

多層パーセプトロンにおける内部情報最適化アルゴリズムと汎化能力の解析

殿村 正延[†] 中山 謙二[†]

An Internal Information Optimum Algorithm for Multilayer Perceptrons and Its Generalization Analysis

Masanobu TONOMURA[†] and Kenji NAKAYAMA[†]

あらまし 多層パーセプトロンの学習アルゴリズムとしてバックプロパゲーション (BP) アルゴリズムが知られている。このアルゴリズムをパターン分類問題に適用した場合、学習が収束しても汎化能力は必ずしも最大にならない。本論文では、BP アルゴリズムの作成した超平面を内部情報に基づいて移動させ汎化能力を向上させるアルゴリズムを提案し、基本的性質について解析を行う。更にこのアルゴリズムを多層パーセプトロンに拡張し、計算機シミュレーションにより汎化能力を検証する。本手法は訓練データが疎な場合に有効である。

キーワード ニューラルネットワーク, パーセプトロン, パターン分類, 汎化能力, BP アルゴリズム

1. ま え が き

多層パーセプトロンの学習アルゴリズムとして BP アルゴリズム (Back-propagation algorithm) [1] がよく知られている。これはフィードフォワード形人工ニューラルネットワークで最も共通に適用される訓練法の一つである。パターン分類問題に適用する場合、パターン分類能力と汎化能力の評価基準が異なるため、ネットワーク出力の教師出力に対する誤差を最小化しても汎化能力は必ずしも最大にならない。

汎化能力は中間層ユニット数により大きな影響を受ける。適切なユニット数を決定するために、情報量基準によるモデル選択の方法 [2] や学習中にユニット数を追加していく方法 [3] や削除していく方法 [4], [5] が提案されている。しかし、同一中間層ユニット数をもつネットワークであっても、結合係数初期値、学習係数により異なる内部表現 (Internal Representation) が構築され汎化能力も違ってくる。そのため応用では、これらの組合せ試行を複数回行い、最も良いパフォーマンスが得られる学習結果を採用する。BP 学習が構

築した超平面の枠内で汎化能力を向上させる方法も検討されており、大きく 3 タイプに分類できる。一つ目は、BP 学習の誤差関数に汎化能力を高める項を追加する方法である [6] ~ [8]。二つ目は、教師入力信号に適当な雑音を付加して学習を行う方法である [9], [10]。一つ目はペナルティ係数の値の設定方法や付加項による学習時間の増大の問題が残されており、二つ目はパターン次元が高くなるにつれて雑音を付加した学習パターン数が増大し指数関数的に計算負荷が大きくなる問題が残されている。三つ目は、BP 学習終了後、ある評価関数に従い分離超平面を移動させる 2 段階学習を行うものである [11] ~ [13]。評価関数の基本的な考え方は更に 2 分類でき、一つは入力ベクトルと分離超平面との距離に着目したものの [12] (重みのチューニング評価関数)、[13] (カルバック [14] 型汎化評価関数)、もう一つはその逆数に着目したものの [11] (対数型汎化評価関数) である。本論文で提案するアルゴリズムは、距離の逆数に着目したもので、多層パーセプトロンへの適用は、BP アルゴリズムと混合したハイブリッド形態をとる。なお、文献 [11] ~ [13] の評価関数との基本的性質の相違点は本文中で示す。

内部表現のうち、特に分類能力を獲得する何らかの学習 (ここでは BP 学習) によって既に決定された入力層 - 第 1 中間層間の超平面と入力ベクトルとの距離関

[†] 金沢大学大学院自然科学研究科, 金沢市
Graduate School of Natural Science and Technology,
Kanazawa University, 2-40-20 Kodatsuno, Kanazawa-shi,
920-8667 Japan

係を内部情報 (Internal Information) と呼び、その超平面を最大の汎化能力が得られる位置に移動することを最適化と表現する。このような学習を行うアルゴリズムを内部情報最適化 (Internal-Information-Optimum algorithm: IIO) アルゴリズムと呼ぶことにする。この学習により超平面位置は一意に決定されるため、BP 学習における学習パラメータの組合せ試行回数を軽減できる。

まず、2. で訓練データが疎 (スパース) な場合の最適境界とその汎化能力の評価方法について述べる。次に、3. で BP アルゴリズムの問題点を明らかにする。更に、4. で IIO アルゴリズムを提案し、理論検討を行う。5. ではこのアルゴリズムの多層への適用方法を述べる。最後に、6. で計算機シミュレーションにより最適化能力を検証する。

本論文では、ネットワークを学習するために使用する入出力データを訓練ベクトルと呼び、学習後のネットワークの能力を評価するために使用する入出力データをテストベクトルと呼ぶ。

2. 最適境界と汎化能力の評価方法

2.1 訓練ベクトルがスパースな場合の最適境界

多次元入力空間における各カテゴリーのデータが正規分布に従う場合、無数の訓練ベクトルと十分な隠れユニットが与えられ学習が理想的に進めば、多層パーセプトロンはベイズ識別関数を統計的な意味で近似することが示されている [15], [16]。これはパターン分類問題に多層パーセプトロンを用いる理論的基礎を与える。ベイズ境界は一般に 2 次超曲面を形成し [17]、全カテゴリーの正解認識率は最大となる。

しかし、これは分布を近似する十分な数の訓練ベクトルが収集可能な場合であり、実際の応用では時間的、コスト的に多数の訓練ベクトルが集められない場合がある。このとき汎化能力について議論する必要がある。この場合、カテゴリーを形成する訓練ベクトルが保証すべき領域が縮退し、カテゴリー間は一般にスパースになるため、BP 学習はベイズ境界を近似しない。そのため、BP 学習によって構築される入力層 - 第 1 中間層間の超平面で他カテゴリーと完全分離可能となり、出力側から見た境界はこの超平面の位置でほとんど規定される。この超平面を分離超平面と呼ぶ [12]。この位置により汎化能力は大きな影響を受ける。入力ベクトルは非線形変換により、第 1 中間層におけるその変換座標は超立方体の頂点近傍に写像される。その

ため、第 1 中間層 - 第 2 中間層間の超平面の位置ずれによる影響は小さい。上層も同様となる。よって、入力層 - 第 1 中間層間の超平面位置が汎化能力に最も大きな影響を与える。

本論文における最適境界とは、分離されるカテゴリーの正解認識率がどちらに偏ることなく等しくなる位置の分離超平面と定義する。この境界はベイズ境界のように正解認識率は最大にならないが、超平面で分離する場合の正解認識率を均等にしている条件を付けた中で最大にするという意味で汎化能力を向上させる。この位置はデータ分布 (発生確率と分布密度) により異なるが、各カテゴリーの訓練ベクトルの発生確率が等しい仮定のもとで次の二つのケースを想定し議論する。

<ケース 1 >

各訓練ベクトルが互いの干渉を無視できるくらい重なりのある小さな正規分布を構成する (混合ガウス分布をなす) 場合、最適境界は各カテゴリー間の最も近距離の訓練ベクトルを結ぶ線分の中点を通り直交する。

<ケース 2 >

各カテゴリーを構成する訓練ベクトルが全体として大きな正規分布をなす場合、最適境界は各カテゴリーの平均ベクトルを結ぶ線分を標準偏差比で内分する点を通り直交する。

ただし、スパースな訓練ベクトルから各カテゴリーの分布を正確に推定することは不可能であるので、簡単のため、共分散行列の固有値はすべて等しい (分布は超球状に発生) と仮定する。

2.2 汎化能力の評価方法

汎化能力の評価方法は 決定論的評価と 集合平均的評価がある。は訓練ベクトルを密度の高いところからの標本値と仮定して汎化性を評価し、は訓練ベクトルを分布に従う標本値とし、その標本値を変えながら学習を繰り返し、得られる集合平均で汎化性を評価する。汎化性の議論としては が一般的であるが、議論を簡潔に行うため、まず、4.1 ~ 4.8 で IIO アルゴリズムが に対して汎化能力を向上させることを示す。次に、4.9 で に対しても汎化能力を向上させることを示す。

