ選択関数 CSF の構造形式

カテゴリ選択関数 (category-selection function) と呼ばれる写像

$$CSF : \Phi \times 2^J \rightarrow 2^J \quad (A.32)$$

は、包含関係 (inclusion relation)

$$\forall \varphi \in \Phi, \forall \gamma \in 2^J, CSF(\varphi, \gamma) \subseteq \gamma \subseteq J \in 2^J \quad (A.33)$$

を満たし、然も、次の axiom 4 を満たすものとして、設定されるとしよう。

Axiom 4 (カテゴリ選択関数 CSF の満たすべき公理)

(i) $\varphi = 0 \vee \gamma = \phi$ の場合

如何なるカテゴリ番号 $k \in \gamma$ も $\langle \varphi, \gamma \rangle$ の有効な候補カテゴリの番号ではない。

(ii) $\varphi \neq 0 \wedge \gamma \neq \phi$ であり、かつ、 $SM(\varphi, \omega_k) = 0 \wedge BSC(\varphi, k) = 0$ の場合

カテゴリ番号 $k \in \gamma$ は、 $\langle \varphi, \gamma \rangle$ の有効な候補カテゴリの番号ではない

(iii) $\varphi \neq 0 \wedge \gamma \neq \phi$ であり、かつ、 $\sum_{k \in \gamma} BSC(\varphi, k) = 0$ の場合

$BSC(\varphi, k) = 0$ であっても、 $SM(\varphi, \omega_k) > 0$ であるようなカテゴリ番号 $k \in \gamma$ は、 $\langle \varphi, \gamma \rangle$ の有効な候補カテゴリの番号である。

(iv) $\varphi \neq 0 \wedge \gamma \neq \phi$ であり、かつ、 $\sum_{k \in \gamma} BSC(\varphi, k) > 0$ の場合

(iv-1) $BSC(\varphi, k) = 0$ なるカテゴリ番号 $k \in \gamma$ は、 $SM(\varphi, \omega_k) > 0$ であっても、 $\langle \varphi, \gamma \rangle$ の有効な候補カテゴリの番号ではない。

(iv-2) $SM(\varphi, \omega_k) = 0$ なるカテゴリ番号 $k \in \gamma$ は、 $BSC(\varphi, k) = 1$ であっても、 $\langle \varphi, \gamma \rangle$ の有効な候補カテゴリの番号ではない。 \square

次の定理 A.2 は、axiom 4 を満たすように、式 (A.32) のカテゴリ選択関数 CSF の構造を決定したものである。この定理 A.2 では、式 (A.32) の写像 CSF は、式 (A.19) の類似度関数 SM、式 (A.21) の大分類関数 BSC を使用する形式で、

その定義域が $\Phi \times 2^J$ であり、その値域が、パターン $\varphi \in \Phi$ のカテゴリ帰属知識 $\langle \varphi, \gamma \rangle \in \langle \Phi, 2^J \rangle$ の有効な候補カテゴリの番号リスト (a list of effective category-numbers) の集合である (A.34) の如く、構成されている。

[定理 A.2] (カテゴリ選択関数 CSF の構成定理)

次のように定義される式 (A.32) の 1 つの写像 CSF は式 (2.36) と上述の axiom 4 を満たす:

(i) $\varphi = 0 \vee \gamma = \phi$ の場合

$$CSF(\varphi, \gamma) = \phi. \quad (A.35)$$

(ii) $\varphi \neq 0 \wedge \gamma \neq \phi$ の場合

$$CSF(\varphi, \gamma) = \begin{cases} \{k \in \gamma \mid SM(\varphi, \omega_k) > 0\} & \text{if } \sum_{k \in \gamma} BSC(\varphi, k) = 0 \end{cases} \quad (\text{A.36})$$

$$\begin{cases} \{k \in \gamma \mid SM(\varphi, \omega_k) > 0 \wedge BSC(\varphi, k) = 1\} & \text{if } \sum_{k \in \gamma} BSC(\varphi, k) > 0 \end{cases} \quad (\text{A.37})$$

(証明) 文献 [B3] の定理E1である。 □

定理A.2のカテゴリ選択関数 CSF について、次のように解釈できる:

処理の対象とするパターン $\varphi \in \Phi$ がカテゴリ $\mathfrak{C}(\gamma)$ 内の何れか 1 つに帰属する可能性があると想定した場合、更に絞り込んで、その内のカテゴリ内 $\mathfrak{C}(CSF(\varphi, \gamma))$ の何れか 1 つに帰属する可能性があると帰納推論 (inductive reasoning) できる機能を備え、その出力 $CSF(\varphi, \gamma)$ はパターン $\varphi \in \Phi$ の有効な候補カテゴリの番号のリストを与えている。 □

