

ブートストラップサンプルを用いた
ニューラルネット識別器の設計
に関する研究

1999年3月

三谷芳弘

目次

1	序論	1
1.1	本研究の目的と意義	1
1.2	歴史的背景	5
1.3	本論文の構成	10
1.4	諸記号	11
2	ブートストラップサンプルを用いた学習	12
2.1	はじめに	12
2.2	ニューラルネット識別器	13
2.3	ブートストラップ法	16
2.4	計算機シミュレーション	22
2.4.1	実験方法	22
2.4.2	パラメータ r と S の影響	26
2.4.3	ブートストラップサンプル数の影響	29
2.4.4	訓練サンプル数の影響	31
2.4.5	次元数の影響	33
2.4.6	Bayes 誤識別率との関係	37
2.5	おわりに	40

3	最適化と分析	41
3.1	はじめに	41
3.2	交差確認法によるパラメータ最適化	42
3.3	実験計画法による分析	45
3.4	おわりに	51
4	性能評価	52
4.1	はじめに	52
4.2	他手法との比較	53
4.3	文字認識問題への適用	58
4.4	おわりに	67
5	結論	68
	謝辞	71
	参考文献	73
	付録	82

第 1 章

序論

1.1 本研究の目的と意義

パターン認識は古くは哲学・心理学の分野で議論されてきた。工学の分野では、計算機の誕生以来、計算機にパターン認識させるという多大な試みがなされてきた。近年の急速な計算機技術の発展に伴い、計算機によるパターン認識も容易に行われるものと期待されたが、いまだに人間のパターン認識には及ばない。それ故、パターン認識は、現在もなお数多くの研究者の興味をひき、活発に研究されている。

さて、パターン認識とは、任意に与えられたパターンを有限個のクラス (概念) のいずれかに対応づける機能である、と定義される [1]。この際、パターンの属すべきクラスが未知の場合と既知の場合がある。前者は教師なしの場合であり、与えられたパターンを類に分類する。一方、後者は教師ありの場合であり、狭義のパターン認識と呼ばれる。本研究はこの場合に主眼を置いている。

一般に、パターン認識過程は、観測系・前処理系・特徴抽出系・識別系の四つの処理系によりモデル化される (図 1.1 参照)。まず、外界に存在するパターンは、観測系による数値化によって計算機内部に取り込まれる。すなわち、外界のパターンが、計算機に入力可能である有限次元ベクトルとして表されることになる。次に前処理系により、パターンに対してノイズ除去・正規化の処理が行われる。続く特徴抽出系に

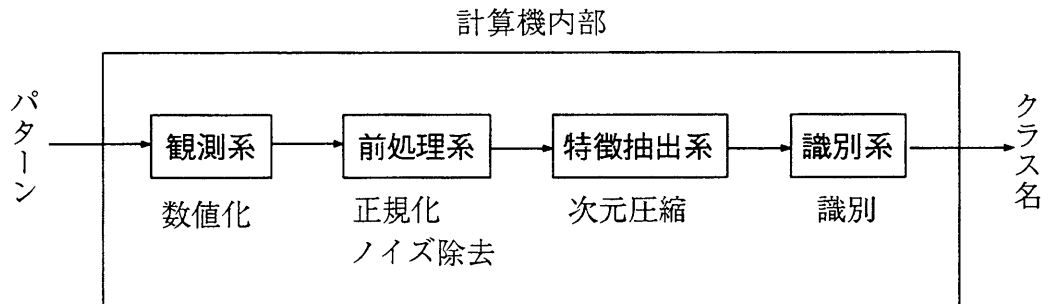


図 1.1: パターン認識過程

より識別に役立つ情報が抽出され、最後の識別系ではパターンの帰属するクラスへの決定が行われる。これら四つの処理系の中で、識別系以外の、観測系・前処理系・特徴抽出系は、それぞれ認識対象の性質に強く依存し、それ故その性質を反映した個別的な手法が数多く提案されている。一般にパターン認識系は、認識対象の性質を解明し、それに基づいて設計されなければならない。しかしながら、認識対象の性質の解明が困難な場合がある。例えば、文字認識においては人間の書く文字パターンの性質は長年の研究にも関わらず、いまだに解明されていない。このような問題に対して、パターンのなす統計的構造に着目し、それに基づいてパターン認識を解く理論がある。これは、統計的パターン認識理論と呼ばれ、これまで数多くの研究がなされている [2]～[7]。本研究では、統計的パターン認識理論のうち特に識別系の設計理論を検討する。

統計的パターン認識理論では、パターンのなす分布の統計的構造が既知の場合、誤識別率が最小という意味で最適な識別器は Bayes 識別器であることが知られている。しかし、実際のパターン認識問題では、パターン分布の統計的構造は未知である。そのため有限個のサンプルを用いて、識別器を設計しなければならない。ここで、識別器の設計に用いるサンプルを訓練サンプルと呼び、また識別器の評価、すなわち誤識別率の推定に用いるサンプルをテストサンプルと呼ぶ。識別器の設計法は、分布を仮

定するパラメトリックな手法と分布を仮定しないノンパラメトリックな手法とに大別される。

識別器の設計において, Jain ら [8] は, 以下に示される, 次元数に対するクラス当たりの訓練サンプル数の比 η が少なくとも 5 から 10 は必要であると指摘している。

$$\eta = \frac{\text{クラス当たりの訓練サンプル数}}{\text{次元数}}$$

しかしながら, 実際のパターン認識問題では, その比 η が 5 よりも小さい場合の方がむしろ多い。これは, パターンが表されるパターン空間の次元数は一般に大きく, その一方で人手によるラベル付けが必要とされるサンプルを入手するコストが高く, 実際には少数のサンプルしか得られない, という事情による。例えば, 手書き漢字認識問題では, 次元数が 500 から 1000 は必要であるのに対し, クラス当たりの訓練サンプル数を数百集めるのが限界となっている。そのような状況下において, 汎化能力の高い識別器をいかに設計するかということは, パターン認識における重要な課題の一つである。ここで, 汎化能力とは, 未学習パターンに対する認識能力を意味するものとする。また, 利用できるサンプル数が少ないという現実的な状況下で, 真に有効なパターン認識系の構築が実際の立場からも強く求められている。そのため実用に耐えうる理論の確立を目指して盛んに研究が行われるようになった [2]。

このような背景の中, 生理学的知見から得られたニューラルネットから, パターン認識の研究が盛んになってきた [9]~[23]。ニューラルネットは, 生体の神経細胞の働きを模倣し数理的にモデル化されたものである [24]。これまで, ニューラルネットは様々な分野において適用されてきた。中でもニューラルネットが識別器として用いられ, パターン認識の分野で, 数多くの成功例が報告された [9]~[13]。ニューラルネット識別器は, 分布の仮定を必要としないためノンパラメトリック識別器の一つであり, 隠れユニット数を増加させることで複雑な識別面が設計可能である。そのため, これまで提案された識別器の中でも有望な識別器の一つとして注目されている。

しかしながら、浜本 [2] が指摘するように、ニューラルネットの分野では従来の統計的パターン認識理論で培われた膨大な知見が活用されていない。これまでのニューラルネットに関する数多くの報告では、作為的とも思われる非現実的な分布が扱われている場合があり、その実用性に疑問が残る。あるいは、誤識別率を推定する場合などにおいて統計的に正当な評価がなされておらず、信頼性が乏しい、という問題もある。そこで、統計的パターン認識理論の枠組みにおいて、ニューラルネット識別器を汎化能力の向上という意味で改良することは、実際的なパターン認識理論の構築に大きな役割を果たすものと考えられる。

本研究の目的は、このような問題設定に基づいて、統計的パターン認識理論で培われた知見を基に、訓練サンプル数が少ないという状況下においても実用に耐えうるニューラルネット識別器の設計法を提案することである。

1.2 歴史的背景

前節では、パターン認識系の基本的構成を述べると共に、近年盛んに研究されているニューラルネットを統計的パターン認識理論の枠組みで議論することの重要性を指摘した。本節では、従来のニューラルネットの研究を概説し、本研究の位置付けを明確にする。

1943年、McCulloch と Pitts [24] が神経細胞 (ニューロンまたはユニット) のモデルを提案したのが、ニューラルネットにおける最初の研究である。生理学的な知見により数理的なモデル化が行われた。つまり、ニューロンの興奮状態を信号 1、非興奮状態を信号 0 と対応付けた。ニューロンが受けた刺激の総和が決められたしきい値を超えると興奮し信号 1 を出力し、それ以下ならば興奮せず信号 0 を出力する、と考え 0, 1 の 2 値入出力関数の機能を有するニューロンのモデル化を行った。1949年、Hebb [25] は、ニューロンが興奮すると入力部のシナプス結合のうち刺激を伝えたいものは結合強度が増加し更に刺激が伝えやすくなる、という説を唱えた。1950年代終り、Rosenblatt [26],[27] は、これらの原理を基にパターン認識する学習機械、いわゆるパーセプトロンを提案した。パーセプトロンは、入力に対して何らかの望ましい出力を得るような対応付けが行えることから、ニューラルネットにおける第 1 期のブームとなった。パーセプトロンの構造は階層型ネットワークであり、2 層あるいは 3 層ネットワークが考えられた。与えられたクラスが線形分離可能であると仮定して、パーセプトロンの学習が正しい分類をするところまで収束するのかという問題も、パーセプトロンの収束定理 [27] によって保証された。ところが、例えば、パターンベクトルの次元数が 2 のとき 2 入力 1 出力の 2 層パーセプトロンでは線形分離不可能な問題を解くことができない、などの問題があった。1969年、Minsky と Papert [28] は、2 層ネットワークが対応付けを行うための条件を解析し、実現できない対応付けが数多くあることを示した。Minsky らは、隠れ層を一つ有する 3 層ネッ

トワークを用いれば、線形分離不可能な問題でも望ましい対応付けが可能である、と指摘した。しかしながら、パーセプトロンの学習は層数に関わらず一つの層間のみ
の単純な学習則であることから、望ましい対応付けは困難であった。また Minsky ら
は、望ましい対応付けを行うには十分に数多くの隠れユニット数が必要である、と指
摘した。Minsky らの発表を機にニューラルネットの研究が下火となったが、1986 年、
Rumelhart ら [9] の誤差逆伝搬法 (以下、BP 法と呼ぶ。) の提案により第 2 期のブー
ムが到来した。BP 法は、訓練サンプルに対するネットワークの出力値と教師信号と
の平均 2 乗誤差の最小化を目指し、誤差を逆伝搬させながらユニット間の重みを更
新する、最急降下法型の学習法である。この学習法は、層数に依存せず、またユニッ
トの入出力関数に非線形性を持たせた点がパーセプトロンとは大きく異なる。BP 法
により、3 層ニューラルネットを用いることで、線形分離不可能な問題でも望ましい
対応付けが実現可能となった。また、3 層ニューラルネットに関して、船橋 [29] ら数
多くの研究者が、関数近似能力に優れていることを報告した。更に、Lippmann [12]
は、ニューラルネットはその構造を変えることにより統計的識別器を実現可能であ
ることを示した。1980 年代、Hopfield [30],[31] は相互結合型ネットワークを、Hinton
と Sejnowski [32] はボルツマンマシンを提案した。最近では、Kohonen [33] が自己組
織化マップ (SOM) や学習型ベクトル量子化法 (LVQ) を提案している。

さて、ニューラルネット識別器の汎化能力については、学習曲線からのアプロ
ーチが Amari ら [34],[35] によりなされている。文献 [34] では確率論的識別器につい
て、一方文献 [35] では決定論的識別器についての学習曲線の諸性質が論じられてい
る。本研究では、決定論的識別器に関心があるので、以下決定論的識別器について
のみ議論していくことにする。

まず、Amari ら [35] は学習曲線の多項式モデルを検討し、それとは独立に Baum [36]
は実験的に多項式モデルの精度を調べた。更に Cohn ら [37] は、指数モデルについ

ても言及し、学習曲線のモデルとして多項式モデル、指数モデルのいずれが良いのかは、与えられた問題に依存する、と結論した。Raudys [38] が指摘するように上記の論文 [35]～[37] では、あまりにも作為的で非現実的な分布を取り扱うため、それらの実際的な問題に対する有効性はほとんど期待できない。そのため、Baum [36] は、より現実的な分布に適用できる理論の構築を目指すべきである、と主張している。更に重要なことに、上記論文 [35]～[37] では、いずれも学習曲線の漸近的なふるまい、すなわち訓練サンプルが多数存在する場合に注目し、訓練サンプル数が少ない実際的な状況下における問題には何も注意を払っていない。

ニューラルネット識別器における少数サンプル問題の重要性を具体的に指摘したのは、Raudys と Jain [39] である。まず、Duin [40] は、ニューラルネット識別器における優れた特性は固有次元によると述べている。ここで、固有次元とはデータの本質的な次元数を意味する [42]。文献 [42] では、カルフーネン・レーベ展開 (K-L 展開) などのように線形変換に基づいた手法では固有次元を見つけるのは困難であることが指摘されている。Duin が予想したように、0,1 の 2 値入出力関数を有する 1 出力 2 層ネットワーク (以下、線形パーセプトロンと呼ぶ。) の汎化能力も固有次元に深く関係することを Raudys [43] は実験的に示した。更に、Raudys [43] は重みの初期化が線形パーセプトロンの汎化能力に影響を与えることを理論的に導出し、それを実験的に示した。また、Raudys [44] は文献 [43] の理論式から、汎化能力と次元数、訓練サンプル数との間の関係について次のような解釈を与えた。次元数の増加に伴い訓練サンプル数を $1/2$ 乗のオーダーで増加させれば、線形パーセプトロンの汎化能力は次元数の影響を受けない。これは、次元数の増加に伴い線形識別関数が線形のオーダーで増加させれば汎化能力が劣化しない、2 次識別関数については 2 乗のオーダーで増加させれば汎化能力が劣化しない、ということに比べ、線形パーセプトロンでは高次元において少ない訓練サンプル数で一定の汎化能力を得ることができることを

意味する。このことは、線形パーセプトロンの優れた特性の一つである。一方で、線形パーセプトロンでは、線形分離不可能な問題は解けないという問題がある。そのため、本研究では、実用上存在する線形分離不可能な問題も解きうる BP 法を用いた 3 層ニューラルネット (以下、単にニューラルネット識別器と呼ぶ。) について議論していく。

さて、あるレベルの誤識別率を達成するためには、パラメトリック識別器よりもノンパラメトリック識別器の方が、より多くの訓練サンプルを必要とすることが知られている [4]。これは、ノンパラメトリック識別器がパラメトリック識別器と異なり、パターン分布に関する確率密度関数の関数形などの情報を活用できないからである。Raudys ら [7] は、あるレベルの誤識別率を達成するためには、Parzen 識別器では次元数の増加と共に訓練サンプル数を指数関数のオーダーで増加させなければならない、という実験結果を示した。これに対し、Hamamoto ら [45] は、ノンパラメトリック識別器の一つであるニューラルネット識別器では、次元数の増加と共に訓練サンプル数を線形のオーダーで増加させても誤識別率が増加しない、という実験結果を報告した。これは、訓練サンプル数が限られ、しかも次元数が大きいという実際的な状況下では、ニューラルネット識別器の方が統計的識別器よりも有効である、ということを示唆している。このことから、ニューラルネット識別器の汎化能力を改善することの重要性が指摘される。

さて、統計学の分野において、Efron [46] は誤識別率の推定精度の向上を直接の目的としてブートストラップアプローチを提案した。以来、ブートストラップアプローチに関する数多くの研究がなされてきた [47]~[51]。ブートストラップアプローチの本質は、少ないサンプルから、分布に関するより多くの情報を積極的に得ようとすることである。さて、パターン認識における識別器は、与えられた訓練サンプルのみを用いて設計される。従って、与えられた訓練サンプルから分布の構造を、いかに抽

出するかが重要となる。ブートストラップアプローチとは、与えられたサンプルから、再標本化により分布に関する情報を積極的に抽出しうる一手法である。そこで、ニューラルネット識別器の設計にブートストラップアプローチを用いることにより汎化能力の改善が期待できる、と考えられる。

このような着想に基づいて、ニューラルネット識別器の汎化能力の改善のために、ブートストラップアプローチを用いることにする。本研究では、ニューラルネット識別器のための新しいブートストラップ法を提案する。計算機シミュレーションを通して、訓練サンプル数が少ない現実的な状況下で本手法の有用性を、BP法、他のブートストラップアプローチとの比較を通して検討し、更に文字認識に関する実験により本手法の実用性をも検討する [52]～[56]。

1.3 本論文の構成

本論文は、以下のように構成されている。

第2章では、ブートストラップサンプルを用いた設計法を提案する。まず、ニューラルネット識別器について説明する。次に、ブートストラップサンプルの生成法を提案する。更に、計算機シミュレーションを通して、訓練サンプル数が少ない状況下で本手法の有効性を様々な観点から検討する。すなわち、本手法のパラメータ r と S 、ブートストラップサンプル数、訓練サンプル数、次元数の誤識別率に与える影響、また誤識別率と Bayes 誤識別率との関係、を実験的に調べる。これらの実験結果より、パラメータ r と S が汎化能力に影響を与えることを明らかにする。

第3章では、パラメータの最適化と分析について述べる。まず、交差確認法によりパラメータ r と S を最適化する。次に、その最適化の妥当性を実験計画法により詳細に検討する。更に、 r と S の主効果、交互作用についても検討する。