3. BP アルゴリズムの問題点

カテゴリー間が疎になる場合、実際の BP 学習では必ずしも 2.1 で述べた最適境界位置に分離超平面は作成されず、一意に定まらない。この原因は、結合係数初期値、学習係数とシグモイド関数の傾きのスケー

ルの差により相対的な結合係数の変化量が変わることにある。傾きを適切な値にすれば、シグモイド関数の線形部分を使用した結合係数の修正が行われ、超平面を近接カテゴリー群から遠ざけようとする性質により、妥当な学習回数でカテゴリー間の中央付近にくる。しかし、傾きが小さいと学習回数が不必要に大きくなる。逆に大きいと1回の更新で修正される相対的な結合係数の変化量が増加し収束回数は小さくなるものの、シグモイド関数の線形部分を使用せず、いきなり結合荷重が飽和領域に達し学習が停止する。このとき、どちらかのカテゴリーに偏った位置に分離超平面がくる。一般に、適切な値はネットワークの規模、扱うパターン群の統計的性質に依存するため、事前に知ることはできない。また、大規模ネットワークでは選んだ値がネットワークのすべての部分に対して適切である保証はない。更に、結合係数初期値により形成される超平面の組合せが異なり汎化能力に大きな影響を与える。このため、最適解を得るにはそれらの組合せ試行を必要とし、費やす時間は膨大となる [18]。

4. 内部情報最適化 (IIO) アルゴリズム

4.1 基本ネットワーク

図1の n 入力, 1 出力の単純パーセプトロンの入力関係を次式に示す。ただし, x_0 はオフセットで常に 1 が与えられる。また, $J \equiv [0, 1]$ とする。 w^t は重みベクトル w の転置を表す。式 (4) の u はシグモイド関数の傾きを制御するパラメータである。

$$x = [x_0, x_1, \dots, x_n]^t, x_0 = 1, x_i \in J \quad (1)$$

$$w = [w_0, w_1, \dots, w_n]^t \quad (2)$$

$$y = f(w^t x) \quad (3)$$

$$f(x) = 1/(1 + e^{-x/u}) \quad (4)$$

4.2 雑音許容半径

図1のネットワークを用いて、訓練ベクトル

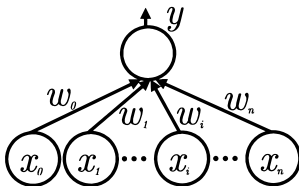


図1 単純パーセプトロン
Fig.1 Single-layer perceptron.

$x^1 \in Class1, x^2 \in Class2$ が分離できたときの重みベクトル w と訓練ベクトル x^1, x^2 からなる多次元ベクトル空間の概念図を図2に示す。テストベクトルが訓練ベクトルを中心として共分散行列の固有値 σ^2 がすべて等しい超球状に発生する場合、入力したテストベクトルが正しく識別されるためには、超球の半径 (例えば 3σ) が超平面を超えないことが必要となる。この超球の半径を雑音許容半径と呼ぶことにする。雑音許容半径は、訓練ベクトルの重みベクトル上への正射影のスカラで表せる。重みベクトルと訓練ベクトルのなす角を θ で表すと、図2から訓練ベクトル x^1 の雑音許容半径 r_1 , 訓練ベクトル x^2 の雑音許容半径 r_2 は次式となる。ただし, $\|x\|$ はベクトル x のユークリッドノルムを表す。

$$r_1 = \frac{|w^t x^1|}{\|w\|} = \|x^1\| |\cos \theta_1| \quad (5)$$

$$r_2 = \frac{|w^t x^2|}{\|w\|} = \|x^2\| |\cos \theta_2| \quad (6)$$

雑音許容半径 $r_1 = r_2$ のとき、訓練ベクトル x^1, x^2 の雑音許容量がどちらに偏ることなく均等になる。

4.3 IIO 評価関数の定義

以下に IIO 評価関数 E_{iio} を定義する。

$$E_{iio} \equiv \frac{1}{2|X|} \sum_{x \in X} \left(\beta(x) \frac{\|w\|}{w^t x} \right)^2 \quad (7)$$

式 (7) で $|X|$ は J^n の有限部分集合 X に属するパターンの個数を表す。また, $\beta(x)$ は各訓練ベクトルの統計的性質を評価関数に反映させる重み関数で、訓練ベクトルのノルムの違いから発生する超平面位置の偏り

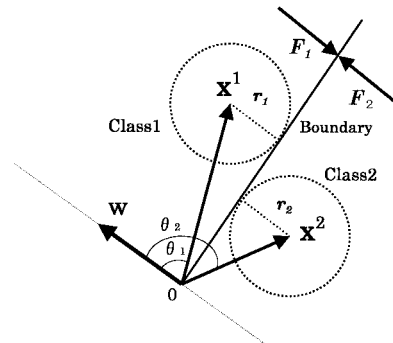


図2 ベクトル空間の概念図
Fig.2 Vector space.

を補正するのにも使用する．これについては，本論文では扱わないため $\beta = 1$ として議論を行う．

4.4 IIO 評価関数の解析

IIO 評価関数の意味について考える．図 2 において，訓練ベクトルが超平面に近いとき雑音許容半径は小さく，離れているとき大きくなる．一方，半径の逆数はそれとは逆の動きとなる．この半径の逆数を訓練ベクトルが超平面を押す力 F と見立てると，訓練ベクトル x^1 が超平面を押す力 F_1 ，訓練ベクトル x^2 が超平面を押す力 F_2 は，それぞれ式 (5)，(6) を用いて

$$F_1 = -1/(\|x^1\| \cos \theta_1) \quad (8)$$

$$F_2 = -1/(\|x^2\| \cos \theta_2) \quad (9)$$

のように表せる．雑音許容半径と符号を反転させるのは，超平面を押す力に見立てるためである．超平面を押す力の 2 乗 F_1^2, F_2^2 及び，その合成力 $F_1^2 + F_2^2 (= 4E_{iio})$ を図 3 に示す．ただし，解析を容易にするため， $\|x^1\| = \|x^2\|$ ， $\theta_2 = \theta_1 + \pi/2$ (訓練ベクトル x^1, x^2 が互いに無相関)， $0 \leq \theta < 2\pi$ の範囲とする．図から，IIO 評価関数 E_{iio} は，4 個の最小値をもつ． $F_1^2 + F_2^2$ を θ_1 で微分し，最小値をとる θ_1 を求めると， $\theta_1 = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$ となる．このとき，式 (5)，(6) から， $r_1 = r_2$ となることがわかる．つまり，IIO 評価関数は最小値をとるとき，雑音許容半径の大きさが等しくなる関数である．

次に，最小値をとるときベクトル空間の状態を図 4 に示す．矢印は重みベクトルを，黒丸は分離したい二つの訓練ベクトルを，太い直線は超平面を表している．図から， $\theta_1 = \pi/4, 5\pi/4$ のとき，超平面は訓練ベクトルの間にあり，ネットワークは分類能力をもつ．しかし， $\theta_1 = 3\pi/4, 7\pi/4$ のとき，超平面は訓練ベクトルの間になく分類能力がない．このことから，超平面が分離したい訓練ベクトルの間に最初からあるとき，IIO 評価関数の最小値探索により雑音許容半径の大きさを等しくすることができるが，間がない場合は，最適状態と全く違う位置で安定状態になる．これは，分類能力を構築する学習とともに使用しなければならないことを示している．

4.5 IIO アルゴリズムの導出

IIO 評価関数 E_{iio} は，重みベクトル w の関数であり，最急降下法を用いて E_{iio} を最小にする重みベクトルを求めることができる．この重みベクトルを時間の関数とし，次の力学系を考える．

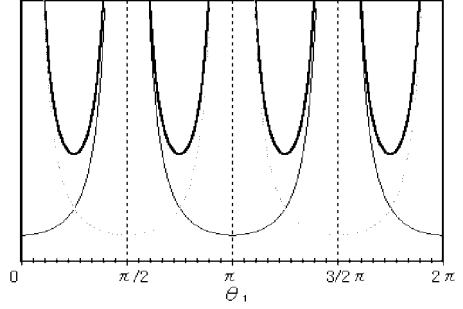


図 3 IIO 評価関数 E_{iio} の描く曲線
Fig. 3 Illustrating the curve of E_{iio} .

