

有限学習標本に基づく 統計的パターン分類器の設計と評価

竹下 鉄夫

有限学習標本に基づく
統計的パターン分類器の設計と評価

和 1253480

竹下 鉄夫

目次

あらまし	1
第1章 序論	4
1.1 統計的パターン認識の概要	4
1.2 本研究の意義	6
1.3 多次元正規分布と識別関数	7
1.3.1 ベイズ分類器	8
(1) 共分散行列が等しい場合	8
(2) 共分散行列が異なる場合	10
1.3.2 標本を用いて設計するパターン分類器	12
(1) 線形識別関数	13
(2) 2次識別関数	14
(3) 2次識別関数の誤り率	17
1.4 本研究に関連する国内外の研究	19
1.5 本論文の構成	21
第2章 マハラノビス距離の推定および Stein 補正の効果	30
2.1 まえがき	30
2.2 平均値ベクトルの推定と Stein の補正	32
2.3 マハラノビス距離とその推定誤差	33
2.3.1 マハラノビス距離の定義	33
2.3.2 マハラノビス距離の推定	33
2.4 Stein 補正の効果	34
2.4.1 シミュレーション手順	34
2.4.2 平均値ベクトルの推定に関する実験結果	35
2.4.3 マハラノビス距離の推定実験	38
(1) 推定誤差	38

目次

(2) Stein 補正の効果	4 0
2.5 むすび	4 3
第3章 マハラノビス距離の推定量の確率分布	4 7
3.1 まえがき	4 7
3.2 マハラノビス距離とその推定量	4 9
3.2.1 マハラノビス距離の定義	4 9
3.2.2 マハラノビス距離の推定量	4 9
3.3 マハラノビス距離の成分毎の推定量	5 0
3.4 考察	5 3
3.5 むすび	5 4
第4章 有限学習標本に基づく統計的パターン分類における認識率の評価 ..	5 9
4.1 まえがき	5 9
4.2 識別関数と認識率との関係	6 0
4.2.1 共分散行列が等しい場合	6 0
4.2.2 共分散行列が異なる場合	6 1
4.2.3 母集団パラメータの推定と認識率の低下	6 1
4.3 誤り率の理論的評価	6 2
4.3.1 線形識別関数の場合	6 2
4.3.2 2次識別関数の場合	6 3
4.3.3 誤り率の増加分の評価	6 4
(1) 線形識別関数の場合	6 4
(2) 2次識別関数の場合	6 5
4.4 実験	6 6
4.4.1 実験方法	6 6
4.4.2 実験結果	6 6
(1) 線形識別関数の場合	6 6
(2) 2次識別関数の場合	6 7
4.4.3 識別実験の誤り率の平均値と分散	7 1

4.4.4 誤り率の増加分の分布（線形識別関数）	71
4.5 むすび	74
第5章 有限学習標本に基づく統計的パターン分類器の改良	76
5.1 まえがき	76
5.2 母集団パラメータと識別関数	78
5.2.1 RDA法に基づく識別関数	78
5.2.2 Bhattacharyya 距離	80
5.3 シミュレーション実験	81
5.3.1 母集団パラメータの決定	81
5.3.2 実験方法	82
5.3.3 実験結果	83
5.4 数字認識実験	87
5.4.1 実験方法	87
5.4.2 実験結果	88
5.5 むすび	91
第6章 むすび	94
6.1 本研究のまとめ	94
6.2 今後の展望と課題	95
謝辞	98

主要記号一覧

クラス

ω_i : クラス i ($i=1$ または 2)

パターン, 標本

$\mathbf{x} = (x_1, x_2, \dots, x_N)^t$: 特徴パターン (N 次元)

$\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$: 標本パターンの集合 (M : 学習標本数)

\mathbf{x}_{ji} : クラス i の第 j 標本

m_i : クラス i の学習標本数

\mathbf{x}_t : テスト標本

g : 総クラス数

m : 総学習標本数

確率, パラメータ

P_i : クラス i の事前確率

$p_i(\mathbf{x})$: クラス i の確率密度関数

$p_t(\mathbf{x})$: テスト標本の確率密度関数

$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: 多次元正規分布

$\boldsymbol{\mu}$: 平均値ベクトル

$\boldsymbol{\Sigma}$: 共分散行列

$N(\eta, \sigma^2)$: 1次元正規分布

η : 平均値

σ^2 : 分散

ε : 誤り率

ε_B : ベイズ誤り率

ε_1 : クラス1の標本をクラス2と分類する誤り率

ε_2 : クラス2の標本をクラス1と分類する誤り率

$\hat{\boldsymbol{\mu}}_i$: クラス i の標本平均値ベクトル

$\hat{\boldsymbol{\mu}}^*$: 平均値ベクトルに関するStein推定量

$\hat{\boldsymbol{\Sigma}}_i$: クラス i の標本共分散行列

$\widehat{\Sigma}$: 併合して推定した標本共分散行列 (pooled covariance matrix)

$E \{ \cdot \cdot \}$: $\cdot \cdot$ の期待値

$V a r \{ \cdot \cdot \}$: $\cdot \cdot$ の分散

関数, 距離

$h(\mathbf{x})$: ベイズ識別関数 (分類器)

$p(h | \omega_i)$: クラス i のパターンの識別関数の条件付き確率密度関数

$\Phi(x)$: 標準正規分布関数

$\phi(x)$: 標準正規分布の密度関数

$\widehat{h}(\mathbf{x})$: 標本に基づく識別関数 (分類器)

$dm^2(\mathbf{x})$: 二乗マハラノビス距離

$d_i^2(\mathbf{x})$: 二乗マハラノビス距離の第 i 主軸成分 ($i=1, 2, \dots, N$)

δ^2 : 共分散行列が等しい 2 クラス間の二乗マハラノビス距離

d_B : Bhattacharyya 距離

d_{B1} : Bhattacharyya 距離の第 1 成分 (主として平均値ベクトルの差による)

d_{B2} : Bhattacharyya 距離の第 2 成分 (主として共分散行列の差による)

$R = d_{B1} / (d_{B1} + d_{B2})$: Bhattacharyya 距離全体に第 1 成分が占める割合

行列

A^t : 行列 A の転置

A^{-1} : 行列 A の逆行列

I : 単位行列

Λ : 対角行列

$\delta(\omega)$: デルタ関数

$u(t)$: ステップ関数

ϕ_i : 第 i 固有ベクトル ($i=1, 2, \dots, N$)

λ_i : 第 i 固有値 ($i=1, 2, \dots, N$)

あらまし

あらまし

コンピュータによって文書から画像と文字の領域を分け、さらに画像の内容を理解したり、文字を読み取ったり、外国語の書籍・雑誌を読み取って翻訳をすること、あるいは人が話す音声を理解して、人とコンピュータが会話を持つこと、外国語の通訳をすること等は、コンピュータが発明された当初からコンピュータ科学者・技術者の夢であった。人工知能に関する研究が始められた当初は、比較的容易であると思われていたこのような分野における研究の困難さは、コンピュータの計算能力が発達し、計算パワーが飛躍的に向上した今日にあっても、大きな壁となって我々の前に立ちただかっている。

人工知能研究の困難さの大きな要因の一つは、人間が行っている知能・知識・記憶・認識等に関する複雑なメカニズムが解明されていないところにあると思われる。このうち、認識に関わるパターン認識の領域では、前述のような困難さにもかかわらず、対象を限定することで郵便番号読み取り装置に代表される手書き数字認識システムや発送仕分けにおける音声認識システム等、かなりの数に上るパターン認識システムが実用化されている。

ところで、パターン認識システムは、商品として開発されていながら、これらの商品はその認識性能を万人が納得する形式で数値化して表示することが難しいという、工業製品としては甚だ不可思議な商品である。この主たる原因は人間が行う情報処理を機械に置き換えるという製品の目的に起因している。すなわち、パターン認識を行う最終目標は明確であるにもかかわらず、人間の行う認識処理が有するロバスト性（robustness）、人間が読み、書き、話す各種の情報に関わるゆらぎ・ばらつきの問題に伴う工学的な処理の困難さに起因している。パターン認識システムを一種の学習機械と見なしたとき、学習に関する理論も完全に解明されているとはいえない。その困難さの原因は学習のための標本が十分に確保できない場合が多いということの他、現実世界を機械処理するために標本化・離散化して得られるパターンの次元数が大きくなり、多次元空間における学習と識別に関して解明されていない問題が数多く存在することも見逃せない。

パターン認識システムの学習と分類に関する手法には、母集団の分布に一定の確率

分布を仮定するパラメトリックな手法と、確率分布を仮定しないニューラルネットワークシステムに代表されるノンパラメトリックな手法の二種類がある。

本来、人間が行っている情報処理機構は非線形かつノンパラメトリックな処理であることが予想され、その処理メカニズムに近いといわれるニューラルネットワークシステムを採用したときには優れたシステムは高い性能を達成することが予想されるが、その数学的な取り扱いを始めとして、理論的に解明されていない部分が多い。

一方、パラメトリックな手法は、多変量解析に代表されるように、母集団の分布に多次元正規分布を仮定する領域では理論的にも数多くの研究が成されている。しかしながら、有限学習標本に基づくパターン分類器の学習と識別の問題については、母集団に多次元正規分布を仮定するとしても解明されていないことが多い。実際のパターン認識システムを設計する場合には、学習標本が十分確保できないケースが数多くあり、さらに、認識対象が画像に代表される2次元パターンや音声のように時間と共に変化する信号であるとき、入力パターンの次元数が理論的解析を著しく困難にするほど大きくなる場合も多い。

このようなパラメトリックな手法における有限学習標本に基づいて設計された分類器の性能評価を行うことは、パターンの次元数や学習標本数が異なるパターン認識システムの性能を比較評価をする際、重要な役割を果たす。また、パターンの次元数に比較して学習標本が十分に確保できない状況におけるパターン分類器の改良もその工学的応用上有意義であり研究意義は大きい。

本論文は、パラメトリックな手法において有限学習標本に基づいて分類器を設計する際に生じる様々な問題とそれに対する対策について検討したものである。以下、第1章で研究の意義、および多次元正規分布と識別関数の基礎について述べる。第2章では、2次識別関数の主要な要素を占めるマハラノビス距離について考察し、学習標本が十分に確保できない状況におけるマハラノビス距離の推定値について、Steinの補正の効果とともに論ずる。第3章では、マハラノビス距離の推定量の確率分布について理論的な解明を行い、マハラノビス距離の推定値においては、標本に基づく共分散行列の推定誤差の影響が大きいことを明らかにする。第4章では有限学習標本に基づくパターン分類器の性能評価を行い、学習標本数が十分に確保できない状況におけるパターン分類器の振る舞いについて述べる。第5章ではFriedmanの提唱するRegularized Discriminant Analysisにより、有限学習標本に基づく標本共分散行列の推定

あらまし

誤差の影響を軽減するためのパターン分類器の改良の試みを検討してその効果を確認し、さらにより優れたパターン分類器の設計指針を得る。第6章では、本論文を総括し、今後の展望と課題について述べる。

第 1 章 序論

1.1 統計的パターン認識の概要

パターン認識とは多次元空間内のパターンをあるカテゴリに割り当てる問題とみなされる。カテゴリ名の決まっていないパターンの集合が与えられ、個々のパターン自体の持つ何らかの意味での類似性（近接性）に基づいていくつかのグループ（カテゴリ）を見出すことをクラスタリングもしくはクラスタ分析と呼ぶ。これに対して、パターン認識は、パターン集合が所属するクラスもしくはカテゴリがあらかじめ与えられていて、入力パターンがどのクラスに所属するかを判断（分類もしくは判別）するプロセスである⁽¹⁾。パターン認識の方法、もしくはそれを具体化するために必要な情報をパターンのサンプルに基づいて取得していくプロセスを（パターン認識における）学習と呼ぶ。学習に用いるサンプルのカテゴリ名が与えられている（既知である）場合とそれも未知である場合がある。通常、前者を教師付き学習、後者を教師無し学習という⁽²⁾。

教師付き学習においてパターンの母集団の確率分布が既知でそのパラメータのみが未知（パラメトリックな問題といわれる）という場合には、サンプルパターンがどのカテゴリに所属するか（正解）がわかっていることを用いてカテゴリ毎のパターンの分布のパラメータを学習する。そして、学習を終えた後、どのカテゴリに所属するか知られていないパターンを分類するというパターン認識本来の機能を実行する段階に入る。工学的見地からいえば、教師付き学習においては、より少ない学習でより良い認識性能に到達できるパターン認識システムは、学習に関してより効率的であるといえる。また、パターン認識システムの最終的な性能はその分類器の識別性能によって評価されるべきである。

実際のパターン認識において、パターン集合が多次元空間内でどのような確率分布を成しているかを正確に知ることはできない。上記のように、パターン集合の確率分布について、分布の形状を決める関数の形などは分かっているとして、その確率分布を決定するパラメータを学習や統計的推定によって求めようとする方式をパラメトリ

第 1 章 序論

ックな方法と呼ぶ。パラメトリックな方法ではパターンの確率分布に正規分布を仮定することがしばしば行われる。しかしながら、実際にはパターンが正規分布に従うという保証はない。このため、パターンの確率分布の形状を仮定しないで分類器の設計をしようとする手法もあり、これをノンパラメトリックな方法と呼ぶ。前者の代表的な例として統計的パターン認識があり、後者のそれにはニューラルネットワークやk-近傍則などがある⁽²⁾⁽³⁾。本論文では前者の統計的パターン認識の場合を扱う。以上の関係を図 1.1 に示す。

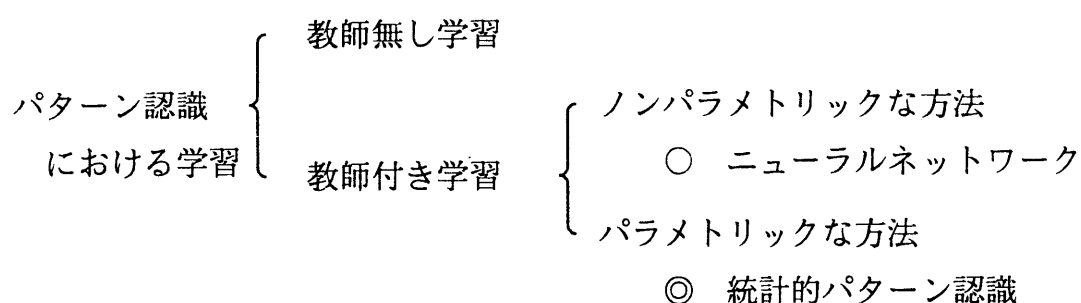


図 1.1 パターン認識における学習（下線は本文で扱う領域）

上に述べるように、パターンの母集団の確率分布としては正規分布を仮定することが非常に多い。この仮定は中心極限定理により正当化されることが多い。そして、実際のパターン認識においてパターンが正規分布に従う保証はないにもかかわらず、確率分布に多次元正規分布を仮定することが多いのは

(1)多次元正規分布の他に数学的に取扱いの容易な分布が見当たらない

という消極的な理由の他に、

(2)厳密に正規分布に従わないパターン集合であっても、関連する統計量が漸近的に正規分布に従うと考えれば、多次元正規分布を仮定して得られた結論が実際のパターン認識においてもそれほど矛盾を生じることなく適用できる場合が多い⁽⁴⁾、

(3)正規分布の仮定に基づいて導かれた分類器でも、識別能力の若干の低下を認めた上で、他の分布に従うパターンにも有効に利用できる。

などの理由による。以下本論文ではパターンの確率分布に多次元正規分布を仮定して、議論を進める。

1.2 本研究の意義

パターン認識の最終性能は分類器の平均誤り率によって評価すべきである。パターン集合の確率分布に関する情報が完全に既知であれば、分類器の平均誤り率を最小にすることが可能である。平均誤り率最小を達成する決定をベイズ決定と呼び、このとき達成される最小平均誤り率をベイズエラーもしくはベイズ誤り率と呼ぶ。また、最小平均誤り率を達成する分類器をベイズ分類器という⁽⁵⁾。

パターンの確率分布に多次元正規分布を仮定するとしても、実際のパターン認識においてはパターン分類器の設計・評価に使用できる標本数は有限であり、パターン集合の確率分布に関する構造も有限個の標本を用いて推定するしかない。このため、有限個の標本を用いて設計された分類器の識別性能はベイズ分類器よりも劣るのが普通である。しかし、有限個の標本から設計された分類器の識別性能も確率的に変動するためその振る舞いに関しては解明されていないことが多い^{(6) - (10)}。また、パターンの次元数が大きくなるに従い、識別率を計算する多重積分が膨大な計算量を要求することから、コンピュータを用いた数値計算も一部の例を除いて成されていない^{(8) (9)}。このため、有限個の学習標本に基づいて設計された分類器に関する性能評価の客観的基準は、最終結果として表れる認識率（もしくは誤り率）しかなく、これを上げるため、パターン認識を行う実際のシステムにおいては、従来は試行錯誤によってシステムを構築するしかなかった。

