

内部情報最適化アルゴリズムを用いた 超平面分割による汎化能力向上法

近畿職業能力開発大学校

附属京都職業能力開発短期大学校 殿 村 正 延

Improving Method of Generalization Ability using to Divide Hyperplanes

with Internal Infomation Optimum Algorithm

Masanobu TONOMURA

要約 バックプロパゲーションアルゴリズムをパターン分類問題に適用した場合、学習が収束しても必ずしも汎化能力は最大にならない。これは、結合係数初期値や学習係数によって学習するごとに異なる内部表現が構築されるためである。この組合せ試行の負担を軽減する内部情報最適化アルゴリズムと、これを多層パーセプトロンに適用するハイブリッドアルゴリズムが提案されている。

本研究では、ハイブリッド学習で構築された入力層 - 第 1 中間層間の分離超平面の組合せの中で、組合せが悪いために低汎化性となる箇所を検出し、その分離超平面を分割し中間層ユニットを増やすことで、汎化性を向上させる方法を提案する。更に、内部情報最適化アルゴリズムの性質を利用して同一の分類能力をもつ第 1 中間層ユニットを検出し削除できることを示す。これらにより、汎化能力を向上させながらネットワークサイズを最適化できる。最後に、計算機シミュレーションにより汎化能力を検証する。

はじめに

BP アルゴリズム (Back-propagation algorithm)¹⁾ は、多層パーセプトロンの学習アルゴリズムとして最も使用される訓練法である。これをパターン分類問題に適用した場合、十分な数の訓練ベクトルが準備でき、テストベクトルが発生する領域が訓練ベクトルによってカバーされ、かつ、学習が理想に進めば、BP 学習のみでも十分な汎化能力が形成される²⁾³⁾。しかし、実際は十分な数の訓練ベクトルが集められない、または、訓練時間の制約から少数の代表ベクトルで学習を行う場合があり、このとき汎化能力について検討する必要性が生じる。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⁴⁾ や学習中にユニット数を追加していく方法⁵⁾ や削除し

ていく方法⁶⁾⁷⁾ が提案されている。しかし、中間層ユニット数が同じ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⁸⁾。この問題に対して多くの汎化能力向上法が検討されている⁹⁾~¹⁷⁾。

この組合せ試行の負担を軽減する IIO アルゴリズム (Internal-Infomation-Optimum algorithm) と、これを多層パーセプトロンに適用するハイブリッドアルゴリズムを筆者は既に提案している⁹⁾。このアルゴリズムは、BP アルゴリズムとは違った評価関数の力学系に従って動作し、BP アルゴリズムが構築した分離超平面を分離したい各カテゴリーからの距離が均等になるように移動することで汎化能力を向上させる。しかし、これは BP アルゴリズムが構築した分離超平面の

組合せの枠内で行われるため、汎化能力の向上には限界がある。汎化能力に大きな影響を与えるもう一つの要因は、各カテゴリーを分離する超平面の組合せであり、これが適切に構築されなければ、訓練ベクトルに対する分類能力は獲得できても汎化能力の小さなネットワークとなる。

本研究では、この問題を対象とし、HIO アルゴリズムを用いた超平面分割 (Division of Hyper-Planes . 以下、DHP) による汎化能力向上法を提案する。まず、で議論するデータ分布を仮定し、でこの分布に対して汎化能力を向上させる HIO アルゴリズムとハイブリッドアルゴリズムについて簡単に述べる。次に、で汎化能力の小さい分離超平面を検出、分割することで、ネットワークの正解認識率を向上させる超平面分割法を提案し基本原理について述べる。更に、でハイブリッドアルゴリズムの性質を利用すれば、同一分類能力をもつ第 1 中間層ユニットを取り除くことができ、ハード的に最小のネットワークを構築できることを示す。では超平面分割を行う全学習プロセスを示す。最後に、で計算機シミュレーションにより、提案方法の有効性を検証する。

データ分布の仮定

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とする。この場合、ベイズ境界は一般に 2 次超曲面を形成する⁽¹⁸⁾。しかし、分布を近似する十分な数の訓練ベクトルが集められない、つまり、訓練ベクトルがスパースな場合、カテゴリーを形成する訓練ベクトルが保証すべき領域は縮退し、カテゴリー間は一般にスパースとなる。そのため、BP 学習はベイズ境界を近似せず、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分離超平面の位置でほとんど規定される。本論文では、各カテゴリーの訓練ベクトルの発生確率が等しく、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する) と仮定する。この場合、最適境界は異クラスのカテゴリー間の最も近距離の訓練ベクトル間の中点を通りかつ直交する分離超平面となる。

既提案アルゴリズム

1 . HIO アルゴリズム

HIO アルゴリズムは、パターン分類能力を構築する

のではなく、式(1)の HIO 評価関数の最小値を探索することにより、BP アルゴリズムが構築したパターン分類能力を最適化する。

$$E_{HIO} = \frac{1}{2|X|} \sum_x \left(\sum_j \left(x_j - \frac{w_{1j}}{w_{1j}^T x} \right)^2 \right) \quad (1)$$

X は $J(J [0 , 1])$ の有限部分集合 X に属するパターンの個数を表す。多次元入力ベクトル x はオフセットを含めて表記する。 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。 \cdot は \cdot のユークリッドノルム、 w_{1j}^T は重みベクトル w_{1j} の転置を表す。また、 (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ある。本論文は超平面分割の基本原理を述べることを目的とするため、 $(x) = 1$ とする。