第4章では、まず本手法と他手法との比較実験を述べ、更に大規模パターン認識問題として文字認識を例にとり、本手法の実用性を検討する。Efron の復元抽出法、Chernick らの凸ブートストラップ法および BP 法を比較の対象とする。実用性を検証するため、手書き数字認識問題に対してタイプの異なる2つの特徴上で本手法を適用し、他手法との比較実験を行う。

第5章は、結論である。本手法について総括し、本論文に関連する今後の課題と展望について述べる。

1.4 諸記号

本論文で用いる主な諸記号を以下に定義する.

p	:	パターン空間の次元数
N	:	クラス当たりの訓練サンプル数
m	:	隠れユニット数
r	:	近傍サンプル数に関するパラメータ
S	:	拡大縮小に関するパラメータ
n	:	クラス当たりのブートストラップサンプル数
ω_c	:	クラス c
C	:	クラス数
\mathbf{x}_i^c	:	ω_c の i 番目の訓練サンプル
\mathbf{y}_i^c	:	ω_c の i 番目のブートストラップサンプル
$X_c = \{\mathbf{x}_i^c\}_{i=1}^N$:	ω_c の訓練サンプル集合
$Y_c = \{\mathbf{y}_i^c\}_{i=1}^N$:	ω_c のブートストラップサンプル集合
$X = \{X^c\}_{c=1}^C$:	訓練サンプル集合
$Y = \{Y^c\}_{c=1}^C$:	ブートストラップサンプル集合
I_ρ	:	ρ 次の単位行列

第 2 章

ブートストラップサンプルを用いた学習

2.1 はじめに

ニューラルネット識別器は、ノンパラメトリック識別器の一つであり、隠れユニット数を増加させることで複雑な識別面が設計可能である。また、ニューラルネット識別器は、統計的パターン認識で提案された識別器と比べ、次元数の誤識別率へ与える影響が小さい、という長所が Hamamoto ら [45] によって報告された。これらからも、ニューラルネット識別器がこれまで提案された多くの識別器の中でも特に有望な識別器であることが分かり、それ故、その汎化能力を改善することに重要な意義がある。

統計学の分野で、Efron [46] は誤識別率の推定精度を改善するためにブートストラップアプローチを提案した。ブートストラップアプローチの本質は、少ないサンプルから、分布に関するより多くの情報を積極的に得ようとすることである。そこで汎化能力の改善のために、ブートストラップアプローチをニューラルネット識別器の設計に用いることが考えられる。

本章では、Efron の生成法とは異なる新しいブートストラップサンプル生成法を提案する。更に、本手法のパラメータ、訓練サンプル数、次元数の誤識別率へ与える影響、また誤識別率と Bayes 誤識別率との関係、を実験的に調べる。

2.2 ニューラルネット 識別器

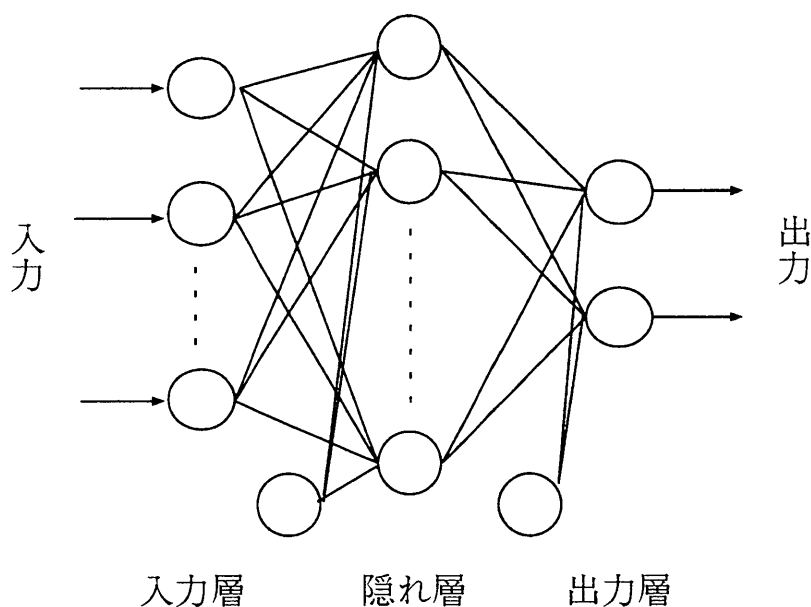


図 2.1: ニューラルネット 識別器

用いたニューラルネット識別器は、一つの隠れ層を有する全結合型の3層階層型ネットワークである(図 2.1参照). 入力層のユニット数はパターンベクトルの次元数に、出力層のユニット数はクラス数に、それぞれ一致する. さて、ニューラルネット識別器では隠れ層のユニット数が汎化能力に対して影響を与えることが知られている [14]. しかしながら、本研究では、隠れユニット数の最適化を議論せず、隠れユニット数を 8, 16 と定めた二種類のニューラルネット識別器をそれぞれ検討する. ニューラルネット識別器の学習には BP 法 [9] を用いる. BP 法は、ニューラルネット識別器のユニット間の重みを決定するための学習法である. 評価関数には、訓練サンプルに対するネットワークの出力値と教師信号との平均 2 乗誤差が用いられる. また、その最適化手法には、最急降下法が用いられる. その学習は、入力された訓練サンプルに対するネットワークの出力値と教師信号との誤差を基に、その誤差を逆伝搬させ

ながらネットワークの重みを更新する。全訓練サンプルがネットワークへ入力されると、学習回数を 1 とカウントする。これを、決められた停止条件となるまで、訓練サンプル集合が繰り返しネットワークへ入力される。この 1 回の学習の際、全訓練サンプルが入力された後に重みが更新される一括修正法と、一つの訓練サンプルが入力される度に重みが更新される逐次修正法とがある。いずれの修正法も、評価関数を最小化するためにネットワークの重みを探索することが可能である、という証明がなされている [15]。本研究では、逐次修正法を用いる。BP 学習において、議論を簡単にするため、慣性項は用いない。ユニットの入出力関数であるシグモイド関数の傾きを 1 とし、重みの初期値は $-0.5 \sim 0.5$ の一様乱数によって定める。学習の停止は、一定の学習回数を超えるとき、ネットワークの出力値と教師信号との平均 2 乗誤差が一定のしきい値以下のとき、あるいは前回の学習での平均 2 乗誤差と今回の学習での平均 2 乗誤差の差の絶対値が一定のしきい値以下のときとする。学習率は予備実験より 0.1 とする。

さて、ニューラルネット識別器を設計する際にいくつかの問題が存在することが知られている。まず過学習の問題がある。ニューラルネット識別器は、学習が進むにつれ、訓練サンプル集合に対して正しく識別するようになる。ところが、訓練サンプル集合に対してネットワークの重みを合わせ過ぎることにより、未学習のパターンを正しく識別できない、いわゆる過学習の問題が生じる。この問題は、学習が進むにつれ、訓練サンプルに対する誤識別率は単調に減少していくが、一方で、未学習のパターンに対する誤識別率は、はじめは減少するものの、ある学習回数を境に、逆に増大してしまう、というものである (図 2.2 参照)。次に局所解の問題がある。BP 法は、最急降下法に基づいているため、重みの初期値によっては局所解へ陥るため評価関数の最適解へ到達できない、という原理的な問題も含んでいる。また、学習スピードに関する問題もある。一般にネットワークの構造が複雑になるにつれ、学習に要する

計算コストが膨大となる。更に、ネットワークの構造をどのように決定すれば妥当であるか、というネットワークの構造決定問題もある。

本研究では、これら問題のうち、本質的と思われる汎化に関わる問題のみを論じることとする。前述したように、BP法にとって理想的な学習は、評価関数を最小にすることであり、すなわち訓練サンプルに対する出力値と教師信号とが一致することである。これは、ニューラルネット識別器が訓練サンプルをできるだけ正しく識別することを意味する。つまり、BP法そのものは、未学習のパターンに対しては何ら注意を与えない学習法である、といえる。訓練サンプルが数多く利用できるときは、多数の訓練サンプルを通してBP法により、未学習のパターンに対する識別能力、すなわち汎化能力が期待される。ところが、実際には、少数の訓練サンプルを用いてニューラルネット識別器を設計しなければならない。本研究では、そのような状況下で、汎化能力を改善するためのニューラルネット識別器の設計法を論じる。

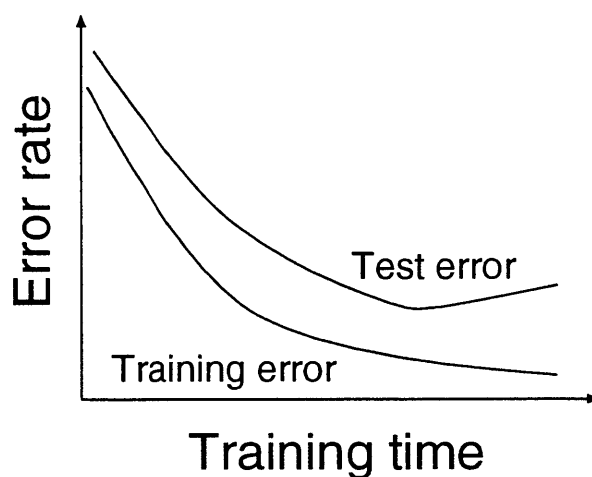


図 2.2: 過学習

2.3 ブートストラップ法

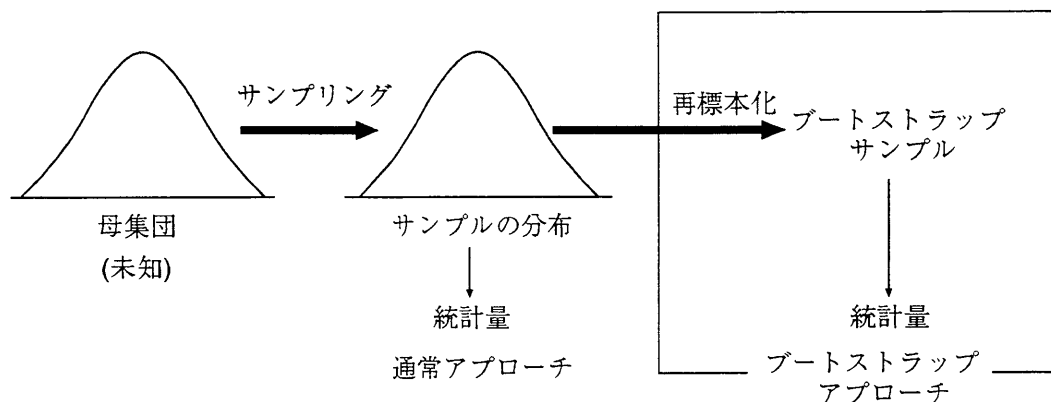


図 2.3: 通常のアプローチとブートストラップアプローチとの違い

Efron は、誤識別率の推定のためにブートストラップアプローチを提案した [46]. ブートストラップアプローチによる誤識別率の推定手順は、まずサンプルを再標本化することでブートストラップサンプルを生成し、次にブートストラップサンプルを用いて誤識別率を推定する、という二つの過程より構成される. このとき、再標本化によるブートストラップサンプルの生成法およびそのサンプルの用い方で様々な変形があり、誤識別率の推定法が数多く提案されている [48]~[51]. Chernick らは、訓練サンプル数が少ない状況下において、ブートストラップアプローチによる推定法を含む様々な誤識別率推定法の優劣を、真の誤識別率と誤識別率の推定値との比較に基づいて実験的に調べた [48]. 実験結果より、ブートストラップアプローチによる推定法の有効性が示された. しかしながら、どの分布に対しても有効なブートストラップアプローチはなく、実際には与えられたパターン認識問題に対して用いる推定法を使い分ける必要がある、ということが明らかにされた. また Jain らは、最近傍識別器に対し、誤識別率の推定値の分散は leave-one-out 法よりもブートストラップアプローチの方が小さいことを実験的に示した [50].

図 2.3はブートストラップアプローチの概略図を表す。サンプルは一般に未知である母集団から無作為に抽出される。通常、この抽出されたサンプルから様々な統計量が求められる。従来のニューラルネット識別器もこのサンプルを直接用いて設計される。さて、抽出されたサンプルは、何らかの分布をなす。ブートストラップアプローチでは、サンプルのなす分布を母集団の分布とみなして、再標本化によりサンプルを抽出する。この再抽出されたサンプルがブートストラップサンプルと呼ばれる。代表的な再標本化手法には、Efron の生成法 [46], [47], Chernick らの凸ブートストラップ法 [48] がある。本研究では、ブートストラップサンプルを生成する再標本化に着目し、Efron, Chernick らとは異なるブートストラップサンプル生成法を新しく提案する。

本研究では各クラスの事前確率は等しいものとし、ブートストラップサンプルはクラス毎に生成する。あるクラス $\omega_c (c = 1, 2, \dots, C)$ に対し N 個の訓練サンプルからなる訓練サンプル集合 $X^c = \{\mathbf{x}_1^c, \mathbf{x}_2^c, \dots, \mathbf{x}_N^c\}$ が与えられたとする。ブートストラップサンプル数を n とすると、ブートストラップサンプル集合 $Y^c = \{\mathbf{y}_1^c, \mathbf{y}_2^c, \dots, \mathbf{y}_n^c\}$ は以下の手順により生成される。但し、 n の値は設計者により任意に設定される。

Step1 訓練サンプル集合 X^c から訓練サンプルをランダムに選択し、それを $\mathbf{x}_{k_0}^c$ と表す。

Step2 ユークリッド距離を用いて、 $\mathbf{x}_{k_0}^c$ の最近傍である $r - 1$ 個の訓練サンプル $\mathbf{x}_{k_1}^c, \mathbf{x}_{k_2}^c, \dots, \mathbf{x}_{k_{r-1}}^c$ を X^c から選択する。ここで、パラメータ r は設計者により与えられる $2 \sim N$ の自然数である。

Step3 r 個の訓練サンプル $\mathbf{x}_{k_j}^c (j = 0, 1, \dots, r - 1)$ の平均ベクトル $\hat{\mathbf{x}}$ を求める。

$$\hat{\mathbf{x}} = \frac{1}{r} \sum_{j=0}^{r-1} \mathbf{x}_{k_j}^c \quad (2.1)$$

Step4 式 (2.2) に従って, ブートストラップサンプル \mathbf{y}_i^c を生成する.

$$\mathbf{y}_i^c = \sum_{j=0}^{r-1} \tau_j (\mathbf{x}_{k_j}^c - \hat{\mathbf{x}}) + \hat{\mathbf{x}} \quad (2.2)$$

ここで, 重み τ_j は式 (2.3) で与えられる.

$$\tau_j = S \frac{\delta_j}{\sum_{h=0}^{r-1} \delta_h}, \quad j = 0, 1, \dots, r-1 \quad (2.3)$$

式 (2.3) における δ_j の値は一様乱数 $[0, 1]$ によって与えられる. またパラメータ S は, 設計者により与えられる非負の実数値であり, $\sum_{j=0}^{r-1} \tau_j = S$ を満たす.

Step5 Step1 ~ Step4 を n 回繰り返すことにより, ブートストラップサンプル集合 Y^c を得る.

本ブートストラップサンプル生成法において, 設計者が設定すべきパラメータは, ブートストラップサンプル数 n , 近傍サンプル数 r , 拡大縮小パラメータ S の三つである. すなわち, n は自然数, r は $2 \sim N$ の自然数, S は非負の実数値の範囲から, 設計者がそれぞれの値を設定することになる.

本ブートストラップサンプル生成法では, 近傍となる r 個の訓練サンプルを線形結合することで, ブートストラップサンプルを人工的に生成する. 線形結合の際, 拡大縮小パラメータ S により, ブートストラップサンプルのなす分布が決定される. つまり, $S < 1$ のとき, 生成されるブートストラップサンプルの分布は, 訓練サンプルの分布に対し縮小された分布が与えられる. また, $S > 1$ のときは拡大された分布が与えられる. なお, $S = 0$ のときは選択された近傍 r 個の訓練サンプルの重心にブートストラップサンプルが生成される. 例として, 図 2.4 に $r = 3$ のときの S の値を変えた場合のブートストラップサンプルの生成領域を示す. 図 2.4 では, ブートストラップサンプルが, S の値によって決定される三角形内に生成される.