さて、実際のパターン認識の過程は、第 1 ステップとして実世界から認識対象をコンピュータで処理できる表現に変換するための「観測系」、観測系を経てシステムに取り込まれたパターンへ適当な補正を施す「前処理」、パターンから後段の決定・分類処理に適した特徴量を抽出する「特徴抽出」、そして最後にベクトル表現されたパターンを各カテゴリに割当て「決定・識別」過程がくるとみると考え易い^{(2) (10)}。この中で最後の「識別・決定」段階は、ベクトル表現されたパターンならば何でも処理できるため、比較的一般的な考察がなされてきている。

もちろん、工学的な見地からシステム全体の性能向上を図るためには、本来、個々の認識対象に強く依存する「観測・前処理系」と分類を行う「決定・識別系」の設計には別々の原理を用いて、それぞれ性能向上をはかるべく努力できることが望ましい。ところが、「決定・識別系」におけるベクトル表現されたパターンの次元数と学習標

第 1 章 序論

本数と識別性能の間の定量的な関係が明らかになっていないため、特徴ベクトルの次元数が異なるような二つ以上の特徴抽出方法を相互に比較することが困難であった。このため常にパターン認識システムは「観測・前処理系」も合わせた、総合的な認識性能によってしか評価をすることが不可能であった。

パターンの特徴を「観測・抽出」する系は明らかに個々の対象パターンに固有の技術によって支えられているので、特徴抽出を経て得られる特徴ベクトルの次元数が異なっているとすると、個々の特徴抽出方法の性能を個別に評価できることが望ましい。そして、システム性能の向上のためには一つ一つの具体的なサブシステムごとに最適化を行い、総合的な性能の向上が図れるようにすべきである。パターン認識システムにおいて、伝統的な諸工学で用いられる古典的かつ基本的な解析と合成に関する手法が採用されていないのは、一つには学習と分類に関する定量的な評価方法が確立していないことに原因がある。この点に関連して、本研究は有限学習標本に基づく分類器の性能に関する定量的な知見を得ることにより、パターン認識システムの評価に有効な解析手段を与えることができるようにすることを目的とするものである。

すなわち、個別のパターン認識システムについて、「観測系」・「前処理」・「特徴抽出」のサブシステム（観測・特徴抽出機構）から導かれる特徴パターンの次元数が異なるとしても、「識別・決定系」のパターン分類器の形、特徴パターンの次元数、学習標本数、および、平均誤り率間の関係が明らかになることにより、同一の対象を有するパターン認識システムの個々の「観測・特徴抽出機構」の性能比較も可能となるであろう。

1.3 多次元正規分布と識別関数

多次元正規分布に従う二つのパターン母集団（クラス 1，および 2）の平均値ベクトルをそれぞれ μ_1 ， μ_2 ，共分散行列を Σ_1 ， Σ_2 とすれば，クラス 1，2 の確率密度関数 $p_1(\mathbf{x})$ ， $p_2(\mathbf{x})$ は

$$p_i(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} |\Sigma_i|^{1/2}} e^{-\frac{1}{2} \{(\mathbf{x} - \mu_i)^t \Sigma_i^{-1} (\mathbf{x} - \mu_i)\}} \quad (1.1)$$

但し， $i=1, 2$ ， N は特徴パターンの次元数となる。これを， $N(\mu_1, \Sigma_1)$ ， $N(\mu_2, \Sigma_2)$ で表す。

クラス 1, 2 の事前確率をそれぞれ P_1, P_2 とし, クラス 1 のパターンをクラス 2 と分類する誤り率を ε_1 , 逆にクラス 2 のパターンをクラス 1 と分類する誤り率を ε_2 とする. このとき両者を合わせた平均誤り率 ε は

$$\varepsilon = P_1 \varepsilon_1 + P_2 \varepsilon_2 \quad (1.2)$$

以下, 特に断わらなければ,

$$P_1 = P_2 = 1/2 \quad (1.3)$$

とする. この場合は

$$\varepsilon = (\varepsilon_1 + \varepsilon_2) / 2 \quad (1.4)$$

以下, 誤り率 $\varepsilon_1, \varepsilon_2$ の求め方について論ずる. なお, 以下の議論の展開は主として文献 (11), (12) ならびに (13) を参考にしたものである.

1.3.1 ベイズ分類器

母集団の分布を規定するパラメータに関する完全な知識がある場合、ベイズ決定を達成するパターン分類器を構成することができる。ベイズ分類器を、それを構成するときに用いた既知のパラメータに従う母集団に対して適用するときには、達成される識別性能はベイズ誤り率となる。ベイズ分類器を設計するときに用いる母集団のパラメータとベイズ誤り率との間には関数関係があり、母集団のパラメータを決定すると、それに伴うベイズ誤り率は一意に決定できる。すなわち、ベイズ誤り率は二つの母集団の分離の程度を示す重要なパラメータであるともいえる。しかし、二つの母集団の分布に多次元正規分布を仮定しても、母集団のパラメータからベイズ誤り率を簡単に求める手法は見つけれられていない。このため、ベイズ誤り率を求めるためには以下に示すような数値積分によらなければならない。また、母集団のパラメータからベイズ誤り率は一意に決まるが、逆は成り立たない。すなわち、ベイズ誤り率を与えても、与えられたベイズ誤り率を示す母集団の組み合わせは無数に存在し、一意に決めることはできない。

以下に二つの母集団のパラメータを与えたときのベイズ誤り率 $\varepsilon_1, \varepsilon_2$ の求め方を示す。

(1) 共分散行列が等しい場合 多次元正規分布に従う二つのパターン母集団の共分散行列が等しい場合、ベイズ決定は次式の線形識別関数による識別になる。

第1章 序論

$$h(\mathbf{x}) = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} \mathbf{x} + \frac{1}{2} (\boldsymbol{\mu}_1^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2) \quad (1.5)$$

ただし $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$

未知パターン \mathbf{x} について、式(1.5)の識別関数を用いて

$$h(\mathbf{x}) \begin{cases} < 0 & : \text{クラス1} \\ > 0 & : \text{クラス2} \end{cases} \quad (1.6)$$

のように分類する。このときのベイズ誤り率 ε_B は次のようにして求められる。クラス1に属する \mathbf{x} が多次元正規分布 $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ に従う場合、式(1.5)が \mathbf{x} に関する1次変換であることから、 $h(\mathbf{x})$ も正規分布 $N(-\eta, \sigma^2)$ に従う（付録1.1参照）。

$$\text{但し, } \eta = \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (1.7)$$

$$\sigma^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (1.8)$$

同様にして、未知パターン \mathbf{x} がクラス2に属するときは $h(\mathbf{x})$ は正規分布 $N(\eta, \sigma^2)$ に従う。 σ^2 は二つのクラス間の2乗マハラノビス距離である。パターン \mathbf{x} がクラス1（もしくはクラス2）に属するときに $h(\mathbf{x})$ の確率密度関数を $ph(h|\omega_1)$ （もしくは $ph(h|\omega_2)$ ）で表すと、 $ph(h|\omega_1)$ は

$$ph(h|\omega_1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left[\frac{h+\eta}{\sigma} \right]^2} \quad (1.9)$$

である。このとき、クラス1（クラス2）のパターンをクラス2（クラス1）と誤分類する確率 ε_1 (ε_2) は、それぞれ次式で与えられる。

$$\varepsilon_1 = \int_0^\infty ph(h|\omega_1) dh \quad (1.10)$$

$$\varepsilon_2 = \int_{-\infty}^0 ph(h|\omega_2) dh \quad (1.11)$$

式(1.10)(1.11)より明らかなように

$$\varepsilon_1 = \varepsilon_2 \quad (1.12)$$

であり、式(1.4)より

$$\varepsilon_B = \varepsilon_1 = \varepsilon_2 \quad (1.13)$$

式(1.10)(1.11)に適当な変数変換を施すことによりベイズ誤り率 ε_B は次に示すようになる。すなわち、二つの分布間のマハラノビス距離 σ （式(1.8)）のみで決まる。

$$\varepsilon_B = \Phi(-\sigma/2) \quad (1.14)$$

ただし、 $\Phi(x)$ は標準正規分布関数、すなわち、

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt \quad (1.15)$$

である。

パターンが1次元のときの母集団の分布と識別境界、ベイズ誤りの関係を図1.2に示す。横軸がクラス1, 2のパターンの値, 縦軸が各パターンの確率密度関数の値を示す。クラス1, 2の事前確率が等しいときには, 識別境界は二つのクラスの確率密度関数が等しくなるAの点を境界とすべきである。そのとき, ハッチングの部分の面積がベイズ誤り率となる。

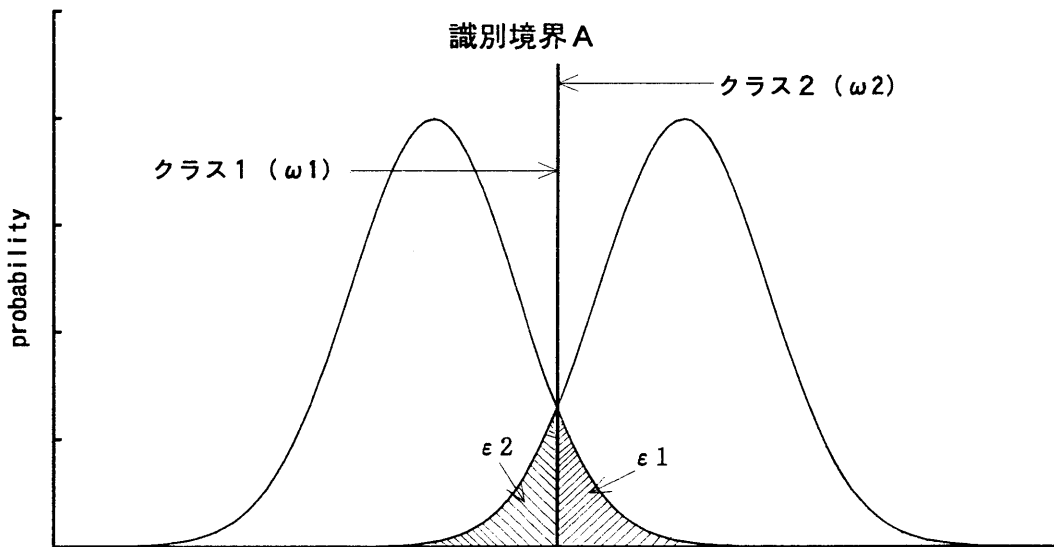


図1.2 ベイズ分類器－共分散行列が等しい場合（線形識別関数）

（2）共分散行列が異なる場合 二つの母集団の共分散行列が異なる場合, ベイズ決定は次式の二次識別関数による識別となる。

$$h(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}_1^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}_2^{-1}(\mathbf{x} - \boldsymbol{\mu}_2) + \frac{1}{2} \ln \frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|} \quad (1.16)$$

未知パターン \mathbf{x} について, 式(1.16)の2次識別関数を用いて式(1.6)と同じ条件によ

第 1 章 序論

りクラス 1 と 2 を分類する. ベイズ誤り率 ε_1 , ε_2 は式(1.17)により求められる.

$$\begin{aligned}\varepsilon_1 &= \int_{h(\mathbf{x}) > 0} p_1(\mathbf{x}) d\mathbf{x} \\ \varepsilon_2 &= \int_{h(\mathbf{x}) < 0} p_2(\mathbf{x}) d\mathbf{x}\end{aligned}\quad (1.17)$$

式(1.17)に基づく 2 次識別関数のベイズ誤り率 ε_1 , ε_2 は式(1.16)が \mathbf{x} に関する 2 次の項を含むことから, 多次元空間における積分境界が複雑な形となり, 一般には解析的に求めることはできない. Fukunaga らは二つの共分散行列の同時対角化とフーリエ変換を用いて巧妙に多次元積分を 1 次元積分に変換している⁽¹¹⁾ が, その 1 次元の定積分も解析的に求めることができない. このためベイズ誤り率を求めるには数値積分によるしかない. この手法は有限個の学習標本から推定したパラメータを用いて設計した分類器の識別性能を求める際に実行される, より一般的な数値積分に帰着されるので, 1.3.2 で詳細に述べる.

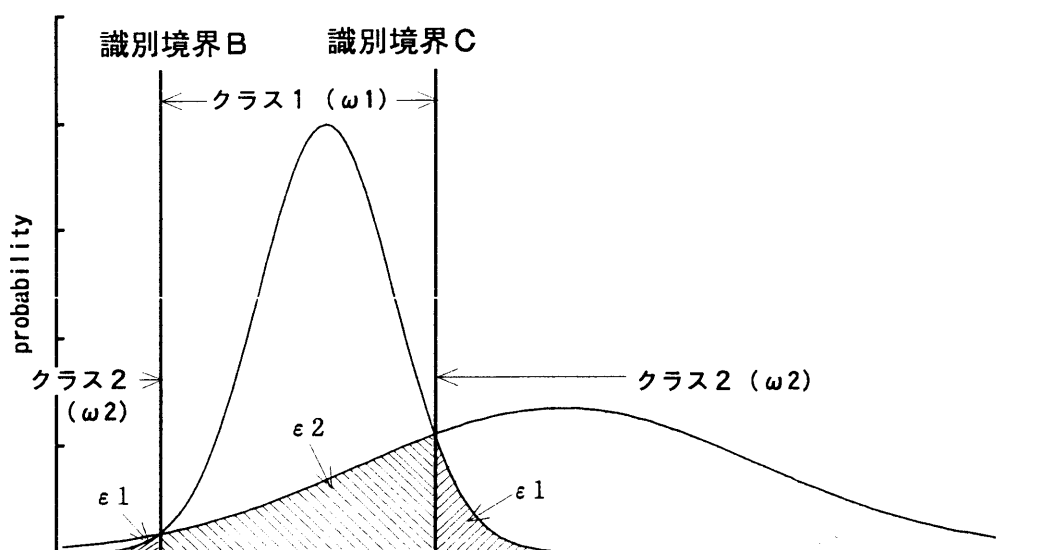


図 1.3 ベイズ分類器—共分散行列が異なる場合 (2 次識別関数)

パターンが 1 次元の場合のベイズ誤り率を図 1.3 に示す. 二つのクラスの共分散行列の違いは, 1 次元の場合には分散の相違に相当し, 2 次識別関数は 2 つの識別境界 (図の B, C) を有する.

1.3.2 標本を用いて設計するパターン分類器

有限学習標本に基づいて推定した母集団パラメータは通常、真の母集団パラメータからずれる。このため、推定パラメータを用いて設計されたパターン分類器の識別性能はベイズ分類器のそれに比べて劣る。従って、パターン分類器の性能はベイズ分類器からの性能の劣化の程度から評価されるべきである。これを実験的に行うものとして次のような方法が考えられる。まず、あるベイズ誤り率を有する母集団パラメータを想定する（前述したように一つのベイズ誤り率が一組の母集団パラメータには対応しないので、母集団パラメータの決め方には工夫を要する）。母集団パラメータが決定されると、ベイズ分類器が設計でき、ベイズ誤り率が決まる。これとは別に、母集団パラメータに従う有限個の学習標本 \mathbf{x}_{ji} （添え字 j が標本番号 $1 \sim m_i$ 、 m_i がクラス i の学習標本の個数、 i がクラス番号 1 または 2）が人工的に生成できる。二つの母集団から有限個の学習標本を生成し、これらの学習標本を用いて母集団パラメータを式(1.18)(1.19)に従い推定し、クラス毎の標本平均値ベクトル $\hat{\mu}_i$ 、標本共分散行列 $\hat{\Sigma}_i$ とする。

$$\hat{\mu}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \mathbf{x}_{ji} \quad (1.18)$$

$$\hat{\Sigma}_i = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (\mathbf{x}_{ji} - \hat{\mu}_i)^t (\mathbf{x}_{ji} - \hat{\mu}_i) \quad (1.19)$$

これらの推定したパラメータ $\hat{\mu}_i$ 、 $\hat{\Sigma}_i$ を用いて、有限個の学習標本に基づく分類器を設計する。ここで、推定する母集団のパラメータ $\hat{\mu}_i$ 、 $\hat{\Sigma}_i$ も確率変数となり、有限個の学習標本の抽出の都度、確率的に変動する。式(1.18)(1.19)の期待値は真の値に一致する不偏推定量であり、かつ学習標本数を無限大にしたときには、母集団パラメータに一致する一致推定量でもある。しかしながら、一組の有限個の学習標本を抽出し、その学習標本から推定したパラメータを用いて設計した分類器は、設計の際に用いるパラメータが真の母集団パラメータに一致しないため、真の母集団パラメータに従うテスト標本を分類したとしても、ベイズ決定は達成されない。このときの学習標本に基づいて設計された分類器が真の母集団パラメータに従うテスト標本を分類する性能は学習標本の有限性に起因するベイズ分類器からの性能の劣化の程度を示しており、学習標本の有限性に起因する分類器の性能評価の手法として妥当であると思われる。

このような有限個の学習標本に基づく分類器のふるまいに関しては、Raudys と

第 1 章 序論

Jain⁽⁸⁾, FukunagaとHayes⁽⁹⁾らによる研究があるが, 分類器の性能評価をする際, 多次元数値積分を行う必要があり, 次元数の増加と共に計算コストが急激に増大し, 系統的かつ包括的に研究された例はない.