式(2)の雑音許容半径 r_s は、訓練ベクトル $x(s = 1, \dots, |X|)$ の w_{1j} 上への正射影のスカラで定義され、汎化能力の評価尺度として使用する。 r_s は w_{1j} と x^s のなす角、 $|\cdot|$ は \cdot の絶対値を示す。

$$r_s = \frac{|w_{1j}^T x^s|}{|w_{1j}|} = |x^s| |\cos \theta_s| \quad (2)$$

HIO アルゴリズムは

$$w_{1j}(n+1) = w_{1j}(n) + \mu \Delta w_{1j} \quad |w = w(n) \quad (3)$$

$$\Delta w_{1j} = \tanh(\Delta w_{1j}^{orig} / T) \quad (4)$$

$$\Delta w_{1j}^{orig} = \frac{1}{|X|} \sum_x \left(\frac{\partial E_{HIO}}{\partial w_{1j}^T x} \right) \left(\frac{w_{1j}}{w_{1j}^T x} x - w_{1j} \right) \quad (5)$$

となる。式(5)は、式(1)に最急降下法を適用して求める。この式は分離したいカテゴリー間の最も近距離の訓練ベクトルの中点を通りかつ直交するよう分離超平面を移動させる。作成された境界は、ベイズ境界のように正解認識率を最大にしないが、超平面で分離する場合の正解認識率を均等ににす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。 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。 \tanh 関数は収束を保証するために使用する。なお、このアルゴリズムは 3 . の超平面分割処理で使用する。

2 . ハイブリッドアルゴリズム

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することで大きな効果が得られる。 $N + 1$ 層パーセプトロンにおけるハイブリッドアルゴリズムは

$$w_{1j}(n+1) = w_{1j}(n) + \Delta w_{1j} \quad |w = w(n) \quad (6)$$

$$\Delta w_{ij} = \begin{cases} \Delta w_{ij}^{bp} + \mu \Delta w_{ij}^{io} & (l=1) \\ \Delta w_{ij}^{bp} & (2 \leq l < N) \end{cases} \quad (7)$$

となる。ただし、 w_{ij} は l 層の全ユニットから $l+1$ 層の第 j 番目ユニットへの重みベクトルである。 Δw_{ij}^{bp} は BP アルゴリズムの修正項、 μ は $0 < \mu < 1$ となる学習係数である。 Δw_{ij}^{io} は式(4)の修正項と同一である。 μ は修正項 Δw_{ij}^{bp} による分類能力の構築に影響を与えない十分小さい値 ($0 < \mu \ll 1$) であることが必要である。なお、このアルゴリズムによる学習は超平面分割処理の前処理、後処理として行う。

超平面分割 (DHP) の基本原理

1. 超平面分割の基本的発想

理解を容易にするため、2次元入力空間について考える。図1(a)は、カテゴリー A (Class 1)、カテゴリー B、C (Class 2)^{注1)}と入力層 - 第1中間層間の分離超平面 (実太線) を示している。 a, b, c, d, e, f, g, h は各カテゴリーを形成する訓練ベクトルである。ここで述べたように、分布を近似する十分な数の訓練ベクトルが集められない場合、カテゴリー領域は縮退し、(この例では一つの) 分離超平面で完全分離可能となる。この場合、雑音許容範囲は点線で示すようになり、汎化能力は小さい。これを図(b)のように二つの分離超平面 (実線) で分離できれば、灰色領域の分だけ汎化能力は大きくなる。BP 学習で分類能力を構築する場合、中間層ユニット数を増やすことで図(b)のケースが構築される可能性が大きくなるが、結合係数初期値によっては図(a)となる。大規模なネットワークでは、このような低汎化能力の分離超平面が各部に構築される可能性があり、分離超平面の最適な組合せを得るために結合係数初期値や学習パラメータを変えながら複数回学習を行うことを強要する。しかし、この試行で得られた最も汎化能力の高いネットワークのすべての箇所で汎化能力が高い保証はない。このような場合に

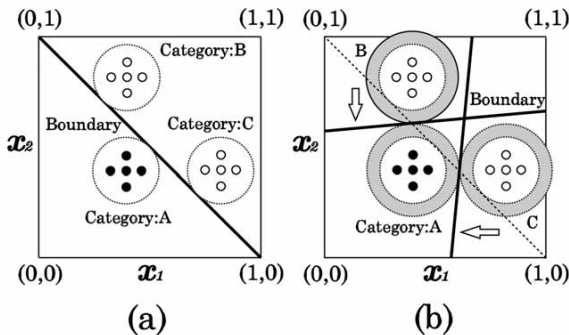


図1 超平面分割による汎化能力向上の例

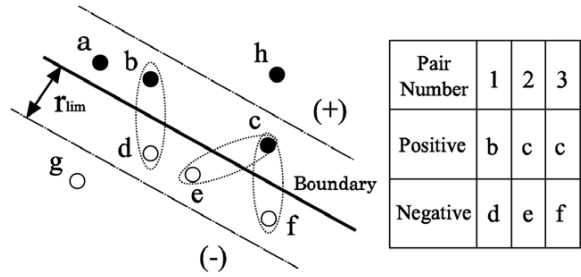


図2 分割対象ベクトルのペアリングの例

有効な汎化能力向上法として、訓練ベクトルに乱数を漸減的に重畳する方法がある⁽¹⁴⁾。しかし、重畳する乱数の減少方法や学習回数によっては認識率を低下させることがある。更に、乱数重畳法には入力パターン次元の上昇に伴い、指数関数的に学習量を増加させる問題がある。そこで、図(a)のような低汎化能力の箇所を検出し、図(b)のように分離超平面を分割して汎化能力を向上させることを考える。提案法では、超平面の位置は一意に決まるため、認識率は学習回数の影響を受けない。

2. 低汎化能力の分離超平面の検出

図2は、分離超平面で分離された二つの領域と訓練ベクトルの様子を示している。この例を使って汎化能力の低い場所の検出方法を説明する。なお、以下の議論は入力層 - 第1中間層間の任意の重みベクトルを対象とするため、 w_{ij} は w と略記する。