(付録A終わり)

付録B. 2次ニューラルネットのon-line学習

本付録Bでは、1つのパターン内の要素同士にあるかもしれない関係を抽出できたりする2次ニューラルネットをon-lineで学習する方法を研究する。

B1. n 次元ユークリッド空間 R^n での2次ニューラルネット

2次ニューラルネットを採用すれば、 n 次元実数値ベクトル

$$\vec{u} = \text{col}(u_1 \quad u_2 \quad \cdots \quad u_n) \in R^n \quad (\text{B.1})$$

を入力したとき、得られる第 $j(=1 \sim q)$ 番目の出力 $\eta_j \in R^q$ は、

$$\eta_j = f_j \left(\sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell) \cdot u_k \cdot u_\ell + \sum_{k=1}^n w(j, k) \cdot u_k + w(j) \right), j = 1 \sim q \quad (\text{B.2})$$

と表される。

学習時刻 $t(= 0, 1, 2, \dots)$ に訓練事例

$$\vec{u}' = \text{col}(u'_1 \quad u'_2 \quad \cdots \quad u'_n) \in R^n \quad (\text{B.3})$$

を入力したとき、各理想出力として

$$\eta_j = \eta'_j \in R^q, j = 1 \sim q \quad (\text{B.4})$$

が得られなければならないとき、更新式

$$w(j, k, \ell; t+1) = w(j, k, \ell; t) + \Delta w(j, k, \ell; t) \quad (\text{B.5})$$

$$w(j, k; t+1) = w(j, k; t) + \Delta w(j, k; t) \quad (\text{B.6})$$

$$w(j; t+1) = w(j; t) + \Delta w(j; t) \quad (\text{B.7})$$

での修正分

$$\Delta w(j, k, \ell; t), \Delta w(j, k; t), \Delta w(j; t) \quad (\text{B.8})$$

は次のようになる。

最急降下法に従い、適応誤差関数

$$E(t) \equiv \frac{1}{2} \cdot \sum_{j=1}^q [\eta_j - \eta'_j]^2 \quad (\text{B.9})$$

を最小にするように求めればよい。計算結果は次の通りである：

$$\begin{aligned} \textcircled{1} \Delta w(j, k, \ell; t) &= -\varepsilon(j, k, \ell; t) \cdot \frac{\partial E(t)}{\partial w(j, k, \ell; t)} \\ &= -\varepsilon(j, k, \ell; t) \cdot [\eta_j - \eta'_j] \frac{\partial}{\partial w(j, k, \ell; t)} \eta_j \\ &= -\varepsilon(j, k, \ell; t) \cdot [\eta_j - \eta'_j] \cdot \frac{df_j(u)}{du} \Big|_{u = \sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell; t) \cdot u_k \cdot u_\ell + \sum_{k=1}^n w(j, k; t) \cdot u_k + w(j; t)} \cdot u_k \cdot u_\ell \end{aligned} \quad (\text{B.10})$$

$$\begin{aligned} \textcircled{2} \Delta w(j, k; t) &= -\varepsilon(j, k; t) \cdot \frac{\partial E(t)}{\partial w(j, k; t)} \\ &= -\varepsilon(j, k; t) \cdot [\eta_j - \eta'_j] \frac{\partial}{\partial w(j, k; t)} \eta_j \\ &= -\varepsilon(j, k; t) \cdot [\eta_j - \eta'_j] \cdot \frac{df_j(u)}{du} \Big|_{u = \sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell; t) \cdot u_k \cdot u_\ell + \sum_{k=1}^n w(j, k; t) \cdot u_k + w(j; t)} \cdot u_k \end{aligned} \quad (\text{B.11})$$

$$\begin{aligned} \textcircled{3} \Delta w(j; t) &= -\varepsilon(j; t) \cdot \frac{\partial E(t)}{\partial w(j; t)} \\ &= -\varepsilon(j; t) \cdot [\eta_j - \eta'_j] \frac{\partial}{\partial w(j; t)} \eta_j \\ &= -\varepsilon(j; t) \cdot [\eta_j - \eta'_j] \cdot \frac{df_j(u)}{du} \Big|_{u = \sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell; t) \cdot u_k \cdot u_\ell + \sum_{k=1}^n w(j, k; t) \cdot u_k + w(j; t)} \cdot (+1) \end{aligned} \quad (\text{B.12})$$