統計的パターン認識においては, 母集団の確率分布関数について仮定をおくが, それ以外は特別な知識を利用しない. 従って, 母集団の共分散行列が等しいか否かということも, パターン分類器を設計するときには知られていない. このため, 有限学習標本に基づいて設計された分類器の性能評価を行う際には, 母集団の共分散行列が等しい場合と, 異なる場合のそれぞれにつき, 分類器として線形識別関数を用いた場合と 2 次識別関数を用いた場合の各々について検討する必要がある. すなわち, 母集団の共分散行列の異同により 2 種類, パターン分類器が 2 種類, 合わせて 4 種類の組み合わせが考えられる. 標本を用いて設計されたパターン分類器の性能を評価することが本研究の主たる目的であるので, ここでは母集団の性質によらず, パターン分類器の形により場合分けをして考える

(1) 線形識別関数 母集団の共分散行列が等しいか否か判明していない状況下で, 二つの母集団の共分散行列が等しいと仮定し, パターン分類をしようとする場合について考えてみる. 図 1. 2 に図示したベイズ誤りに対応する部分について, 線形識別関数の場合を図 1. 4 に示す. 母集団の分散が等しい場合も異なる場合も単一の識別境界 A' で分類する. 当然誤り率は図 1. 2 もしくは 1. 3 のベイズ分類器に比較し増加する.

学習標本に基づいて推定した二つのクラスの平均値ベクトルを $\hat{\mu}_1, \hat{\mu}_2$ とする. 分類器設計時点では共分散行列は等しいと仮定しているので, 二つのクラスの学習標本について併合して推定した標本共分散行列 (pooled covariance matrix) を $\hat{\Sigma}$ とする. 推定したパラメータに基づく線形識別関数は式(1.5)の μ_1, μ_2, Σ の代わりに推定したパラメータを形式的に代入して設計され, 式(1.20)となる.

$$\hat{h}(\mathbf{x}) = (\hat{\mu}_2 - \hat{\mu}_1)^t \hat{\Sigma}^{-1} \mathbf{x} + \frac{1}{2} (\hat{\mu}_1^t \hat{\Sigma}^{-1} \hat{\mu}_1 - \hat{\mu}_2^t \hat{\Sigma}^{-1} \hat{\mu}_2) \quad (1.20)$$

式(1.5)と同様に, 式(1.20)も \mathbf{x} に関する 1 次式であることから, \mathbf{x} が正規分布に従えば式(1.20)も正規分布に従う.

母集団の平均値ベクトルを μ_1, μ_2 とし, 二つのクラスの共分散行列が Σ に等しいとき, 式(1.20)のクラス毎の平均は式(1.21)(1.22)となり, 分散はクラスによらず等しく, 式(1.23)となる.

$$E \{ \hat{h}(\mathbf{x}) | \omega_1 \} = (\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)^t \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} (\hat{\boldsymbol{\mu}}_1^t \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2^t \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}}_2) \quad (1.21)$$

$$E \{ \hat{h}(\mathbf{x}) | \omega_2 \} = (\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)^t \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}_2 + \frac{1}{2} (\hat{\boldsymbol{\mu}}_1^t \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2^t \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}}_2) \quad (1.22)$$

$$Var \{ \hat{h}(\mathbf{x}) | \omega_1 \} = Var \{ \hat{h}(\mathbf{x}) | \omega_2 \} = (\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)^t \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1) \quad (1.23)$$

二つの母集団の共分散行列が異なり、それぞれ $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ のとき、式(1.20)の分散はクラス毎に異なり、式(1.24)(1.25)となる。

$$Var \{ \hat{h}(\mathbf{x}) | \omega_1 \} = (\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)^t \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma}_1 \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1) \quad (1.24)$$

$$Var \{ \hat{h}(\mathbf{x}) | \omega_2 \} = (\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)^t \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma}_2 \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1) \quad (1.25)$$

これらの正規分布の平均、分散の値を求めることができれば、誤り率 ε_1 , ε_2 は式(1.10)(1.11)と同様な 1 次元数値積分に帰着される。

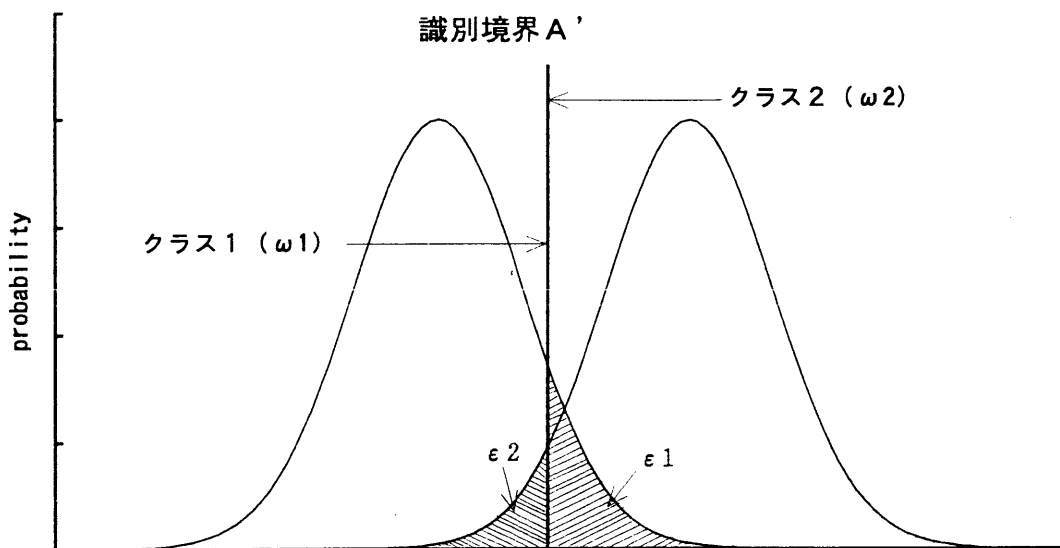
(2) 2 次識別関数 母集団の共分散行列に関する知識を用いず、二つの母集団が異なる共分散行列を有していると仮定して分類器を設計するときには、2 次識別関数を用いなくてはならない。

図 1.3 に対応するベイズ誤り率をグラフ表現したものを図 1.5 に示す。図 1.3 に示すように、2 次識別関数ではパターン之母集団の真の分布においては共分散行列が等しい場合でも、識別境界が B' , C' の 2 か所となる (分類器設計の時点では共分散行列が異なると仮定していることに注意)。このときの誤り率がベイズ誤り率より大きな値となることは明白である。

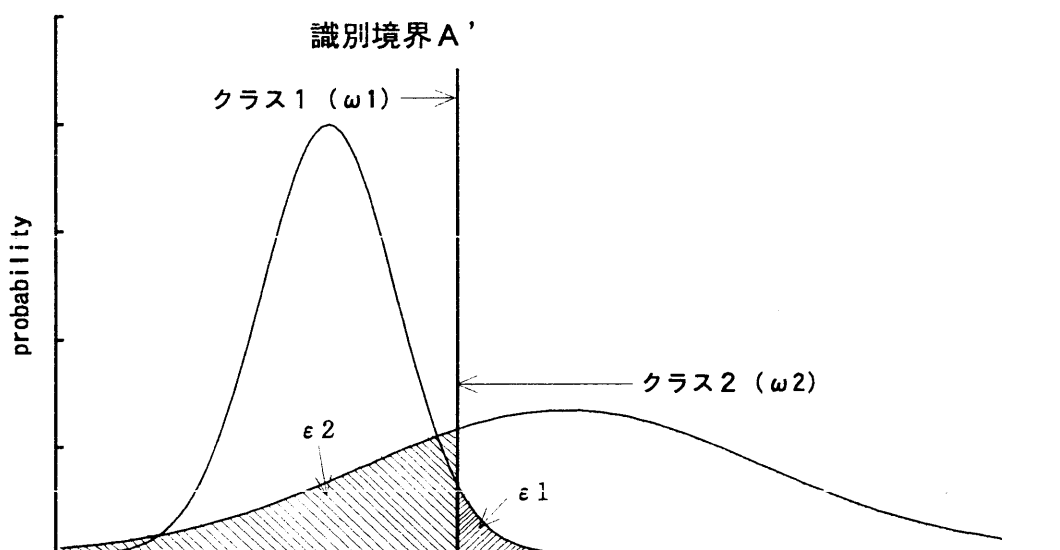
学習標本に基づいて推定した平均値ベクトルと共分散行列を、 $\hat{\boldsymbol{\mu}}_1$, $\hat{\boldsymbol{\mu}}_2$, $\hat{\boldsymbol{\Sigma}}_1$, $\hat{\boldsymbol{\Sigma}}_2$ とする。学習標本に基づいて設計する 2 次識別関数は、真のパラメータに基づく式(1.16)の識別関数の $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ の代わりに上記の推定パラメータを用いて式(1.26)となる。

$$\hat{h}(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \hat{\boldsymbol{\mu}}_1)^t \hat{\boldsymbol{\Sigma}}_1^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_1) - \frac{1}{2} (\mathbf{x} - \hat{\boldsymbol{\mu}}_2)^t \hat{\boldsymbol{\Sigma}}_2^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_2) + \frac{1}{2} \ln \frac{|\hat{\boldsymbol{\Sigma}}_1|}{|\hat{\boldsymbol{\Sigma}}_2|} \quad (1.26)$$

このときの誤り率は、式(1.26)が \mathbf{x} に関する 2 次の項を含むことから、前記線形識別関数のように簡単には求められないことは、前節で述べた通りである。次節でこの点をもう少し詳しく述べる。

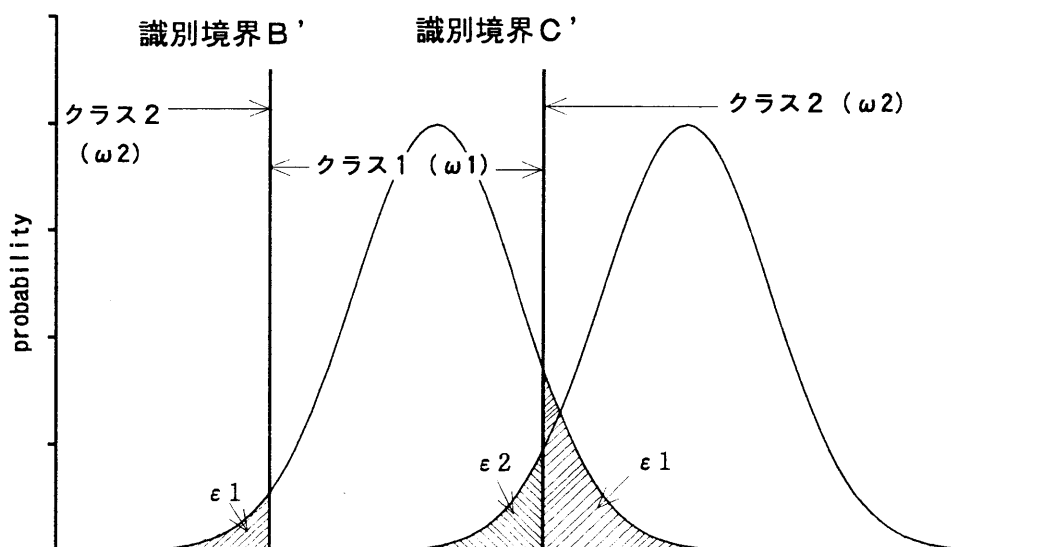


(1) 母集団の分散が等しい場合

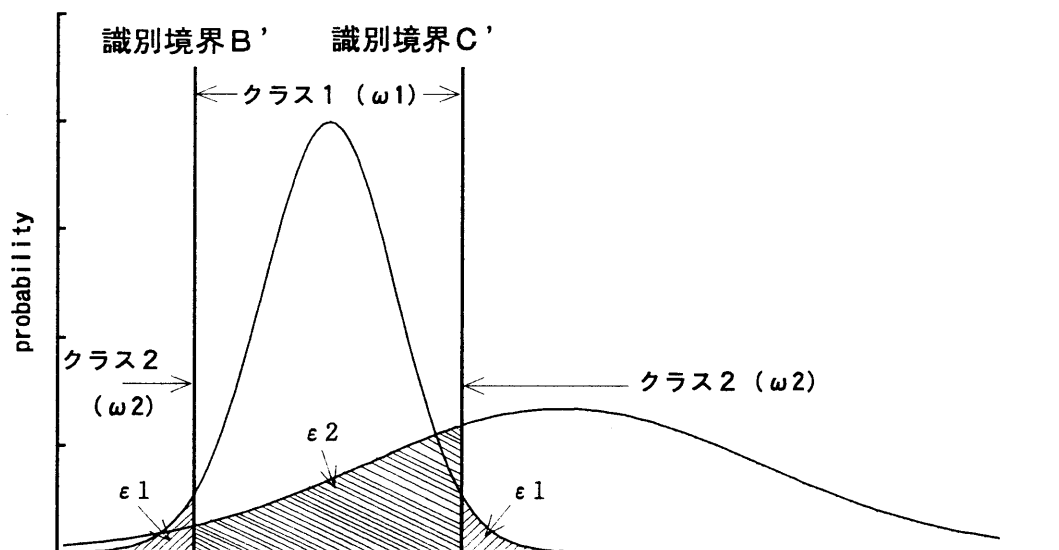


(2) 母集団の分散が異なる場合

図 1. 4 標本に基づく識別関数の決定境界－線形識別関数の場合
(曲線は真の確率密度関数を示す)



(1) 母集団の分散が等しい場合



(2) 母集団の分散が異なる場合

図 1.5 標本に基づく識別関数の決定境界－2 次識別関数の場合

(曲線は真の確率密度関数を示す)

第 1 章 序論

(3) 2次識別関数の誤り率 2次識別関数の誤り率を一般的に求めるため、2次識別関数を構成する母集団のパラメータの標本からの推定値を平均値ベクトル $\hat{\mu}_1, \hat{\mu}_2$, 共分散行列 $\hat{\Sigma}_1, \hat{\Sigma}_2$ とする。これにテストのための標本 \mathbf{x}_t を加えたとき、それがどちらのクラスに分類されるかテストする。テスト標本の確率密度関数を $p_t(\mathbf{x})$ とし、平均値ベクトル μ_t , 共分散行列 Σ_t の多次元正規分布に従うとする。このときテスト標本 \mathbf{x}_t の分類は有限学習標本に基づく分類器である式(1.26)により、

$$\hat{h}(\mathbf{x}_t) \begin{cases} < 0 & : \text{クラス 1} \\ > 0 & : \text{クラス 2} \end{cases} \quad (1.27)$$

という分類規則に従って行われる。標本 \mathbf{x}_t をクラス1（クラス2）に分類する確率を q_1 （ q_2 ）とすれば、 q_1 （ q_2 ）は式(1.28)に従う。

$$\begin{aligned} q_1 &= \int_{\hat{h}(\mathbf{x}) < 0} p_t(\mathbf{x}) d\mathbf{x} \\ q_2 &= \int_{\hat{h}(\mathbf{x}) > 0} p_t(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (1.28)$$

これは式(1.17)と同じ形であり、一般には解析的に求めることができない。

平均値ベクトルが μ_1, μ_2 , 共分散行列が Σ_1, Σ_2 に従う母集団を想定し、この母集団から有限個の学習標本を取り出し、式(1.18)(1.19)に基づいて母集団のパラメータを推定する。推定したパラメータ $\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}_1, \hat{\Sigma}_2$ を用い、標本に基づく2次識別関数 $\hat{h}(\mathbf{x})$ を設計する。この分類器の分類性能を評価するためには、テスト標本 \mathbf{x}_t としてもう一度クラス1もしくはクラス2の母集団に従う標本を抽出し、分類テストを行えばよい。すなわち、テスト標本 \mathbf{x}_t としてクラス1（クラス2）の標本を用い、式(1.28)に従い q_2 （ q_1 ）を求めれば、クラス1（クラス2）の標本をクラス2（クラス1）に分類する誤り確率 ε_1 （ ε_2 ）が求められる。以下、一般的にテスト標本 \mathbf{x}_t は平均値ベクトル μ_t , 共分散行列 Σ_t の多次元正規分布に従うものとする。

行列 Σ_t および $(\hat{\Sigma}_1 - \hat{\Sigma}_2)$ を同時対角化するため、次の条件を満たす変換行列 \mathbf{A} を用いる。具体的な \mathbf{A} の求め方は付録1.3による。

$$\begin{cases} \mathbf{A}^t \Sigma_t \mathbf{A} = \mathbf{I} \\ \mathbf{A}^t (\hat{\Sigma}_1^{-1} - \hat{\Sigma}_2^{-1})^{-1} \mathbf{A} = \Lambda \end{cases} \quad (1.29)$$

変換行列 \mathbf{A} を用い、次の線形変換を行う。

$$\mathbf{y} = \mathbf{A}^t (\mathbf{x} - \boldsymbol{\mu}_t) \quad (1.30)$$

式(1.26)の標本に基づく 2 次識別関数 $\widehat{h}(\mathbf{x}_t)$ は

$$g(\mathbf{y}) = \frac{1}{2} \mathbf{y}^t \boldsymbol{\Lambda}^{-1} \mathbf{y} - \mathbf{v}^t \mathbf{y} + c \quad (1.31)$$