○ Training sample

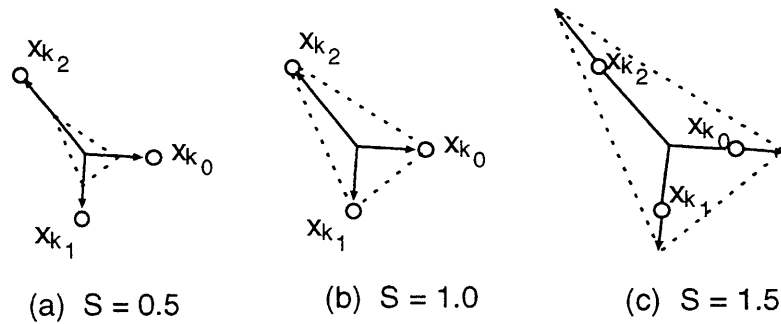
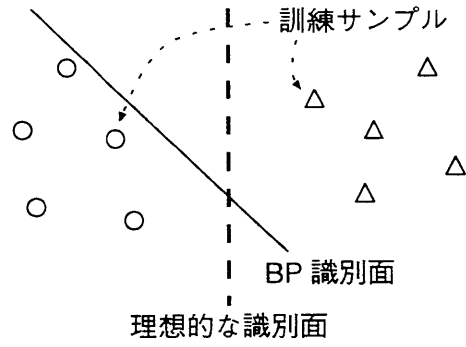
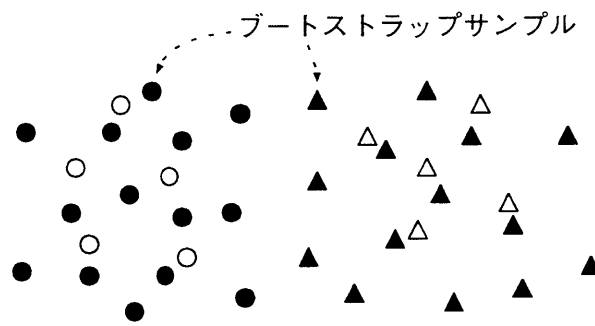


図 2.4: ブートストラップサンプルの生成領域例

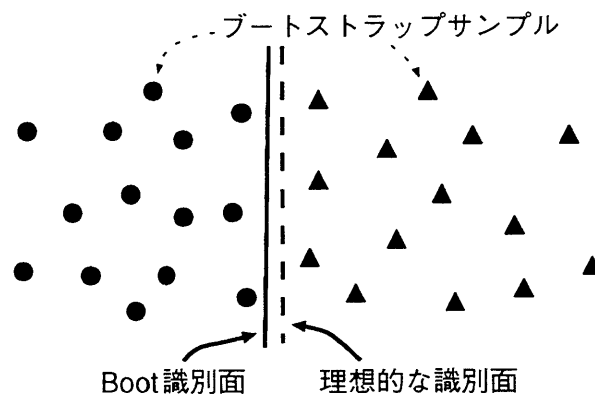
さて、本手法が提案された背景について述べる。BP 法では、訓練サンプルのみをできるだけ正しく識別しようとし、未学習パターンに対しては何の注意も与えない。それに対して、本手法のブートストラップサンプルは、訓練サンプルの背後に存在する母集団、すなわち未学習パターンをも想定しようとして試みている。例えば、ニューラルネット識別器の汎化能力を改善させるために次のことを考えた。図 2.5 に示されている例を用いて説明する。BP 法では、訓練サンプルのみを正しく識別しようとするため、例えば (a) に示される BP 識別面が設計されることがある。これに対し、(b) に示されるように、訓練サンプルからブートストラップサンプルを生成させ、これらのブートストラップサンプルを用いることで、BP 法により (c) の Boot 識別面が設計されることがある。将来入力されるパターンに対しては、(a) の BP 識別面より (c) の Boot 識別面の方が理想的な識別面に近いと思われる。すなわち、(b) に示すブートストラップサンプルを用いることができれば、汎化能力を改善させる可能性がある。そこで、本論文ではブートストラップサンプルの生成式である式 (2.2) に、拡大・縮小の機能をもたせ、また訓練サンプルを乱数により線形結合させることによってブートストラップサンプルを任意に生成させることを考えた。このように式 (2.2) には、パターン分布の構造を反映させることが考慮に入れられている。



(a) BP 法による識別面



(b) ブートストラップサンプルの生成



(c) ブートストラップ法による識別面

図 2.5: ブートストラップ法を用いる背景

Efron のブートストラップサンプル生成法では、ブートストラップサンプルは訓練サンプルから復元抽出により選択される。それゆえ、Efron の生成法を復元抽出法と呼ぶ。この復元抽出法では、元の訓練サンプルと同数か、あるいはそれより実質的に少数のサンプルしか得られず、汎化能力の改善が期待されにくい。これに対して、本ブートストラップサンプル生成法では、パラメータの値を変えることでパターン分布の構造がブートストラップサンプルに反映され、識別に有用なブートストラップサンプルを生成することができ、これらのブートストラップサンプルを用いることで結果的に汎化能力の改善が期待される。

従来の BP 学習では訓練サンプル集合 $X = \{X^c | c = 1, 2, \dots, C\}$ を用いてニューラルネット識別器を設計する (図 2.6 参照)。それに対して、本論文では X から生成されるブートストラップサンプル集合 $Y = \{Y^c | c = 1, 2, \dots, C\}$ を用いてニューラルネット識別器を設計する。この方法をブートストラップ法と呼ぶことにする。

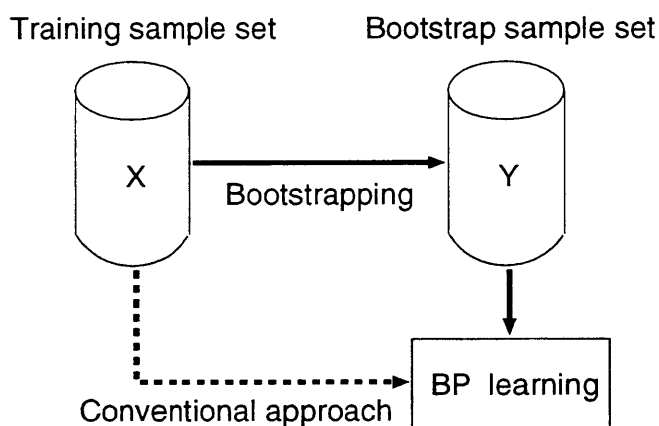


図 2.6: ブートストラップ法と従来法との比較

2.4 計算機シミュレーション

2.4.1 実験方法

本節では, 人工データを用いて, 訓練サンプル数が少ない現実的な状況下で本手法の有効性を様々な観点から検討する. 人工データでは, サンプル数を任意に変えられるので, 真の誤識別率を求めることができる. ここで, 真の誤識別率とは限られた訓練サンプルを用いて設計された識別器に, 大量のテストサンプルを入力して得られる誤識別率を意味する. このようにすることで, 訓練サンプル数が少ないという現実的な状況下で設計された識別器が, 将来入力されるパターンに対していかなる汎化能力を有するか, ということが正確に調べられることになる [45]. 前述したように, 識別器の設計には, 次元数に対するクラス当たりの訓練サンプル数の比 η が少なくとも 5 から 10 は必要である [8]. しかしながら, 実際のパターン認識問題には, その比 η が 5 よりも小さい場合の方がむしろ多い. 本研究では, そのような状況を設定し, ニューラルネット識別器の汎化能力を調べることにする.

用いた人工データを説明する.

- Marill-Green データ [57] のクラス 1 と 2

これは 8 次元正規分布の人工データである. 2 クラスの母平均ベクトルと母共分散行列が共に異なり, 誤識別率の理論的な限界である Bayes 誤識別率は 1.9% である [5].

この人工データは, 手書き文字を基にして作成されたものである. つまり, 手書き文字 A, B, C および D から抽出された 8 つの特徴に基づいている. 40 人の被験者が, 2 インチ×2 インチの枠内に, それぞれ 5 例ずつ手書きし, 一文字当たり 200 個のサンプルが得られる. 一つの文字パターンに対し, あるパスに沿って, 枠から文字のエッジまでの距離を特徴とする. すなわち, Marill-Green データとは, 側面距離特徴抽出法 [58] からの特徴を基にしたデータである. ここで, 側面距離特徴抽出法で

は、パスから図形に線を伸ばしたときに最初に交差する点までの距離を特徴とする。各標本点から線を伸ばすので、図形を側面から見たときの距離が特徴として得られる。つまり、側面距離特徴とは、側面からの図形の凹凸を表す特徴である。パスの取り方は、垂直軸あるいは水平軸が一般的であるが、図 2.7ではパスを中心から8方向に取っている。このように、一つの文字パターンに対し、予め決められたパスを用いることで、8つの特徴が求められる。つまり文字パターンは8次元パターンベクトルとして記述される。8次元パターンベクトルの標本平均ベクトルと標本共分散行列を求め、これらを正規分布の母平均ベクトルと母共分散行列とみなして、乱数により人工データが得られることになる。

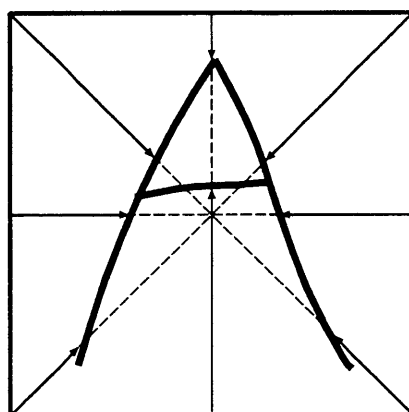


図 2.7: 側面距離特徴抽出

本論文では、クラス 1(文字 A) とクラス 2(文字 B) を用いる。このデータは、文献 [5] で標準データとして用いられ、パターン認識の分野ではよく知られたデータである。分布を定める母平均ベクトルと母共分散行列については、付録を参照されたい。

- Ness データ [59]

これも正規分布から抽出されたデータである。このデータは、文献 [45] で用いら

表 2.1: Δ および p の値と Bayes 誤識別率 (%) との関係

p		2	4	6	8
Δ	2	21.7	20.6	19.6	18.6
	4	6.2	5.9	5.7	5.5
	5	2.7	2.6	2.5	2.4
	6	1.06	1.02	0.98	0.95
	8	0.106	0.103	0.099	0.096

れ, 高次元空間におけるニューラルネット識別器の振る舞いが調べられている. この分布を定める母平均ベクトルと母共分散行列は次のとおりである.

$$\begin{aligned} \boldsymbol{\mu}_1 &= [0, \dots, 0]^T & \boldsymbol{\mu}_2 &= [\Delta/2, 0, \dots, 0, \Delta/2]^T \\ \Sigma_1 &= I_p & \Sigma_2 &= \begin{bmatrix} I_{p/2} & 0 \\ 0 & \frac{1}{2}I_{p/2} \end{bmatrix} \end{aligned}$$

ここで, クラス ω_i の母平均ベクトルを $\boldsymbol{\mu}_i$, クラス ω_i の母共分散行列を Σ_i と表す. Δ はパターン分布間の距離に関するパラメータであり, 擬似マハラノビス距離を表すと解釈される. このデータでは Δ および p の値を変えることによって, Bayes 誤識別率を変えることができる. ここで, Bayes 誤識別率とは, 誤識別率の理論的境界を表す. つまり, いかなる識別器も Bayes 誤識別率を下回る誤識別率を与えることはできない, ということである. Bayes 誤識別率を変えることは, パターン認識問題の難易度を自由に設定できる, ということを意味する. 具体的には, Δ の値が小さいときは, パターン分布が大きく重なり, Bayes 誤識別率が大きくなることで問題が難しくなる. 一方, Δ の値が大きいたときは, パターン分布が離れ, Bayes 誤識別率が小さくなることで問題が易しくなる. 表 2.1に, 実験で用いた Δ および p の値と Bayes 誤識別率との一例を示す. なお Bayes 誤識別率は文献 [60] の手法で計算された.

本実験では、モンテカルロ解析により、図 2.8に示すように、試行毎に母集団から訓練サンプルおよびテストサンプルをそれぞれ独立に生成し、真の誤識別率を推定した。実験では、テストサンプル数を 1000/クラスとした。また、モンテカルロの試行回数を 10 として、真の誤識別率の平均値 E を求めた。以降簡単のため、誤識別率の平均値を単に誤識別率と呼ぶことにする。

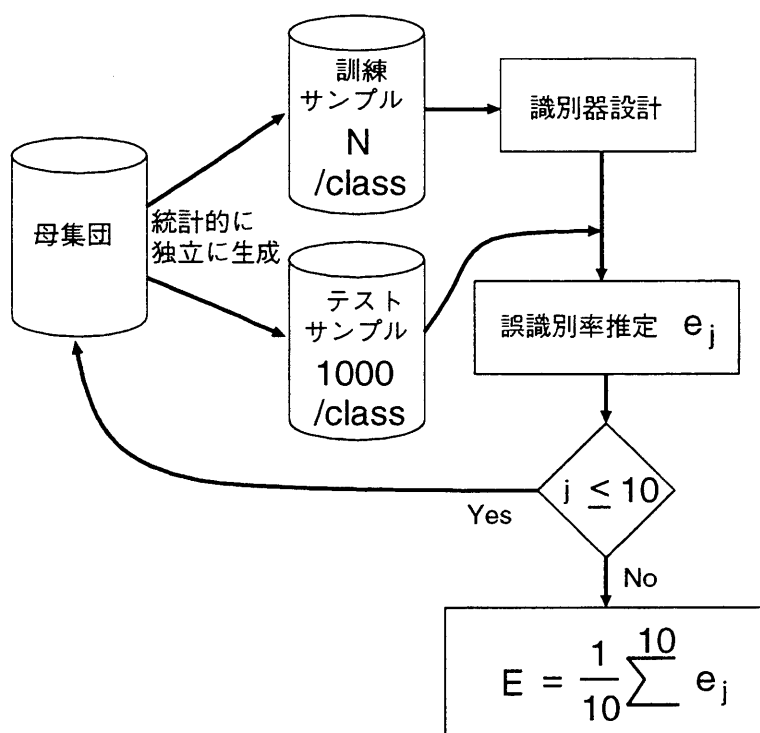


図 2.8: 真の誤識別率の推定

2.4.2 パラメータ r と S の影響

本小節の実験の目的は、パラメータ r と S が誤識別率に与える影響を調べることである。Marill-Green データを用いて、パラメータ r と S の値を変え、本手法の誤識別率を求めた。但し、ブートストラップサンプル数 n は予備実験から定めた。また、比較のため訓練サンプルを直接用いる従来法 (BP 法) の誤識別率も求めた。実験内容を整理すると次のとおりである。

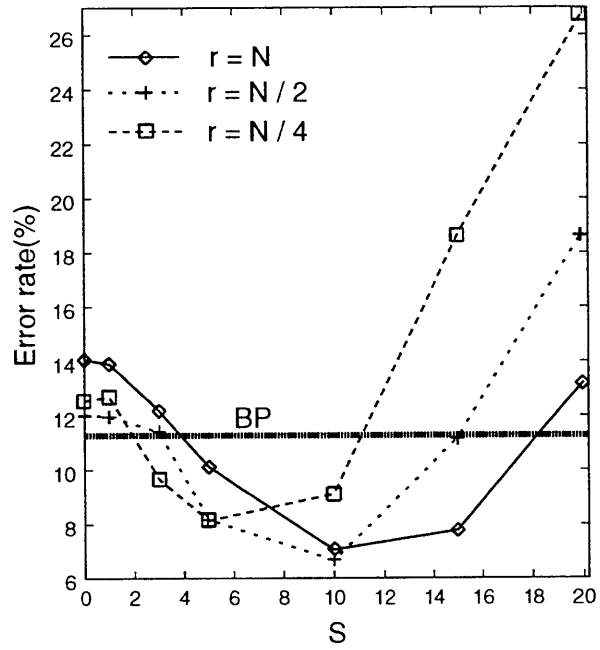
データ	: Marill-Green データ
訓練サンプル数 N	: 32/クラス
ブートストラップサンプル数 n	: 30 N /クラス
r の値	: $N/4, N/2, N$
S の値	: 0, 1, 3, 5, 10, 15, 20

図 2.9 は、様々な r の値に対して、パラメータ S の関数としての誤識別率を表す。(a) と (b) は、それぞれ隠れユニット数 m が 8 と 16 のときの結果を表す。図中の BP は直接訓練サンプルを用いて設計されたニューラルネット識別器の誤識別率を表す。これより、 r および S の値が適当に選択された場合、隠れユニット数 m に依存せず、本手法では BP 法よりも低い誤識別率を与えることができた。

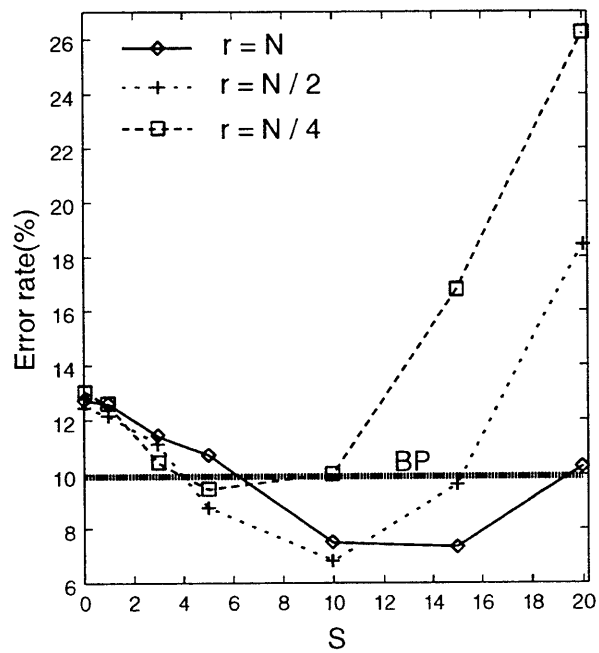
さて、実験結果から、 $r = N$ のとき、隠れユニット数 m に依存せず、誤識別率を改善するという意味で S の選択範囲が最大となることが分かる。すなわち、 $r = N$ のとき、 S の値を定めるのに最大のマージンが与えられていることになる。従って、以降では、 $r = N$ として議論を進めることにする。

前述したように、 S の値が小さいとブートストラップサンプルが縮小された領域で生成され、また S の値が大きいと拡大された領域で生成されることになる。図 2.9 では、小さな、あるいは大きな S の値によってブートストラップサンプルが不適切な位置に生成され、それらのブートストラップサンプルを用いることで結果的に BP

法よりも汎化能力が劣化している. 適切に S の値が定められると, 誤識別率が最小となることが (a), (b) 共に見られる.