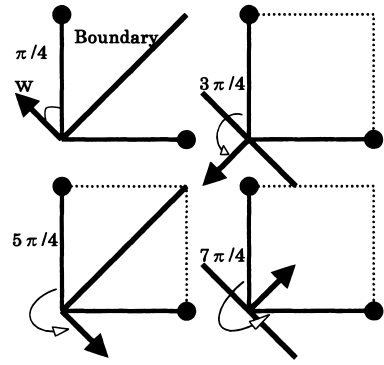


図 4 IIO 評価関数 E_{iio} が最小値をとるときの超平面
Fig. 4 Hyperplane with the minimum of E_{iio} .

$$\frac{dw_i}{dt} = -\frac{\partial E_{iio}}{\partial w_i} \quad (i=0, \dots, n) \quad (10)$$

[定理]

上記力学系において，IIO 評価関数 E_{iio} は時間 t に関して非増加である．

$$\frac{dE_{iio}}{dt} \leq 0 \quad (11)$$

(証明)

$$\frac{dE_{iio}}{dt} = \sum_{i=0}^n \frac{\partial E_{iio}}{\partial w_i} \frac{dw_i}{dt} = -\sum_{i=0}^n \left(\frac{\partial E_{iio}}{\partial w_i} \right)^2 \leq 0$$

よって，評価関数は時間 t に関して非増加である．また，評価関数が不変になるのは $\partial E_{iio} / \partial w_i = 0$ が成り立つときで，このときに限る．

(証明終)

E_{iio} の w による偏微分を求めると，次式を得る．また，式の導出は付録で行う．

$$\frac{\partial E_{iio}}{\partial w} = \frac{-1}{|X|} \sum_{x \in X} \frac{\beta^2(x)}{(w^t x)^2} \left(\frac{\|w\|^2 x}{w^t x} - w \right) \quad (12)$$

以上から、HIO アルゴリズムは次式となる。

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \Delta \mathbf{w} |_{\mathbf{w}=\mathbf{w}(n)} \quad (13)$$

$$\Delta \mathbf{w} = -\mu \frac{\partial E_{iio}}{\partial \mathbf{w}} \quad (14)$$

ただし、 μ は $0 < \mu \ll 1$ となる学習係数である。

式 (12) からわかるように、このアルゴリズムは、学習する全訓練ベクトルを提示後に 1 回修正を行い、超平面を押しよす力の不均衡量を結合係数の修正量としている。つまり、このアルゴリズムでは、個別の訓練ベクトル提示ごとに修正を行う逐次修正法（オンライン修正法）は不可能である。更新は一括修正法（バッチ修正法）によりのみ行われる。以下に、学習プロセスを示す。

[学習プロセス]

- step1. 重みベクトルの成分ごとの修正項 Δw_i を初期化する。
- step2. 訓練ベクトルの提示ごとに、修正量を計算し Δw_i に加算する。
- step3. 全訓練ベクトル提示後、修正量の総和が誤差上限値を下回った場合、収束と判定し終了する。
- step4. Δw_i を w_i に加算する。
- step5. step1 に戻る。

4.6 HIO アルゴリズムの領域形成の性質

図 5 (a) は、2 個の訓練ベクトル x^1, x^2 が分離超平面で別クラスに分類されている様子を表している。に超平面があるとき、2 個の訓練ベクトルから超平面までの距離はともに等しいが、雑音許容半径が小さい。パターン空間において最大の雑音許容半径を得る

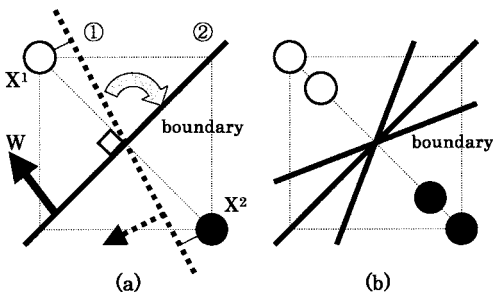


図 5 (a) HIO アルゴリズムによる超平面の移動 (b) 重みのチューニング評価関数、カルバック型汎化評価関数の超平面位置

Fig. 5 (a) A movement of hyperplane by HIO algorithm. (b) Hyperplane position of weight tuning estimate function or Kulback-Leibler measure.

のは、超平面が θ にあるときである。つまり、2 個の訓練ベクトルを通る直線と超平面が垂直になるときで、このとき、直線は重みベクトルと平行になる。HIO アルゴリズムは、2 個の訓練ベクトルの雑音許容半径を等しくするだけではなく、BP アルゴリズムが θ に超平面を作成した場合でも θ に超平面を移動する。このことを以下に述べる。

訓練ベクトルが 1 個の場合、式 (7) の HIO 評価関数を訓練ベクトルと重みベクトルのなす角 θ で微分すると

$$\dot{E}_{iio} = \tan \theta E_{iio} \quad (15)$$

となる。式 (15) が 0 となるのは、 $\theta = n\pi$ (ただし、 n は整数) のときである。つまり、超平面を訓練ベクトルに対して直交するように移動する性質があり、結果的に HIO 評価関数は雑音許容半径を最大にする。2 パターン 2 分類の場合も超平面を 2 個の訓練ベクトルに直交するように移動する。ただし、両方の訓練ベクトルに対して同時に直交することはできないため、雑音許容半径がなるべく大きくなる均衡点で安定状態となる。このことを式で示す。学習の修正量は、式 (14) より

$$\Delta \mathbf{w} = \left(\frac{\|\mathbf{w}\|^2 \mathbf{x}^1}{(\mathbf{w}^t \mathbf{x}^1)^3} - \frac{\mathbf{w}}{(\mathbf{w}^t \mathbf{x}^1)^2} \right) + \left(\frac{\|\mathbf{w}\|^2 \mathbf{x}^2}{(\mathbf{w}^t \mathbf{x}^2)^3} - \frac{\mathbf{w}}{(\mathbf{w}^t \mathbf{x}^2)^2} \right) \quad (16)$$

となる。ただし、簡単のため μ 及び $|X|$ は省略している。 $\Delta \mathbf{w} = 0$ として学習が終了するときの \mathbf{w} を式 (16) から求めると次式となる。

$$\mathbf{w} = \frac{\|\mathbf{w}\|^2 ((\mathbf{w}^t \mathbf{x}^2)^3 \mathbf{x}^1 + (\mathbf{w}^t \mathbf{x}^1)^3 \mathbf{x}^2)}{(\mathbf{w}^t \mathbf{x}^1)^3 (\mathbf{w}^t \mathbf{x}^2) + (\mathbf{w}^t \mathbf{x}^2)^3 (\mathbf{w}^t \mathbf{x}^1)} \quad (17)$$

雑音許容半径の大きさは等しくなる ($r_1 = r_2$) ので

$$\mathbf{w}^t \mathbf{x}^1 = -\mathbf{w}^t \mathbf{x}^2 \quad (18)$$

となる。これを、式 (17) に代入すると

$$\mathbf{w} = \frac{\|\mathbf{w}\|^2}{2(\mathbf{w}^t \mathbf{x}^2)} (\mathbf{x}^2 - \mathbf{x}^1) \quad (19)$$

となる。式 (19) から、HIO アルゴリズムによる学習の結果、 \mathbf{w} が訓練ベクトル x^1, x^2 を結ぶ直線と平行となることがわかる。

実際は、4.8 で述べるように、偏りによって雑音許

容半径は正確に等しくならないため、 w が訓練ベクトル x^1, x^2 を結ぶ直線と正確に平行にはならない。しかし、ここで解析したアルゴリズムの性質の一般性を失うものではない。