となる。ただし、

$$\left\{ \begin{array}{l} \mathbf{v} = \mathbf{K}_1^{-1} \mathbf{D}_1 - \mathbf{K}_2^{-1} \mathbf{D}_2 \\ c = \frac{1}{2} (\mathbf{D}_1^t \mathbf{K}_1^{-1} \mathbf{D}_1 - \mathbf{D}_2^t \mathbf{K}_2^{-1} \mathbf{D}_2) \\ \mathbf{K}_1 = \mathbf{A}^t \widehat{\boldsymbol{\Sigma}}_1 \mathbf{A}, \quad \mathbf{K}_2 = \mathbf{A}^t \widehat{\boldsymbol{\Sigma}}_2 \mathbf{A} \\ \mathbf{D}_1 = \mathbf{A}^t (\widehat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_t), \quad \mathbf{D}_2 = \mathbf{A}^t (\widehat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_t) \end{array} \right. \quad (1.32)$$

これらの変換をすると式(1.28)は

$$\begin{aligned} q_1 &= \int_{\widehat{h}(\mathbf{x}) < 0} p_t(\mathbf{x}) d\mathbf{x} \\ &= \int_{\widehat{g}(\mathbf{y}) < 0} r_t(\mathbf{y}) d\mathbf{y} \end{aligned} \quad (1.33)$$

$$\text{ただし, } r_t(\mathbf{y}) = \frac{1}{(2\pi)^{N/2}} e^{-\frac{1}{2} \mathbf{y}^t \mathbf{y}}$$

線形識別関数の場合と同じように、ステップ関数のフーリエ変換を利用して、式(1.33)を解くと、

$$\begin{aligned} q_1 &= \int u(-\widehat{g}(\mathbf{y})) r_t(\mathbf{y}) d\mathbf{y} \\ &= \frac{1}{2\pi} \iint_{-\infty}^{+\infty} (\pi \delta(\omega) - \frac{1}{j\omega}) e^{j\omega \widehat{g}(\mathbf{y})} r_t(\mathbf{y}) d\omega d\mathbf{y} \\ &= \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{F(\omega)}{j\omega} d\omega \end{aligned} \quad (1.34)$$

となる。ここで、 $F(\omega)$ は、

$$F(\omega) = \prod_{i=1}^N \left[\frac{1}{\sqrt{1 - \frac{j\omega}{\lambda_i}}} e^{-\frac{1}{2} \left\{ \frac{\omega^2 v_i^2}{1 - \frac{j\omega}{\lambda_i}} \right\}} \right] \cdot e^{j\omega c} \quad (1.35)$$

である。ただし、

第 1 章 序論

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N), \quad \mathbf{v} = (v_1, v_2, \dots, v_N)^T \quad (1.36)$$

である。式(1.34)は $F(\omega)$ を実部と虚部に分けると、実部が偶関数、虚部が奇関数となり、積分区間が正負対称であることから、分母の ω と合成され、実部は積分値が0となる。一方、虚部は積分範囲が半無限区間である1次元積分に帰着される。すなわち、

$$q_1 = \frac{1}{2} - \frac{1}{\pi} \int_0^{+\infty} \frac{|F(\omega)|}{\omega} \sin \{ \angle F(\omega) \} d\omega \quad (1.37)$$

$$\text{ただし, } |F(\omega)| = \prod_{i=1}^N \frac{1}{\left\{ 1 + \left(\frac{\omega}{\lambda_i} \right)^2 \right\}^{1/4}} e^{-\frac{1}{2} \left\{ \frac{\omega^2 v_i^2}{1 + \left(\frac{\omega}{\lambda_i} \right)^2} \right\}}$$

$$\angle F(\omega) = \sum_{i=1}^N \left\{ \frac{1}{2} \tan^{-1} \left(\frac{\omega}{\lambda_i} \right) - \frac{1}{2} \cdot \frac{\omega^2 v_i^2}{1 + \left(\frac{\omega}{\lambda_i} \right)^2} \cdot \frac{\omega}{\lambda_i} \right\} + \omega c$$

式(1.37)は解析的に求めることはできないが、数値積分であれば実行することができる。

1.4 本研究に関連する国内外の研究

多次元正規分布における有限学習標本の問題はかなり古くから統計的パターン認識の課題として数多くの研究が成され、優れた調査研究もいくつか報告されている。Toussaintは有限学習標本に基づく誤り率の推定に関する幅広い調査研究を行っている⁽¹⁴⁾。また、Lachenbruchはその著書「判別分析」の巻末において広範な文献調査を行っている⁽¹³⁾。最近では浜口が統計的パターン認識における識別理論と特徴抽出理論の最近の研究成果を解説している⁽¹⁰⁾。この中で、特に誤り率の推定を行うブートストラップ法について詳しく調査している。また、DudaとHart⁽⁵⁾、Fukunaga⁽¹²⁾、McLachlan⁽¹⁵⁾らの著書は統計的パターン認識全般の優れた解説を与え、多数の文献も含んでいる。

有限学習標本に基づくパターン分類器の識別性能を識別実験によって評価する方法として、再代入法(Resubstitution method)⁽¹²⁾、分割法(Holdout method)⁽¹⁶⁾、一つ取って置き法(Leave-one-out method)⁽⁷⁾などが広く知られている。一般に再代入法は

正しい識別性能に比較して楽観的な評価を示す（評価の良い方への偏りが大きい）という欠点を有し、分割法は標本が少ないとき性能評価のためのテスト標本を確保すると設計に用いる学習標本数が減少してしまうというジレンマを有している。分割法の一変形である一つ取って置き法は平均的にはパターン分類器の性能をほぼ正しく評価できるが、推定する誤り率の分散が大きくなるという欠点を有している。Efronはブートストラップ法⁽¹⁷⁾⁻⁽¹⁹⁾と呼ぶ評価方法を提案し、識別性能の評価に関する分散を小さくすることに成功している。これらの性能評価法は母集団の分布として特別な分布を仮定せず、分類器の設計と評価を同一セットの標本を用いて行うことができる。

一方、あらかじめ母集団のパラメータを設定しておけば、パターン分類器の設計のための学習標本と識別性能の評価のためのテスト標本を独立に発生させることができ、ベイズ決定からの性能の低下の程度は識別実験を行うことにより可能となる。しかしながら、この方法はベイズ誤り率が小さいところでは、膨大なテスト標本を用意する必要が生じる上、テスト標本の有限性に起因する推定誤差要因を除くことができないため、学習標本の有限性のみに起因する認識システムの識別性能の評価ができない。母集団の分布に正規分布を仮定すると、前節で述べたようにパターン分類器の性能評価に数値積分を導入することにより正確な識別性能の評価ができる。しかしながら、特徴パターンの次元数が増加するとともに計算量の増加を招き、さらに、識別性能を示す誤り率自身が学習標本の選び方によって確率的に変動するため、誤り率を確率変数として考えなければならなくなり、計算量の増大は甚だしいこととなる。Fukunagaは数値積分により誤り率を求める方法を示しているが、学習標本数に関係しない多次元正規分布において共分散行列が異なる場合のベイズ誤り率を求めているに過ぎない⁽¹¹⁾。同時にFukunagaは学習標本の有限性に起因するベイズ誤り率からの識別性能の低下の程度を示す近似式の導出に関連して、共分散行列が等しい場合について数値積分により誤り率を求めている⁽⁹⁾。この他に、誤り率を確率変数としてとらえ、パターン分類器の性能評価を数値積分により包括的かつ系統的に実施した例は見当たらない。

パターンの確率分布の未知パラメータを標本パターンから推定する問題は統計的推定の一問題である。しかし、推定の良否が認識システムの識別性能に及ぼす影響については、前述したように識別性能を示す誤り率自身が確率的に変動するため、その評価は困難な問題である。平均値ベクトルの推定量として標本の単純平均がよく用いられているがSteinは二乗誤差最小という基準のもとでは最良でない(admissibleになら

第 1 章 序論

ない)ことを示し⁽²⁰⁾, 統計学の分野でセンセーションを巻き起こした。したがって, 母集団の分布に正規分布を仮定して設計する式(1.20)(1.26)に示す線形識別関数や 2 次識別関数についても, 平均値ベクトルや共分散行列の推定値として式(1.18)(1.19)に示す標本平均値ベクトル, 標本共分散行列を用いることが一般的に行われているがこれが最良の方法であるという保証はない。

学習標本が少ないとき, 特に共分散行列の推定精度が認識システムの識別性能に悪影響を与えていることは従来から多くの研究者によって指摘されてきた^{(20) - (28)}。しかしながら, それらの研究はいずれも共分散行列それ自体の推定精度を検討するというよりも, 認識システムの識別性能との関連で共分散行列の推定精度についての評価をしている。最終的には, 推定の良否は認識システムの識別性能により評価することは必要であるが, 一般的には前述したように膨大な計算を必要とする。従って, 従来共分散行列の推定精度を上げるための試みは個別的な認識システムの識別性能の実験的な評価^{(29) - (33)}にとどまっており, 理論的もしくは系統的に研究された例は見当たらない。本研究においては, 2 次識別関数の主要な部分を成すマハラノビス距離に着目し, この推定精度の向上が認識システムの識別性能の向上につながると予想した。すなわち, 共分散行列の推定精度をマハラノビス距離と関連付けることにより, 2 クラス以上の認識システムを構成することなく, マハラノビス距離の推定精度を評価することにより, 共分散行列の推定の良否を評価するようにした点に特徴がある。

1.5 本論文の構成

本論文では, パターンの母集団の確率分布として多次元正規分布を仮定し, 有限学習標本に基づく統計的パターン分類器の設計と評価に関して論ずる。統計的パターン分類器を設計する際に必要となる主要なパラメータとしては, 平均値ベクトル, 共分散行列がある。これらのパラメータ自身の推定に関しては統計学を通して研究がなされているものの, これらの推定パラメータを用いて構成されるマハラノビス距離や統計的パターン分類器の誤り確率という諸統計量は, その振る舞いが複雑であり, 十分に調べられているとは言い難い。本論文は, 標本平均値ベクトルや標本共分散行列などの推定精度や, それらを用いて構成される上記諸統計量について, 主としてコンピュータによりモンテカルロ・シミュレーションを実施し, それに基づいて統計的パタ

ーン分類器の性能を評価したものである。以下に、各章の概要を述べる。

第 2 章では、2 次識別関数の主要な要素を占めるマハラノビス距離の推定誤差について論ずる。Stein が示した、” 3 次元以上のパターンでは、単純標本平均が平均二乗誤差最小という基準のもとでは最良の推定を与えない” という⁽²⁰⁾現象について確認したのち、さらに、Stein の提唱した平均値ベクトルに関する補正の効果も確認する。

さらに、Stein の補正は平均値ベクトルのみではなく、標本共分散行列の推定精度の向上をももたらし、その推定値を用いて構成されるマハラノビス距離についても、母集団の平均値ベクトルの絶対値が小さいほど Stein の補正の効果が大きいことを示す。

第 3 章では、マハラノビス距離の推定量の確率分布を理論的に明らかにし、有限学習標本を用いて推定した標本平均値ベクトルと標本共分散行列がマハラノビス距離の推定量に与える影響の程度を定量化する。マハラノビス距離の推定量の確率分布を前記二つの推定パラメータの成分に分解し、各々の成分毎の平均値と分散に関する理論式を導出する。その結果、標本共分散行列の推定誤差がマハラノビス距離の推定に大きな影響を与えていることを示す。

第 4 章では、有限学習標本から母集団のパラメータを推定し、その推定パラメータを用いて、線形識別関数と 2 次識別関数を構成し識別関数の性能を評価する。特徴パターンの次元数、学習標本数、および平均認識率の間の関係を明らかにし、Raudys⁽⁸⁾ および Fukunaga⁽⁹⁾ らの提案する近似認識率と比較し、それらの評価式の適用限界を明らかにする。さらに、認識率の低下分に関する評価を行い、ベイズ誤り率が高く、特徴ベクトルの次元数が高くなるに従い、認識率の低下分の分布は χ^2 分布に近くなることを示す。

2 次識別関数の値を実質的に支配するマハラノビス距離の推定値について、有限学習標本に起因して母集団のパラメータが正確に推定できないことでマハラノビス距離の推定誤差が増大する主たる要因は標本共分散行列の推定誤差にあることを第 3 章で明らかにする。第 5 章では、標本共分散行列の推定誤差に起因して、識別関数の性能が低下すること防止するため Friedman が提案した Regularized Discriminant Analysis⁽²⁸⁾ (以下 R D A 法と呼ぶ) により、有限学習標本に基づくパターン分類器の識別性能の低下を評価する。その結果、2 次識別関数が学習標本数に敏感なパターン分類器であることを確認する。また R D A 法の有効性を確認したのち、ベイズ認識率の高い状況では R D A 法の方のパラメータがあまり有効に作用しないことを示す。R D

第 1 章 序論

A法が有している二つのパラメータの自由度の内，より有効なパラメータを指摘し，より優れた統計的パターン分類器を設計できる見通しを得る．

第 6 章では，本論文を総括し，今後の展望と課題について述べる．

文 献

- (1)渡辺 慧：認識とパターン，第 1 章，岩波書店（1978）
- (2)鳥脇純一郎：認識工学－パターン認識とその応用－，第 3 章，コロナ社（1993）
- (3)大津展之：“パターン認識における特徴抽出に関する数理的研究”，電総研研究報告，第818号（1981）
- (4)竹村彰通：多変量推測統計の基礎，第 1 章，共立出版（1991）
- (5)Duda, R. O. and Hart, P. E.：“Pattern Classification and Scene Analysis” John Wiley & Sons, Inc., New-York,(1973)
- (6)Jain, A. K. and Chandrasekaran, B.：“Dimensionality and sample size consideration in pattern recognition practice”, in Handbook of Statistics, vol. 2, Krishnaiah, P. R. and Kanal, L. N. (Eds), North Holland, pp. 835-855, (1982)
- (7)Lachenbruch, P. A. and Mickey, M. R.：“Estimation of error rates in discriminant analysis”, Technometrics, Vol. 10, No. 1, pp. 1-11 (1968)
- (8) Raudys, S. J. and Jain, A. K.：“Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners”, IEEE Trans. Pattern Anal. and Machine Intell., Vol. 13, No. 3, pp. 252-264 (March 1991)
- (9)Fukunaga, K. and Hayes, R. R.：“Effects of sample size in classifier design”, IEEE Trans. Pattern Anal. and Machine Intell., Vol. 2, No. 8, pp. 873-885 (Aug. 1989)
- (10)浜本義彦：“パターン認識理論の最近の動向”，信学誌, Vol. 77, NO. 8, pp. 853-864, (1994)
- (11)Fukunaga, K. and Krile, T. F.：“Calculation of Bayes' Recognition Error for Two Multivariate Gaussian Distributions”, IEEE Trans. Computers, Vol. C-18, No. 3, pp. 220-229 (March 1969)
- (12)Fukunaga, F.：“Introduction to Statistical Pattern Recognition” 2nd ed.,

Academic Press, New-York, (1990) (Chapter 3)

(13) Lachenbruch, P. A., 鈴木義一郎・三宅章彦訳：判別分析, 現代数学社, (1979)

(14) Toussaint, G. T. : "Bibliography on Estimation of Misclassification",
IEEE Trans. Inform. Theory, Vol. IT-20, No. 4, pp. 472-479 (July 1974)

(15) McLachlan, G. J., : "Discriminant Analysis and Statistical Pattern Recognition", John Wiley & Sons, Inc., New-York (1992)

(16) Jain, A. K., : "Advances in statistical pattern recognition", Pattern Recognition Theory and Applications, Devijver, P.A. and Kittler, J., Eds., Springer-Verlag, pp. 1-19 (1987)

(17) Efron, B. : "Bootstrap methods: Another look at the jackknife", The Annals of Statist., Vol. 7, No. 1, pp. 1-26 (1979)

(18) Efron, B. : "Estimating the error rate of a prediction rule: Improvement on cross-validation", J. of American Statist. Assoc., Vol. 78, No. 382, pp. 316-331 (June 1983)

(19) Efron, B. : "How biased is the apparent error rate of a prediction rule?", J. of American Statist. Assoc., Vol. 81, No. 394, pp. 461-470 (June 1986)

(20) W. James, and C. Stein, "Estimation with Quadratic Loss", Proc. 4th Berkeley Symp. on Math. Statist. and Prob., pp. 361-379 (1961)

(21) Stein, C., Efron, B., Morris, C. : "Improving the usual estimator of a normal covariance matrix", Technical Report No. 37, Stanford: Department of Statistics, Stanford University (1972)

(22) Stein, C. : "Estimation of the mean of a multivariate normal distribution", Proc. Prague Symp. on Asymptotic Statistics. Karlova, Czechoslovakia: Prague University Press, pp. 345-381 (1973)