ここに、 $\varepsilon(j, k, \ell; t), \varepsilon(j, k; t), \varepsilon(j; t)$ は十分小さい正の関数（学習率）である。

求める $w(j, k, \ell), w(j, k), w(j)$ は、

$$w(j, k, \ell) = \lim_{t \rightarrow \infty} \Delta w(j, k, \ell; t), w(j, k) = \lim_{t \rightarrow \infty} w(j, k; t), w(j) = \lim_{t \rightarrow \infty} w(j; t) \quad (\text{B.13})$$

と与えられる。

B2. 一般抽象実ヒルベルト空間 \mathfrak{H} での2次ニューラルネット

$\{\phi_j\}_{j=1 \sim m}$ を \mathfrak{H} の1次独立な系とする。

$$w(j, k, \ell), w(j, k), w(j) (j=1 \sim m; k, \ell=1 \sim n): \text{実定数} \quad (\text{B.14})$$

$$P_{k\ell}, P_k, P_0: \mathfrak{H} \rightarrow \mathfrak{H} (k, \ell=1 \sim n) \quad (\text{B.15})$$

$$f_j: R \rightarrow R (j=1 \sim m): \text{ニューロン発火関数} \quad (\text{B.16})$$

を導入して、

$$\varphi_i \in \mathfrak{H}, 1 \sim q \quad (\text{B.17})$$

を入力したとき、得られる第 $i (= 1 \sim q)$ 番目の出力 $\eta_i \in \mathfrak{H}$ は

$$\eta_i = \sum_{j=1}^m f_j \left(\sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell) \cdot P_{k\ell} \varphi_i + \sum_{k=1}^n w(j, k) \cdot P_k \varphi_i + w(j) \cdot P_0 \varphi_i \right) \cdot \phi_j, i=1 \sim q \quad (\text{B.18})$$

と表される。学習時刻 $t (= 0, 1, 2, \dots)$ に訓練事例

$$\varphi'_i \in \mathfrak{H}, 1 \sim q \quad (\text{B.19})$$

を入力したとき、各理想出力として

$$\eta_i = \eta'_i \in \mathfrak{H}, i=1 \sim q \quad (\text{B.20})$$

が得られなければならないとき、更新式

$$w(j, k, \ell; t+1) = w(j, k, \ell; t) + \Delta w(j, k, \ell; t) \quad (\text{B.21})$$

$$w(j, k; t+1) = w(j, k; t) + \Delta w(j, k; t) \quad (\text{B.22})$$

$$w(j; t+1) = w(j; t) + \Delta w(j; t) \quad (\text{B.23})$$

での修正分

$$\Delta w(j, k, \ell; t), \Delta w(j, k; t), \Delta w(j; t) \quad (\text{B.24})$$

は次のようになる。

最急降下法に従い、適応誤差関数

$$E(t) \equiv \frac{1}{2} \cdot \sum_{p=1}^q \|\eta_p - \eta'_p\|^2 = \frac{1}{2} \cdot \sum_{p=1}^q (\eta_p - \eta'_p, \eta_p - \eta'_p) \quad (\text{B.25})$$

を最小にするように求めればよい。計算結果は次の通りである：

$$\begin{aligned}
 \textcircled{1} \Delta w(j, k, \ell; t) &= -\varepsilon(j, k, \ell; t) \cdot \frac{\partial E(t)}{\partial w(j, k, \ell; t)} \\
 &= -\varepsilon(j, k, \ell; t) \cdot \sum_{p=1}^q \left(\frac{\partial}{\partial w(j, k, \ell; t)} \eta_p, \eta_p - \eta'_p \right) \\
 &= -\varepsilon(j, k, \ell; t) \cdot \sum_{p=1}^q \left(\frac{df_j(u)}{du} \Big|_{u=\sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell; t) \cdot P_{k\ell} \varphi_p + \sum_{k=1}^n w(j, k; t) \cdot P_k \varphi_p + w(j; t) \cdot P_0 \varphi_p} \cdot P_{k\ell} \varphi_p, \eta_p - \eta'_p \right)
 \end{aligned} \tag{B.26}$$