2 パターン 2 分類の場合 (図 (a)) や分類したいパターンが同一直線上にある場合 (図 (b)), カルバック型汎化関数では図 (b) に示すように各パターンからの距離は等しくなるが一意に定まらない。つまり、雑音許容半径を最大にする保証はない。重みのチューニング評価関数はカルバック・ライブラー情報を非線形変換前 (入力側) に使用したものであり、同様の性質がある。対数型汎化評価関数は HIO 評価関数と同様一意に定まる。なお、このような特殊なケースは、応用において訓練ベクトルが十分集められない場合や複数の観測ベクトルを少数の代表ベクトルに代えて学習を行う場合に発生することが考えられる。

4.7 HIO アルゴリズムの修正

2 次元 2 パターンで $\|x^1\| = \|x^2\|, \theta_2 = \theta_1 + \pi/2$ の場合、区間 $[0, \pi/2]$ における式 (14) の修正量 Δw を図 6(a) に示す。図から θ_1 が 0 あるいは $\pi/2$ に近づくにつれて、幾何学的に見れば、超平面が訓練ベクトルに近づくにつれて、修正量は指数関数的に増大することがわかる。もし、BP 学習により作成された超平面が訓練ベクトルの近傍にあると、式 (14) の学習係数 μ をかなり小さな値としても、 Δw は学習が発散するくらい大きくなる可能性がある。式 (14) を用いたアルゴリズムでは、定数 μ の使用は収束を保証しない。仮に、 $\mu \ll 1$ として発散の可能性を小さくしても、なお収束の保証はなく、むしろ収束のスピードを奪うことになる。

超平面が訓練ベクトルの近傍にあるときでも収束を保証できるように、修正量のある範囲に拘束する飽和形関数を使用する。ここでは、修正量を傾きで制御できる \tanh 関数を使用する。次式のように Δw_{new} を新しく定義する。

$$\Delta w_{new} = \mu \tanh\left(-\frac{\partial E_{iio}}{\partial w} / T\right) \quad (20)$$

ここで、 μ は学習係数、 T は \tanh 関数の傾きを制御するパラメータである。 Δw_{new} を図 6 の (b) に示す。今後は式 (14) の Δw を Δw_{new} に置き換えて議論を進める。

< 収束の保証 >

分離したいカテゴリー間の最短距離を超えない範囲で学習係数 μ を決定することで、収束を保証する

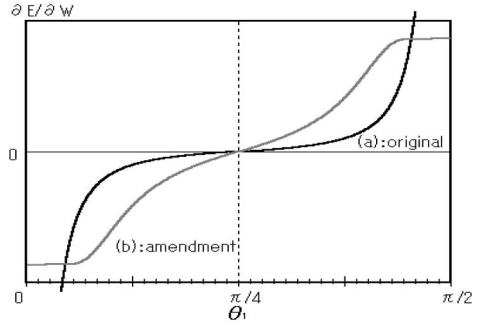


図 6 HIO アルゴリズムの区間 $[0, \pi/2]$ における修正量 Δw
Fig. 6 Illustration of Δw for interval $[0, \pi/2]$.

ことができる。その成分間の距離を $d[x_i^p, x_i^q]$, ($i = 1, \dots, n$), $p, q \in X$ とすると、次式の条件が成り立つ必要がある。

$$|\max[\Delta w]| = \mu < \min\{d[x_i^p, x_i^q]\} \quad (21)$$

実際は、振動を抑えるために、式 (21) の条件を満たす十分小さな値を使用する。

4.8 ベクトルノルムと分布密度の違いによる超平面位置の偏り

HIO アルゴリズムを用いてノルムの等しい 2 個の訓練ベクトルの雑音許容半径を等しくできることを示した。実際の応用では、分離したい訓練ベクトルのノルムや各カテゴリーの分布密度が違うことが考えられる。このことが、HIO アルゴリズムの動作にどのような影響を与えるか述べ、更に、他の評価関数との比較を行う。

まず、ノルムが違う 2 個の訓練ベクトルの分類を考える。訓練ベクトル x^1, x^2 のなす角は、 $\pi/2$ とする。このとき、式 (7) をノルムとなす角で表示すると

$$E_{iio} = \frac{1}{\|x^1\|^2 \cos^2 \theta_1} + \frac{1}{\|x^2\|^2 \sin^2 \theta_1} \quad (22)$$

となる。式 (22) を θ_1 で微分し、0 となるときの重みベクトルと訓練ベクトル x^1 のなす角の正接を求めると、次式となる。

$$\tan \theta_1 = \left(\frac{\|x^1\|}{\|x^2\|}\right)^{\frac{1}{2}} \quad (23)$$

雑音許容半径 r_1, r_2 の比を求めると

$$\frac{r_2}{r_1} = \frac{\|x^2\| \sin \theta_1}{\|x^1\| \cos \theta_1} = \frac{\|x^2\|}{\|x^1\|} \tan \theta_1 \quad (24)$$

となる．式 (24) に式 (23) を代入し，次式を得る．

$$\frac{r_2}{r_1} = \left(\frac{\|x^2\|}{\|x^1\|} \right)^{\frac{1}{2}} \quad (25)$$

式 (25) は，ノルムが小さい訓練ベクトル側に超平面が寄る傾向があることを示している．この偏りは理想的にはないのが望ましい．パターン空間において高精度の汎化能力が期待されるのは一般に近接カテゴリ間であり，近接しているほどノルム比は 1 に近づくため，ノルムの違いによる偏りの影響は小さくなる．一方，上記と同一仮定のもとで対数型汎化関数の雑音許容半径比を求めると

$$\frac{r_2}{r_1} = \frac{\|x^2\|}{\|x^1\|} \quad (26)$$

となる．式 (26) は式 (25) と比較して，ノルム比による偏りが 2 乗倍大きくなる．よって，両者を比較した場合，IIO 評価関数の方が優れているといえる．なお，カルバック型汎化関数と重みのチューニング評価関数はノルム比の影響は受けない．

次に，各カテゴリの訓練ベクトルの発生確率は等しい仮定のもとで，分布密度の違う二つのカテゴリの分類を考える．図 7 はその簡単な例である（パターン次元：2 次元）．黒点は訓練ベクトルを，直線は各評価関数に従って移動した分離超平面を表す．a-IIO 評価関数，b-対数型汎化関数，c-カルバック型汎化関数，d-重みのチューニング評価関数．更に，点線は 2.1 のケース 1 の最適境界を，点線はケース 2 の最適境界を示している．ケース 1 の場合，IIO 評価関数を用いたものが最も最適境界に近い．ケース 2 の場合，他の評価関数に比べ偏りが小さいものの十分とは

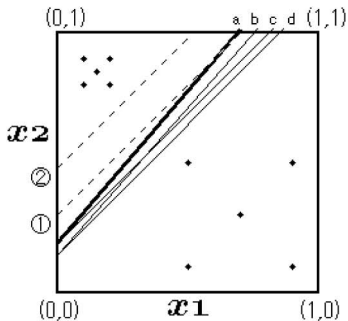


図 7 異なる広がりのカテゴリを分離する超平面位置の比較

Fig. 7 Compare of hyperplane position between two different expand categories.