(23) Efron, B., Morris, C. : "Multivariate empirical Bayes and estimation of covariance matrices, The Annals of Statist., Vol. 4, No. 1, pp. 22-32 (1976)

(24) Olkin, I., Sellian, J.B. : "Estimating covariances in a multivariate normal distribution", Statistical Decision Theory and Related Topics II, S.S. Gupta and D. Moore (Eds.). New York: Academic Press, pp. 313-326 (1977)

(25) Haff, L.R. : "Empirical Bayes estimation of the multivariate normal covari-

第 1 章 序論

- ance matrix", The Annals of Statist., Vol. 8, pp. 586-597 (1980)
- (26) Lin, S. P., Perlman, M. D. : "A Monte Carlo comparison of four estimators for a covariance matrix", Multivariate Analysis, Vol. 6, P. R. Krishnaiah (Ed.), Amsterdam North-Holland, pp. 415-429 (1966)
- (27) Dey, D. K., Srinivasan, D. : "Estimation of a covariance matrix under Stein's loss", The Annals of Statist., Vol. 13, pp. 1581-1591 (1985)
- (28) Friedman J. H., "Regularized Discriminant Analysis", Journal of American Statistical Association, 84, No. 405, pp. 165-175 (1989)
- (29) Di Pillo, P. J., "Biased discriminant analysis: evaluation of the optimum probability of misclassification", Commun. Statist. - Theory Meth., Vol. A8, pp. 1447-1457 (1979)
- (30) Campbell, N. A. : "Shrunken estimators in discriminant and canonical variate analysis", Appl. Statist., Vol. 29, pp. 5-14 (1980)
- (31) Peck, R., Van Ness, J., : "The use of shrinkage estimators in linear discriminant analysis", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-4, pp. 530-537 (1982)
- (32) Kimura, F., Takashina, K., Tsuruoka, S., Miyake, Y. : "Modified quadratic discriminant functions and the application to Chinese character recognition", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-9, pp. 149-153 (1987)
- (33) Rodriguez, A. F. : "Admissibility and unbiasedness of the ridge classification rules for two normal populations with equal covariance matrices", Statistics, Vol. 19, pp. 383-388 (1988)

付録1.1 式(1.7)(1.8)の導出

\mathbf{x} が多次元正規分布 $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ に従うとき、 $h(\mathbf{x})$ の \mathbf{x} に関する期待値 $(-\eta)$ と分散 σ^2 を求める。

まず、期待値 $(-\eta)$ を求める。

$$\begin{aligned}
 E\{h(\mathbf{x}) \mid \omega_1\} & (\equiv -\eta) \\
 &= E\{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} \mathbf{x} + \frac{1}{2}(\boldsymbol{\mu}_1^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2) \mid \omega_1\} \\
 &= (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} E\{\mathbf{x} \mid \omega_1\} + \frac{1}{2}(\boldsymbol{\mu}_1^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2) \\
 &= (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \frac{1}{2}(\boldsymbol{\mu}_1^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2) \\
 &= \boldsymbol{\mu}_2^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \frac{1}{2}(\boldsymbol{\mu}_1^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2) \\
 &= -\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \tag{1.38}
 \end{aligned}$$

次に分散 σ^2 を求める。

$$\begin{aligned}
 E\{ (h(\mathbf{x}) - E\{h(\mathbf{x}) \mid \omega_1\})^2 \mid \omega_1\} & (\equiv \sigma^2) \\
 &= E\{ (h(\mathbf{x}) - \eta)^2 \mid \omega_1\} \\
 &= E\{ \left((\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} \mathbf{x} + \frac{1}{2}(\boldsymbol{\mu}_1^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2) \right. \\
 &\quad \left. + \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \right)^2 \mid \omega_1\} \\
 &= E\{ \left((\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} \mathbf{x} - (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 \right)^2 \mid \omega_1\} \\
 &= E\{ \left((\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) \right)^2 \mid \omega_1\} \\
 &= E\{ (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) (\mathbf{x} - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \mid \omega_1\} \\
 &= (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} E\{ (\mathbf{x} - \boldsymbol{\mu}_1) (\mathbf{x} - \boldsymbol{\mu}_1)^t \mid \omega_1\} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \\
 &= (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \\
 &= (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \tag{1.39}
 \end{aligned}$$

付録1.2 別の方法による式(1.10)(1.11)の導出

$h(\mathbf{x}) < 0$ の範囲で \mathbf{x} をクラス1と分類することから、クラス1をクラス2と誤る確率 ε_1 は

第1章 序論

$$\varepsilon_1 = \int_{h(\mathbf{x}) > 0} p_1(\mathbf{x}) d\mathbf{x} \quad (1.40)$$

と表し得る. 今, $\mathbf{y} = \Sigma^{-1/2}(\mathbf{x} - \mu_1)$ なる変換を施すと, \mathbf{y} は $N(\mathbf{0}, I)$ に従い, 式(1.40)は

$$\varepsilon_1 = \int_{g(\mathbf{y}) > 0} r_1(\mathbf{y}) d\mathbf{y} \quad (1.41)$$

となる. ただし,

$$r_1(\mathbf{y}) = \frac{1}{(2\pi)^{N/2}} e^{-1/2 \mathbf{y}^t \mathbf{y}} \quad (1.42)$$

$$g(\mathbf{y}) = (\mu_2 - \mu_1)^t \Sigma^{-1/2} \mathbf{y} - \eta. \quad (1.43)$$

式(1.41)の積分をステップ関数のフーリエ変換を用いて求める. $u(t)$ はステップ関数である. すなわち,

$$u(t) = \begin{cases} -1 & t < 0 \\ 1 & t > 0 \end{cases} \quad (1.44)$$

ステップ関数 $u(t)$ のフーリエ変換対は式(1.45)である.

$$u(t) \longleftrightarrow \pi \delta(\omega) + \frac{1}{j\omega} \quad (1.45)$$

ただし, $\delta(\omega)$ はデルタ関数であり, j は虚数単位である. ステップ関数を用いれば式(1.41)は \mathbf{y} に関する全空間の積分で次のように書かれる.

$$\varepsilon_1 = \int u(g(\mathbf{y})) r_1(\mathbf{y}) d\mathbf{y} \quad (1.46)$$

ステップ関数の逆フーリエ変換

$$u(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\pi \delta(\omega) + \frac{1}{j\omega}) e^{j\omega t} d\omega \quad (1.47)$$

を使えば, 式(1.46)は

$$\begin{aligned} \varepsilon_1 &= \frac{1}{2\pi} \iint_{-\infty}^{+\infty} (\pi \delta(\omega) + \frac{1}{j\omega}) e^{j\omega g(\mathbf{y})} r_1(\mathbf{y}) d\omega d\mathbf{y} \\ &= \frac{1}{2} \int r_1(\mathbf{y}) d\mathbf{y} + \frac{1}{2\pi} \iint_{-\infty}^{+\infty} \frac{1}{j\omega} e^{j\omega g(\mathbf{y})} r_1(\mathbf{y}) d\omega d\mathbf{y} \\ &= \frac{1}{2} + \frac{1}{2\pi} \iint_{-\infty}^{+\infty} \frac{1}{j\omega} e^{j\omega g(\mathbf{y})} r_1(\mathbf{y}) d\omega d\mathbf{y} \end{aligned} \quad (1.48)$$

式(1.48)の第2項の分子を \mathbf{y} について積分すれば,

$$\begin{aligned}
F(\omega) &= \int e^{j\omega g(\mathbf{y})} \cdot \frac{1}{(2\pi)^{N/2}} e^{-1/2 \mathbf{y}^t \mathbf{y}} d\mathbf{y} \\
&= \int e^{j\omega (\mathbf{W}^t \mathbf{y} - \eta)} \cdot \frac{1}{(2\pi)^{N/2}} e^{-1/2 \mathbf{y}^t \mathbf{y}} d\mathbf{y} \\
&= e^{-j\omega \eta} \prod_{i=1}^N \int \frac{1}{\sqrt{2\pi}} e^{-1/2 (y_i^2 - 2j\omega \omega_i y_i)} dy_i \\
&= e^{-j\omega \eta} \prod_{i=1}^N \int \frac{1}{\sqrt{2\pi}} e^{-1/2 (y_i^2 - 2j\omega \omega_i y_i - \omega^2 \omega_i^2 + \omega^2 \omega_i^2)} dy_i \\
&= e^{-1/2 \omega^2 \sum_{i=1}^N \omega_i^2 - j\omega \eta} \prod_{i=1}^N \int \frac{1}{\sqrt{2\pi}} e^{-1/2 (y_i^2 - j\omega \omega_i)^2} dy_i \\
&= e^{-1/2 \omega^2 \sigma^2 - j\omega \eta} \quad (1.49)
\end{aligned}$$

ただし,

$$\mathbf{W}^t = (\mu_2 - \mu_1)^t \Sigma^{-1/2}, \quad \mathbf{W} = (\omega_1, \omega_2, \dots, \omega_N)^t, \quad \sigma^2 = \mathbf{W}^t \mathbf{W}. \quad (1.50)$$

式(1.48)(1.49)より, 誤り率 ε_1 は

$$\varepsilon_1 = \frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1}{j\omega} e^{-1/2 \omega^2 \sigma^2 - j\omega \eta} d\omega \quad (1.51)$$

となる. オイラーの公式を用いて式(1.51)は,

$$\begin{aligned}
\varepsilon_1 &= \frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1}{j\omega} e^{-1/2 \omega^2 \sigma^2} (\cos \omega \eta - j \sin \omega \eta) d\omega \\
&= \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1}{\omega} e^{-1/2 \omega^2 \sigma^2} \cdot \sin \omega \eta d\omega. \quad (1.52)
\end{aligned}$$

式(1.53) (1.54) の正規分布密度関数と窓関数のフーリエ変換対

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{t}{\sigma}\right)^2} \longleftrightarrow e^{-1/2 \omega^2 \sigma^2} \quad (1.53)$$

$$G_{2\eta} \begin{cases} 1 & ; |t| < \eta \\ 0 & ; |t| > \eta \end{cases} \longleftrightarrow 2 \frac{\sin \omega \eta}{\omega} \quad (1.54)$$

を用いて, 式(1.52)は

$$\varepsilon_1 = \frac{1}{2} - \frac{1}{2} \int_{-\eta}^{\eta} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{t}{\sigma}\right)^2} dt$$

第1章 序論

$$= \frac{1}{2} - \int_0^{\pi} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2} dt \quad (1.55)$$

となり、本文の式(1.10)と等しくなる。式(1.11)も同様にして求められる。

付録1.3 式(1.29)の変換行列Aの求め方

テスト標本の共分散行列 Σ_t の固有ベクトル行列を Φ ，固有値行列を Θ とする。すなわち，

$$\Sigma_t \Phi = \Phi \Theta. \quad (1.56)$$

固有ベクトル行列 Φ ，固有値行列 Θ を用い，Kを求める。

$$K = \Theta^{-1/2} \Phi^t (\widehat{\Sigma}_1^{-1} - \widehat{\Sigma}_2^{-1})^{-1} \Phi \Theta^{-1/2} \quad (1.57)$$

$\widehat{\Sigma}_1$ ， $\widehat{\Sigma}_2$ が実対称行列なので，Kは一般には対角行列にはならないが，実対称行列となる。このKを固有値解析して，Kの固有ベクトル行列を Ψ ，固有値行列を Λ とする。

$$K \Psi = \Psi \Lambda \quad (1.58)$$

$$\Psi^t \Theta^{-1/2} \Phi^t (\widehat{\Sigma}_1^{-1} - \widehat{\Sigma}_2^{-1})^{-1} \Phi \Theta^{-1/2} \Psi = \Lambda \quad (1.59)$$

式(1.29)と(1.59)より，変換行列Aは

$$A = \Phi \Theta^{-1/2} \Psi \quad (1.60)$$

となる。

第2章 マハラノビス距離の推定および Stein補正の効果

誤り確率に密接に関係する統計量としてマハラノビス距離がある。マハラノビス距離の推定に用いる学習標本の数パターン次元に比較して十分でないと有意の推定誤差が生じる。マハラノビス距離の推定誤差は、対象とするパターンのパラメータ、具体的には平均値ベクトル、および、共分散行列が正しく推定できないために生ずる。

James & Stein (以下、単にSteinという) は、単純平均 (標本平均) が二乗誤差最小という基準ではadmissibleにならないことを示し、具体的に単純平均を補正する係数を提案している⁽¹⁾。本章では平均値ベクトルにSteinの補正を施し、計算機によるシミュレーション実験を行う。その結果、Steinの補正により平均値ベクトルの推定精度のみならず、マハラノビス距離の推定精度も向上すること、さらにパターンの各成分の分散が一定であれば、平均値ベクトルの絶対値が小さいほどSteinの補正による効果が大きいことを示す。

2.1 まえがき

パターンの確率分布の未知パラメータを標本パターンから推定することは統計的推定の一問題とみなされる。しかし、推定の良否が認識システムの性能に及ぼす影響については、古典的な問題でありながら十分に調べられているとは言いがたい。例えば、学習標本による認識率の評価法、特徴量の選択と増加に伴う認識率への効果、などはその代表例である⁽²⁾⁻⁽⁸⁾。これらの問題の難しさの主要な要因は認識率が確率分布のパラメータの複雑な関数になって、理論的解析が極めて困難な点にある。従って、認識に用いる諸量 (例えばパターン間距離など) がパターン集合の確率分布のパラメータの推定とどのように関連づけられるかを、シミュレーションによって明らかにしていくことは、上記諸問題の解明のために重要な意義をもつと考えられる。

そこで、本章では、パターン間の距離尺度の一つとして重要なマハラノビス距離に着目し、平均値ベクトルの推定精度との関係をシミュレーションによって実験的に検討する。

第2章 マハラノビス距離の推定およびStein補正

周知の如く、マハラノビス距離は、パターン集合（母集団）の平均値ベクトルと共分散行列に基づいて計算される。筆者らは、これまで、共分散行列の固有値、固有ベクトルによるマハラノビス距離の固有値・固有ベクトル表現を用いて、これらの固有値・固有ベクトルの推定誤差に伴うマハラノビス距離の推定誤差、およびそれと標本数などとの関係を計算機シミュレーションにより明らかにしてきた⁽⁹⁾。これは、いわば共分散行列の推定誤差（その固有値・固有ベクトルの推定と等価）の効果であるが、その他に平均値ベクトルの推定誤差もマハラノビス距離の誤差を引き起こす。

平均値ベクトルの推定量としては標本平均がよく用いられる。しかるに、Steinは、3次元以上のベクトルパターンに対しては標本の単純平均が二乗誤差最小という基準のもとでは最良ではない（admissibleにならない）ことを示し、具体的に単純平均を補正する係数を提案するとともに、改善量の理論値を示している^{(1) (10) (11)}。また、この補正が実際のデータに対して有効であることの例をEfron等が示している⁽¹²⁾。

そこで本章ではまず、平均値ベクトルの推定量として標本平均にSteinの補正を施したときのマハラノビス距離の推定誤差を明らかにするため、計算機によるシミュレーション実験を行う。その結果、Steinの補正を施すことにより平均値ベクトルの推定精度のみならず、マハラノビス距離の推定精度も向上すること、さらに分散が一定であれば、平均値ベクトルの絶対値が小さいほどSteinの補正による効果が大きいことを明らかにする。

次に、マハラノビス距離の推定誤差の理論値を示すYoungの式⁽¹³⁾とマハラノビス距離の固有値・固有ベクトル表現とを比較検討する。その結果、マハラノビス距離の推定誤差には、①固有値、固有ベクトルの推定誤差による成分と、②平均値ベクトルの推定誤差による成分とがあり、これら2つの推定誤差の成分がYoungの式と明解に対応すること、および、Steinの補正により前記①②の両成分共に推定誤差が減少すること、を明らかにする。

これらの結果はパターンの次元数に比較して標本数が少ないときの距離関数の改良に応用でき、パターン認識の認識性能の向上をもたらし得ると期待される。

以下2.2および、2.3で平均値ベクトルの推定とSteinの補正および、マハラノビス距離とその推定誤差に関する理論的成果を要約した後、2.4でSteinの補正に基づく推定量の効果をシミュレーション実験によって検証する。

2.2 平均値ベクトルの推定とSteinの補正

N 次元パターン \mathbf{x} の母集団の平均値ベクトルを μ , 共分散行列を Σ とする. また, M 個の標本パターン $\{\mathbf{x}_i; i=1, 2, \dots, M\}$ が与えられたとき, その標本平均値ベクトルを $\hat{\mu}$ とする. すなわち,

$$\hat{\mu} = \sum_{j=1}^M \mathbf{x}_j / M \quad (2.1)$$

特に, パターンの各成分の分散がすべて等しい(σ^2 とおく)と仮定すると, $\hat{\mu}$ に対する推定誤差(平均二乗誤差)は

$$E \{ (\mu - \hat{\mu})^t (\mu - \hat{\mu}) \} = \frac{N\sigma^2}{M} \quad (2.2)$$

これに対して, Stein は, パターンの次元数 N が3以上($N \geq 3$)のときには $\hat{\mu}$ のかわりに次式の $\hat{\mu}^*$ (以下Stein推定量とかく)を用いることを提案した⁽¹⁾.

$$\hat{\mu}^* = \left(1 - \frac{(N-2) \cdot \hat{\sigma}^2}{M(F+2) \cdot \mu^t \mu} \right) \hat{\mu} \quad (2.3)$$

ただし,

$$\hat{\sigma}^2 = \sum_{j=1}^M (\mathbf{x}_j - \hat{\mu})^t (\mathbf{x}_j - \hat{\mu}) / F \quad (2.4)$$

$$F = N(M-1) \quad (2.5)$$

そして, これによって推定量の平均2乗誤差が改善され, 式(2.6)のようになることも示した⁽¹⁾⁽¹⁰⁾.

$$E \{ (\mu - \hat{\mu}^*)^t (\mu - \hat{\mu}^*) \} = \frac{\sigma^2}{M} \cdot \left\{ N - \frac{F}{F+2} (N-2)^2 \cdot E \left\{ \frac{1}{N-2+2K} \right\} \right\} \leq \frac{N\sigma^2}{M} \quad (2.6)$$

$N \geq 3$

但し, F は式(2.5)を用いる. さらに, K は平均 $(M \cdot \mu^t \mu)/2$ のポアソン分布に従う変数である.