(a) 隠れユニット数 $m = 8$



(b) 隠れユニット数 $m = 16$

図 2.9: 近傍パラメータ r と 拡大縮小パラメータ S の影響

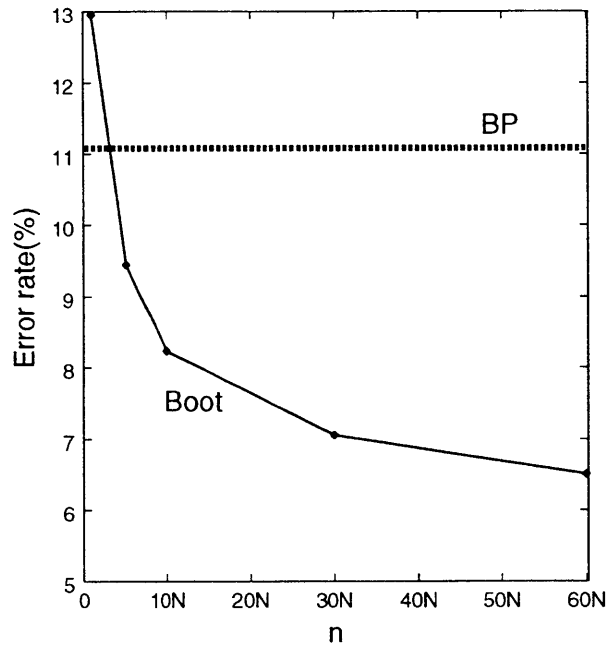
2.4.3 ブートストラップサンプル数の影響

本小節の実験の目的は、ブートストラップサンプル数 n が誤識別率に与える影響を調べることである。Marill-Green データを用いて、 n の値を変え、本手法の誤識別率を求める。また、BP 法の誤識別率も比較のため求めた。実験内容を整理すると次のとおりである。

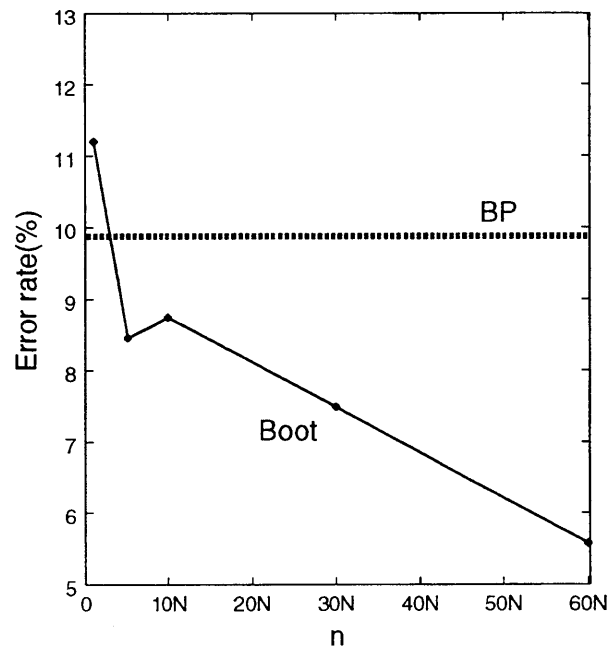
データ	: Marill-Green データ
訓練サンプル数 N	: 32/クラス
ブートストラップサンプル数 n	: $N, 5N, 10N, 30N, 60N$ /クラス
r の値	: N
S の値	: 10

但し、 S の値は前小節の結果を基に定めた。

図 2.10 は、Marill-Green データについて、ブートストラップサンプル数 n に対する誤識別率を表す。図中の *Boot* は本手法の誤識別率を表す。これより、隠れユニット数に依存せず、 n の値の増加と共に、本手法では誤識別率が減少していることが分かる。設計コストが許される限り、一般論としてはブートストラップサンプル数 n を多く取ることが望ましい。しかしながら、ブートストラップサンプル数 n の増加と共に計算コストも明らかに増大する。従って、本研究では、 $n = 30N$ として以降の議論を進めることにする。



(a) 隠れユニット数 $m = 8$



(b) 隠れユニット数 $m = 16$

図 2.10: ブートストラップサンプル数 n の影響

2.4.4 訓練サンプル数の影響

本小節の実験の目的は、訓練サンプル数 N が誤識別率に与える影響を調べることである。Marill-Green データを用いて、 N の値を変え、本手法と BP 法の誤識別率をそれぞれ求めた。実験内容を整理すると次のとおりである。

データ : Marill-Green データ
訓練サンプル数 N : 16, 24, 32, 40/クラス
 r の値 : N
 S の値 : 0, 1, 3, 5, 10, 15, 20

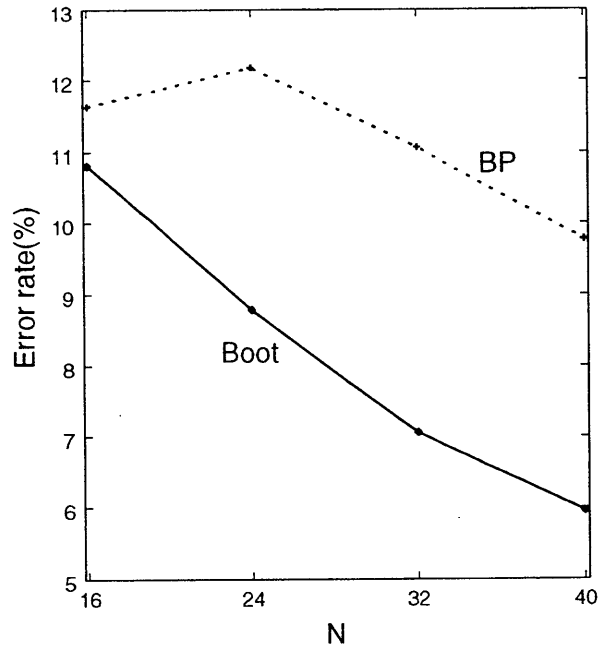
但し、 S の値は、テストサンプルに対する誤識別率を最小にするという観点から選択されている。

図 2.11 は、Marill-Green データについて、訓練サンプル数 N に対する誤識別率を表す。これより、 N の値とは無関係に、本手法では BP 法よりも低い誤識別率を与えた。また、この結果は、隠れユニット数に依存しないことが判明した。表 2.2 は、訓練サンプル数 N に対する S の最適値を表す。これより、概ね S の最適値は 10 であると考えられる。

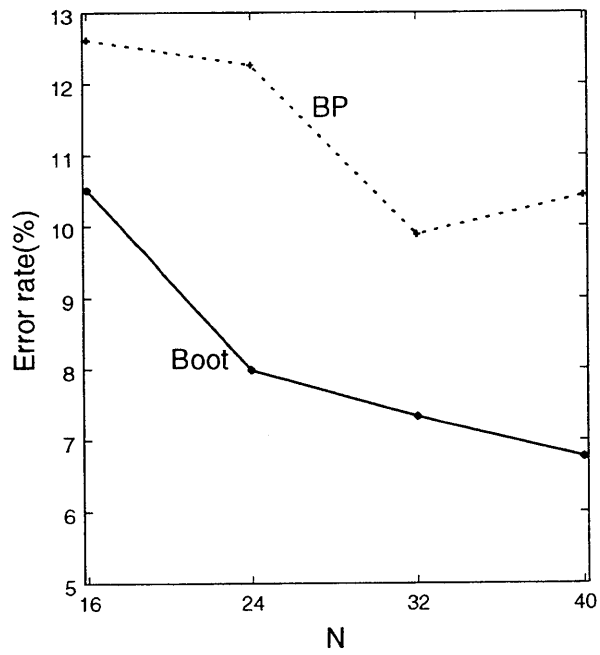
訓練サンプル数が少ない、すなわち次元数に対するクラス当たりの訓練サンプル数の比 η が小さい状況下において、本手法により誤識別率が改善することが示されたことは、パターン認識理論の意味において重要であると言える。

表 2.2: 訓練サンプル数 N と S の最適値との関係

N		16	24	32	40
S	$m = 8$	10	10	10	10
	$m = 16$	10	10	15	10



(a) 隠れユニット数 $m = 8$



(b) 隠れユニット数 $m = 16$

図 2.11: 訓練サンプル数 N の影響

2.4.5 次元数の影響

本小節の実験の目的は、次元数 p が誤識別率に与える影響を調べることである。訓練サンプル数が一定の下で次元数を増加させると汎化能力が劣化することが知られているため、高次元空間における振る舞いを調べ、BP 法に対する優位性を検討する。次元数を任意に変えることのできる Ness データを用いて、 p の値を変え、本手法と BP 法の誤識別率をそれぞれ求める。実験内容を整理すると次のとおりである。

データ : Ness データ

Δ の値 : 2, 8

次元数 p : 2, 4, 6, 8

訓練サンプル数 N : $4p/\text{クラス}$

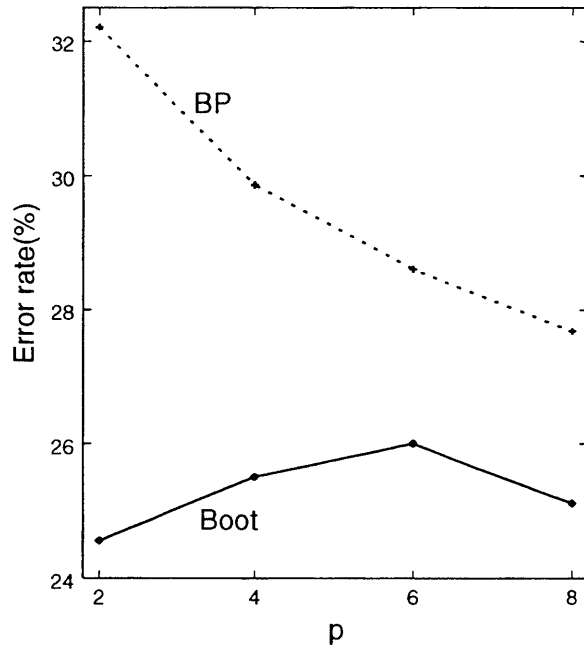
r の値 : N

S の値 : 0, 1, 3, 5, 10, 15, 20

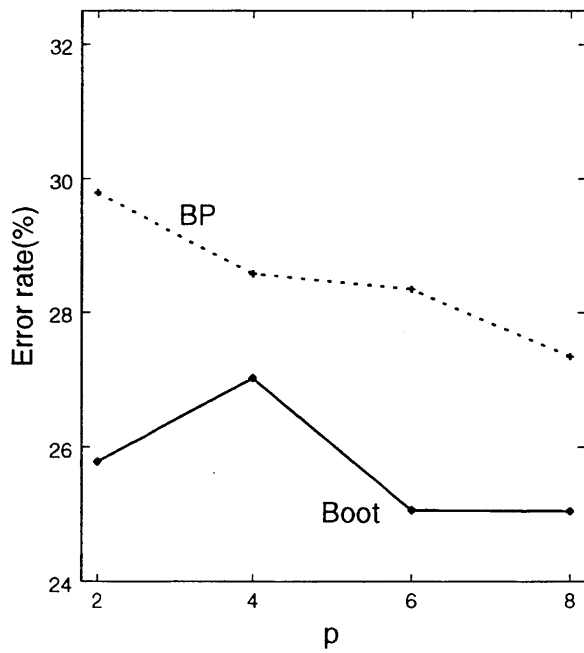
但し、 S の値は、テストサンプルに対する誤識別率を最小にするという観点から選択されている。

図 2.12 と図 2.13 は、Ness データについてそれぞれ $\Delta = 2$ および $\Delta = 8$ としたときの、次元数 p と誤識別率との関係を表す。これらより、 p の値に依存せず、本手法では BP 法よりも低い誤識別率を与えることができる。また、これらの結果は隠れユニット数 m に依存しない。表 2.3 は次元数 p に対する S の最適値を表す。これより、 $\Delta = 2$ のとき、0, 1 あるいは 3 が S の最適値として選択された。一方、 $\Delta = 8$ のとき、5 あるいは 10 が S の最適値として選択された。これより、Bayes 誤識別率が大きい (18.6%~21.7%) 程 S の値を小さく、一方 Bayes 誤識別率が小さい (0.096%~0.106%) 程 S の値を大きくすれば、汎化能力が改善する傾向にある、と考えられる。

パターン認識では一般に誤識別率に与える次元数の影響は大きい。BP 法と比較して次元数に依存せず誤識別率が改善されることは実用上意義があると言える。

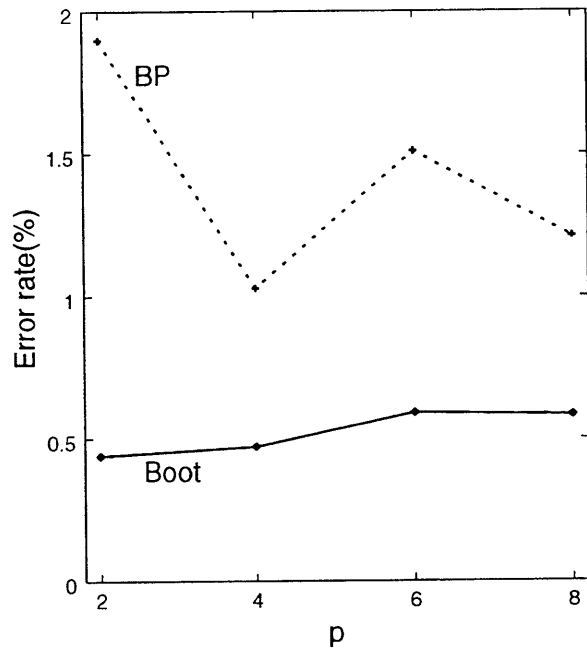


(a) 隠れユニット数 $m = 8$

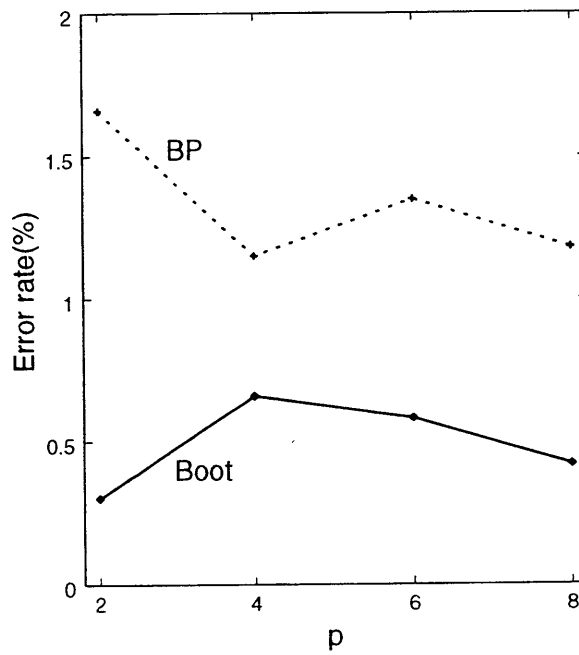


(b) 隠れユニット数 $m = 16$

図 2.12: $\Delta = 2$ のときの次元数 p の影響



(a) 隠れユニット数 $m = 8$



(b) 隠れユニット数 $m = 16$

図 2.13: $\Delta = 8$ のときの次元数 p の影響

表 2.3: 次元数 p と S の最適値との関係

		p				
		2	4	6	8	
S	$\Delta = 2$	$m = 8$	1	0	0	1
		$m = 16$	0	0	0	3
	$\Delta = 8$	$m = 8$	5	5	10	10
		$m = 16$	5	5	10	10

2.4.6 Bayes 誤識別率との関係

本小節の実験の目的は, Bayes 誤識別率と本手法との関係を調べることである. Bayes 誤識別率を自由に設定できる Ness データを用いて, 本手法と BP 法の誤識別率をそれぞれ求める. 実験内容を整理すると次のとおりである.

データ	: Ness データ
Δ の値 (Bayes 誤識別率)	: 2(18.6%), 4(5.5%), 5(2.4%), 6(0.95%), 8(0.096%)
次元数 p	: 8
訓練サンプル数 N	: 32/クラス
r の値	: N
S の値	: 0, 1, 3, 5, 10, 15, 20

但し, S の値は, テストサンプルに対する誤識別率を最小にするという観点から選択されている.

改善度 F を式 (2.4) で定義する. これは, 本手法が BP 法と比較して誤識別率をどの程度に改善するのかを表す.

$$F = \frac{BP}{BP + Boot} \times 100 \quad (2.4)$$

ここで, BP および $Boot$ は BP 法および本手法の誤識別率の平均値をそれぞれ表す. F は, $F > 50$ のとき本手法は BP 法よりも汎化能力の意味で優れ, 一方 $F < 50$ のとき本手法は BP 法よりも汎化能力の意味で劣る, ということを意味する. 更に, F の値が大きいく程, 本手法では BP 法よりも汎化能力が改善されることを意味する.

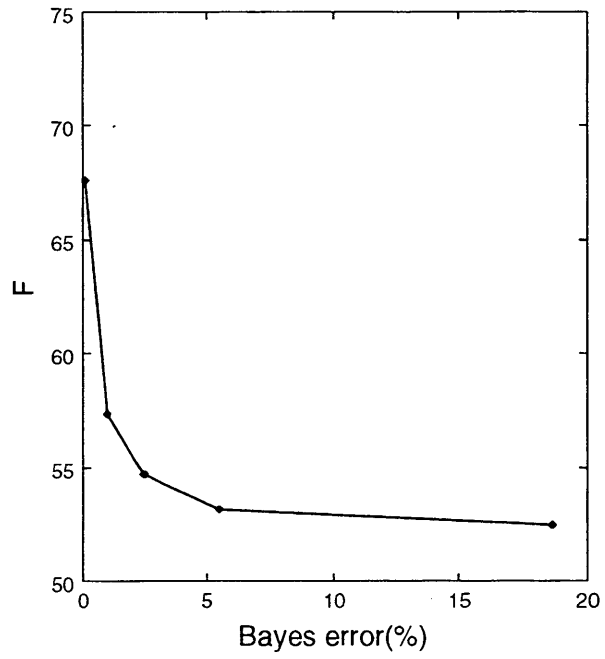
図 2.14 は, Ness データについて, Bayes 誤識別率に対する改善度 F を表す. これより, Bayes 誤識別率に依存せず, 本手法では BP 法よりも低い誤識別率を与える. また, Bayes 誤識別率が小さい程, 本手法の改善度は大きい. 更に, この結果は隠れユ

ニット数に依存しない. 表 2.4は, Bayes 誤識別率に対する S の最適値を表す. これより, Bayes 誤識別率が大きい程 S の値を小さく, 一方 Bayes 誤識別率が小さい程 S の値を大きくすれば, 汎化能力が改善する傾向にある, という前小節の予想が検証されることになった.