(1) 分割対象訓練ベクトルの抽出

極端に分離超平面に近接している訓練ベクトルを探すために、式(2)の r_s が分割対象半径 $r_{lim} (> 0)$ より小さい訓練ベクトルを抽出する。

$$r_s < r_{lim} < (I+1)^{\frac{1}{2}} \quad (8)$$

ただし、 I はパターン次元である。 $r_{lim} = (I+1)^{\frac{1}{2}}$ とすると、すべての訓練ベクトルが対象となるが、極端に分離超平面に近接している訓練ベクトルを探すことが目的のため、適当な小さな定数を用いる。抽出した訓練ベクトルが超平面のどちらの側 ($w \cdot x^s$ の符号) にあるかで2クラス (正、負) に分類する。図2の例では、正クラスに a, b, c 、負クラスに d, e, f が分類され、 g, h は対象外となる。また、同一クラスの訓練ベクトルが正負両クラスに存在する場合は、分割による汎化性向上は期待できないため、分割処理は行わない。

(2) ペアリング

2クラスに分類された訓練ベクトルの中から選択したあるペアを直径の端点とする超球内部に(1)で抽出した (そのペアを構成する訓練ベクトルを除く) 訓練ベ

クトルが入らない場合、分割ペアとして選択する。これにより、超平面を挟んで最近傍の訓練ベクトルの組合せが求まる。図2の例では、点線で囲んだ訓練ベクトルがペアを作る(図2の右表)。ペアリングは組合せ計算を行うが、不要な組合せの計算量を減らすために式(8)の r_{lim} の値は訓練ベクトルと全体の分布の関係を考えて選ぶ必要がある。

(3) 分割ペアの判定

(2)で求めた分割ペアは単に超平面を挟んで最近傍の訓練ベクトルを求めただけであり、この中には分割することで汎化能力が向上するものとししないものがある。例えば、そのペアを結んだ線分が超平面と直交する場合には分割する効果はない。そこで判定基準を導入する。

式(5)の Δw^{org} は、次式のように書き直せる。

$$\Delta w^{org} = \frac{1}{|X|} \sum_{x^s \in X} a_s \left(x^s - \frac{w^T x^s}{w^T w} w \right) \quad (9)$$

$$= \frac{1}{|X|} \sum_{x^s \in X} \Delta w^s \quad (10)$$

$$a_s = \frac{1}{w^T x^s} \left(\frac{x^s}{w^T x^s} w \right)^2 \quad (11)$$

式(9)は、訓練ベクトル x^s の重みベクトル w の直交補空間 w^\perp 上への射影ベクトルを a_s 倍し、総和したものに他ならない。HIO アルゴリズムが収束したとき、つまり、 $\Delta w^{org} = 0$ のとき、式(10)のベクトル Δw^s は均衡している。これを図1(a)の例で考えると、 $\Delta w^s (s = a, b, c)$ は図3のようになる。ただし、簡単のためにカテゴリー A、B、C は分布の中心から標本された訓練ベクトル a、b、cのみからなるとする。この図で Δw^s の分割ペアの外積について考える。これは超平面を回転させる力とみなすことができる。 Δw^s が直交補空間上のベクトルであることから、外積方向はすべて重みベクトル w と等しい。よってスカラーで考えてもさしつかえない。次にその符号についてみると、例えば Δw^a と Δw^b 、及び Δw^a と Δw^c の外積には次式の関係がある。

$$\Delta w^a \times \Delta w^b = - \Delta w^a \times \Delta w^c \quad (12)$$

これは、全体で考えると超平面を回転させる力は0であるが、部分的に見ると保持されているとみることができる。

以上を踏まえ、超平面分割を行う効果の指標として回転力 $M_{pq} (> 0)$ を導入する。

$$\begin{aligned} M_{pq} &= | \Delta \hat{w}^p \times \Delta \hat{w}^q | \\ &= \Delta \hat{w}^p \quad \Delta \hat{w}^q \quad | \sin \theta_{pq} | \\ &= \Delta \hat{w}^p \quad \Delta \hat{w}^q \quad (1 - \cos^2 \theta_{pq})^{\frac{1}{2}} \end{aligned} \quad (13)$$

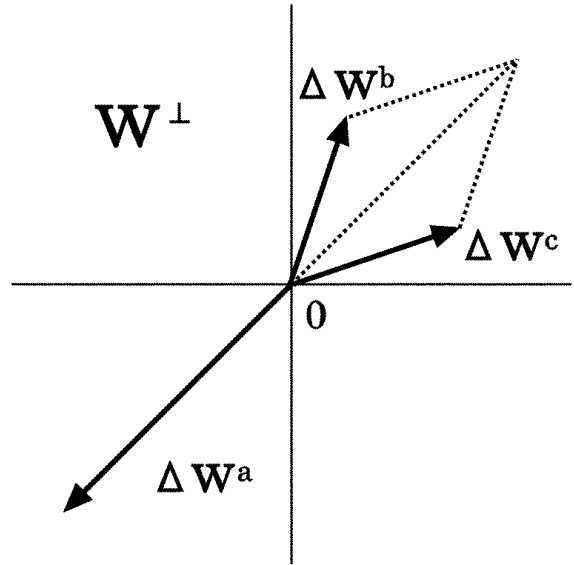


図3 直交補空間上の修正量 Δw^s の均衡

$$\Delta \hat{w}^s = \frac{1}{\hat{w}^T x^s} \left(\frac{x^s}{\hat{w}^T x^s} \right) \left(x^s - \hat{w}^T x^s \hat{w} \right) \quad (14)$$

$$\hat{w} = \frac{w}{|w|} \quad (15)$$

$$\cos \theta_{pq} = \frac{\Delta \hat{w}^p \cdot \Delta \hat{w}^q}{|\Delta \hat{w}^p| |\Delta \hat{w}^q|} \quad (16)$$

ただし、 p, q は分割ペアを構成する訓練ベクトルの番号を表す。 $\Delta \hat{w}^s (s=p, q)$ 、 \hat{w} は式(14)、(15)となる。この正規化は回転力 M_{pq} が重みベクトル w のノルムの影響を受けずに一意に決まるようにするためである。