しかしながら, 上式による誤差の値は数式的には無限級数の形でしか与えられず, 数値例も示されていないため, この方法による改善量が実際にどの程度のものであるかは必ずしも明らかではない. この点に関して, 2.4節で実験的検討を行なう.

なお, この結果はパターンが正規分布以外の(適当な条件を満たす)分布に従う場合にも成り立つ⁽¹⁾.

第2章 マハラノビス距離の推定およびStein補正

2.3 マハラノビス距離とその推定誤差

2.3.1 マハラノビス距離の定義

N 次元パターン \mathbf{x} の母集団の平均値ベクトルを μ 、共分散行列を Σ とすると、 μ と \mathbf{x} との間の(二乗)マハラノビス距離 $dm^2(\mathbf{x})$ は次式で与えられる.

$$dm^2(\mathbf{x}) = (\mathbf{x} - \mu)^t \Sigma^{-1} (\mathbf{x} - \mu) \quad (2.7)$$

また、共分散行列 Σ の固有値とそれに対応する固有ベクトルを、それぞれ λ_i 、 ϕ_i ($i=1, 2, \dots, N$)で表わせば、マハラノビス距離は次のように書かれる(マハラノビス距離の固有値・固有ベクトル表現)⁽⁹⁾

$$\begin{aligned} dm^2(\mathbf{x}) &= \sum_{i=1}^N [\{\phi_i^t (\mathbf{x} - \mu)\}^2 / \lambda_i] \\ &= \sum_{i=1}^N d_i^2(\mathbf{x}), \quad d_i^2(\mathbf{x}) = \{\phi_i^t (\mathbf{x} - \mu)\}^2 / \lambda_i \end{aligned} \quad (2.8)$$

2.3.2 マハラノビス距離の推定

式(2.8)中の μ の代わりに標本平均値ベクトル $\hat{\mu}$ 、 ϕ_i 、 λ_i の代わりに標本共分散行列の固有ベクトル $\hat{\phi}_i$ と固有値 $\hat{\lambda}_i$ を用いたものは、マハラノビス距離の1つの推定値となる. このとき第 i 主軸成分(式(2.8)の右辺の i 番目の項)の推定値は次式

$$\hat{d}_i^2(\mathbf{x}) = \{\hat{\phi}_i^t (\mathbf{x} - \hat{\mu})\}^2 / \hat{\lambda}_i \quad (2.9)$$

で与えられ、その \mathbf{x} に関する平均は

$$E\{\hat{d}_i^2(\mathbf{x})\} = E[\{\hat{\phi}_i^t (\mathbf{x} - \hat{\mu})\}^2 / \hat{\lambda}_i] \quad (2.10)$$

$$= \left\{ \sum_{j=1}^N \lambda_j (\hat{\phi}_i^t \phi_j)^2 \right\} / \hat{\lambda}_i + \{\hat{\phi}_i^t (\mu - \hat{\mu})\}^2 / \hat{\lambda}_i \quad (2.11)$$

で与えられる(付録2.1参照).

式(2.11)の第1項は固有値と固有ベクトルの推定誤差の指標を、第2項は推定固有ベクトル、推定固有値空間で計測した平均値ベクトルの推定誤差を示しているとみなされる.

ここで、標本数 M を無限大にすると、 $\mu = \hat{\mu}$ となるので、式(2.11)右辺の第2項はゼロであり、さらに、固有ベクトルの直交性から、

$$\begin{aligned} \lim_{M \rightarrow \infty} E\{\hat{d}_i^2(\mathbf{x})\} &= 1/\lambda_i \cdot \left\{ \sum_{j=1}^N \lambda_j (\phi_i^t \phi_j)^2 \right\} \\ &= 1 \end{aligned} \quad (2.12)$$

となる.

一方、マハラノビス距離は

$$dm^2(\mathbf{x}) = \sum_{i=1}^N d_i^2(\mathbf{x}) \quad (2.13)$$

であり、式(2.11)に式(2.13)を適用すると

$$E\{\widehat{dm}^2(\mathbf{x})\} = \sum_{i=1}^N \left[\left\{ \sum_{j=1}^N \lambda_j (\widehat{\phi}_i' \phi_j)^2 \right\} / \widehat{\lambda}_i \right] + \sum_{i=1}^N \left[\left\{ \widehat{\phi}_i' (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \right\}^2 / \widehat{\lambda}_i \right] \quad (2.14)$$

となる。式(2.11)と同様に、式(2.14)の第1項は固有値と固有ベクトルの推定誤差の影響を、第2項は推定固有ベクトル、推定固有値空間で計測した平均値ベクトルの推定誤差を示していると考えられる。

一方、Youngによれば、パターンが正規分布に従うと仮定すると推定マハラノビス距離の期待値の理論値 $E\{\widehat{dm}^2(\mathbf{x})\}$ は、パターンの次元数を N 、標本数を M として、次の式(2.15)のように表現できる⁽¹³⁾。

$$E\{\widehat{dm}^2(\mathbf{x})\} = N \cdot \left[\frac{M+1}{M} \right] \left[\frac{M-1}{M-N-2} \right] \quad (2.15)$$

さらに、式(2.15)を2つの項に分解すると、次の結果を得る。

$$\frac{E\{\widehat{dm}^2(\mathbf{x})\}}{N} = \frac{M-1}{M-N-2} + \frac{1}{M} \cdot \frac{M-1}{M-N-2} \quad (2.16)$$

ここで、式(2.1)の推定量 $\widehat{\boldsymbol{\mu}}$ のかわりに式(2.3)のStein推定量 $\widehat{\boldsymbol{\mu}}^*$ を用いた場合、マハラノビス距離にどのような効果を生じるかはこれまで全く調べられておらず、以下2.4.3節で検討する。

2.4 Stein補正の効果

本節では、平均値ベクトルおよびマハラノビス距離の推定のために前記のStein推定量を用いた時の効果を実験的に検討する。

2.4.1 シミュレーションの手順

シミュレーションの手順の概略を以下に示す。

- (1) 母集団のパラメータ $\boldsymbol{\mu}$ 、および、 $\boldsymbol{\Sigma}$ を決める N 個の固有値 λ_i 、固有ベクトル ϕ_i ($i = 1, 2, \dots, N$) を、以下の式(2.17)以後に述べるような形で与える。
- (2) N 次元正規分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従う正規乱数 \mathbf{x} を M 個発生させ、大きさ M の標本パターン $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ を得る。

この標本パターンを用いて、推定量 $\widehat{\boldsymbol{\mu}}$ 、 $\widehat{\boldsymbol{\Sigma}}$ を求め、さらに、 $\widehat{\boldsymbol{\Sigma}}$ の固有値、固有ベクトルを求めて、 $\widehat{\lambda}_i$ 、および $\widehat{\phi}_i$ ($i = 1, 2, \dots, N$) とする。

第2章 マハラノビス距離の推定およびStein補正

(3) 平均値ベクトルの二乗誤差 $(\mu - \hat{\mu})^2$, および, 推定マハラノビス距離 $\hat{d}m^2(\mathbf{x})$ を求める.

(4) (2), (3)をL回繰り返して(3)で求められる諸量のL回にわたっての平均を求める. これを, 平均値ベクトルの推定実験における平均二乗誤差 $E\{(\mu - \hat{\mu})^2\}$, および, 推定マハラノビス距離 $E\{\hat{d}m^2(\mathbf{x})\}$ の実験値とみなす.

今回の実験では $L=1000$ とした. 実験に用いる正規乱数の発生方法は付録2.2に示す.

ここで, 上記(1)における母集団のパラメータ μ , λ_i , ϕ_i は次のように与えた.

$$\mu = (a, a, \dots, a)^t \quad (2.17)$$

但し, a は実験により異なる定数である. 以下の実験では $-1.0 \leq a \leq 1.0$ の範囲で0.1間隔で変化させた.

$$\lambda_i = \sigma^2 \cdot e^{-\alpha(i-1)} \quad (2.18)$$

α は, \mathbf{x} の各成分の分散の大きさを与えるパラメータで, それが小さければ分散は成分によらずほぼ一定であり, 大きくなるに従って成分毎に分散が大きく異なる. 具体的に, 以下の実験では $\alpha = 0.2, 0.1, 0.05, 0.01$, $\sigma^2 = 0.1, 1.0, 10.0$ とした.

さらに, ϕ_i は N 次元直交座標系の各基底ベクトルに等しくした. すなわち,

$$\left\{ \begin{array}{l} \phi_1 = (1, 0, 0, \dots, 0, 0)^t \\ \phi_2 = (0, 1, 0, \dots, 0, 0)^t \\ \phi_3 = (0, 0, 1, \dots, 0, 0)^t \\ \vdots \\ \vdots \\ \phi_n = (0, 0, 0, \dots, 0, 1)^t \end{array} \right. \quad (2.19)$$

2.4.2 平均値ベクトルの推定に関する実験結果

前節のようにして生成した標本パターンに対して, 平均値ベクトルを標本の単純平均により求める場合とSteinの補正(式(2.3)による)を施した場合とを比較する実験を行なった. 平均値ベクトルの真値と推定値との二乗誤差の期待値,

$$E\{(\mu - \hat{\mu})^t (\mu - \hat{\mu}) / (N \cdot \sigma^2)\} \quad (2.20)$$

を1000回の反復実験の結果の平均値により評価した. 結果を図2.1~2.3に示す.

第2章 マハラノビス距離の推定およびStein補正

実験毎の標本の抽出に伴うばらつきを評価するために図中, Steinの補正を加えた場合については, その理論値⁽¹⁾ (式(2.6)による) もプロットした.

図2.1は次元数 $N=10$, 標本数 $M=15\sim 100$ の場合, 図2.2は次元数 $N=10$, 標本数 $M=15$ および, $M=100$ の場合である. 図2.3は次元数 $N=10$, 標本数 $M=15$ の場合であり, いずれも $\sigma^2=1$ である. 図2.2, 2.3の横軸は式(2.17)で示した平均値ベクトル μ の各成分の大きさ a である. 図2.1の縦軸は式(2.20)の値を示す. 補正を施さない単純平均の場合, 式(2.20)の理論値は $1/M$ である. 図2.2, 2.3では式(2.20)にさらに標本数 M を掛けて, 正規化した. したがって, 補正がない場合には, 平均値ベクトルの各成分の大きさ a によらず, 評価式の理論値は1である.

Steinの補正を施した場合には, 式(2.20)中の $\hat{\mu}$ の代わりに $\hat{\mu}^*$ を用いた. 図中のパラメータ α は式(2.18)に示す固有値の減衰を表わすパラメータである.

図2.1より, 補正を施さない場合の平均値ベクトルの二乗誤差は標本数に反比例し, 理論値の $1/M$ を示している. Steinの補正を施すとその効果があらわれ, 平均二乗誤差は著しく改善される. また, その改善量も Steinの理論値とよく一致している.

図2.2では等分散に近い $\alpha=0.01$ の場合であり, 標本数 $M=15$ のときも $M=100$ のときもSteinの理論値とよく一致している. 平均値ベクトルの大きさ a が大きくなると共に補正の効果が減少すること, さらに標本数が大きくなるに従って, その傾向が強くなることが読み取れる.

L 回の反復に伴うマハラノビス距離の推定値の二乗誤差の分散は, マハラノビス距離の推定値の二乗誤差の期待値の二乗に比例し, その比例定数は学習標本数の増加に伴いほぼ $2/N$ に漸近する.

Steinの補正の式(2.3)は等分散を仮定しているが, 図2.3では, 等分散でない場合($\alpha=0.2$)でもSteinによる補正の効果があることを示している.

以上を要約すれば, 図2.1~2.3より Steinの補正により平均値ベクトルの推定精度が向上し, またその改善量は理論値とよく一致していることが知られる. さらに, 図2.2, 2.3より, 推定精度の向上は分散一定の条件のもとでは平均値ベクトルの絶対値(式(2.17)の a)が小さい時ほど顕著であること, 分散が一定でなくても補正の効果はあること, がわかる.

第2章 マハラノビス距離の推定およびStein補正

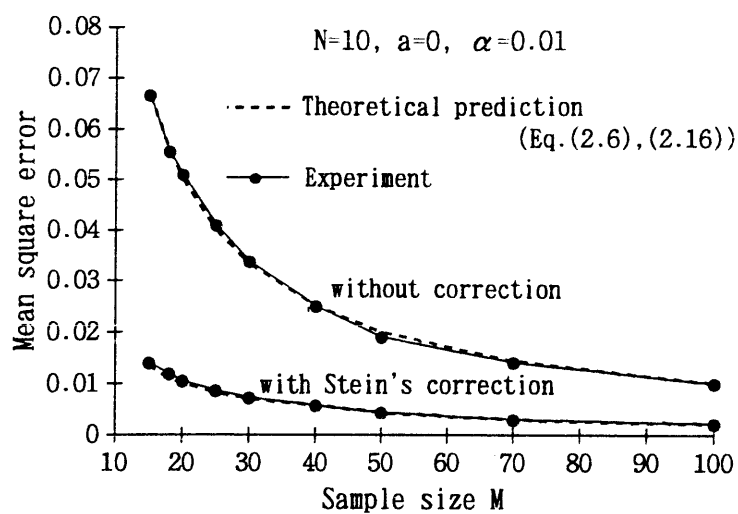


図2.1 平均値ベクトルの推定実験の結果(1)

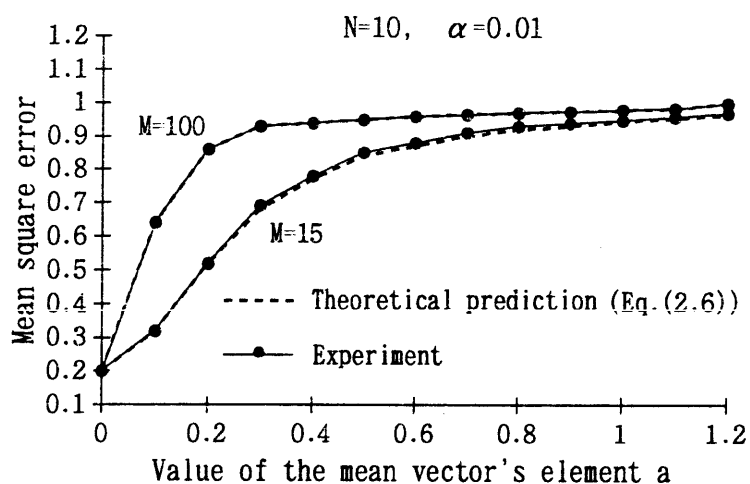


図2.2 平均値ベクトルの推定実験の結果(2)

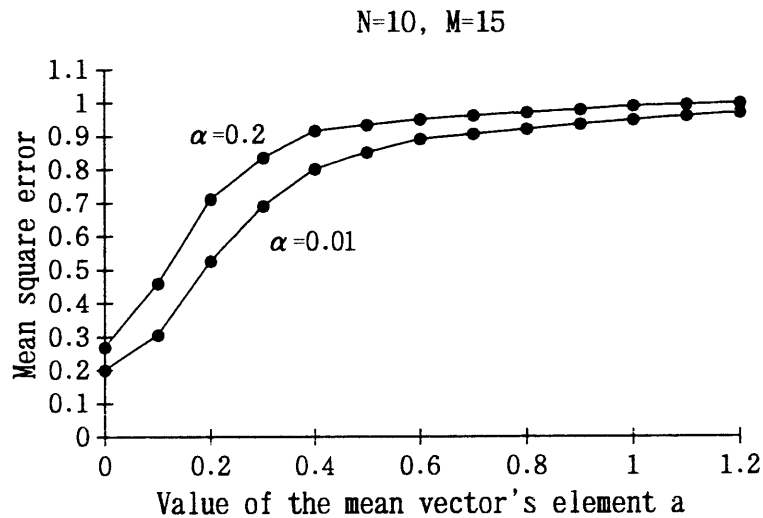


図 2. 3 平均値ベクトルの推定実験の結果(3)

2. 4. 3 マハラノビス距離の推定実験

(1) 推定誤差 2. 4. 1 の標本パターンに対して, マハラノビス距離の推定誤差を式(2.14)の第1項と第2項とに分けて求め, さらに式(2.16)の第1項と第2項と各々対応させ比較した. 式(2.14)の評価の際には, 式(2.16)の評価と一致させるため, 式(2.14)を次元数 N で割り, 正規化をしている. 図 2. 4, 2. 5 に次元数 $N=10$, 標本数 $M=15\sim 100$ の結果を示す. 図 2. 4 が式(2.14)(2.16)の第1項の固有値, 固有ベクトルの推定誤差に基づくマハラノビス距離の推定誤差, 図 2. 5 が前記2式の第2項をプロットしたものである.