$$\begin{aligned}
 \textcircled{2} \Delta w(j, k; t) &= -\varepsilon(j, k; t) \cdot \frac{\partial E(t)}{\partial w(j, k; t)} \\
 &= -\varepsilon(j, k; t) \cdot \sum_{p=1}^q \left(\frac{\partial}{\partial w(j, k; t)} \eta_p, \eta_p - \eta'_p \right) \\
 &= -\varepsilon(j, k; t) \cdot \sum_{p=1}^q \left(\frac{df_j(u)}{du} \Big|_{u=\sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell; t) \cdot P_{k\ell} \varphi_p + \sum_{k=1}^n w(j, k; t) \cdot P_k \varphi_p + w(j; t) \cdot P_0 \varphi_p} \cdot P_k \varphi_p, \eta_p - \eta'_p \right)
 \end{aligned} \tag{B.27}$$

$$\begin{aligned}
 \textcircled{3} \Delta w(j; t) &= -\varepsilon(j; t) \cdot \frac{\partial E(t)}{\partial w(j; t)} \\
 &= -\varepsilon(j; t) \cdot \sum_{p=1}^q \left(\frac{\partial}{\partial w(j; t)} \eta_p, \eta_p - \eta'_p \right) \\
 &= -\varepsilon(j; t) \cdot \sum_{p=1}^q \left(\frac{df_j(u)}{du} \Big|_{u=\sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell; t) \cdot P_{k\ell} \varphi_p + \sum_{k=1}^n w(j, k; t) \cdot P_k \varphi_p + w(j; t) \cdot P_0 \varphi_p} \cdot P_0 \varphi_p, \eta_p - \eta'_p \right)
 \end{aligned} \tag{B.28}$$

ここに、 $\varepsilon(j, k, \ell; t), \varepsilon(j, k; t), \varepsilon(j; t)$ は十分小さい正の関数（学習率）である。

求める $w(j, k, \ell), w(j, k), w(j)$ は、

$$w(j, k, \ell) = \lim_{t \rightarrow \infty} \Delta w(j, k, \ell; t), w(j, k) = \lim_{t \rightarrow \infty} w(j, k; t), w(j) = \lim_{t \rightarrow \infty} w(j; t) \tag{B.29}$$

と与えられる。

B 3. パターンから抽出される特徴量の組を使った一般抽象実ヒルベルト空間 \mathfrak{H} での 2 次ニューラルネット

$\{\phi_j\}_{j=1 \sim m}$ を \mathfrak{H} の 1 次独立な系とする。

$$w(j, k, \ell), w(j, k), w(j) (j=1 \sim m; k, \ell=1 \sim n) : \text{実定数} \tag{B.30}$$

$u : \Phi \times \{1, 2, \dots, n\} \rightarrow R$: 特徴抽出写像

$$f_j : R \rightarrow R (j=1 \sim m) : \text{ニューロン発火関数} \quad (\text{B.31})$$

を導入して、

$$\varphi_i \in \mathfrak{S}, 1 \sim q \quad (\text{B.32})$$

を入力したとき、得られる第 $i (= 1 \sim q)$ 番目の出力 $\eta_i \in \mathfrak{S}$ は

$$\begin{aligned} \eta_i \\ = \sum_{j=1}^m f_j \left(\sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell) \cdot u(\varphi_i, k) \cdot u(\varphi_i, \ell) + \sum_{k=1}^n w(j, k) \cdot u(\varphi_i, k) + w(j) \right) \cdot \phi_j, i = 1 \sim q \end{aligned} \quad (\text{B.33})$$

と表される。 $u(\varphi, \ell) \in R$ はパターン $\varphi \in \Phi$ から抽出された第 $\ell (= 1, 2, \dots, n)$ 番目の実数値特徴量である。

学習時刻 $t (= 0, 1, 2, \dots)$ に訓練事例

$$\varphi'_i \in \mathfrak{S}, 1 \sim q \quad (\text{B.34})$$

を入力したとき、各理想出力として

$$\eta_i = \eta'_i \in \mathfrak{S}, i = 1 \sim q \quad (\text{B.35})$$

が得られなければならないとき、更新式

$$w(j, k, \ell; t+1) = w(j, k, \ell; t) + \Delta w(j, k, \ell; t) \quad (\text{B.36})$$

$$w(j, k; t+1) = w(j, k; t) + \Delta w(j, k; t) \quad (\text{B.37})$$

$$w(j; t+1) = w(j; t) + \Delta w(j; t) \quad (\text{B.38})$$

での修正分

$$\Delta w(j, k, \ell; t), \Delta w(j, k; t), \Delta w(j; t) \quad (\text{B.39})$$