いえない^(注1)．よって，IIO 評価関数は，ケース 1 の場合に大きな汎化能力を実現できることがわかる．重みのチューニング評価関数を用いた場合，広がり大きなカテゴリ側に最も寄っている．これは各カテゴリの重心間の距離を均等にする性質により，結果的に広がり大きなカテゴリ側に寄るためである．カルバック型汎化関数にも同様の性質があるが，シグモイド関数による非線形変換を受けることにより広がり大きなカテゴリからやや離れる．対数型汎化関数は IIO 評価関数と同様に雑音許容半径の逆数に着目した関数であるが，広がり大きなカテゴリ側に寄る．これは対数をとることに原因がある．つまり，入力ベクトルが超平面に近づくと，雑音許容半径の逆数は指数関数的に増大するが，その対数をとることで IIO 評価関数に比べて相対的に値が小さくなるためである．IIO 評価関数は雑音許容半径の逆数の総和をとるため，超平面から遠い入力ベクトルの影響を受けにくく，超平面に最も近い訓練ベクトルによりほとんど位置が決まるため中央寄りとなる^(注2)．

以上の各評価関数の比較から，IIO 評価関数はベクトルノルムの違いによる偏りはあるが，分布密度の違いによる偏りが他の評価関数と比較して小さいという点，4.6 で示したように超平面が分離したいパターンに直交するように一意に定まる点で有用性が高い．また，2.1 のケース 1 で特に汎化能力を向上させる評価関数といえる．

4.9 統計的な意味での汎化能力

本節では，分布に従う標本値を訓練ベクトルとして BP 学習と IIO 学習（BP 学習後に行う）を行い，統計的な意味での汎化性を比較する．2 次元 2 クラス分類を考える．2 クラスのカテゴリの中心座標は $x^1 = (0.3, 0.7)$ ， $x^2 = (0.7, 0.3)$ ，同一標準偏差 $\sigma = 0.1$ とする．テストベクトルは分布の中心から σ を使用して 1000 個発生させる．訓練ベクトルを変えながら 100 回の試行を行い，平均正解認識率と各クラスの正解認識率のばらつき（分散）を求める．その結果を図 8 に示す．図の横軸は訓練ベクトルの数を示す．数が増えるに従って認識率が向上し，ばらつきが小さくなる．

(注1): ケース 2 の最適境界をより近似するには，前処理で入力空間の分布からカテゴリの中心と分布の広がり（標準偏差）を適当なクラスタリング法により求め，これを代表ベクトルとして学習を行う（分布の広がりには式 (7) の β に重み付けする）方法があるが，本論文では扱わない．

(注2): 文献 [11] の対数型汎化評価関数と IIO 評価関数の性質の違いは，導関数でみれば式 (A.4) の $(\|w \cdot x\|/w \cdot x)^2$ の有無による．

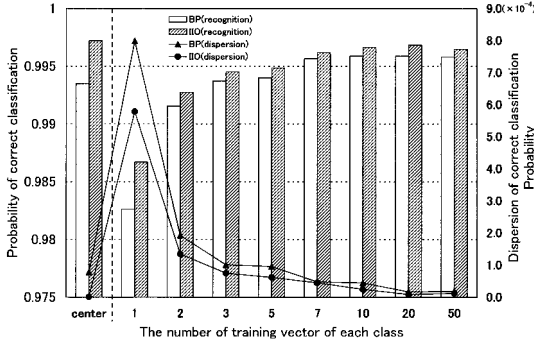


図 8 統計的な意味での汎化能力の比較

Fig. 8 Compare of statistical generalization ability of BP learning and IIO learning.

どの場合も IIO 学習が BP 学習を上回っている。これから統計的な意味においても IIO 学習により汎化能力が向上することがわかる。横軸の最も左の center は、訓練ベクトルを中心座標に固定した場合である。分布に従って発生させた 1 個の訓練ベクトルの場合に比べ正解認識率 (0.99721) が高い。これは分布の中心付近から標本できるほど認識率が高くなることを示している。また、この例は同一標準偏差で行っているためベイズ境界は超平面となるが、IIO 学習はその最適認識率 (理論値: 0.99766) をよく近似している。

5. 多層パーセプトロンへの適用方法

まず、BP アルゴリズムについて簡単に述べ、次に BP アルゴリズムと IIO アルゴリズムを混合したハイブリッドアルゴリズムを提案する。

5.1 BP アルゴリズム

n_0 入力, n_N 出力の $N + 1$ 層パーセプトロンにおける BP アルゴリズムの誤差関数を

$$E_{bp} \equiv \frac{1}{2|X|} \sum_{\mathbf{x} \in X} \alpha(\mathbf{x}) \|\mathbf{t}(\mathbf{x}) - \mathbf{h}(\mathbf{x})\|^2 \quad (27)$$

とする。 $\mathbf{h}(\mathbf{x})$ は訓練ベクトル \mathbf{x} に対する出力層の出力ベクトル, $\mathbf{t}(\mathbf{x})$ はその出力教師ベクトル, α は J^{n_0} で定義される \mathbf{h} と \mathbf{t} の近さの重要さを定める関数である。 p 層第 i ユニットと $p+1$ 層第 j ユニット間の結合係数 w_{pij} ($p=1, \dots, N, i=0, \dots, n_{p-1}, j=1, \dots, n_p$) の修正は

$$w_{pij}(n+1) = w_{pij}(n) + \Delta w_{pij}^{bp} |_{\mathbf{w}=\mathbf{w}(n)} \quad (28)$$

$$\Delta w_{pij}^{bp} = \eta \frac{1}{|X|} \sum_{\mathbf{x} \in X} \delta_{pj}(\mathbf{x}) g_{p-1,i}(\mathbf{x}) \quad (29)$$

により行われる。ただし、 η は $0 < \eta < 1$ となる学習係数である。

$$\delta_{pj}(\mathbf{x}) = \begin{cases} \alpha(\mathbf{x}) [t_q(\mathbf{x}) - h_q(\mathbf{x})] \dot{g}_{pq}(\mathbf{x}) & (p=N) \\ \left\{ \sum_{r=1}^{n_{p+1}} \delta_{p+1,r}(\mathbf{x}) w_{p+1,qr} \right\} \dot{g}_{pq}(\mathbf{x}) & (1 \leq p < N) \end{cases}$$

$$\dot{g}_{pq}(\mathbf{x}) = g_{pq}(\mathbf{x}) (1 - g_{pq}(\mathbf{x}))$$

$$g_{pq}(\mathbf{x}) = f \left(\sum_{i=0}^{n_{p-1}} w_{piq} g_{p-1,i}(\mathbf{x}) \right)$$

$$g_{p0}(\mathbf{x}) = 1 \quad (p=0, 1, \dots, N)$$

$$g_{0q}(\mathbf{x}) = x_q \quad (q=1, \dots, n_0)$$

5.2 ハイブリッドアルゴリズム

2.1 で示したように、訓練ベクトルがスパースな場合において汎化能力を向上させるには、カテゴリーの広がりの影響を最も受ける入力層 - 第 1 中間層間の分離超平面を適切な位置に移動させることで大きな効果が得られる。そこで、式 (30) に示すように BP アルゴリズムの入力層 - 第 1 中間層間の結合係数の修正に IIO アルゴリズムの修正を付加したハイブリッドアルゴリズムを提案する。

$$\Delta w_{pij} = \begin{cases} \Delta w_{pij}^{bp} + \Delta w_{pij}^{iio} & (p=1) \\ \Delta w_{pij}^{bp} & (2 \leq p < N) \end{cases} \quad (30)$$

$$\Delta w_{1ij}^{iio} = \mu \tanh \left(\frac{1}{|X|} \sum_{\mathbf{x} \in X} \frac{\beta^2(\mathbf{x})}{(\mathbf{w}_{lj}^t \mathbf{x})^2} \left(\frac{\|\mathbf{w}_{lj}\|^2}{\mathbf{w}_{lj}^t \mathbf{x}} x_i - w_{ij} \right) \right) / T \quad (31)$$

ただし、式 (31) は式 (20) を成分表示したもので、 \mathbf{w}_{1j} は入力層から第 1 中間層の j 番目ユニットへの重みベクトルである。 μ は式 (21) を満たすと同時に、BP 学習による分類能力の構築に影響を与えない十分小さい値 ($0 < \mu/\eta \ll 1$) であることが必要である。形式的には IIO アルゴリズムの全層への適用も可能であるが、計算量と効果の観点から入力層 - 第 1 中間層間の適用が妥当である。