θ_{pq} は、 w 上の Δw^p と Δw^q のなす角を表し、 $\cos \theta_{pq}$ は式(16)より求まる。また、回転力を求めるために実際の外積計算を行わないのは、外積のベクトル方向が重みベクトル方向と一致することを利用している。訓練ベクトル x^a を基準として超平面を挟んで反対側の空間における回転力を3段階に色分けしたものを図4(a)に示す。 x^a と各座標を結んだ線と超平面との直交性が崩れるに従い、また反対側の座標が超平面に近づくに従い回転力が大きくなるのがわかる。このことは超平面分割を行う効果の良い指標となる。

分割することで効果が得られるとみなす下限値を分

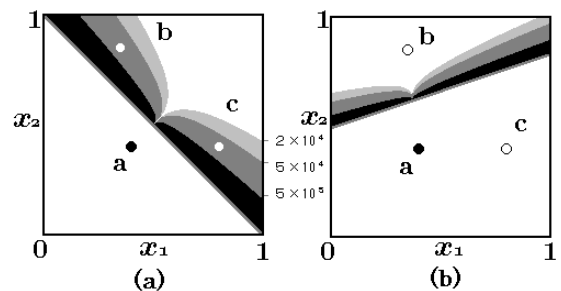


図4 超平面分割の範囲

割下限値 M_B とする。回転力 $M_{pq} > M_B$ のとき、超平面を回転させる。図4(a)の例において $M_B = 5 * 10^4$ とすれば訓練ベクトル x^b, x^c は分割対象となる。もし、訓練ベクトル x^c がなければ、つまり、訓練ベクトル x^a, x^b 以外の訓練ベクトルがないものとしてHIO学習を行った場合、HIOアルゴリズムの領域形成の性質⁽⁹⁾により、 x^a, x^b の雑音許容半径が最大となるように超平面は回転する(図4(b))。このとき、 $\Delta w^a + \Delta w^b = 0$ であることに注意すれば、 $\Delta w^a, \Delta w^b$ のなす角 θ_{ab} の余弦は

$$\cos \theta_{ab} = \frac{\Delta w^a \cdot \Delta w^b}{|\Delta w^a| |\Delta w^b|} = -1 \quad (17)$$

となり、式(13)の回転力は0となる。以上から、ある超平面について2.(3)で得られた分割ペアの回転力を計算し、分割下限値 M_B を下回るペアが存在する場合は分割の効果が小さいと考えられるため、その超平面を分割しない。それ以外は次の3.の過程で超平面分割を行う。

3. 超平面の分割

2.(3)の条件を満足した超平面について、その超平面を作る重みベクトルをもつユニットを分割ペアの数だけ追加する。その後、分割ペアを入力として入力層 - 第1中間層間のHIO学習を行う。

4. 分割対象半径 r_{lim} と分割下限値 M_B の関係の目安

分割対象半径 r_{lim} と分割下限値 M_B の関係を厳密に議論することは難しいが、その適切な関係を近似的に求めることはできる。

図5に示すように、ある分割ペアを構成する訓練ベクトル x^p, x^q は分割対象半径 r_{lim} 内の超平面の近傍にある。超平面からの距離を均等にするHIOアルゴリズムの性質により

$$r_p \approx r_q \leq r_{lim} = r_{\hat{p}} = r_{\hat{q}} \quad (18)$$

$$p \approx \frac{\pi}{2} - \quad (19)$$

$$q \approx \frac{\pi}{2} + \quad (20)$$

$$\hat{p} \approx \frac{\pi}{2} - \quad (21)$$

$$\hat{q} \approx \frac{\pi}{2} + \quad (22)$$

$$x^p \approx x^{\hat{p}}, \quad x^q \approx x^{\hat{q}} \quad (23)$$

が近似的に成り立つ。 \hat{p}, \hat{q} は訓練ベクトル x^p, x^q

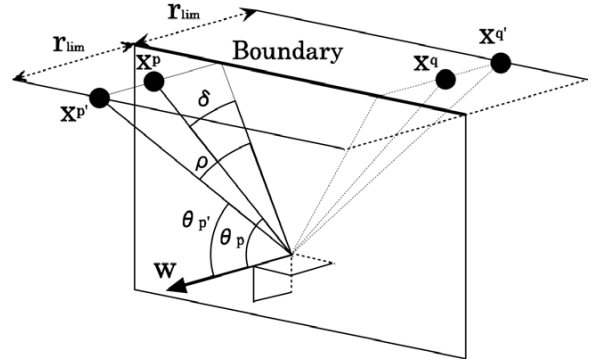


図5 分割対象半径 r_{lim} と分割ペアを構成する訓練ベクトル x^p, x^q との位置関係

を超平面からの距離が r_{lim} になるように移動したベクトル $x^{\hat{p}}, x^{\hat{q}}$ と重みベクトル w とのなす角を表し、 $0 < \theta_p \leq \theta_q < \pi/2$ の関係がある。式(14)より

$$\begin{aligned} \Delta \hat{w}^s &= \frac{1}{(w^s x^s)^3} (x^s - \hat{w}^s x^s \hat{w}) \\ &= \frac{1}{r^3} |x^s| \sin \theta_s | (s=p, q) \end{aligned} \quad (24)$$