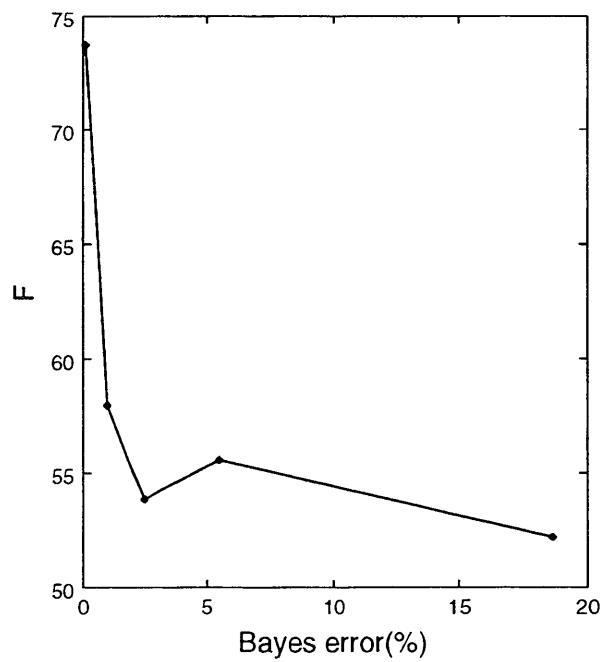
本実験から, S の値と Bayes 誤識別率との関係が明らかとなり, 設計者にとって S の値を決定する際の一つの指針が与えられたと言える.

表 2.4: Bayes 誤識別率と S の最適値との関係

Bayes 誤識別率		0.096%	0.95%	2.44%	5.49%	18.64%
S	$m = 8$	10	10	5	3	1
	$m = 16$	10	10	5	5	3



(a) 隠れユニット数 $m = 8$



(b) 隠れユニット数 $m = 16$

図 2.14: Bayes 誤識別率と改善度 F との関係

2.5 おわりに

第2章では, 汎化能力の改善のためのブートストラップ法によるニューラルネットワーク識別器の設計法を提案した. 計算機シミュレーションを通して, 訓練サンプル数が少ない状況下で本手法の有効性を様々な観点から検討した. その結果, 本手法のパラメータ r と S は, 汎化能力に対して重大な影響を与えることが明らかになった. 更に, 本手法のパラメータ r と S が適当に定められると, 訓練サンプル数, 次元数, Bayes 誤識別率に関わらず, BP 法よりも汎化能力が改善されることが分かった.

結局, パラメータ r と S は, 誤識別率を最小にするという意味で最適化されなければならない. 次章では, パラメータの最適化およびその妥当性について論じることにする.

第 3 章

最適化と分析

3.1 はじめに

前章より, パラメータ r と S がニューラルネット識別器の汎化能力に影響を与えることが明らかとなった. 汎化能力を問題とするため, パラメータの最適値は訓練サンプルのみを用いて推定されなければならない.

本章では, 誤識別率を最小にするという意味で, パラメータ r と S の最適化を行う. まず, 交差確認法 [61] によるパラメータの最適化について説明する. 次に, 実験計画法 [62] を用いて, 交差確認法を用いることの妥当性を検討する. 更に, パラメータ r と S が誤識別率に与える影響も, 実験計画法を用いて主効果, 交互作用の両面から分析する.

なお分析では, 真の誤識別率を用いる必要があるため, 人工データである Marill-Green データ [57] を用いることにする.

3.2 交差確認法によるパラメータ最適化

パラメータ最適化については、様々なアプローチが考えられる。ここでは、誤識別率を最小にするという意味で最適化を行う。また、パラメータの最適値はデータに依存するため、与えられたデータ毎にパラメータを最適化することにする。この誤識別率は訓練サンプルのみを用いて推定しなければならない、という制約がある。限られた訓練サンプルを有効に活用でき、しかも実行可能な手法として交差確認法 [61](Cross-Validation 法, 以下 CV 法と呼ぶ) を用いることにする。本節では、CV 法によるパラメータ r と S の最適化について論じる。CV 法による誤識別率の推定手順は次のとおりである。

Step1 すべてのクラスに対して、 N 個/クラスの訓練サンプルを改めて訓練用と評価用のサンプルにランダムに二分割する。すなわち、 $N - N_T$ 個/クラスを訓練用サンプル、 N_T 個/クラスを評価用サンプルとする。

Step2 訓練用サンプルを用いて識別器を設計する。

Step3 評価用サンプルを用いてその識別器の誤識別率 e_j を推定する。

Step4 Step1~3 の試行を独立に J 回繰り返す、CV 法の誤識別率 $E_{CV} = \frac{1}{J} \sum_{j=1}^J e_j$ を得る。

上記の推定量 E_{CV} は N_T と J の値に依存する。サンプルの利用効率を高めるため N_T の値を小さく設定することが考えられる。本実験では $N_T = 1$ とした。一方、 J については E_{CV} の推定精度を高めるため J の値を大きく設定する方が望ましい。本実験では $J = 20$ とした。

図 3.1 に最適化すべきあるパラメータ α と誤識別率との関係図を表す。実線は CV 法の誤識別率 $E_{CV}(\alpha)$ を、また破線は真の誤識別率 $E_{True}(\alpha)$ を表す。ここで、 $E_{CV}(\alpha)$

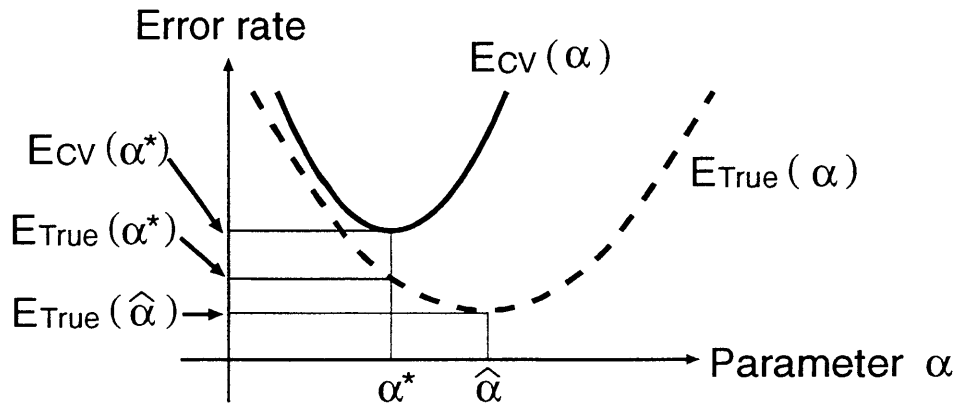


図 3.1: パラメータに対する誤識別率

は訓練サンプルのみを用いて得られたこと, また $E_{True}(\alpha)$ は未知であることに注意されたい. なお, 未知とは, 設計時にコンピュータにとってクラス名が未知である, ということの意味する. また, パラメータ α に対して $E_{CV}(\alpha)$ と $E_{True}(\alpha)$ の推定で用いる N 個/クラスの訓練サンプルは, 試行毎に同一とした.

最適解 $\hat{\alpha}$ は次式を満たす.

$$E_{True}(\hat{\alpha}) = \min\{E_{True}(\alpha)\} \quad (3.1)$$

一方, 式(3.2)に示されているように, CV法によるパラメータ最適化とは, $\min\{E_{CV}(\alpha)\}$ を与えるパラメータ値 α^* を解とすることである. すなわち,

$$E_{CV}(\alpha^*) = \min\{E_{CV}(\alpha)\} \quad (3.2)$$

である.

α^* が $\hat{\alpha}$ と一致すれば, CV法は最適化法として問題がない. 実際には α^* と $\hat{\alpha}$ とは一致しない場合があり, α^* を用いて設計された識別器は, 将来入力されるパターンに対して $E_{True}(\alpha^*)$ なる誤識別率を与えることになる. このとき, $E_{True}(\alpha^*)$ と $E_{True}(\hat{\alpha})$ との差が許容できるか否かが問題となる. また, $E_{True}(\hat{\alpha})$ (未知) はこのモ

デルにおける最高性能を表す。さて、最小値 $E_{CV}(\alpha^*)$ が複数個存在する場合については、候補パラメータ α^* 内でネットワークの出力値と教師信号との平均二乗誤差が最小となるパラメータ値を解とする。実際にはパラメータは r と S の二つであるから、パラメータ空間は2次元である。結局2次元パラメータ空間内で解 (r^*, S^*) が求められることになる。

3.3 実験計画法による分析

CV法により求められた解 (r^*, S^*) が、最適解 (\hat{r}, \hat{S}) と一致するとは限らない。そこで実験計画法 [62] を用いて CV 法の妥当性を検証することにする。

以下、実験計画法を概説する。実験計画法は、数理統計学の一応用手法であり、一般には効率の良い実験計画の立て方と、実験データの分析法の両輪から成り立つ [62], [63]。Fisher [64] によって、その基礎が築かれ、はじめは農業試験に主として用いられ、ついで工場実験にも取り入れられ、最近では官能検査や市場調査研究などの分野でも用いられるようになった。

まず、実験計画の立て方について述べる。いかなる実験においても、その実験の成否は、比較のために取り上げた処理条件 (因子と水準) が適切であったか否かによってほとんど決まる。それゆえ、因子と水準の選び方は、実験計画において重要であり、可能なら予備実験などを行ない周到に実験計画を立てなければならない。ここで、因子とは原因系のうち実験で意識的に取り上げて比較されるものである。また、水準とは因子のとり種々の条件である。例えば、因子として識別器を取り上げたとすれば、その水準は最近傍識別器、ユークリッド距離識別器、ニューラルネット識別器などとなる。実験の際には、それぞれの水準の下で、その他の原因系の条件は全て一定に保たなければならない。実験計画法により、因子の効果を表す主効果や、因子同士の組合せにより生じる効果を表す交互作用が、分析できる。主効果と交互作用を合わせて要因と呼ぶ。因子数、反復数が共に 2 以上では、因子同士の交互作用が求められる。例えば、3 因子 (A, B, C) の場合では、二因子交互作用 $A \times B, A \times C, B \times C$ 、さらに三因子交互作用 $A \times B \times C$ が求められる。但し、3 因子以上の交互作用は技術的に解釈が難しいとされる。

実験精度を向上させるため、誤差を減少させる必要がある。それゆえ、Fisher は、反復・無作為化・局所管理の三原則に基づいて実験を行うことを提唱している。代

表的な実験計画の一手法である乱塊法は、この Fisher の三原則を満たす。乱塊法では、ブロックが構成され、各ブロック内を局所管理することで系統誤差がブロック因子に吸収され、実験精度が高められる。ここで、系統誤差とは偶然誤差 (偶然的に起こって自ら統計的分布法則に従うような誤差) 以外の誤差のことである。本研究では、実験計画に乱塊法を用いる。

次に、実験データの分析法について述べる。実験で得られた測定値は、一般平均・主効果・交互作用 (因子数, 反復数共に 2 以上の場合に得られる)・ブロック効果 (乱塊法の場合に得られる)・誤差の和としてモデル化される。実験計画法では、誤差は正規分布 $N(0, \sigma^2)$ に従うものと仮定される。すると、測定値は、その平均は一般平均・主効果・交互作用・ブロック効果の和であり、その分散は σ^2 である正規分布に従うことになる。実験計画法では、データの分析のために、測定値から誤差が抽出される。抽出された誤差を基に、2 つの平均値差の検定や分散分析などによりデータが分析されることになる [62]。

では、実際に実験計画法を用いて CV 法の妥当性を検証することにする。 $E_{True}(r^*, S^*)$ と $E_{True}(\hat{r}, \hat{S})$ との差が統計上有意味な差でなければ、CV 法は最適化法として用いることができる。有意な差であるか否かを、実験計画法における最小有意差法により検証する。実験内容を整理すると表 3.1 のとおりであり、パラメータの組毎に真の誤識別率の 5 回の平均値を求めた。一方 CV 法により、各パラメータの組 (r, S) に対し $E_{CV}(r, S)$ を推定した。

表 3.2 は、CV 法によって求められた解 (r^*, S^*) を試行毎に表す。表 3.3 は各パラメータの組に対する真の誤識別率の 5 回の平均値を昇順に表す。表 3.2, 3.3 より、隠れユニット数 m が 8 のとき、5 回の独立試行中 5 回共 5%水準で 1 位の誤識別率と有意差がなかった。一方 $m = 16$ では、5 回の独立試行中 3 回は 5%水準で、2 回は 1%水準で有意差がなかった。以上の実験結果から、隠れユニット数 m に依らず CV 法に

より最適化は可能である, といえる.

また, パラメータ r と S が誤識別率に与える影響を調べるために, 表 3.1 の実験内容に従い分散分析を行った. 実験配置は, 反復毎に N 個/クラスの訓練サンプルと 1000 個/クラスのテストサンプルをブロック因子とし, パラメータ r と S をそれぞれ因子とした, 5 反復の乱塊法による二因子要因実験を実行した.

分散分析表を表 3.4 に示す. 表 3.4 より, 隠れユニット数 m に依らず, パラメータ r と S の主効果は共に 1% 有意であり, 更に Marill-Green データについては S の効果の方が r よりも大であった. また, r と S の交互作用も 1% 有意であることも分かる. 結局, r と S は個別に最適化できず, それらの組合せの効果をも考慮に入れなければならない, ということが判明した.

更に, CV 法による最適化の限界を誤識別率の観点から検討する. 上述の実験条件と同じく, 表 3.1 の実験内容に従い, CV 法により得られた誤識別率 $E_{True}(r^*, S^*)$ と最高性能を表す $E_{True}(\hat{r}, \hat{S})$ を推定した. 但し, 試行回数は 5 であり, 誤識別率の平均値をそれぞれ求めた.

表 3.5 は最適化の限界を表す. これより, $E_{True}(r^*, S^*)$ と $E_{True}(\hat{r}, \hat{S})$ の差は, $m = 8$ のとき 1.32%, $m = 16$ では 1.11 % であり, その差は僅少であるとは言えない. 最小有意差法によって CV 法の妥当性が示されているものの, パラメータ最適化については更なる検討が必要である. しかしながら, 実際には少数の訓練サンプルのみを用いるパラメータ最適化は基本的にかなり困難であり, 実行可能な手法として CV 法を用いることは現実的であると考えられる.

表 3.1: 最適性の検証のための実験条件

データ集合	Marill-Green
訓練サンプル数 N /クラス	32
テストサンプル数 /クラス	1000
r の候補値	8, 16, 30
S の候補値	0, 1, 3, 5, 10, 15, 20

表 3.2: CV 法により最適化されたパラメータの組

	$m = 8$		$m = 16$	
試行	r^*	S^*	r^*	S^*
1	30	15	30	10
2	16	10	8	5
3	16	10	30	15
4	16	10	16	15
5	30	10	30	5

表 3.3: 各パラメータの組に対する誤識別率 (%) (上位 12 位)

順位	$m = 8$				$m = 16$			
	r	S	誤識別率	*	r	S	誤識別率	*
1	30	10	6.07	*	30	10	6.38	*
2	16	10	6.99	*	16	10	7.15	*
3	30	15	7.96	*	30	15	7.28	*
4	16	5	8.64	*	8	5	8.47	*
5	8	5	8.65	*	16	5	8.81	*
6	8	10	9.40	*	30	5	9.98	**
7	30	5	9.72	**	16	15	10.48	**
8	8	3	11.13		8	3	10.60	**
9	16	3	11.33		8	10	10.73	**
10	30	20	12.10		16	3	11.16	
11	30	3	12.14		30	20	11.47	
12	16	0	12.21		30	3	11.93	

*, ** は, それぞれ 5%, 1% で 1 位の誤識別率と有意差がないことを表す

太字は, 各試行で CV 法により求められた解を示す

表 3.4: 分散分析表

	要因	自由度	平方和	分散	分散比	
$m = 8$	全体	104	2707.23			
	反復	4	13.67	3.42	0.42	
	r	2	194.06	97.03	11.98	**
	S	6	1269.28	211.55	26.12	**
	$r \times S$	12	582.32	48.53	5.99	**
	誤差	80	647.89	8.10		
$m = 16$	全体	104	1859.45			
	反復	4	14.00	3.50	0.45	
	r	2	116.36	58.18	7.42	**
	S	6	755.75	125.96	16.07	**
	$r \times S$	12	346.11	28.84	3.68	**
	誤差	80	627.24	7.84		

** は, 1%有意であることを表す

表 3.5: 誤識別率 (%) による最適化の限界

	$m = 8$	$m = 16$
$E_{True}(r^*, S^*)$	7.20	7.72
$E_{True}(\hat{r}, \hat{S})$	5.78	6.11

3.4 おわりに

本章では, 汎化能力に重大な影響を与える二つのパラメータ r と S に対して, 誤識別率を最小にするという意味で最適化を試みた. パラメータの最適化には CV 法を用い, 実験計画法によりその妥当性を検討した. その結果, パラメータ最適化手法として CV 法を用いることが妥当であることを明らかにした. これにより, 以下, パラメータ最適化には CV 法を用いることにする. 更に, パラメータ r と S が誤識別率に与える影響も実験計画法を用いて分析した. 分析結果より, r と S は誤識別率に影響を与えることが明らかにされた. つまり, パラメータ r と S の主効果は共に 1% 有意であり, 更に本実験で用いたデータに対しては S の方が r よりも誤識別率に大きな影響を与えることが示された. また, r と S の交互作用も 1% 有意であることから, r と S は, 個別に最適化できず, それらの組合せの効果をも考慮に入れなければならないことが判明した. また, 最適化の限界を示し, 更なる改良の必要性があることも判明した.