ハイブリッドアルゴリズムによる入力層 - 第 1 中間層間の結合係数 w_{1ij} の修正量の変化について考える。図 9 に示すように、実矢印方向から Δw_{1ij}^{bp} が逆伝搬され、点線矢印方向から Δw_{1ij}^{iio} が順伝搬される。両修正量の和 Δw_{1ij} が結合係数の更新に使用される。図 10 は、 Δw_{1ij}^{iio} (太実線) と Δw_{1ij}^{bp} (太点線) の相対関係を示したものである。分類能力が未構築の段階では、 μ は η に比し非常に小さいため、 Δw_{1ij}^{iio} はその符号にかかわらず BP 学習に影響を与えない (図のケース

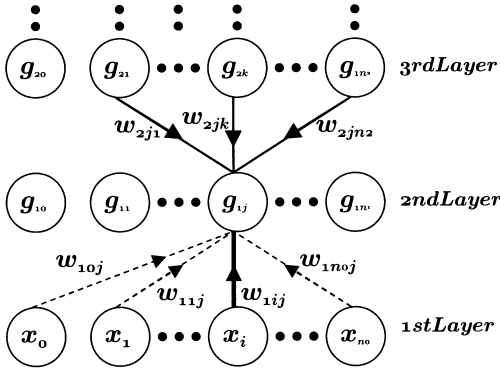


図9 結合係数 w_{1ij} の修正量の伝播
Fig. 9 The spread of the amount of modification of the combination coefficient w_{1ij} .

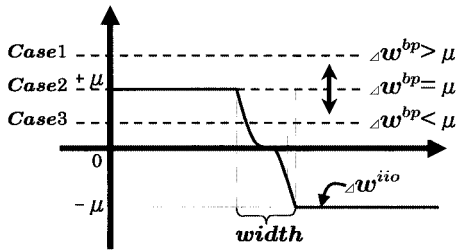


図10 修正量 Δw^{bp} と Δw^{iio} の関係
Fig. 10 Relations between the correction Δw^{bp} and Δw^{iio} .

1). ただし、分類能力の構築が進むに従い Δw_{1ij}^{bp} は小さくなり、 Δw_{1ij}^{bp} と Δw_{1ij}^{iio} が同符号のときは学習を加速し、異符号のときは減速するように働く。ケース2や3の場合、 Δw_{1ij}^{bp} と Δw_{1ij}^{iio} が異符号のとき図の細点線で示す位置で $\Delta w_{1ij} = 0$ となり、学習が停止してしまう。理想的な分離超平面は $\Delta w_{1ij}^{iio} = 0$ の位置であるから、ずれが生じることがわかる。これを最小限に抑えるためには、式(31)の T を小さくして \tanh を急峻にすることで、 $|\Delta w_{1ij}^{iio}| < \mu$ の幅を小さくすればよい。実際には Δw_{1ij}^{bp} は分類能力が構築された後、上層から逆伝搬される修正量は一定の符号で単調減少するため、分離超平面は最適位置に漸近する。また、逆に分類能力の構築が不十分でBP学習が阻害される場合は、逆伝搬量が増加して $|\Delta w_{1ij}^{bp} + \Delta w_{1ij}^{iio}| > 0$ となり学習が促進する。よって、学習回数に応じた精度で内部情報を最適化できる。

入力層 - 第1中間層間の超平面で他カテゴリーと完全分離可能な場合は、BP学習終了後IIO学習を行う

2段階学習でも汎化能力向上は可能である。しかし、実際の応用においては分離したいカテゴリー間が疎と密(カテゴリー間にオーバーラップあり)混在したパターン空間を想定する必要がある。カテゴリー間が密な場合、BP学習で構築されるベイズ境界を近似した2次超曲面は、単に超平面と訓練ベクトルの距離関係ではとらえられない。そのため、2段階学習で超平面を移動させると2次超曲面に悪影響を与える可能性がある。ハイブリッドアルゴリズムは、分離したいパターン間が疎な箇所では Δw_{1ij}^{iio} の働きにより超平面が移動し汎化能力が向上する。密な部分では、仮に Δw_{1ij}^{iio} がBP学習を阻害する方向に働いても $|\Delta w_{1ij}^{bp} + \Delta w_{1ij}^{iio}| > 0$ となるように逆伝搬量が増えるため、その働きが無視され2次超曲面に影響を与えない。

6. 計算機シミュレーション

6.1 2クラス分類問題

4層パーセプトロン(入力層:3ユニット,第1中間層:10ユニット,第2中間層:5ユニット,出力層:1ユニット。ただし、出力層以外はオフセットユニットを含む)を用いてBP学習とハイブリッド学習を同一の結合係数初期値を用いて比較実験を行う。2クラスの教師出力は0.99,0.01を使用する。各例題に応じた標準偏差の雑音を加えたテストベクトルを各訓練ベクトルごとに1000個発生させ、正解認識率を計算する。また、教師出力 ≥ 0.5 に対して出力 ≥ 0.5 , 教師出力 < 0.5 に対して出力 < 0.5 を正解とする。その正解認識率の平均を求め、平均からの各パターンの正解認識率のばらつき(分散)を求める。結果は表1にまとめて示す。正解領域の表示において、出力 ≥ 0.5 となる領域を黒色で、出力 < 0.5 となる領域を白色で示す。また、入力層 - 第1中間層間の超平面を灰色直線で示す。座標は図7と同一である。なお、パターン次元を2次元にするのは、分離超平面の移動と正解領域を視覚的に観察するためである。

6.1.1 例題1

まず、最もシンプルな各カテゴリーの訓練ベクトルが1個の場合を示す。その一例を図11に示す。×は訓練ベクトルを表し、色の違い(黒色と灰色)はクラスの違いを表す。各パターンから等距離の補助線を灰色点線で示す。ハイブリッド学習により異クラスの訓練ベクトル間の距離が等距離になるように分離超平面が移動し、正解領域も妥当な形状になっている。表1から、平均正解認識率が向上し、ばらつきが小さく

表 1 各例題における BP 学習とハイブリッド学習の正解認識率とばらつきと比較

Table 1 Compare of correct classification probability and dispersion of BP learning and hybrid learning in each example.

	BP Learning		Hybrid Learning	
	mean	dispersion	mean	dispersion
Example 1	0.953	81.1×10^{-5}	0.993	2.5×10^{-5}
Example 2	0.939	197.7×10^{-5}	0.949	106.6×10^{-5}
Example 3	0.950	87.0×10^{-5}	0.950	87.0×10^{-5}
Example 4	0.898	643.0×10^{-5}	0.901	622.1×10^{-5}

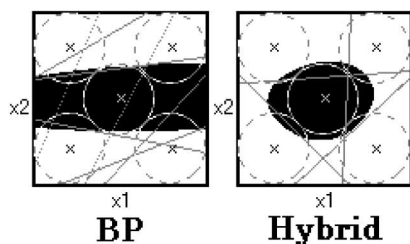


図 11 例題 1 の正解領域と分離超平面
Fig. 11 Classification domain and separation hyperplane in Example 1.