となる。ただし、 $(x^s) = 1$ とする。式(13)と式(24)、及び式(18)~(20)より、回転力 M_{pq} は

$$M_{pq} = \frac{1}{r_p^6} |x^p| |x^q| \cos^2 \theta_{pq} \quad (25)$$

となる。超平面からの距離を固定すると、式(25)が最大となるのは、 $|\sin \theta_{pq}| = 1$ のとき、つまり、 $\Delta \hat{w}^p$ と $\Delta \hat{w}^q$ が直交するときである。これを

$$\max[M_{pq}] = \frac{1}{r_p^6} |x^p| |x^q| \cos^2 \theta_{pq} \quad (26)$$

と置く。同様にして、式(13)、(24)、(18)、(21)、(22)より、回転力 $M_{\hat{p}\hat{q}}$ 、及び $\max[M_{\hat{p}\hat{q}}]$ は

$$M_{\hat{p}\hat{q}} = \frac{1}{r_{\hat{p}}^6} |x^{\hat{p}}| |x^{\hat{q}}| \cos^2 \theta_{\hat{p}\hat{q}} \quad (27)$$

$$\max[M_{\hat{p}\hat{q}}] = \frac{1}{r_{\hat{p}}^6} |x^{\hat{p}}| |x^{\hat{q}}| \cos^2 \theta_{\hat{p}\hat{q}} \quad (28)$$

となる。分割対象半径 r_{lim} が

$$\begin{aligned} r_{lim} &= |w^s x^s| = |x^s| |\cos(\frac{\pi}{2} \pm \theta_s)| \\ &= |x^s| \sin \theta_s \quad (s=p, q) \end{aligned} \quad (29)$$

となること、及び式(23)から、式(28)は近似的に

$$\max[M_{\hat{p}\hat{q}}] \approx \frac{1}{r_{lim}^6} (|x^p| |x^q| - r_{lim}^2) \quad (30)$$

となる。一方、 x^p, x^q と $x^{\hat{p}}, x^{\hat{q}}$ の位置関係から

$$\max[M_{pq}] \geq \max[M_{\hat{p}\hat{q}}] \quad (31)$$

が成り立つ。上式は分割下限値 $M_B = \max[M_{\hat{p}\hat{q}}]$ とすれば、 $|\sin \theta_{pq}| = 1$ の極限において分割対象半径 r_{lim} の範囲にあるすべての分割ペアの回転力が分割下限値

を上回ることを意味している。この関係は、ある r_{lim} に対して M_B を決める際の目安となる。しかし、これは極限における議論のため、実際はこの値をやや下回る。そこで条件を緩めるために分割下限値を

$$M_B = \frac{1}{r_{MB}} (x^p \quad x^q) - r_{MB}^2 \tag{32}$$

$$r_{MB} = r_{lim} \tag{32}$$

として与える。 r_{lim} は分割対象の範囲を制御するパラメータである。

5 . 超平面分割に必要な層数

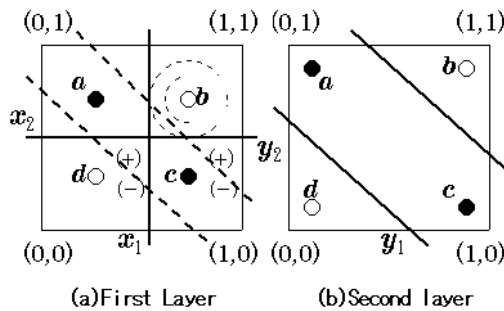


図6 XOR問題の超平面の例

図6(a)は XOR 問題の訓練ベクトル(点 a b c d) と入力層 - 第1中間層間で構築される分離超平面(直線)を示している。図の点線のように分離超平面が構築されることで、XOR 問題は3層パーセプトロンで分類可能である。しかし、汎化能力という点では実線のように分離超平面が構築されたほうが大きくなる。超平面分割法によって強制的に点線の超平面を実線のように分割した場合、第1中間層における各訓練ベクトルの変換座標は図(b)の点になる。これから3層では実現不可能であることがわかる。しかし、もう1層追加することによってこの問題を解くことができる^(注2)。このことから、一般に任意のパターンを分類するには最低3層必要であるが、超平面を分割することで汎化能力を向上させるには最低4層必要となる。

同一分類能力をもつ第1中間層ユニットの削除

スパースな訓練ベクトルを用いて BP 学習を行う場合、同一分類能力をもつ超平面であっても、全く同じ位置に構築されるとは限らない。これは、スパースな訓練ベクトルでは超平面の位置が一意に決まらないことと同一理由による⁽⁹⁾。このため、構築された微妙に違う位置の同一分類能力を有する第1中間層ユニットの検出は難しい。

ハイブリッドアルゴリズムは式(7) Δw_{ij}^{bp} によって構

築された超平面の組合せの枠内で超平面位置を一意に決定する。この性質を利用すれば、同一分類能力を有するユニットの検出が可能となる。具体的には、ハイブリッド学習中において、式(33)の判定を行い、許容値 ($\ll 1$) より小さいとき、同一分類能力をもつ第1中間層ユニットとみなす。

$$\sum_{i=0}^{i=I} \left| \frac{w_{1ji}}{w_{1j}} - \frac{w_{1ki}}{w_{1k}} \right| < \tag{33}$$

($j, k = 1, \dots, J, j \neq k$)

ただし、 I はオフセットユニットを除く入力ユニット数、 J はオフセットを除く第1中間層ユニット数を表す。式(33)を満たす第1中間層ユニットの中でノルムの大きな重みベクトル(第1中間層ユニットと1対1対応する)を残し、それ以外は不要ユニットとして削除する。

同様な方法が既に提案されている⁽⁷⁾。しかし、これは中間層ユニットの出力の相関により不要ユニットを判定しており、ノルムの大きさのみが異なる重みベクトル(同一位置の超平面)を検出できない場合がある。よって、本論文ではハイブリッドアルゴリズムの性質を利用した式(33)を使用する。