式(2.12)より明らかなように, 第1項は標本数の増大と共に1に漸近し, 第2項は0に漸近してゆく. 図では標本の分散を与えるパラメータ α は0.01の場合のみを示している. なお図 2. 4, 2. 5 の実験では Steinの補正を施していない.

第2章 マハラノビス距離の推定およびStein補正

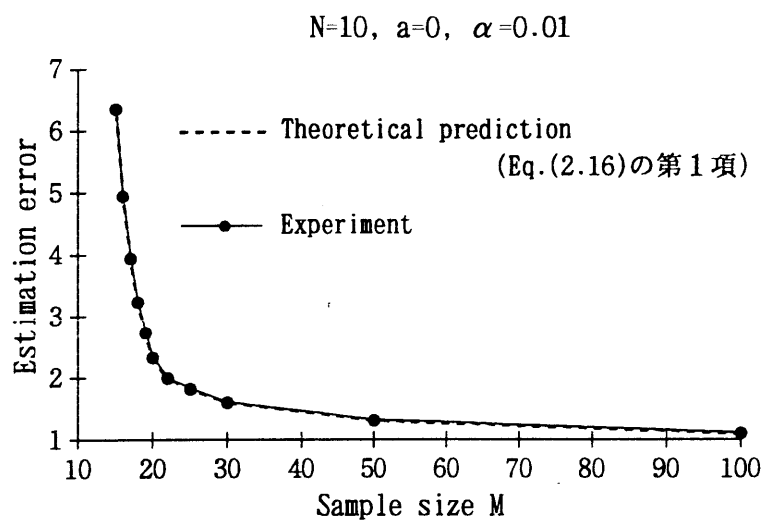


図2.4 マハラノビス距離の推定誤差(1)－共分散行列の推定誤差の効果

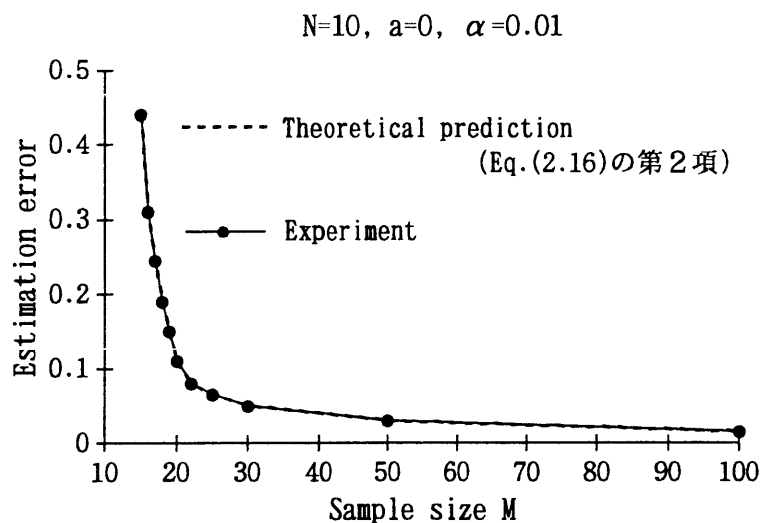


図2.5 マハラノビス距離の推定誤差(2)－平均値ベクトルの推定誤差の効果

第2章 マハラノビス距離の推定およびStein補正

図2.4, 2.5より明かなように式(2.14)の第1項, 第2項がYoungの理論式(2.16)と極めてよく対応することが分かる. すなわちこれは, マハラノビス距離の推定誤差が式(2.16)のように表現すると第1項は共分散行列の推定誤差に起因するものとし, 第2項は平均値ベクトルの推定誤差に起因するものを推定共分散行列で重み付けをして測定したものであることを示すものと考えられる. 結果的には共分散行列の推定誤差の方がマハラノビス距離の推定誤差に大きな影響を与えていることがわかる.

マハラノビス距離の推定誤差を引き起こすものとして共分散行列と平均値ベクトルの2つの要因があり, 今回の実験でその影響の程度を分離し定量化できた. 従って, 今後のSteinの補正を施さない場合の実験ではYoungの理論式を採用することとする.

(2) **Stein 補正の効果** 平均値ベクトルの推定にStein推定量を用いることの効果がマハラノビス距離の推定誤差に及ぼす影響を調べるため以下に示す実験を行った.

(1) 平均値ベクトルを式(2.3)に従い補正する.

(2) 推定共分散行列を次の式(2.21)に従い求める.

$$\hat{\Sigma} = \sum_{j=1}^M (\mathbf{x}_j - \hat{\boldsymbol{\mu}}^*)(\mathbf{x}_j - \hat{\boldsymbol{\mu}}^*)^t / (M-1) \quad (2.21)$$

(3) 式(2.21)により求めた推定共分散行列の固有値, 固有ベクトルを求めた後, 式(2.14)の第1項と第2項とにこの値を代入して, マハラノビス距離の推定を行う. 但し, 式(2.14)中の $\hat{\boldsymbol{\mu}}$ の代わりにStein推定量 $\hat{\boldsymbol{\mu}}^*$ を用いる.

図2.4, 2.5の場合と同様に, 次元による正規化を行うため, 式(2.14)で求めた値を次元数Nで割る.

実験結果を図2.6~2.9に示す. 図2.6は次元数N=10, 標本数M=15, 固有値の減衰を示す定数 $\alpha=0.01$ の場合, 図2.7は次元数N, 標本数Mは図2.6と同じ条件で $\alpha=0.2$ の場合である. 図中には, Youngの式(補正なしの場合に相当する), Steinの補正ありの各場合について,

①共分散行列(固有値, 固有ベクトルと等価)の推定誤差に起因する成分,

②平均値ベクトルの推定誤差に起因する成分,

③前記①と②との和(マハラノビス距離自身),

が示されている.

第2章 マハラノビス距離の推定およびStein補正

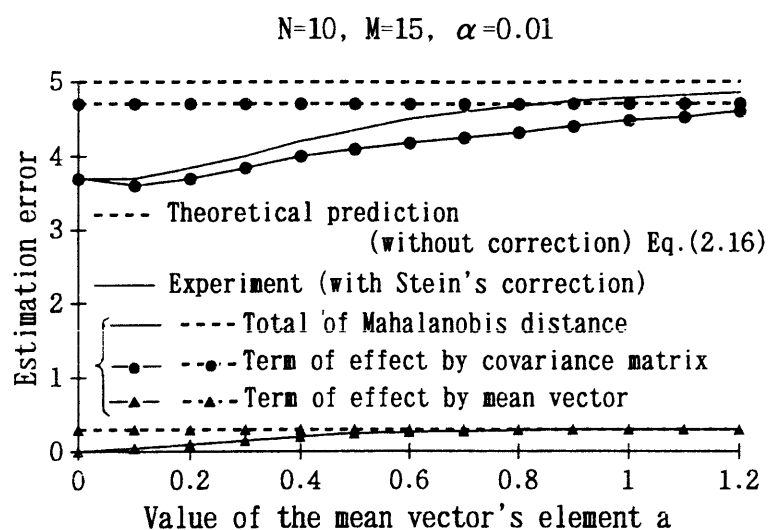


図 2. 6 Stein 推定量を用いたマハラノビス距離の推定実験(1)

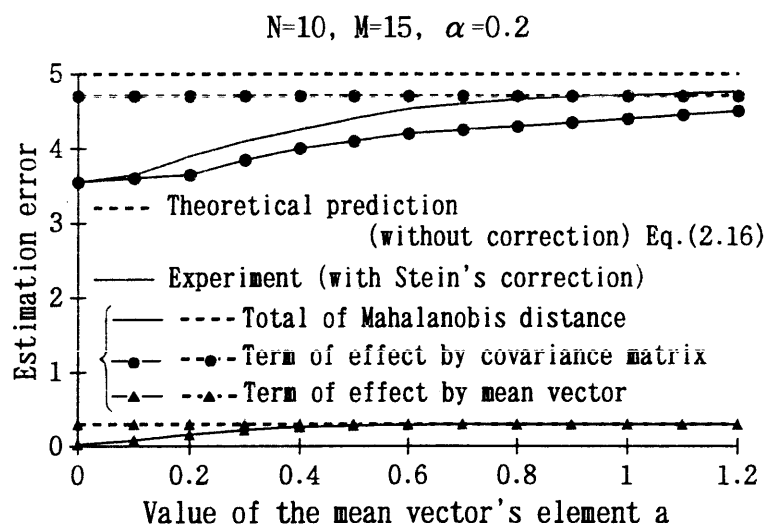


図 2. 7 Stein 推定量を用いたマハラノビス距離の推定実験(2)

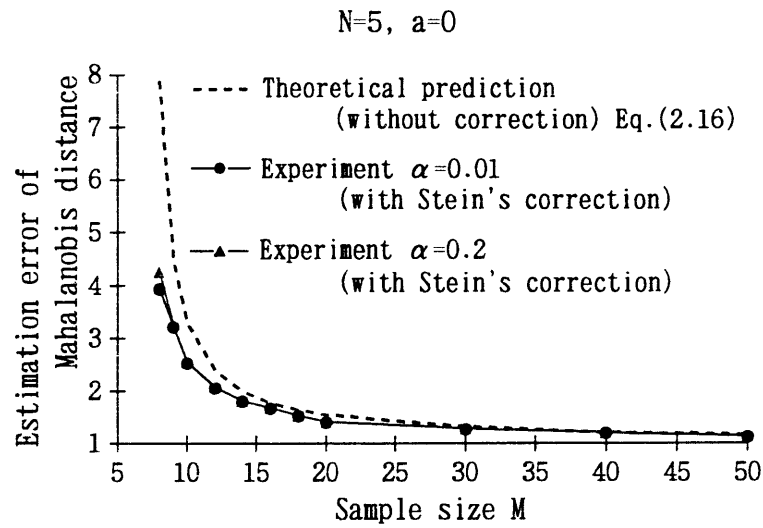


図2.8 Stein 推定量を用いたマハラノビス距離の推定実験(3)

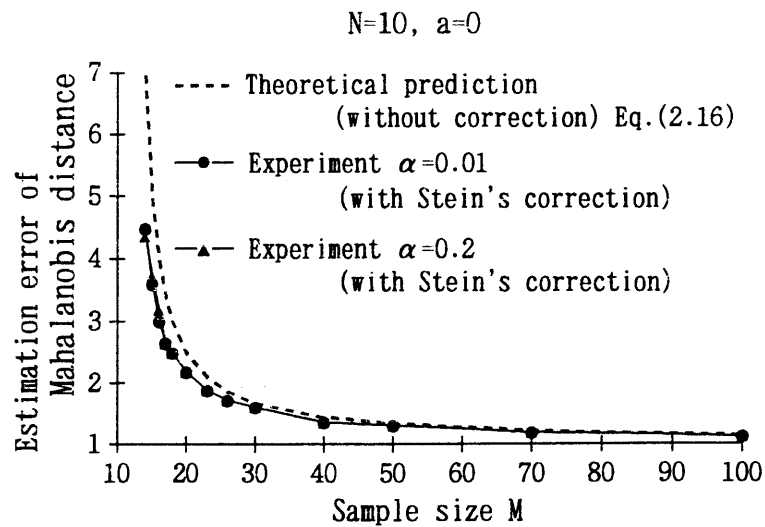


図2.9 Stein 推定量を用いたマハラノビス距離の推定実験(4)

第2章 マハラノビス距離の推定およびStein補正

この結果から次の事柄が知られる.

- (1) 平均値ベクトルの補正を行うことにより, マハラノビス距離の推定誤差は平均値ベクトルに起因する式(2.14)の第2項の推定精度が向上する(図2.6, 2.7中の”補正あり▲印”)のみでなく, 共分散行列の推定精度も併せて向上することから式(2.14)の第1項の成分(図2.6, 2.7中の”補正あり●印”)も誤差の減少を実現でき, 全体としてもマハラノビス距離の推定精度の向上(図2.6, 2.7中の”補正あり実線”)が達成されている.
- (2) 平均値ベクトルの大きさ a が小さいほどマハラノビス距離の推定精度の向上に効果がある.
- (3) 等分散を仮定した補正を行なっているにもかかわらず, 等分散とはいえない図2.7の $\alpha=0.2$ の場合においても, 平均値ベクトルの補正の効果とともに, マハラノビス距離の推定精度の向上に, Steinの補正は効果がある.

図2.8は次元数 $N=5$, 標本数 $M=8\sim 50$, 平均値ベクトルの各要素の大きさ $a=0$, パターンの各成分間の大きさの相違の度合を示すパラメータ $\alpha=0.01$ と 0.2 の場合のマハラノビス距離の推定誤差である. 破線で示すものが補正を施さない場合であり, 実線がSteinの補正をした場合である. 図2.9は次元数 $N=10$, 標本数 $M=15\sim 100$ の場合の同じ実験の結果である.

これらから, 等分散に近い $\alpha=0.01$ の場合も, 等分散とは言えない $\alpha=0.2$ の場合も同じ程度に, マハラノビス距離の推定精度は向上していることがわかる. 標本数 M の増大と共にマハラノビス距離の推定誤差は補正を加えなくても1に漸近するため, 補正による推定精度の向上の効果は, さほど顕著ではなくなる.

2.5 むすび

本章では平均値ベクトルの推定におけるStein推定量の推定精度を実験的に評価した. また, その結果がマハラノビス距離に及ぼす効果についても検討した. すなわち, パターンの次元数が3以上の場合につき, 平均値ベクトルの推定精度を向上させるためSteinの補正を行い, その効果を確認するため計算機によるシミュレーション実験を行なった. その結果以下のことが判明した.

- (1) Steinの補正により平均値ベクトルの推定精度が向上する. その効果は理論的

結果とよく一致している。

(2) Steinの補正は、パターンの各成分の分散が一定の場合、平均値ベクトルの絶対値が小さいときほど、また、標本数が小さいときほど、補正の効果が大きい。

(3) マハラノビス距離の推定誤差を示す Youngの理論式は共分散行列（固有ベクトル、固有値と等価）に起因する成分と平均値ベクトルに起因する成分とに分解することができることを導き、各成分の大きさの定量的評価を与えた。

(4) Steinの補正はマハラノビス距離の推定精度の向上にも効果がある。平均値ベクトルの推定誤差に起因する成分のみならず、共分散行列の推定精度の向上にも効果がある。

(5) パターンの各成分の分散が等しくない場合に形式的に Steinの補正を適用しても、補正の効果がみられる。

今後、検討すべき事項として、パターンの等分散を仮定しない場合の補正の効果を理論的検討も含めてさらに確認すること、極端に標本数が小さいために標本共分散行列が正則でなくなる場合の取扱いを可能にする方法を明らかにすること、マハラノビス距離の推定誤差の理論的評価、などが残されている。最後の問題については次の第3章で検討する。

文 献

- (1) James, W. and Stein, C. : "Estimation with Quadratic Loss", Proc. 4th Berkeley Symp. on Math. Statist. and Prob., pp. 361-379 (1961)
- (2) Duda, R. O. and Hart, P. E. : "Pattern Classification and Scene Analysis"
John Wiley & Sons, Inc., New-York, 1973 (Chapter 2)
- (3) Fukunaga, K. and Krile, T. F. : "Calculation of Bayes' Recognition Error for Two Multivariate Gaussian Distributions", IEEE Trans. Computers, Vol. C-18, No. 3, pp. 220-229 (March 1969)
- (4) Trunk, G. V. : "A problem of dimensionality : A simple example", IEEE Trans. Pattern Anal. and Machine Intell., Vol. PAMI-1, No. 3, pp. 306-307 (July 1979)
- (5) Cover, T. M. and Campenhout, J. M. V. : "On the Possible Orderings in the Measurement Selection Problem", IEEE Trans. Systems, Man, and Cybernetics,

第2章 マハラノビス距離の推定およびStein補正

Vol. SMC-7, No. 9, pp. 657-661(Sep. 1977)

(6)Foley, D. H. : "Considerations of Sample and Feature Size", IEEE Trans. Inform. Theory, Vol. IT-18 No. 5, pp. 618-626, (Sept. 1972)

(7)Lachenbruch, P. A. and Mickey, M. R. : "Estimation of error rates in discriminant analysis", Technometrics, Vol. 10, No. 1, pp. 1-11(1968)

(8)Fukunaga, K. : Introduction to Statistical Pattern Recognition, 2nd ed., Academic Press, New-York, 1990 (Chapter 8)