次章では, まず他手法との比較実験を行い, 更に本手法を文字認識問題へ適用し, 実用性も検討する.

第 4 章

性能評価

4.1 はじめに

本章では, 汎化能力の観点から本手法の性能を評価する. まず, 他手法との比較実験を述べる. 比較の対象は, Efron の復元抽出法, Chernick らの凸ブートストラップ法および BP 法である. 実験データは, 人工データとして Marill-Green データを, また実在データとして Iris データを用いる. 更に本手法を文字認識問題へ適用し, その実用性も検討する. 用いられる特徴としては, Zernike 特徴と Gabor 特徴を用いる. タイプの異なる 2 つの特徴上で, 本手法と他手法との比較を行う.

4.2 他手法との比較

本節では, 訓練サンプル数が少ない状況下において, 本手法と Efron の復元抽出法, Chernick らの凸ブートストラップ法および BP 法との比較を汎化能力の観点から行う. 図 4.1は, 比較実験の内容を表す.

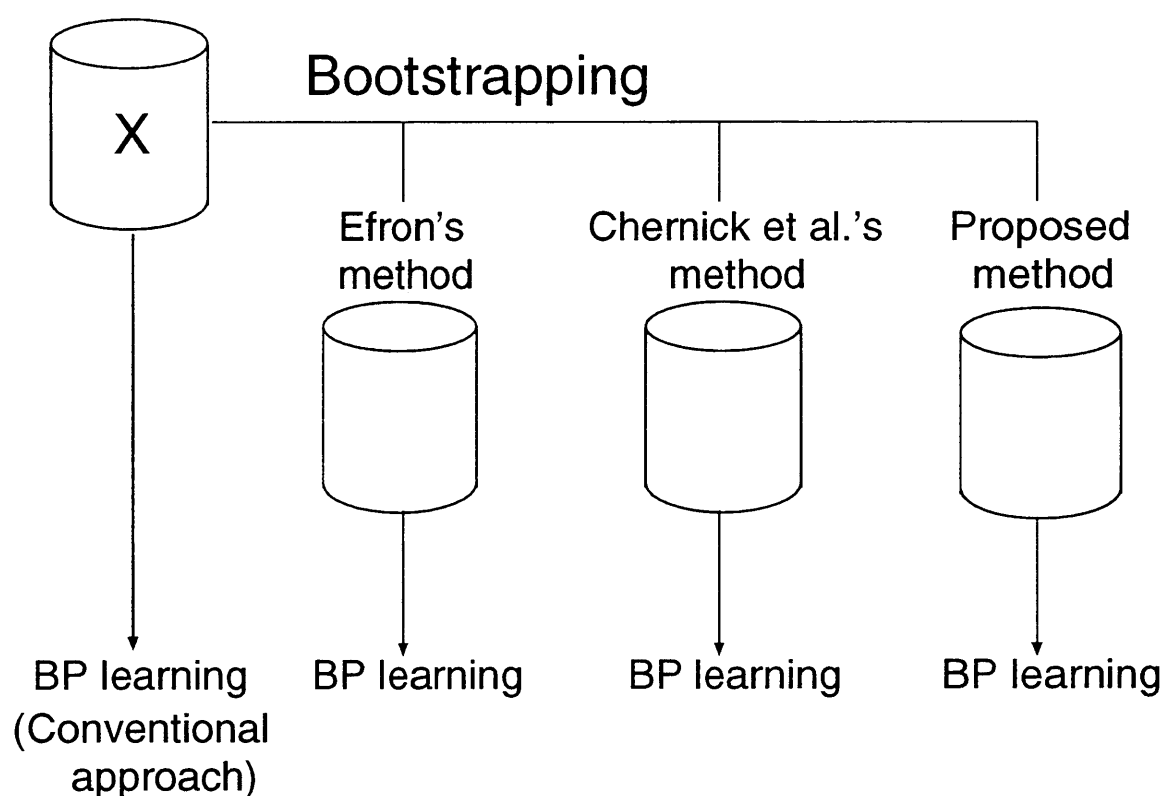


図 4.1: 比較実験の内容

Efron の復元抽出法, Chernick らの凸ブートストラップ法について述べる. 前述の通り, 各クラスの事前確率は等しいものとし, ブートストラップサンプルはクラス毎に生成する. あるクラス $\omega_c (c = 1, 2, \dots, C)$ に対し N 個の訓練サンプルからなる訓練サンプル集合 $X^c = \{\mathbf{x}_1^c, \mathbf{x}_2^c, \dots, \mathbf{x}_N^c\}$ が与えられたとする. ブートストラップサンプル数を n とすると, ブートストラップサンプル集合 $Y^c = \{\mathbf{y}_1^c, \mathbf{y}_2^c, \dots, \mathbf{y}_n^c\}$ は以

下の手順により生成される。但し、 n の値は設計者により任意に設定される。

A. Efron の復元抽出法 [46]

Step1 訓練サンプル集合 X^c から訓練サンプルをランダムに選択し、それを \mathbf{y}_i^c と表す。

Step2 Step1 を復元抽出で n 回繰り返すことにより、ブートストラップサンプル集合 Y^c を得る。

B. Chernick らの凸ブートストラップ法 [48]

Step1 訓練サンプル集合 X^c から訓練サンプルをランダムに選択し、それを \mathbf{x}_1^c と表す。

Step2 更に X^c から \mathbf{x}_1^c とは異なる訓練サンプルをランダムに選択し、それを \mathbf{x}_2^c と表す。

Step3 式 (4.1) に従って、ブートストラップサンプル \mathbf{y}_i^c を生成する。

$$\mathbf{y}_i^c = \lambda \mathbf{x}_1^c + (1 - \lambda) \mathbf{x}_2^c \quad (4.1)$$

ここで、 λ の値は一様乱数 $[0, 1]$ によって与えられる。

Step4 Step1 ~ Step3 を n 回繰り返すことにより、ブートストラップサンプル集合 Y^c を得る。

両手法共、訓練サンプル集合 $X = \{X^c | c = 1, 2, \dots, C\}$ から生成されるブートストラップサンプル集合 $Y = \{Y^c | c = 1, 2, \dots, C\}$ を用いてニューラルネット識別器が設計されることになる。

前述した通り, Efron の復元抽出法では, ブートストラップサンプルは訓練サンプルから復元抽出により選択される. 従って, 元の訓練サンプルと同数か, あるいはそれより実質的に少数のサンプルしか得られず, 汎化能力の改善が期待されにくい. 一方, 凸ブートストラップ法では, 二つの訓練サンプルがランダムに選択され, それらの線形結合によりブートストラップサンプルが生成される. そのため, 例えば, 図 4.2(a) に示されているように, ブートストラップサンプルが分布外に生成される可能性もある. このようなブートストラップサンプルを識別器の設計に用いると, 汎化能力がかえって劣化することが予想される. これに対し, 本ブートストラップサンプル生成法では, パターン分布の構造がブートストラップサンプルに反映され, 結果的に汎化能力が改善されることが期待される. 図 4.2(b) では, $r = 5, S = 1$ でのブートストラップサンプル生成領域を示す.

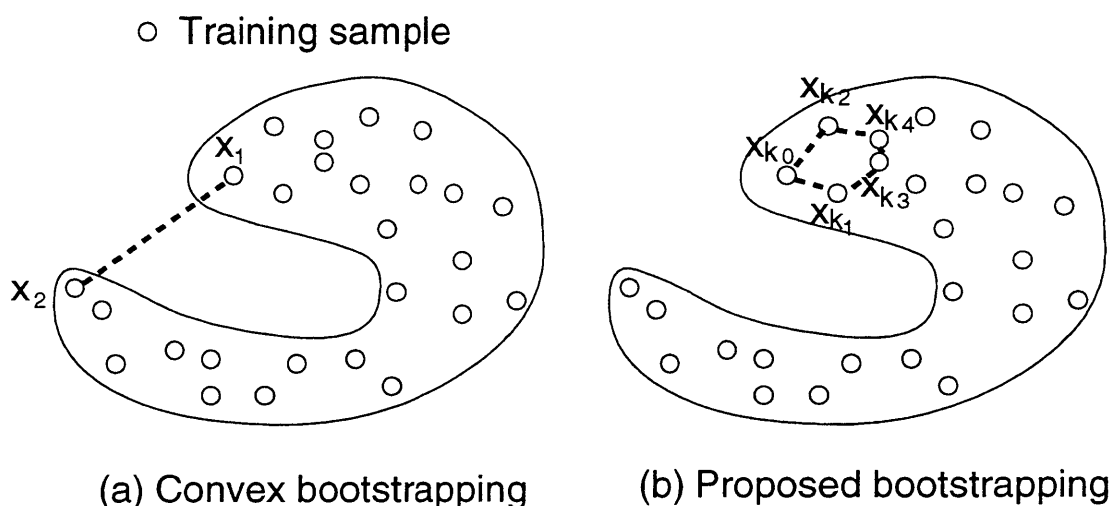


図 4.2: 凸ブートストラップ法と本手法との違い

本節の比較実験では, 人工データと実在データの両方を用いてシミュレーションを行う. 人工データには Marill-Green データを, 実在データには Iris データを用いる. 以下, Iris データについて説明する. このデータは, ニューラルネット識別器と他

の統計的識別器との比較実験においても引用されている [65].

- Iris データ [64]

これは, アイリス科の 3 品種, 各 50 本の花の, がく片および花弁の, 長さと幅をそれぞれ求めることによって得られたものである. つまり Iris データは, 3 クラスから成り, 1 クラス当たり 50 個のサンプルを有する 4 次元の实在データである.

Iris データでは, 利用できるサンプル 50 個/クラスを訓練サンプルとテストサンプルにそれぞれ 25 個/クラスずつランダムに二分割して用いるという試行を独立に 5 回行い, 誤識別率の平均値を求めた. 但し, 特徴毎に平均 0 分散 1 に正規化 [67] してから実験を行った.

各データ集合に対する実験内容を表 4.1 に示す. また, CV 法による誤識別率の推定において, Marill-Green データ, Iris データ共に, 前章と同じく $N_T = 1, J = 20$ とする. 表 4.1 の実験内容に従い, 試行回数を 5 とし, 各手法の誤識別率を求めた. 実験結果を表 4.2 に示す. Efron の復元抽出法, Chernick らの凸ブートストラップ法および BP 法に対する本手法の優位性が示されている. これより, 隠れユニット数 m にも, データ集合にも依らず, 本手法は比較した中で最小の誤識別率を与えることが分かる.

表 4.1: 各データ集合に対する実験条件

データ集合	Marill-Green	Iris
訓練サンプル数 N /クラス	32	25
テストサンプル数/クラス	1000	25
r の候補値	8, 16, 30	5, 20
S の候補値	0, 1, 3, 5, 10, 15, 20	

表 4.2: 誤識別率 (%) による各手法の比較

各手法	$m = 8$		$m = 16$	
	Marill-Green	Iris	Marill-Green	Iris
本手法	7.20	3.46	7.72	2.93
Convex	10.86	6.13	10.61	6.13
Efron	10.94	6.39	10.92	6.13
BP	10.82	5.60	9.96	5.60

4.3 文字認識問題への適用

本節では、文字認識問題へ適用し、本手法の実用性を検討する。手書き数字認識問題に対し、Zernike 特徴および Gabor 特徴を用いて、本手法の実用性を明らかにする。

手書き数字データは、電子技術総合研究所が提供する ETL-1 [66] の自由手書き数字 (10 字種, 1445 文字/字種) を用いる。1 文字は、64(横) × 63(縦) ピクセルで、1 ピクセル当たり 4 ビットの階調画像である。本論文では、数字 10 字種、各字種 1400 文字を選択し、これらに二値化処理 (しきい値 7) を行う。更に、文字の縦横比を同じにするため白画素を下部に付加し、64 × 64 ピクセルとして用いる。図 4.3 に本データの一例を示す。

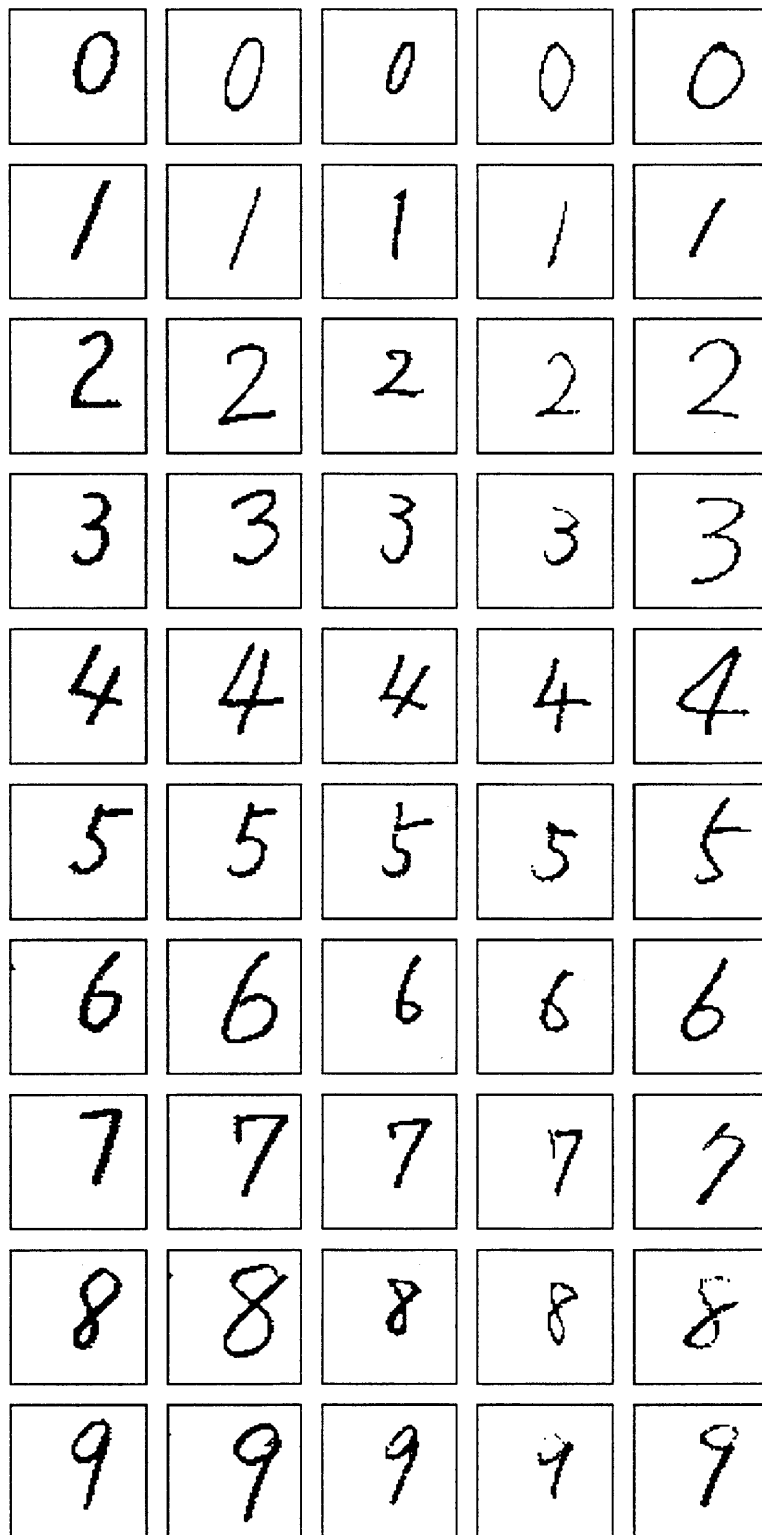


図 4.3: 文字パターンの一例

このデータに対し Zernike 特徴および Gabor 特徴をそれぞれ抽出する。以下、Zernike 特徴, Gabor 特徴について説明する。

- Zernike 特徴 [67] [68]

Zernike 特徴は, 文献 [67], [68] では大文字アルファベット認識問題, 文献 [69], [70] では手書き数字認識問題に用いられている。Zernike 特徴は, 画像の全体的な情報を抽出するモーメントに基づいた特徴である。そこでは, 正規直交基底 V_{nm} が定義され, 画像 $I(x, y)$ をその基底へ射影させると Zernike モーメント A_{nm} が得られる (式 (4.2)~式 (4.4))。

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm\theta) \quad (4.2)$$

ここで, n は非負の整数であり, m は, $n - |m|$ が偶数かつ $|m| \leq n$ を満たす整数である。ベクトル (x, y) の大きさを ρ とし, それに対する x 軸となす角度を θ と表す。また, R_{nm} は次のように定義される。

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|/2} (-1)^s \cdot \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s} \quad (4.3)$$

但し, $R_{n,-m}(\rho) = R_{nm}(\rho)$ である。

$$A_{nm} = \frac{n+1}{\pi} \int \int_{x^2+y^2 \leq 1} I(x, y) V_{nm}^*(\rho, \theta) dx dy \quad (4.4)$$

但し, $A_{nm}^* = A_{n,-m}$ である。更に, $x^2 + y^2 \leq 1$ に関しては, 一つの画像がその範囲内に収まるように定める必要がある。表 4.3 に, 次数と Zernike モーメントとの関係を示す。

この Zernike モーメントから, 必要な次数 n_{max} までの Zernike 特徴を得ることができる。すなわち, Zernike 特徴は式 (4.5) の通り n_{max} までの特徴の組で表される。

表 4.3: 0 次から 12 次までの次数と Zernike モーメントとの関係

次数	Zernike モーメント	モーメントの数
0	A_{00}	1
1	A_{11}	1
2	A_{20}, A_{22}	2
3	A_{31}, A_{33}	2
4	A_{40}, A_{42}, A_{44}	3
5	A_{51}, A_{53}, A_{55}	3
6	$A_{60}, A_{62}, A_{64}, A_{66}$	4
7	$A_{71}, A_{73}, A_{75}, A_{77}$	4
8	$A_{80}, A_{82}, A_{84}, A_{86}, A_{88}$	5
9	$A_{91}, A_{93}, A_{95}, A_{97}, A_{99}$	5
10	$A_{10,0}, A_{10,2}, A_{10,4}, A_{10,6}, A_{10,8}, A_{10,10}$	6
11	$A_{11,1}, A_{11,3}, A_{11,5}, A_{11,7}, A_{11,9}, A_{11,11}$	6
12	$A_{12,0}, A_{12,2}, A_{12,4}, A_{12,6}, A_{12,8}, A_{12,10}, A_{12,12}$	7

$$\{|A_n| | n = 0, \dots, n_{max}\} \quad (4.5)$$

但し, $|z|$ は z の絶対値を表す.