全学習プロセスの概要

全学習プロセスを図7に示す。

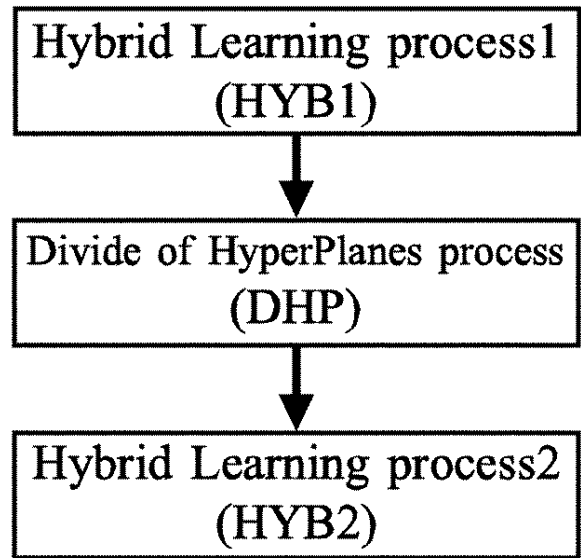


図7 全学習プロセス

(HYB1) ハイブリッド学習プロセス1

ハイブリッドアルゴリズムで分離超平面を移動し汎化能力を向上させる。ただし、⁽⁷⁾で示した方法で学習中に不要ユニットを削除する。このプロセスを HYB1 と表記する。本プロセスは、以下の理由で DHP プ

ロセスの前処理として必要な過程である。1. 超平面分割は、分割対象半径 r_{lim} を用いて分割対象訓練ベクトルの抽出を行うが、BP 学習のみでは超平面位置に偏りが生じる可能性があるため正しく判定できない。2. 分割対象の超平面が重複していると DHP プロセス中で重複する超平面分割が不必要に行われる。(DHP) DHP プロセス

で述べた方法で低汎化能力の分離超平面の検出と超平面分割を行う。

(HYB2) ハイブリッド学習プロセス2

DHP プロセスで入力層 - 第1中間層間の分離超平面位置は適切なものとなるが、層間の分類能力は構築されていない。これを構築するために、再度ハイブリッド学習を行う。HYB1と同様、学習中に不要ユニットを削除する。このプロセスをHYB2と表記する。

計算機シミュレーション

1. 重複ユニット削除による正解認識率への影響

の同一分類能力をもつ第1中間層ユニット(以下、重複ユニット)の削除による正解認識率への影響を検証するため、以下の4クラス分類を行う。第1中間層ユニット数が可変な4層パーセプトロン(入力層: 7ユニット、第1中間層: 50ユニット、第2中間層: 7ユニット、出力層: 5ユニット。ただし、各層のユニット数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8個を適当に発生させ、各クラスに割り当てる(1クラス2カテゴリー)。割り当てたクラスの出力値: 0.99、他クラスの出力値: 0.01とする。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。テストベクトルは標準偏差: 0.15の雑音を使用する。式(33)の $\epsilon = 0.01$ とする。学習過程における正解認識率の推移を図8に示す。黒は正解認識率、灰色は第1中間層ユニット数を示す。

図(a)は、第1中間層における初期ユニット数を30個に固定し、同一の結合係数初期値を用いて、ユニット削除を行わないBP学習とハイブリッド学習、及びユニット削除を行うハイブリッド学習の結果である。ユニット削除の有無に関わらずハイブリッド学習はBP学習と比較して正解認識率が高い。学習が進むに従ってユニット削除を行うハイブリッド学習の活動ユニットが減少するが、正解認識率は削除しない学習と

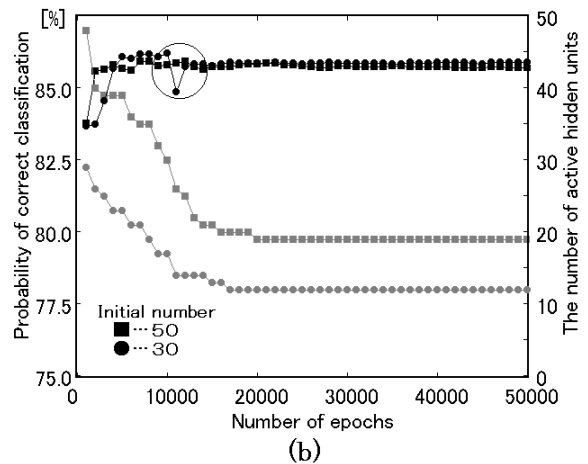
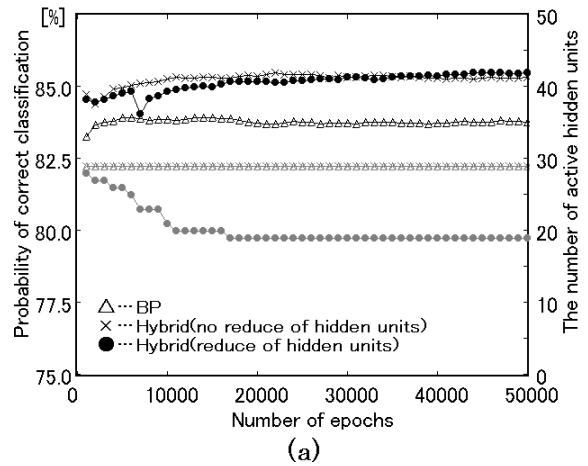


図8 重複ユニット削除による正解認識率の推移

変わらないことが確認できる。

図(b)は、初期ユニット数: 50個と30個(同一の結合係数初期値を使用)のユニット削除を行うハイブリッド学習の結果である。初期ユニット数によって最終ユニット数は多少異なるが、必要とするユニット数はほぼ一定(この例では、ユニット数は12~20個)である。また、正解認識率は初期ユニット数に関わらず変わらない。

図(b)の初期ユニット数が30個の例で、ユニット削除により正解認識率が一時的に低下している(図の円中)。これは式(33)を満たすユニットのうち、ノルムの大きなものが分類能力を形成するユニットではなかったため、ユニット削除により分類能力が破壊されたためであるが、学習が進むに従って分類能力は回復している。以上から、このユニット削除方法は正解認識率に影響を与えず、ネットワークサイズを小さくできることがわかる。

2. 2クラス分類問題

入力層 - 第1中間層間の超平面の移動と正解領域の

表1 各手法の正解認識率の比較

	Recognition rate [%]		
	BP	HYB	HYB 2
Example 1	96.928	98.520	99.312
Example 2	94.576	96.516	98.752