なっているのが確認できる (テストベクトルは標準偏差 0.07 の雑音を使用)。

6.1.2 例題 2

次に、各カテゴリーの訓練ベクトルが混合ガウス分布をなす場合を示す。その一例を図 12 に示す。BP 学習では左側のカテゴリーに偏って作成された分離超平面が、ハイブリッド学習ではカテゴリー間の中央に移動し、黒領域が妥当な形状になっている。表 1 から、平均正解認識率が向上し、ばらつきが小さくなっているのが確認できる (テストベクトルは標準偏差 0.02 の雑音を使用)。

6.1.3 例題 3: カテゴリー間が密な例

訓練ベクトルを $x_1 = 0.4, x_2 = 0.6$ を中心とする標準偏差: 0.05 の雑音で 200 個発生させ、同様に他クラスの訓練ベクトルを $x_1 = 0.6, x_2 = 0.4$ を中心とする標準偏差: 0.15 の雑音で 200 個発生させる。テストベクトルは訓練ベクトルと同方法で 1000 個発生させる。結果を図 13 に示す (訓練ベクトルは点で表示。色の違いはクラスの違いを表す)。ベイズ識別関数を近似した境界が作成されているのが確認できる。BP 学習とハイブリッド学習で作成される超平面位置が異なるが、正解領域は同一のものが作成され、表 1 から正解認識率も等しくなっているのがわかる。これはハイブリッド学習の途中で分類能力に貢献してない超平面が

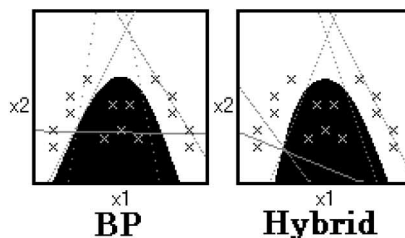


図 12 例題 2 の正解領域と分離超平面
Fig. 12 Classification domain and separation hyperplane in Example 2.

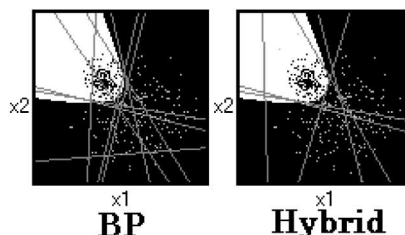


図 13 カテゴリー間が密な場合の正解領域と超平面
Fig. 13 Classification domain and hyperplane in the case that the interval of the category is dense.

移動し、貢献している超平面は分類能力を構築する位置にとどまるためと考えられる。この例から、カテゴリー間が密な場合、ハイブリッド学習は BP 学習の構築する 2 次超曲面に影響を与えないことが確認できる。

6.1.4 例題 4: カテゴリー間が疎と密が混合する例

訓練ベクトルを $x_1 = 0.4, x_2 = 0.6$ を中心とする標準偏差: 0.1 の雑音で 100 個 (カテゴリー A) 発生させ、同様に他クラスの訓練ベクトルを $x_1 = 0.9, x_2 = 0.1$ を中心とする標準偏差: 0.02 の雑音で 100 個 (カテゴリー B) と $x_1 = 0.3, x_2 = 0.7$ を中心とする標準偏差: 0.05 の雑音で 100 個発生させる (1 クラス 2 カテゴリー)。また、テストベクトルは訓練ベクトルと同方法で 1000 個発生させる。結果を図 14 に示す。カテゴリー間が疎と密が混合する場合は、密な場所の正解領域には無影響で、疎な場所の分離超平面は移動し、正解領域の形状も変化することが確認できる。この例の場合、BP 学習ではカテゴリー A とカテゴリー B の間の境界がカテゴリー A に大きく入り込んでいるが、ハイブリッド学習ではその形状が改善している。また、表 1 から正解認識率も改善している。

6.2 4 クラス分類問題

マルチクラス分類問題への適用を検証するため、

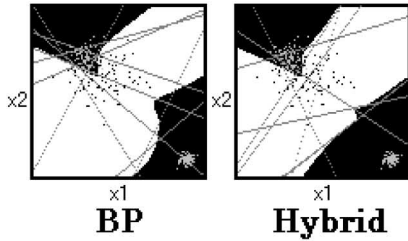


図 14 カテゴリー間が疎と密が混合する場合の正解領域と超平面

Fig. 14 Classification domain and hyperplane in the mixed case that the interval of the category is dense and sparse.

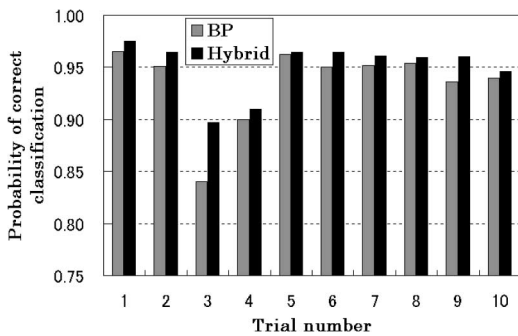


図 15 BP 学習とハイブリッド学習の正解認識率の比較

Fig. 15 Compare of correct classification probability of BP learning and hybrid learning.

以下の 4 クラス分類を行う。4 層パーセプトロン（入力層：7 ユニット，第 1 中間層：20 ユニット，第 2 中間層：7 ユニット，出力層：4 ユニット。ただし，出力層以外はオフセットユニットを含む）を使用する。パターン次元は 6 次元で，距離がほぼ 1 になる訓練ベクトル 4 個を適当に発生させ，各クラスに割り当てる（割り当てたクラスの出力値：0.99，他クラスの出力値：0.01）。同一の結合係数初期値を用いて BP 学習とハイブリッド学習を行う。学習回数は 5 万回で打ち切りとしているが，学習誤差を十分小さくする回数である。結合係数初期値を変えながらこれを 10 回行い，正解認識率を比較する。テストベクトルは標準偏差：0.2 の雑音を使用する。結果を図 15 に示す。3 回目の試行で，ハイブリッド学習の正解認識率が大きく改善している。BP 学習では分離超平面が大きく偏った位置に作成されたためである。5 回目の試行は改善度は小さい。これは BP 学習だけでも偏りの少ない位置に作成されたためである。この原因は 3. で述べたとおりで

あるが，どの場合もハイブリッド学習の正解認識率が BP 学習を上回ることが確認できる。3, 4 回目の試行は他の試行と比べハイブリッド学習を行っても正解認識率が際立って低い。この違いは，BP アルゴリズムによって構築される分離超平面の組合せの違いにより発生する。HIO アルゴリズムはこの枠組みの中で分離超平面を移動させ，汎化能力を向上させるものである。

7. む す び

カテゴリー間が疎な場合に，分離超平面を移動させ汎化能力を向上させる HIO アルゴリズムを提案し，この基本的性質について解析を行った。更に，多層パーセプトロンへの適用方法として，BP アルゴリズムと HIO アルゴリズムを混合したハイブリッドアルゴリズムを提案し，例題を用いてその基本的性質と汎化性を検証した。これにより，BP 学習で最適解を得るために学習パラメータの値を変えながら試行錯誤を繰り返す回数を削減できることを明らかにした。

今後の課題として，4. 8 で示した偏りの解析，制御方法の検討や，2. 1 のケース 2 の最適境界を近似する方法の検討（具体的には，式 (7) の β を導入）がある。また，実際の識別課題への適用や他の手法との比較を行い，得失を明確にする必要がある。更に，本論文ではハイブリッドアルゴリズムにおいて HIO アルゴリズムの学習係数 μ と BP アルゴリズムの学習係数 η の比 $\mu/\eta = 10^{-3} \sim 10^{-4}$ としてシミュレーションを行ったが，この値の範囲を理論的に明らかにする必要がある。

文 献

- [1] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing vol.1*, eds. J.L.McClelland, D.E.Rumelhart, and The PDP Research group, MIT press, 1986.
- [2] 栗田多喜夫, "情報量基準による 3 層ニューラルネットの隠れ層のユニット数の決定法," 信学論 (D-II), vol.J73-D-II, no.11, pp.1872-1878, Nov. 1990.
- [3] M.R. Azimi-Sadjadi, S. Sheedvash, and F.O. Trujillo, "Recursive dynamic node creation in multilayer feed-forward networks," *IEEE Trans. Neural Networks*, vol.4, no.2, pp.242-256, 1993.
- [4] A.S. Weigend, D.E. Rumelhart, and B.A. Huberman, "Generalization by weight elimination applied to currency exchange rate prediction," *Proc. International Joint Conference on Neural Networks*, vol.3, pp.2374-2379, Singapore, 1990.
- [5] K. Nakayama and Y. Kimura, "Optimization of activation functions in multilayer neural network,"

Proc.IEEE ICNN'94, pp.431-436, Orlando, June 1994.