Zernike 特徴は, それ自身では回転についての不変性しか持たない [68]. 移動・拡大縮小の不変性を得るためには, 式 (4.6) を用いて予め文字パターンを正規化しておく必要がある.

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q I(x, y) dx dy \quad (4.6)$$

ここで, m_{pq} は $(p+q)$ 次の画像 $I(x, y)$ のモーメントである. 本論文では Khotanzad

らの方法 [67] に従い, 1 次のモーメントまでを用いて移動・拡大縮小の正規化を行う. 正規化は次式 (4.7) で行われる.

$$\hat{I}(x, y) = I\left(\frac{x}{a} + \bar{x}, \frac{y}{a} + \bar{y}\right) \quad (4.7)$$

$$\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}} \quad (4.8)$$

ここで, \bar{x}, \bar{y} は $I(x, y)$ の中心をなし, $a = \sqrt{\beta/m_{00}}$ である. 式中の m_{00} は画像 $I(x, y)$ の黒画素の数を表す. また, m_{10} は画像 $I(x, y)$ の x 方向に関する黒画素とその距離の積和を, 同様に m_{01} は y 方向に関するそれを表す. \bar{x}, \bar{y} については移動の不変性を得るため, また a については拡大縮小の不変性を得るために導入された. 正規化された画像 $\hat{I}(x, y)$ に対して m_{10} と m_{01} を計算すると, 両方共その値は 0 となることに注意されたい.

本論文では, 拡大縮小に関する正規化パラメータ β は文献 [70] より $\beta = 300$ と設定する. また, Zernike 特徴では, 正規化後の $|A_{00}|$ はどの画像でも同じ数値となり, また $|A_{11}|$ は 0 となる [68]. それ故, 0 次と 1 次の Zernike 特徴は実際には冗長であり, 無視して良い [68]. 本論文もこれらの次数を用いず, 2 次以上の次数を用いることにした. 更に, Zernike 特徴は次数 12 (次元数 47) より大きい次数を用いても誤識別率にほとんど変化がない [70]. 従って, 本論文では, Zernike 特徴の次元数を 47 とする.

- Gabor 特徴 [69]

Gabor 特徴は, 著者の所属する研究室で開発され, 手書き漢字認識問題 [71] や手書き数字認識問題 [69], [70] に用いられ, 従来的特徴抽出法に対する優位性が示されている. その特長は, 生理学的基盤と数理的基盤を併せもち, 特徴抽出系の設計者にとって見通しの良い特徴抽出法である, と報告されている [71]. Gabor 特徴は, Gabor フィルタと文字パターンの畳み込みから得られる. Gabor フィルタは 2 次元ガウス曲

面と2次元平面上を一方向に伝わる平面波とを掛け合わせたフィルタである。Gaborフィルタは、ガウス曲面における標準偏差、平面波の進行方向および平面波の波長により決定される。標準偏差はガウス曲面の広がりをも定めるもので、波長と密接な関係がある。平面波の波長に合わせてガウス曲面の標準偏差を変えるため、標準偏差は波長の関数となる [69]。このことは、文献 [72] から、標準偏差と波長との相互作用が大きいことが実験計画法により示されている。波長を λ 、 x 方向および y 方向の標準偏差をそれぞれ σ_x, σ_y とし、波の進行方向と x 軸がなす角度を θ_u とすると、Gaborフィルタは式 (4.9) で表される。

$$f(x, y, \theta_u, \lambda, \sigma_x, \sigma_y) = \exp \left[-\frac{1}{2} \left\{ \frac{(x \cos \theta_u + y \sin \theta_u)^2}{\sigma_x^2} + \frac{(-x \sin \theta_u + y \cos \theta_u)^2}{\sigma_y^2} \right\} \right] \times \exp \left[i \frac{2\pi(x \cos \theta_u + y \sin \theta_u)}{\lambda} \right] \quad (4.9)$$

但し、方向 θ_u の数を n_θ とすると、方向 $\theta_u (u = 1, \dots, n_\theta)$ は式 (4.10) で表される。

$$\theta_u = \frac{\pi(u-1)}{n_\theta}, u = 1, \dots, n_\theta \quad (4.10)$$

図 4.4に、Gabor フィルタの一例を示す。但し、パラメータは、 $\theta_u = 0, \lambda = 4\sqrt{2}, \sigma_x = 2\sqrt{2}, \sigma_y = 2\sqrt{2}$ である。

Gabor 特徴は、文字パターン上のサンプリング点において Gabor フィルタと文字パターンとを畳み込むことで得られる。文字パターンを大きさ $L \times L$ の二値画像 $I(i, j) (0 \leq i, j \leq L-1)$ とすると、サンプリング点は以下に示す外接枠サンプリング法 [71] により定められる。

Step1 文字パターンの外接枠を求める。

Step2 外接枠の周りに α ピクセル付け加えた枠を求める。これを拡大外接枠と呼ぶ。

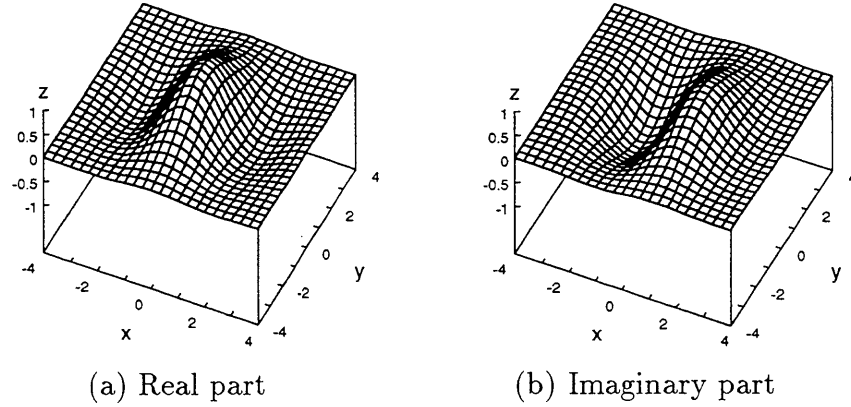


図 4.4: Gabor フィルタ

Step3 拡大外接枠内において, サンプルング点を縦横それぞれ等間隔に全部で M 個のサンプルング点を置く.

サンプルング点 (X, Y) における Gabor 特徴 $g(\cdot)$ は, Gabor フィルタ $f(\cdot)$ と文字パターンとを畳み込むことで得られる (式 (4.11)).

$$g(X, Y, \theta_u, \lambda, \sigma_x, \sigma_y) = \left| \sum_{x=-X}^{L-X-1} \sum_{y=-Y}^{L-Y-1} I(X+x, Y+y) f(\cdot) \right| \quad (4.11)$$

但し, $|z|$ は z の絶対値を表す. サンプルング点毎に, すべての方向 $\theta_u (u = 1, \dots, n_\theta)$ について $g(\cdot)$ が計算される. 従って, 一つの文字パターンから得られる次元数は $M \times n_\theta$ となる. あるサンプルング点上で Gabor 特徴が得られる様子を図 4.5 に示す.

本論文では, $n_\theta = 2, M = 64$ とする. 従って, Gabor 特徴の次元数は $M \times n_\theta = 128$ である. このとき, 各パラメータの値は文献 [69] より, $\alpha = 2, \lambda = 4\sqrt{2}, \sigma_x = 2\sqrt{2}, \sigma_y = 2\sqrt{2}$ と設定する.

前小節の比較実験と同じく, Efron の復元抽出法, Chernick らの凸ブートストラップ法, BP 法を比較の対象とする. 表 4.4 に, 実験で用いた各特徴に対する本手法のパラメータ値を表す. Zernike 特徴, Gabor 特徴共に, 利用できるサンプル 1400 個/ク

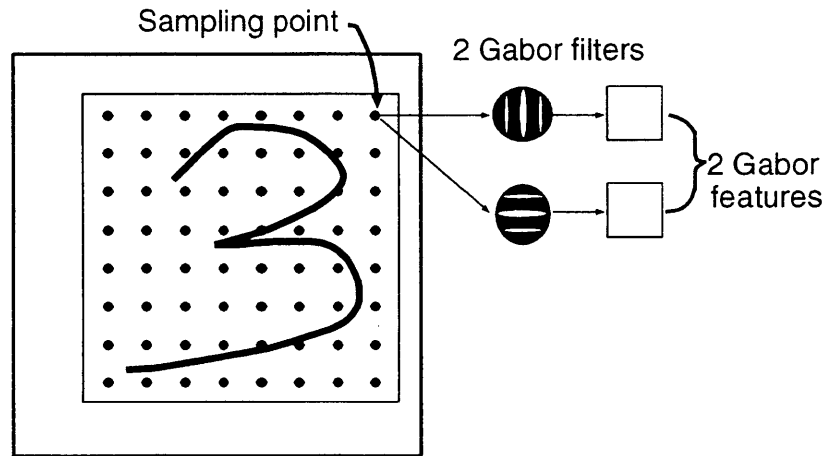


図 4.5: Gabor フィルタによる特徴抽出 ($n_\theta = 2$ の場合)

ラスを訓練サンプルとテストサンプルにそれぞれ 100 と 1300 個/クラスずつランダムに二分割して用いるという試行を独立に 5 回行い、誤識別率の平均値を求める。但し、いずれの特徴に対しても特徴毎に平均 0 分散 1 に正規化 [67] してから実験を行う。また、CV 法による誤識別率の推定において、Zernike 特徴、Gabor 特徴共に大規模データであるため $J = 10$ とする。一方、 N_T については前章と同じく $N_T = 1$ とする。

表 4.5は実験結果である。これより、本手法は、特徴の種類に依存せず、他手法よりも誤識別率の点で優れていることが分かる。例えば、Zernike 特徴に対し、 $m = 8$ のとき、本手法が 26.63%に対し、BP 法では 31.25%であり、約 5%の誤識別率の低減が見られる。一方、Gabor 特徴では、 $m = 8$ のとき、本手法は Efron の復元抽出法に比べ 0.56%ほど誤識別率が低減されることが分かる。このように、タイプの異なる 2 つの特徴上で他手法に対する本手法の優位性が示された。

表 4.4: 各特徴に対する実験条件

パラメータ	Zernike 特徴	Gabor 特徴
r の候補値	10, 50	
S の候補値	0, 1, 3, 5, 10, 15, 20	5, 10, 15

表 4.5: 誤識別率 (%) による各手法の比較

	$m = 8$		$m = 16$	
	Zernike 特徴	Gabor 特徴	Zernike 特徴	Gabor 特徴
本手法	26.63	8.79	25.28	6.10
Convex	30.06	9.10	28.74	6.21
Efron	33.04	9.35	29.56	6.27
BP	31.25	9.17	28.73	6.34

4.4 おわりに

本章では、汎化能力の観点から本手法の性能を評価し、ニューラルネット識別器の設計に本手法を用いることの有効性を明らかにした。まず、訓練サンプル数が少ないという現実的な状況下において、Efron の復元抽出法、Chernick らの凸ブートストラップ法および BP 法との比較を行い、汎化能力の観点から本手法の優位性を明らかにした。更には、文字認識への適用により、本手法の実用性も確認した。自由手書き数字認識問題に対し、Zernike 特徴、Gabor 特徴というタイプの異なる二つの特徴上で、本手法の他手法に対する優位性を示した。

次章では、本論文の結論を述べ、今後の課題と展望についても言及する。

第 5 章

結論

これまで、ニューラルネットの分野では従来の統計的パターン認識理論で培われた膨大な知見が積極的に活用されていなかった。しかし、統計的パターン認識理論の枠組みにおいて、ニューラルネット識別器を汎化能力の向上という意味で改良することは、実地的なパターン認識理論の構築に大きな役割を果たすものと考えられる。このような問題設定に基づいて、ニューラルネット識別器の設計法を論じた。

本論文では、ニューラルネット識別器の汎化能力を改善するために、ブートストラップサンプル生成法を提案した。計算機シミュレーションを通して、訓練サンプル数が少ない現実的な状況下で本手法の有効性を、BP 法、他のブートストラップアプローチとの比較を通して検討した。更に文字認識に関する実験により本手法の実用性も検討した。以下に各章を要約し、本論文のまとめとする。

第 2 章では、新しいブートストラップサンプル生成法を用いたニューラルネット識別器の設計法を論じた。まず、ニューラルネット識別器について説明した。次に、ブートストラップサンプル生成法を提案した。更に、計算機シミュレーションを通して、訓練サンプル数が少ない状況下で本手法の有効性を様々な観点から検討した。すなわち、本手法のパラメータ r と S 、ブートストラップサンプル数、訓練サンプル数、次元数の誤識別率に与える影響、また誤識別率と Bayes 誤識別率との関係、を実験的

に調べた。これらの実験結果より、パラメータ r と S が汎化能力に影響を与えることが明らかにされた。

第 3 章では、パラメータの最適化と分析について述べた。まず、交差確認法によるパラメータ r と S の最適化を論じた。次に、その最適化の妥当性を実験計画法により詳細に検討した。実験結果から、交差確認法を用いることが妥当であることを明らかにした。

第 4 章では、まず本手法と他手法との比較実験を述べ、更に大規模パターン認識問題として文字認識を例にとり、本手法の実用性を検討した。比較実験では、訓練サンプル数が少ない状況下において、Efron の復元抽出法、Chernick らの凸ブートストラップ法および BP 法を比較対象とし、本手法の有効性を明らかにした。また、文字認識では手書き数字認識問題に対し、タイプの異なる 2 種類の特徴上で、他手法に対する本手法の優位性が示された。

以上、ニューラルネット識別器設計のための新しいブートストラップ法を提案し、計算機シミュレーションにより様々な観点からその有用性を明らかにした。本研究で得られた最も大きい知見は、原サンプルを直接用いることよりも、何らかの手法によりパターン分布の構造を反映させたサンプルを用いる方が、より汎化能力の高い識別器を設計することができる、ということを示した点である。

最後に、本研究に関連する今後の課題、また展望をいくつか述べ、本論文のむすびとする。

本論文の目的は、ニューラルネット識別器の汎化能力を改善することにあった。そのため、ニューラルネット識別器の学習に要する計算コストについては特に考慮しなかった。しかし、工学的立場からは計算コストの問題は実際的である。従って、計算コストを削減することについては今後の課題である。

また、パラメータの最適化については、実験結果から、最小有意差法により交差確

認法を用いることの妥当性が示されたものの、改良の余地が見受けられた。従って、パラメータ最適化については、更なる検討が必要であり、今後の課題である。

更に、ニューラルネット識別器を設計するためには、予め数多くのパラメータを定める必要がある。例えば、層数、隠れユニット数、学習率、非線形性の度合い、重みの初期値、などがある。更に、学習の停止条件、慣性項を付けるべきか否か、逐次修正か一括修正か、といった側面やあるいは学習法をも含めてその設定には検討する必要がある。どの要素もニューラルネット識別器の汎化能力に影響を与えるものであり、その決め方はアドホックなところがある。それ故、ニューラルネット識別器の設計者は、予め定めるべき要素を注意深く設定しなければならない。本論文においては、これまで公表された数多くの文献を参考にし、実際には予備実験を繰り返して検討し、これらの要素を設定した。この設定法に関する詳細な考察については、ニューラルネットにおける重要な課題であろう。

最後に、提案したブートストラップサンプル生成法は再標本化手法の一つであるといえる。それ故、工学分野はもちろん数学・統計学分野も含めた本手法の適用範囲は大きいものと期待される [73]~[75]。例えば、本論文ではニューラルネット識別器を対象としたが、他の識別器への適用も考えられる。あるいは、クラスタリング、誤識別率の推定、などにおいても活用されることが期待される。工学的ツールとして、データの補間や補外、関数や画像などの近似、あるいは平滑化など擬似データとしての活用法としても適用可能であると考えられる。また統計分野では、平均ベクトル、共分散行列などの推定問題への適用がある。このように、本論文で提案された方法論は、様々な方面へ用いられることができるものとする。