表2 提案法の第1中間層ユニット数の推移

	First hidden units number		
	HYB 1	DHP	HYB 2
Example 1	8	11	5
Example 2	7	13	5

形状を視覚的に観察するために、2次元入力、2クラス分類問題を行う。

2.1 BP学習とハイブリッド学習との比較

第1中間層ユニット数が可変な4層ネットワーク(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:2ユニット。ただし、各層のユニット数はオフセットユニットを含む。)を用いてBP学習、ハイブリッド学習、及び図7の提案法を同一の結合係数初期値を用いて比較実験する。2クラスの教師出力は0.99、0.01を使用する。標準偏差0.08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:0.99に対して出力 ≥ 0.5 、教師出力:0.01に対して出力 < 0.5 を正解とし、その正解認識率の平均を求める。と は訓練ベクトルを表し、色の違い(黒色と白色)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。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。入力層-第1中間層間の超平面を灰色直線で示す。なお、DHPプロセスにおいて $r_{lim}=0.2$, $\alpha=1.2$ を使用する。

典型的な2例を図9に示す。BP学習と比較して、ハイブリッド学習(HYBと表記する)は例1、2とともに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、正解領域の形状が変化している。表1から正解認識率の向上も確認できる。しかし、例2は例1に比べ認識率が低い。ハイブリッド学習は、式(7)の Δw_{ij}^{bp} によって構築された超平面の組合せの枠内でその超平面を移動させるものであるため、汎化能力の改善度がその組合せに依存することを示している。一方、本論文提案法(HYB2)は例1、2ともに正解領域が妥当な形状となり、正解認識率も大きく改善している。つまり、超平面を分割することで結合係数初期値に依存した超平面の組合せを超えて

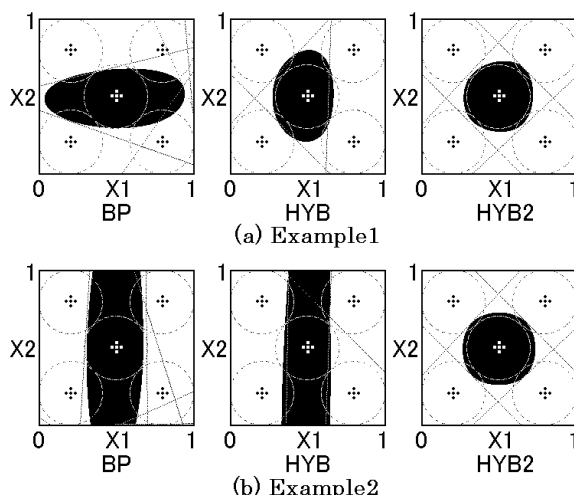


図9 各手法による正解領域と超平面の比較

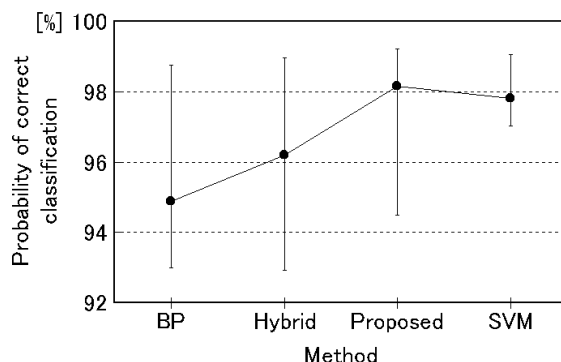


図10 各手法による平均正解認識率の比較

汎化能力が改善できる。表2は、提案法の各プロセスごとの第1中間層ユニット数の推移を示している。10個の初期ユニット数が最終的に5個にまで減少している。

図10に、結合係数初期値を変えながら100回の試行を行い、各手法の平均正解認識率を求めた結果を示す。は平均正解認識率を、縦棒は100回の試行の最大値と最小値を表す。提案法は平均正解認識率が大きく改善し、認識率のばらつきも小さくなっているのが確認できる。なお、第1中間層ユニット数の平均値の推移は、初期:10 HYB1:5.66 DHP:10.68 HYB2:5.26となった。

2.2 SVMとの比較

SVM⁽¹⁹⁾の解は式(34)より得られる。ただし、 λ_i はラグランジュ乗数、 C はソフトマージンの弛緩パラメータ、 y_i ($= \pm 1$)はクラスラベルを表す。

$$\min \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \lambda_i \lambda_j y_i y_j K(x_i, x_j) - \sum_{i=1}^l \lambda_i \sum_{j=1}^l y_i y_j K(x_i, x_j) \leq C \lambda_i \lambda_j = 1, \dots, l \quad (34)$$

RBFカーネル $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$ を使用し、 $\gamma =$

[0 .00025 0 .0005 0 .001 0 .002 0 .004 0 .008 0 .016 0 .032 0 .064 0 .128 0.256 0 .512 ,1 .024 2 .048] $C = [1 \ 2 \ 4 \ 8 \ 16 \ 32 \ 64 \ 128 \ 256 \ 512]$ の140通りのパラメータ組合せ試行によって、2.1と同様の学習を行った。分類能力を獲得したのは17.9% (25/140)である。それ以外は分類能力を構築できなかった。結果の一例を図11に、その正解認識率とパラメータを表3に示す。パラメータによって正解領域の形状が異なることが確認できる。最も使用される汎化評価尺度の *loo rate* (leave-one-out cross validation) の最大値は1.0となり、その割合は17.1% (24/140)である。(b)、(c)はともに *loo rate* = 1.0であるが、正解領域の形状や正解認識率に差がある。*loo rate* = 1.0となった24個の正解認識率の平均、最大値、最小値を図10に示す。*loo rate* が高い結果だけで見ると、正解認識率が高くばらつきも小さい。