[6] 木村義政, “誤差及び出力変動を最小化する学習アルゴリズムとその低品質手書き漢字認識への応用,” 信学論 (D-II), vol.J73-D-II, no.6, pp.840-847, June 1990.

[7] 石川眞澄, “コネクションモデルの忘却を用いた構造化学習,” 信学技報, MBE88-144, 1988.

[8] 渡辺栄治, “パターン認識問題に対する階層型ニューラルネットワークの汎化能力改善学習法,” 信学論 (D-II), vol.J79-D-II, no.5, pp.917-923, May 1996.

[9] S. Akaho, “Regularization learning of neural networks for generalization,” Proc. Workshop on Algorithmic Learning Theory, pp.99-110, 1992.

[10] 鹿山昌宏, 阿部重夫, “汎化能力向上を目的としたクラスタリング用ニューラルネットの学習方式,” 信学論 (D-II), vol.J76-D-II, no.4, pp.863-872, April 1993.

[11] 大堀隆文, 西村昭彦, 道見忠範, 林 勇, 渡辺一央, “追加学習による単純パーセプトロンの汎化,” 信学論 (D-II), vol.J74-D-II, no.10, pp.1472-1474, Oct. 1991.

[12] 阿部重夫, 鹿山昌宏, 武長 寛, “パターン認識用ネットの学習の高速化と汎化能力の向上方式,” 信学論 (D-II), vol.J76-D-II, no.3, pp.647-652, March 1993.

[13] 大堀隆文, 星野 健, 渡辺一央, “局所表現型パーセプトロンの汎化能力向上法,” 信学論 (D-II), vol.J78-D-II, no.4, pp.670-678, April 1995.

[14] S. Kullback, Information Theory and Statistics, New York: Dover, 1968.

[15] D.W. Ruck, S. Rogers, M. Kabrisky, H. Oxley, and B. Suter, “The multilayer perceptron as an approximator to a Bayes optimal discriminant function,” IEEE Trans. Neural Networks, vol.1, no.4, pp.296-298, 1990.

[16] K. Funahashi, “Multilayer neural networks and Bayes decision theory,” Neural Networks, vol.11, pp.209-213, 1998.

[17] R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis, John Wiley & Sons, 1973.

[18] P.D. Wasserman, Advanced Methods In Neural Computing, Van Nostrand Reinhold, New York, 1993.

付 録

式 (12) の導出

スカラ場 E_{iio} は, 訓練ベクトル \mathbf{x} を定数, 重みベクトル \mathbf{w} を変数とする関数である. こう配ベクトルを

$$\nabla = \left[\frac{\partial}{\partial w_0}, \dots, \frac{\partial}{\partial w_n} \right] \quad (\text{A.1})$$

と表す. 重み空間が直交基底をなすことから

$$\begin{aligned} \nabla(\mathbf{w} \cdot \mathbf{w}) &= 2((\mathbf{w} \cdot \nabla)\mathbf{w} + \mathbf{w} \times (\nabla \times \mathbf{w})) \\ &= 2\left((\mathbf{w} \cdot \nabla)\mathbf{w} + \left[\mathbf{w} \cdot \frac{\partial \mathbf{w}}{\partial w_0}, \dots, \mathbf{w} \cdot \frac{\partial \mathbf{w}}{\partial w_n}\right] - (\mathbf{w} \cdot \nabla)\mathbf{w}\right) \end{aligned}$$

$$= 2\mathbf{w} \quad (\text{A.2})$$

$$\begin{aligned} \nabla(\mathbf{w} \cdot \mathbf{x}) &= (\mathbf{x} \cdot \nabla)\mathbf{w} + \mathbf{x} \times (\nabla \times \mathbf{w}) \\ &= (\mathbf{x} \cdot \nabla)\mathbf{w} + \left[\mathbf{x} \cdot \frac{\partial \mathbf{w}}{\partial w_0}, \dots, \mathbf{x} \cdot \frac{\partial \mathbf{w}}{\partial w_n}\right] - (\mathbf{x} \cdot \nabla)\mathbf{w} \\ &= \mathbf{x} \quad (\text{A.3}) \end{aligned}$$

となる. ただし, \cdot は内積を, \times は外積を表す. この結果を用いて

$$\begin{aligned} \nabla\left(\frac{\|\mathbf{w}\|}{\mathbf{w} \cdot \mathbf{x}}\right)^2 &= \nabla \frac{\mathbf{w} \cdot \mathbf{w}}{(\mathbf{w} \cdot \mathbf{x})^2} \\ &= \frac{\nabla(\mathbf{w} \cdot \mathbf{w})(\mathbf{w} \cdot \mathbf{x})^2 - \mathbf{w} \cdot \mathbf{w} \nabla(\mathbf{w} \cdot \mathbf{x})^2}{(\mathbf{w} \cdot \mathbf{x})^4} \\ &= \frac{\nabla(\mathbf{w} \cdot \mathbf{w})(\mathbf{w} \cdot \mathbf{x})^2 - 2(\mathbf{w} \cdot \mathbf{w})(\mathbf{w} \cdot \mathbf{x})\nabla(\mathbf{w} \cdot \mathbf{x})}{(\mathbf{w} \cdot \mathbf{x})^4} \\ &= 2\left(\frac{\mathbf{w}(\mathbf{w} \cdot \mathbf{x})^2 - \|\mathbf{w}\|^2(\mathbf{w} \cdot \mathbf{x})\mathbf{x}}{(\mathbf{w} \cdot \mathbf{x})^4}\right) \\ &= -2\left(\frac{\|\mathbf{w}\|}{\mathbf{w} \cdot \mathbf{x}}\right)^2 \left(\frac{\mathbf{x}}{\mathbf{w} \cdot \mathbf{x}} - \frac{\mathbf{w}}{\|\mathbf{w}\|^2}\right) \quad (\text{A.4}) \end{aligned}$$

となる. よって, 式 (A.4) から次式を得る.

$$\begin{aligned} \nabla E_{iio} &= \frac{1}{2|X|} \sum_{\mathbf{x} \in X} \beta(\mathbf{x})^2 \nabla\left(\frac{\|\mathbf{w}\|}{\mathbf{w} \cdot \mathbf{x}}\right)^2 \\ &= \frac{-1}{|X|} \sum_{\mathbf{x} \in X} \left(\frac{\beta(\mathbf{x}) \|\mathbf{w}\|}{\mathbf{w} \cdot \mathbf{x}}\right)^2 \left(\frac{\mathbf{x}}{\mathbf{w} \cdot \mathbf{x}} - \frac{\mathbf{w}}{\|\mathbf{w}\|^2}\right) \quad (\text{A.5}) \end{aligned}$$

(平成 12 年 7 月 31 日受付, 12 月 11 日再受付)



殿村 正延 (正員)

平 1 金沢大・工・電気情報卒. 平 3 同大学院修士課程了. 同年 (株)北陸銀行入行. 主に汎用機動定系オンラインシステムの管理・開発, 及び情報系 C/S システムの開発に従事. 平 11 金沢大学院・自然科学研究科・博士後期課程入学. ニューラルネットワークによる連想記憶, パターン分類の研究に従事.



中山 謙二 (正員)

昭 46 東工大・工・電子卒, 昭 46~47 同大研究生, 昭 58 工博(東工大), 昭 47 日本電気(株)入社, 伝送通信事業部及び C & C システム研究所に勤務し, 通信用各種フィルター及びデジタル信号処理の研究に従事. 昭 63 金沢大・工・電気情報工学科助教授, 平 2 教授, 平 12 情報システム工学科(新設)教授. 最近の研究テーマは, 適応信号処理及びニューラルネットワーク. 昭 62.9 IEEE Circuits & Devices Mag. 論文賞受賞. 著書「SC 回路網の設計と応用」(東海大学出版)他. IEEE シニア, INNS 会員.