謝 辞

学部生の頃から本論文をまとめる現在に至るまで、本学教授 浜本義彦先生には終始懇切なる御指導とさまざまな有益なる御助言および温かい御激励を賜りました。ここに心より感謝の意を表しますと共に謹んで御礼申し上げます。また、本論文をまとめるにあたり、有益な御教示を頂きました本学教授 井上克司先生、教授 宮本文穂先生、教授 和田憲造先生、助教授 大林正直先生には深く謝意を表します。

また、日頃からさまざまな御助言と御討論を賜りました、本学助手 内村俊二先生ならびに技術専門職員 渡邊政典氏に深く感謝致します。徳山工業高等専門学校助教授 三木幸先生、大島商船高等専門学校助教授 岡村健史朗先生ならびに宇部工業高等専門学校講師 武藤義彦先生には温かい御激励を頂きました。また、ご退官された本学名誉教授 富田眞吾先生には終始一貫温かい御指導を頂きました。ここに諸先生方に謹んで御礼申し上げます。

実験に用いた手書き文字データベース ETL-1 を作成、提供いただいた電子技術総合研究所の関係諸氏に深く感謝します。

本研究を遂行するにあたり御協力頂きました、三菱電機株式会社 石原裕士氏、NEC エンジニアリング株式会社 長谷俊徳氏ならびに富士ゼロックス株式会社 油布憲治氏に厚く御礼申し上げます。また、日頃より御討論頂きました、積水ハウス株式会社 政水克典氏、株式会社富士通山口情報 鈴木一雄氏、本学ベンチャービジネスラボラトリ非常勤研究員 水上嘉樹氏ならびに株式会社安川電機 福間博彰氏に厚く御礼申

上げます。最後に、御協力頂きました本学工学部知能情報システム工学科情報認識工学講座の学部生、大学院生ならびに卒業生の関係諸氏に心より御礼申し上げます。

参考文献

- [1] 飯島泰蔵, “パターン認識理論,” 森北出版, 1989.
- [2] 浜本義彦, “パターン認識理論の最近の動向,” 電子情報通信学会誌, Vol. 77, No. 8, pp.853–864, Aug. 1994.
- [3] P.A. Devijver and J. Kittler, “Pattern recognition: A statistical approach,” Prentice / Hall, 1982.
- [4] R. O. Duda and P. E. Hart, “Pattern classification and scene analysis,” John Wiley & Sons, 1973.
- [5] K. Fukunaga, “Introduction to statistical pattern recognition,” Second Edition, Academic Press, 1990.
- [6] 鳥脇純一郎, “認識工学,” コロナ社, 1993.
- [7] S. Raudys and A. K. Jain, “Small sample size effects in statistical pattern recognition: Recommendations for practitioners,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-13, No. 3, pp.252–264, Mar. 1991.
- [8] A. K. Jain and B. Chandrasekaran, “Dimensionality and sample size considerations in pattern recognition practice,” in Handbook of Statistics 2, P. R. Krishnaiah and L. N. Kanal, Eds., North Holland, pp.835–855, 1982.

- [9] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, ch.8, MIT Press, 1986.
- [10] D. C. Plaut, S. J. Nowlan, and G. E. Hinton, "Experiments on learning by back propagation," Tech. Rep. CMU-CS-86-126, Carnegie-Mellon Univ., 1986.
- [11] J. Sietsma and R. J. F. Dow, "Neural net pruning - Why and How," *Proc. IEEE Int. Conf. Neural Networks*, pp.1325-1333, Jul. 1988.
- [12] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE Acoustics, Speech and Signal Processing Magazine*, 4(2), pp.4-22, Apr. 1987.
- [13] R. P. Lippmann, "Pattern classification using neural networks," *IEEE Communications Magazine*, Vol. 27, No. 11, pp.47-64, Nov. 1989.
- [14] R. Reed, "Pruning algorithms - A survey," *IEEE Trans. Neural Networks*, Vol. 4, No. 5, pp.740-747, Sep. 1993.
- [15] 甘利俊一, "神経回路網の数理," 産業図書, 1978.
- [16] 長尾真, "情報・システム研究グループの活動状況," (甘利俊一, "ニューロコンピューティングの展望"), *電子情報通信学会誌*, Vol. 72, No. 3, pp.324-330, Mar. 1989.
- [17] 上坂吉則, "ニューロン回路網の可能性," *電子情報通信学会誌*, Vol. 71, No. 11, pp.1241-1247, Nov. 1988.
- [18] 三宅誠, "ニューラルコンピューテーション," *電子情報通信学会誌*, Vol. 70, No. 12, pp.1255-1262, Dec. 1987.

- [19] 麻生英樹, “ニューラルネットワーク情報処理,” 産業図書, 1988.
- [20] 倉田耕治, 麻生英樹, “神経回路網の理論的研究における最近の動向,” 電子情報通信学会論文誌 (D-II), Vol. J73-D-II, No. 8, pp.1103–1110, Aug. 1990.
- [21] 中野肇, “ニューロコンピュータの基礎,” コロナ社, 1990.
- [22] 安居院猛, “ニューラルプログラム,” 昭晃堂, 1993.
- [23] 坂和正敏, 田中雅博, “ニューロコンピューティング入門,” 森北出版, 1997.
- [24] W. S. McCulloch and W. H. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bulletin of Mathematical Biophysics*, 5, pp.115–133, 1943.
- [25] D. O. Hebb, “The organization of behavior,” Wiley, 1949.
- [26] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain,” *Psychological Review*, Vol. 65, pp.386–408, 1958.
- [27] F. Rosenblatt, “Principles of neurodynamics – Perceptrons and the theory of brain mechanisms,” Spartan, 1961.
- [28] M. Minsky and S. Papert, “Perceptrons,” Cambridge, MA, MIT Press, 1969.
- [29] 船橋賢一, “3層ニューラルネットワークによる恒等写像の近似的実現についての理論的考察,” 電子情報通信学会論文誌 (A), Vol. J73-A, No. 1, pp.139–145, Jan. 1990.
- [30] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proc. of the National Academy of Sciences U. S. A.*, Vol. 79, pp.2554–2558, Apr. 1982.

- [31] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," Proc. of the National Academy of Sciences U. S. A., Vol. 81, pp.3088–3092, May 1984.
- [32] G. E. Hinton and T. J. Sejnowski, "Learning and relearning in Boltzmann machines," in D. E. Rumelhart, J. L. McClelland and the PDP Research Group, Parallel Distributed Processing: Explorations in the Microstructure of Cognition," Vol. 1: Foundations, The MIT Press, Cambridge, pp.282–317, 1986.
- [33] T. Kohonen, "Self-organizing maps," Springer, 1995.
- [34] S. Amari and N. Murata, "Statistical theory of learning curves under entropic loss criterion," Neural Computation, 5, pp.140–153, 1993.
- [35] S. Amari, N. Fujita, and S. Shinomoto, "Four types learning curves," Neural Computation, 4, pp.605–618, 1992.
- [36] E. B. Baum, "When are k -nearest neighbor and back propagation accurate for feasible sized sets of examples?," Proc. of the EURASP Workshop on Neural Networks, L. B. Almeida and C. J. Wellekens Eds., Delft University of Technology, Delft, 1990.
- [37] D. Cohn and G. Tesauro, "How tight are the Vapnik-Chervonenkis bounds?," Neural Computation, 4, pp.249–269, 1992.
- [38] S. Raudys, "Why do multilayer perceptrons have favorable small sample properties?," Pattern Recognition in Practice 4, E. S. Gelsema and L. N. Kanal, Eds., Elsevier Science, B. V., pp.287–298, 1994.

- [39] S. Raudys and A. K. Jain, "Small sample size problems in designing artificial neural networks," in *Artificial Neural Networks and Pattern Recognition: Old and New Connections*, I.Sethi and A.K. Jain(eds.), pp.33–50, Elsevier, 1991.
- [40] R. P. W. Duin, "Superlearning capabilities of neural networks?," *Proc. of the 8th Scandinavian Conf. Image Analysis, Tromso*, pp.547–554, 1993.
- [41] R. P. W. Duin, "Superlearning and neural network magic," *Pattern Recognition Letters*, Vol. 15, pp.215–217, Mar. 1994.
- [42] K. Fukunaga, D. R. Olsen, "An algorithm for finding intrinsic dimensionality of data," *IEEE Trans. Computers*, Vol. C-20, No. 2, pp.176–183, Feb. 1971.
- [43] S. Raudys, "On shape of pattern error function, initializations and intrinsic dimensionality in ANN classifier design," *Informatika*, Vol. 4, No. 3-4, pp.360–383, 1993.
- [44] S. Raudys, "Unexpected small sample properties of the linear perceptrons," Technical report on research activities in LAFORRIA-IBP, LAFORRIA 95/17, 1995.
- [45] Y. Hamamoto, S. Uchimura, and S. Tomita, "On the behavior of artificial neural network classifiers in high-dimensional spaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-18, No.5, pp.571–574, May 1996.
- [46] B. Efron, "Bootstrap methods: Another look at the jackknife," *The Annals of Statistics*, Vol.7, No.1, pp.1–26, 1979.
- [47] B. Efron, "Estimating the error rate of a prediction rule: Improvement on cross-validation," *J. American Statistical Association*, Vol. 78, No. 382, pp.316–331,

Jun. 1983.

- [48] M. R. Chernick, V. K. Murthy, and C. D. Nealy, "Application of bootstrap and other resampling techniques: Evaluation of classifier performance," *Pattern Recognition Letters*, 3, pp.167–178, May 1985.
- [49] D. J. Hand, "Recent advances in error rate estimation," *Pattern Recognition Letters*, 4, pp.335–346, Oct. 1986.
- [50] A. K. Jain, R. C. Dubes, and C.-C. Chen, "Bootstrap techniques for error estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 5, pp.628–633, Sep. 1987.
- [51] P. Hall, T. J. DiCiccio, and J. P. Romano, "On smoothing and the bootstrap," *The Annals of Statistics*, Vol. 17, No. 2, pp.692–704, 1989.
- [52] Y. Mitani, Y. Hamamoto, and S. Tomita, "Use of bootstrap samples in designing artificial neural network classifiers," *Proc. 1995 IEEE Int. Conf. Neural Networks, Perth*, Vol. 4, pp.2103–2106, Nov. 1995.
- [53] Y. Mitani, Y. Hamamoto, and S. Tomita, "Back-propagation learning with bootstrap samples," *Proc. Second Asian Conf. Computer Vision, Singapore*, Vol. 2, pp.90–93, Dec. 1995.
- [54] Y. Mitani, Y. Hamamoto, and S. Tomita, "A consideration on the bootstrapping in artificial neural network classifier design," *Proc. Fifth Conf. Int. Federation of Classification Societies, Kobe*, Vol. 2, pp.271–274, Mar. 1996.
- [55] Y. Mitani and Y. Hamamoto, "A bootstrap-based learning with the cross validation," *Proc. Fifth Int. Conf. Neural Information Processing, Kitakyushu*, Vol.

- 1, pp.55–58, Oct. 1998.
- [56] 三谷芳弘, 浜本義彦, “ニューラルネット識別器設計のためのブートストラップ法,” 電子情報通信学会論文誌 (D-II), Vol. J82-D-II, No. 2, pp.268–275, Feb. 1999.
- [57] T. Marill and D. M. Green, “On the effectiveness of receptors in recognition systems,” IEEE Trans. Information Theory, Vol. IT-9, pp.11–17, 1963.
- [58] 森俊二, 坂倉梅子, “画像認識の基礎 I,” オーム社, 1986.
- [59] John Van Ness, “On the dominance of non-parametric Bayes rule discriminant algorithms in high dimension,” Pattern Recognition, Vol. 12, pp.355–368, May 1980.
- [60] K. Fukunaga and T.F. Krile, “Calculation of Bayes’ recognition error for two multivariate Gaussian distributions,” IEEE Trans. Computers, Vol. C-18, No. 3, pp.220–229, Mar. 1969.
- [61] 竹内 啓 (編), “統計学辞典,” 東洋経済新報社, 1989.
- [62] 奥野忠一, 芳賀敏郎, “実験計画法,” 培風館, 1969.
- [63] 大村平, “実験計画と分散分析のはなし,” 日科技連, 1984.
- [64] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” Ann. Eugenics, 7, pp.179–188, 1936.
- [65] W. F. Schmidt, D. F. Levelt, and R. P. W. Duin, “An experimental comparison of neural classifiers with ’traditional’ classifiers,” in Pattern Recognition

in Practice IV, E. S. Gelsema and L. N. Kanal (eds.), Elsevier Science B. V., pp.391–402, 1994.

- [66] 齊藤泰一, 山田博三, 森俊二, “手書き文字データベースの解析 (III),” 電子技術総合研究所, Vol. 42, No. 5, May 1978.
- [67] A. Khotanzad and J.-H. Lu, “Shape and texture recognition by a neural network,” in Artificial Neural Networks and Statistical Pattern Recognition: Old and New Connections, I. K. Sethi and A. K. Jain, Eds., Elsevier Science Publishers B. V., pp.109–131, 1991.
- [68] A. Khotanzad and Y. H. Hong, “Invariant image recognition by Zernike moments,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-12, No. 5, pp.489–497, May 1990.
- [69] Y. Hamamoto, S. Uchimura, M. Watanabe, T. Yasuda, Y. Mitani, and S. Tomita, “A Gabor filter-based method for recognizing handwritten numerals,” Pattern Recognition, Vol. 31, No. 4, pp.395–400, Apr. 1998.
- [70] 渡辺政典, 安田哲也, 浜本義彦, 富田眞吾, “Gabor 特徴を用いた手書き数字認識,” 電子情報通信学会, パターン認識・メディア理解研究会, PRMU96-27, pp.9–16, Jun. 1996.
- [71] 浜本義彦, 政水克典, 内村俊二, 富田眞吾, “手書き漢字認識のための Gabor 特徴,” 電子情報通信学会論文誌, Vol.J79-D-II, No.2, pp.202–209, Feb. 1996.
- [72] 内村俊二, 水野圭, 浜本義彦, 富田眞吾, “実験計画法による Gabor 特徴の分析,” 電子情報通信学会, ヒューマン情報処理研究会, HIP97-3, pp.17–24, Jul. 1997.

- [73] 北川敏男, “統計情報論 I,” 共立出版, 1987.
- [74] 北川敏男, “統計情報論 II,” 共立出版, 1987.
- [75] G. A. Young, “Bootstrap: More than a stab in the dark?,” *Statistical Science*, Vol.9, No.3, pp.382-415, 1994.

付録

Marill-Green データ [57] は正規分布をなし, 2 クラスの母平均ベクトル, 母共分散行列が共に異なる. ここでは, クラス ω_i の母平均ベクトル μ_i , クラス ω_i の母共分散行列 Σ_i を以下に示す.

$$\mu_1 = [7.825 \ 6.750 \ 5.835 \ 8.525 \ 6.615 \ 7.065 \ 7.865 \ 4.435]^T$$

$$\mu_2 = [5.760 \ 5.715 \ 5.705 \ 4.150 \ 6.225 \ 6.960 \ 6.750 \ 3.910]^T$$

$$\Sigma_1 = \begin{pmatrix} 1.034 & 1.281 & 0.351 & -0.293 & 0.098 & 0.301 & 0.141 & 1.336 \\ 1.281 & 1.967 & 0.664 & -0.219 & 0.259 & 0.556 & 0.276 & 2.094 \\ 0.351 & 0.664 & 7.138 & 1.192 & 2.726 & 1.116 & 0.678 & 2.097 \\ -0.293 & -0.219 & 1.192 & 2.269 & 1.367 & 0.146 & 0.201 & -0.308 \\ 0.098 & 0.259 & 2.726 & 1.367 & 5.727 & 1.280 & 0.933 & 2.107 \\ 0.301 & 0.556 & 1.116 & 0.146 & 1.280 & 2.941 & 1.949 & 2.197 \\ 0.141 & 0.276 & 0.678 & 0.201 & 0.933 & 1.949 & 1.577 & 1.229 \\ 1.336 & 2.094 & 2.097 & -0.308 & 2.107 & 2.197 & 1.229 & 6.606 \end{pmatrix}$$

$$\Sigma_2 = \begin{pmatrix} 4.792 & 4.417 & 4.244 & 2.406 & 1.798 & 0.790 & 0.785 & 2.993 \\ 4.417 & 5.074 & 4.636 & 2.798 & 1.824 & 0.639 & 0.644 & 2.799 \\ 4.244 & 4.636 & 5.428 & 3.224 & 2.111 & 0.903 & 1.131 & 2.943 \\ 2.406 & 2.798 & 3.224 & 5.287 & 3.006 & 1.326 & 1.897 & 2.648 \\ 1.798 & 1.824 & 2.111 & 3.006 & 3.574 & 2.229 & 2.471 & 1.915 \\ 0.790 & 0.639 & 0.903 & 1.326 & 2.229 & 4.008 & 2.405 & 1.106 \\ 0.785 & 0.644 & 1.131 & 1.897 & 2.471 & 2.405 & 4.507 & 1.727 \\ 2.993 & 2.799 & 2.943 & 2.648 & 1.915 & 1.106 & 1.727 & 3.972 \end{pmatrix}$$