表3 SVMのパラメータと正解認識率

Example	(a)	(b)	(c)
	2 .048	0 .256	0 .512
C	2	256	32
loo rate	0 .8	1 .0	1 .0
recognition rate	0 .947	0 .970	0 .991

SVMの現実的な問題は、カーネル関数やパラメー

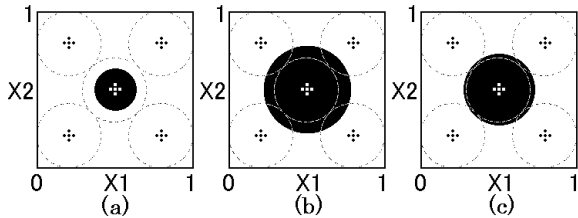


図11 SVMによる正解領域の例

表4 提案法の第1中間層ユニット数の推移

	First hidden units number		
	HYB 1	DHP	HYB 2
4 class problem	24 .14	42 .81	21 .99
8 class problem	24 .16	38 .19	21 .44

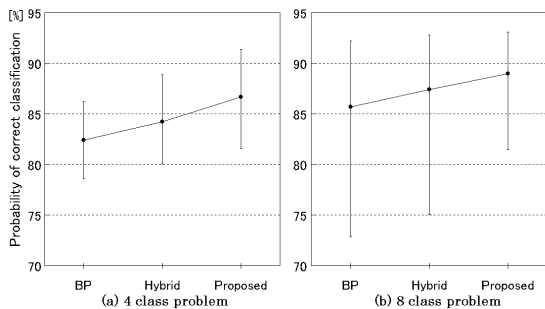


図12 マルチクラス分類問題における各手法の平均正解認識率の比較

タ選択によって汎化性が大きく異なることである。高い認識率を実現するネットワークを獲得するためには、精度の要求に従い、試行すべきパラメータの組合せ数は増加する。また、本シミュレーションではRBFカーネルを用いたが、例えば多項式カーネルを使用した場合、更に多項式次元パラメータが追加され、パラメータの組合せ数は膨大となる。

3. 多クラス分類問題

マルチクラス分類問題に対する提案法の有効性を検証するため、1. で用いた4層ネットワークを用いてBP学習、ハイブリッド学習、及び提案法の比較実験を行う。なお、DHPプロセスにおいて $r_{lim} = 0.2$ 、 $= 1.2$ を使用する。1. と同条件の学習を結合係数初期値を変えながら100回行い、平均正解認識率を求める。結果を図12(a)に示す。提案法によって正解認識率が向上している。1. で用いた4層ネットワークの出力ユニットを9個に変更し、8個の訓練ベクトルを各クラスに割り当てる(1クラス1カテゴリー)。他の条件は上記の4クラス分類問題と同様とし、平均正解認識率を求める。結果を図12(b)に示す。1クラスに1出力ユニットを割り当てる場合は、本手法が有効であることが確認できる。第1中間層ユニット数の推移を表4に示す。30個の初期ユニット数が最終的に22個程度まで減少している。

まとめ

スパースな訓練データを用いてパターン分類能力を学習する場合に、既提案のハイブリッドアルゴリズムによって構築された入力層 - 第1中間層間の分離超平面の組合せの中で、低汎化性の箇所を検出し、その分離超平面を分割し汎化性を向上させる超平面分割法を提案した。更に、このアルゴリズムの性質を利用して同一分類能力をもつ第1中間層ユニットを検出し、削除できることを示した。提案学習プロセスによって汎化能力を向上させながらネットワークサイズを最適化できることを明らかにした。

今後の課題として、実際の識別課題へ適用し有効性を検証する必要がある。

[注]

(注1) マルチクラス分類問題において、1クラスに1出力ユニットを割り当てる場合には、カテゴリーB (Class 2)、カテゴリーC (

Class 3) の場合でも、同様の議論ができる。

(注2) 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中間層間の分離超平面の位置ずれによる汎化能力への影響は小さい。上層も同様となる。よって、IIの仮定を満たすスパースな訓練ベクトルを用いる場合、汎化能力に最も大きな影響を与える入力層 第1中間層間の超平面位置を適切すればよい。このことから、入力層 第1中間層間の分離超平面は図6(a)の実線のようになり、第1中間層 第2中間層間の分離超平面は図6(b)の実線のようになる。

[参考文献]

- (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, pp. 318 - 362, 1986.
- (2) 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.
- (3) K. Funahashi, "Multilayer neural networks and Bayes decision theory," *Neural Networks*, vol. 11, pp. 209 - 213, 1998.
- (4) 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.
- (5) M. R. Azimi-Sadjadi, S. Sheedvash, and F. O. Trujillo, "Recursive dynamic node creation in multilayer feedforward networks," *IEEE Trans. Neural Networks*, vol. 4 no. 2, pp. 242 - 256, 1993.
- (6) 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*, Singapore, vol. 3, pp. 2374 - 2379, 1990.
- (7) K. Nakayama and Y. Kimura, "Optimization of activation functions in multilayer neural network applied to pattern classification," *Proc. IEEE ICNN 94, Orlando*, vol. 1, pp. 431 - 436, June 1994.
- (8) P. D. Wasserman, *Advanced Methods In Neural Computing*, Van Nostrand Reinhold, New York, 1993.
- (9) 殿村正延、中山謙二, "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" *信学論 (D-II)* vol. J84 - D-II, no. 5, pp. 830 - 842, May 2001.
- (10) 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.
- (11) 石川眞澄, "コネクショニストモデルの忘却を用いた構造化学習" *信学技報*, MBE88 - 144, 1988.
- (12) 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.
- (13) S. Akaho, "Regularization learning of neural networks for generalization," *Proc. Workshop on Algorithmic Learning Theory*, pp. 99 - 110, 1992.
- (14) 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.
- (15) 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.
- (16) 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.
- (17) 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.
- (18) R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, pp. 10 - 38, 1973.
- (19) V. Vapnic, *Statistical Learning Theory*, Wiley, pp. 25 - 42, 1998.