は次のようになる。

最急降下法に従い、適応誤差関数

$$E(t) \equiv \frac{1}{2} \cdot \sum_{p=1}^q \|\eta_p - \eta'_p\|^2 = \frac{1}{2} \cdot \sum_{p=1}^q (\eta_p - \eta'_p, \eta_p - \eta'_p) \quad (\text{B.40})$$

を最小にするように求めればよい。計算結果は次の通りである：

$$\begin{aligned} \textcircled{1} \Delta w(j, k, \ell; t) &= -\varepsilon(j, k, \ell; t) \cdot \frac{\partial E(t)}{\partial w(j, k, \ell; t)} \\ &= -\varepsilon(j, k, \ell; t) \cdot \sum_{p=1}^q \left(\frac{\partial}{\partial w(j, k, \ell; t)} \eta_p, \eta_p - \eta'_p \right) \end{aligned}$$

$$\begin{aligned}
 &= -\varepsilon(j, k, \ell; t) \cdot \\
 &\sum_{p=1}^q u(\varphi_p, k) \cdot u(\varphi_p, \ell) \cdot \left(\frac{df_j(u)}{du} \right) \Big|_{u=\sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell; t) \cdot u(\varphi_p, k) \cdot u(\varphi_\ell) + \sum_{k=1}^n w(j, k; t) \cdot u(\varphi_p, k) + w(j; t) \cdot P_0 \varphi_p} \cdot \phi_j \\
 &\quad \cdot (\eta_p - \eta'_p)
 \end{aligned} \tag{B.41}$$

$$\begin{aligned}
 \textcircled{2} \Delta w(j, k; t) &= -\varepsilon(j, k; t) \cdot \frac{\partial E(t)}{\partial w(j, k; t)} \\
 &= -\varepsilon(j, k; t) \cdot \sum_{p=1}^q \left(\frac{\partial}{\partial w(j, k; t)} \eta_p, \eta_p - \eta'_p \right) \\
 &= -\varepsilon(j, k; t) \cdot \\
 &\sum_{p=1}^q u(\varphi_p, k) \cdot \left(\frac{df_j(u)}{du} \right) \Big|_{u=\sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell; t) \cdot u(\varphi_p, k) \cdot u(\varphi_\ell) + \sum_{k=1}^n w(j, k; t) \cdot u(\varphi_p, k) + w(j; t) \cdot P_0 \varphi_p} \cdot \phi_j, \eta_p - \eta'_p)
 \end{aligned} \tag{B.42}$$

$$\begin{aligned}
 \textcircled{3} \Delta w(j; t) &= -\varepsilon(j; t) \cdot \frac{\partial E(t)}{\partial w(j; t)} \\
 &= -\varepsilon(j; t) \cdot \sum_{p=1}^q \left(\frac{\partial}{\partial w(j; t)} \eta_p, \eta_p - \eta'_p \right) \\
 &= -\varepsilon(j; t) \cdot \sum_{p=1}^q \left(\frac{\partial}{\partial w(j; t)} \eta_p, \eta_p - \eta'_p \right) \\
 &= -\varepsilon(j, k; t) \cdot \\
 &\sum_{p=1}^q (+1) \cdot \left(\frac{df_j(u)}{du} \right) \Big|_{u=\sum_{k=1}^n \sum_{\ell=1}^n w(j, k, \ell; t) \cdot u(\varphi_p, k) \cdot u(\varphi_\ell) + \sum_{k=1}^n w(j, k; t) \cdot u(\varphi_p, k) + w(j; t) \cdot P_0 \varphi_p} \cdot \phi_j, \eta_p - \eta'_p)
 \end{aligned} \tag{B.43}$$

ここに、 $\varepsilon(j, k, \ell; t)$, $\varepsilon(j, k; t)$, $\varepsilon(j; t)$ は十分小さい正の関数（学習率）である。

求める $w(j, k, \ell)$, $w(j, k)$, $w(j)$ は、

$$w(j, k, \ell) = \lim_{t \rightarrow \infty} \Delta w(j, k, \ell; t), w(j, k) = \lim_{t \rightarrow \infty} w(j, k; t), w(j) = \lim_{t \rightarrow \infty} w(j; t) \tag{B.44}$$

と与えられる。 (付録B終わり)

(著者 鈴木昇一、論文題目 一般抽象実ヒルベルト空間でのニューラルネットのon-line学習、文教大学情報学部・情報研究no. 投稿論文、投稿年月日2006年4月17日 (月))