(9) 竹下, 木村, 三宅 : " マハラノビス距離の推定誤差に関する考察" , pp. 567-573, 信学論J70-D, No. 3 (1987)

(10)Stein, C. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution" Proc. 3rd Berkeley Symp. on Math. Statist. and Prob., pp. 197-206(1956)

(11)Anderson, T. W. : An Introduction to Multivariate Statistical Analysis 2nd. ed., John Wiley & Sons, Inc., New-York, 1971 (Sec. 3.5)

(12)Efron, B. and Morris, C. : "Stein's Paradox in statistics", Scientific American, Vol. 236 No. 5, pp. 119-127(1977)

(13)Young, I. T. : " Further Consideration of Sample and Feature Size" , IEEE Trans. Inf. Theory, IT-24, pp. 773-775(Nov. 1978)

付録2.1 式(2.11)の導出

$$\begin{aligned}
 E\{\widehat{d}_i^2(\mathbf{x})\} &= E[\{\widehat{\phi}_i^t(\mathbf{x} - \widehat{\mu})\}^2]/\widehat{\lambda}_i \\
 &= \widehat{\phi}_i^t E\{(\mathbf{x} - \widehat{\mu})(\mathbf{x} - \widehat{\mu})^t\} \widehat{\phi}_i / \widehat{\lambda}_i \\
 &= \widehat{\phi}_i^t [E\{(\mathbf{x} - \mu)(\mathbf{x} - \mu)^t\} + (\mu - \widehat{\mu})(\mu - \widehat{\mu})^t] \widehat{\phi}_i / \widehat{\lambda}_i \\
 &= \widehat{\phi}_i^t \{\Sigma + (\mu - \widehat{\mu})(\mu - \widehat{\mu})^t\} \widehat{\phi}_i / \widehat{\lambda}_i \\
 &= [\sum_{j=1}^N \lambda_j (\widehat{\phi}_i^t \phi_j)^2 + \{\widehat{\phi}_i^t (\mu - \widehat{\mu})\}^2] / \widehat{\lambda}_i
 \end{aligned}$$

付録2.2 正規乱数の発生方法

N 次元正規分布 $N(\mu, \Sigma)$ に従う N 次元正規乱数 \mathbf{x} は N 次元正規乱数 $\mathbf{u} \sim \{N(0, 1)\}$ を次式で変換することによって得られる.

$$\mathbf{x} = \Phi^t \Lambda^{1/2} \mathbf{u} + \mu$$

ここで Φ は Σ の固有ベクトルを列ベクトルとする行列, すなわち

$$\Phi = [\phi_1 \ \phi_2 \ \phi_3 \ \cdots \ \phi_N]$$

Λ は共分散行列 Σ の固有値を対角要素とする対角行列である. すなわち,

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_N)$$

第3章 マハラノビス距離の推定量の確率分布

多次元正規分布に従う母集団から抽出された有限個の標本を用いて、母集団の共分散行列と平均値ベクトルを推定することはしばしば行われる。本章ではその推定パラメータを用いて、同じ母集団から抽出されたパターンに関するマハラノビス距離を計算する際の推定誤差において、共分散行列と平均値ベクトルの影響を分離して評価する。

3.1 まえがき

二つのクラスのパターンの集合（母集団）が多次元正規分布に従うとき、母集団のパラメータが正確に推定できれば、一般に誤り確率を最小にする分類法であるベイズ識別は2次識別関数によって達成できる⁽¹⁾。しかしながら、実際にはパターン集合の確率分布のパラメータが未知であり、それを標本（学習標本）から推定して識別関数を設計し、その能力も標本パターン（テスト標本）で統計的に評価する。このとき、2次識別関数の設計に使用できる学習標本数や、識別器の性能評価に利用することのできるテスト標本の数には制約があり、実際の応用においては、ベイズ識別は達成できないことが多い。学習標本に基づいて設計した識別関数の能力は複雑な振舞いを示し、解明されていない事柄が多い。たとえば、特徴ベクトルの次元数に比較して学習標本数が十分でないときには、母集団のパラメータが正確に推定できないため理想的につくれば線形識別関数よりも識別能力が高い筈の2次識別関数の方が識別能力が劣ることがある⁽²⁾ことが指摘されている。また、Jain & Chandrasekaran はこの分野における幅広い調査を行っている⁽³⁾。

このことは、具体的なパターン識別においても極めて重要である。例えば、Kimuraらは2次識別関数の設計に際し、主として共分散行列の推定誤差の影響を減らすため修正2次識別関数を提案し、これを手書き文字認識に応用し成果を上げている⁽⁴⁾。ま

第3章 マハラノビス距離の推定量の確率分布

た、中川らは音韻認識に前記修正2次識別関数を適用し、学習標本より推定した母集団の推定パラメータをそのまま2次識別関数に適用した場合より良い認識率を得ている⁽⁵⁾。

ところで、2次識別関数の値を実質的に支配する部分は、入力パターンベクトルの各成分のつくる一種の2次形式であり、マハラノビス距離と呼ばれる(2章3節参照)。本章では、2次識別関数の性質を正確に知るため、このマハラノビス距離に着目し、マハラノビス距離の性質を明らかにする。

マハラノビス距離の計算をする際、そのパラメータとして共分散行列と平均値ベクトルが必要である。しかしながら学習標本数が有限であることに起因して、母集団のそれらのパラメータは正しく推定できない。このことは、特徴ベクトルの次元数に比較して学習標本数が少ないときより顕著にあらわれる。そして、母集団パラメータの推定量が確率的に変動するため、マハラノビス距離の推定量も学習標本の値、特徴パターンの次元数や学習標本数に依存する確率分布に従う。

Youngはマハラノビス距離の推定量に関し、特徴ベクトルの次元数、および学習標本数という2種類のパラメータ間の関係に関する理論的評価式を導いている⁽⁶⁾。しかしながら、推定誤差を引き起こす要因としての共分散行列と平均値ベクトルの効果を区別することなく、両者を合わせたもののマハラノビス距離への影響のみを評価したにとどまる。

以上のような観察に基づいて、第2章ではマハラノビス距離の推定誤差の振る舞いについて、シミュレーションに基づいて実験的に検討した。そこで、本章ではマハラノビス距離の性質を理論的に考察する。基本的には、マハラノビス距離の推定誤差には、①共分散行列の固有値、固有ベクトルの推定誤差による成分と、②平均値ベクトルの推定誤差による成分とが含まれることを示し、これら2つの成分の確率分布を明らかにすると共に各成分毎の期待値、分散を導く。

これらの結果は標本共分散行列の固有値、固有ベクトルの推定誤差に関する理論的解析の有力な手段として利用できる。さらにパターンの次元数に比較して標本数が少ないときの距離関数の改良に応用でき、パターン認識の認識性能の向上をもたらし得ると期待される。

以下3.2でマハラノビス距離の推定量について要約した後、3.3でマハラノビス距離の推定誤差に影響する成分を共分散行列による部分と、平均値ベクトルによる部

第3章 マハラノビス距離の推定量の確率分布

分とに分解し、その各々の確率分布を求め、平均値と分散を与える。

3.2 マハラノビス距離とその推定量

3.2.1 マハラノビス距離の定義

N 次元パターン \mathbf{x} の母集団の平均値ベクトルを $\boldsymbol{\mu}$ 、共分散行列を $\boldsymbol{\Sigma}$ とするとき、 $\boldsymbol{\mu}$ と \mathbf{x} の間の（二乗）マハラノビス距離 $dm^2(\mathbf{x})$ は次式で与えられる（式(2.7)と同じ）。

$$dm^2(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad (3.1)$$

今、 \mathbf{x} が N 次元正規分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従うときには、式(3.1)は自由度 N のカイ二乗分布に従い、期待値は N に、分散は $2N$ になることが知られている⁽⁷⁾。しかるに、一般のパターン認識の問題では、母集団の平均値ベクトル $\boldsymbol{\mu}$ や、共分散行列 $\boldsymbol{\Sigma}$ を正しく求められないことが多い。その場合には式(3.1)中の平均値ベクトル $\boldsymbol{\mu}$ や、共分散行列 $\boldsymbol{\Sigma}$ のかわりに、標本から推定した標本平均値ベクトル $\hat{\boldsymbol{\mu}}$ や、標本共分散行列 $\hat{\boldsymbol{\Sigma}}$ を使わねばならない。有限個の標本を用いて、標本から推定した標本平均値ベクトル $\hat{\boldsymbol{\mu}}$ や、標本共分散行列 $\hat{\boldsymbol{\Sigma}}$ を用いた場合のマハラノビス距離は式(3.1)中の $\boldsymbol{\mu}$ 、 $\boldsymbol{\Sigma}$ のかわりに $\hat{\boldsymbol{\mu}}$ 、 $\hat{\boldsymbol{\Sigma}}$ を用い、式(3.2)のように表現できる。

$$\hat{dm}^2(\mathbf{x}) = (\mathbf{x} - \hat{\boldsymbol{\mu}})^t \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}) \quad (3.2)$$

式(3.2)においては、標本平均値ベクトル $\hat{\boldsymbol{\mu}}$ や、標本共分散行列 $\hat{\boldsymbol{\Sigma}}$ もある確率分布に従って変動し、 \mathbf{x} が $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従うときでも式(3.2)は、もはやカイ二乗分布には従わない。

3.2.2 マハラノビス距離の推定量

$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従う N 次元正規分布母集団から、 M 個の標本 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$ を取り出す。この $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$ を用い、標本平均値ベクトル $\hat{\boldsymbol{\mu}}$ および、標本共分散行列 $\hat{\boldsymbol{\Sigma}}$ を求める。平均値ベクトルや共分散行列を推定するのに用いた標本とは独立に、同じ母集団から標本 \mathbf{x} を抽出し、式(3.2)によりマハラノビス距離を求める、すなわち、有限個の学習標本を用いて、母集団のパラメータを推定し、しかる後、その求めた推定パラメータを用いて、母集団の各パターンについてマハラノビス距離の計算を行なう場合を考える。このとき、2度目に選ぶ標本の数を増やしていけば、結局母集団のパラメータとしてのマハラノビス距離に関する推定の精度を評価していることになる。

YoungはHotellingの T^2 統計量を用いて理論的解析を行い、式(3.2)の \mathbf{x} に関する期待値、分散を次のように求めた⁽⁶⁾。

$$E \{ \widehat{dm^2}(\mathbf{x}) \} = N \cdot \left[\frac{M+1}{M} \right] \cdot \left[\frac{M-1}{M-N-2} \right] \quad (3.3)$$

$$\text{Var} \{ \widehat{dm^2}(\mathbf{x}) \} = 2 \cdot N \cdot \left[\frac{M-2}{M-N-4} \right] \left[\frac{M+1}{M} \right]^2 \left[\frac{M-1}{M-N-2} \right]^2 \quad (3.4)$$

また、筆者らは、先に次の式(3.5)

$$E \{ \widehat{dm^2}(\mathbf{x}) \} = \sum_{i=1}^N \left[\left\{ \sum_{j=1}^N \lambda_j (\widehat{\phi}_i^t \phi_j)^2 \right\} / \widehat{\lambda}_i \right] + \sum_{i=1}^N \left[\{ \widehat{\phi}_i^t (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \}^2 / \widehat{\lambda}_i \right] \quad (3.5)$$

を導き(2.3.2の式(2.11)参照)、さらに、 $E \{ \widehat{dm^2}(\mathbf{x}) \}$ を次の式(3.6)のように二つの項に分解して考えると上式(3.5)の各項とよく対応することを実験的に示した⁽⁸⁾(2章4節)。

$$E \{ \widehat{dm^2}(\mathbf{x}) \} = N \cdot \left[\frac{M-1}{M-N-2} \right] + N \cdot \left[\frac{1}{M} \right] \cdot \left[\frac{M-1}{M-N-2} \right] \quad (3.6)$$

しかしながら、この式を理論的に直接に導くことはなされていない。

3.3 マハラノビス距離の成分毎の推定量

Youngはマハラノビス距離の推定量に関する期待値・分散に関する式(3.3)(3.4)を導出したが、以下本節ではYoungによっても言及されていない式(3.5)に示すようなマハラノビス距離の推定量を成分毎に分解し、それぞれの期待値・分散を求め、次の関係式(3.7)(3.8)を証明する。その結果はYoungの導出した理論式(3.3)(3.4)と等しくなる。

$$\begin{aligned} E \{ \widehat{dm^2}(\mathbf{x}) \} &= E \{ (\mathbf{x} - \boldsymbol{\mu})^t \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \} + E \{ (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})^t \widehat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \} \\ &= N \cdot \left[\frac{M+1}{M} \right] \cdot \left[\frac{M-1}{M-N-2} \right] \end{aligned} \quad (3.7)$$

$$\begin{aligned} \text{Var} \{ \widehat{dm^2}(\mathbf{x}) \} &= \text{Var} \{ (\mathbf{x} - \boldsymbol{\mu})^t \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \} \\ &\quad + \text{Var} \{ (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})^t \widehat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \} + \text{Var} \{ 2 \cdot (\mathbf{x} - \widehat{\boldsymbol{\mu}})^t \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \} \\ &= 2 \cdot N \cdot \left[\frac{M-2}{M-N-4} \right] \left[\frac{M+1}{M} \right]^2 \left[\frac{M-1}{M-N-2} \right]^2 \end{aligned} \quad (3.8)$$

一方、式(3.2)を変形すると、次のようになる。

$$\begin{aligned} \widehat{dm^2}(\mathbf{x}) &= (\mathbf{x} - \boldsymbol{\mu} + \boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})^t \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \boldsymbol{\mu} + \boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \\ &= (\mathbf{x} - \boldsymbol{\mu})^t \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \\ &\quad + (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})^t \widehat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) + 2 \cdot (\mathbf{x} - \boldsymbol{\mu})^t \widehat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \end{aligned} \quad (3.9)$$

第3章 マハラノビス距離の推定量の確率分布

ここで、式(3.9)の第2項はHotellingの T^2 統計量を用いると、次のように書き直すことができる⁽⁹⁾。なお以後の議論では、式(3.9)の各項に対応して T^2 統計量が出現するため、混同を避ける意味で第1項に対応するものを T_1^2 、第2項のそれを T_2^2 で示す。すなわち、第2項に対応するHotellingの T^2 は

$$T_2^2 = M(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})^t \widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \quad (3.10)$$

ここで、 $[T_2^2/(M-1)] \cdot [(M-N)/N]$ は自由度 $(N, M-N)$ の F 分布に従う⁽⁶⁾。

一方、確率変数 \mathbf{x} が自由度 (n_1, n_2) の F 分布に従うとき、 \mathbf{x} の期待値と分散は、それぞれ次のようになる。

$$E\{\mathbf{x}\} = \frac{n_2}{n_2 - 2} \quad (\text{ただし, } n_2 > 2) \quad (3.11)$$

$$\text{Var}\{\mathbf{x}\} = 2 \cdot n_2^2 \frac{n_1 + n_2 - 2}{n_1(n_2 - 2)^2(n_2 - 4)} \quad (3.12)$$

(ただし, $n_2 > 4$)

式(3.10)～(3.12)より、式(3.9)の右辺の第2項の期待値はつぎのようになる。

$$\begin{aligned} E\{(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})^t \widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})\} \\ &= E\{T_2^2/M\} \\ &= \left\{ \frac{N \cdot (M-N)}{M-N-2} \right\} \left\{ \frac{M-1}{M-N} \right\} \left\{ \frac{1}{M} \right\} \\ &= N \cdot \frac{M-1}{(M-N-2) \cdot M} \end{aligned} \quad (3.13)$$

同様にして、分散も次のように求められる。

$$\begin{aligned} \text{Var}\{(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})^t \widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})\} \\ &= \text{Var}\{T_2^2/M\} \\ &= 2 \cdot (M-N)^2 \frac{M-2}{N \cdot (M-N-2)^2(M-N-4)} \cdot \left\{ \frac{M-1}{M-N} \right\}^2 \cdot N^2 \cdot \left\{ \frac{1}{M^2} \right\} \\ &= 2 \cdot N \cdot \left\{ \frac{M-2}{M-N-4} \right\} \cdot \left\{ \frac{M-1}{M-N-2} \right\}^2 \left\{ \frac{1}{M^2} \right\} \end{aligned} \quad (3.14)$$

次に、式(3.9)の右辺の第1項について考えてみよう。式(3.9)の右辺の第1項において、 $(\mathbf{x} - \boldsymbol{\mu})$ を新しく \mathbf{y} という、 N 次元正規分布 $N(\mathbf{0}, \boldsymbol{\Sigma})$ に従う確率変数で置き換えることが可能である。すると、確率変数 \mathbf{y} と、第2項の確率変数 $(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})$ とは、共に、平均は $\mathbf{0}$ であり、共分散行列について、第1項が第2項の M 倍である点のみが異なる。したがって、式(3.9)の右辺の第1項の分布もHotellingの T^2 分布を用いると、次のように書き直すことができる(付録3.1)。

第3章 マハラノビス距離の推定量の確率分布

$$T_1^2 = \mathbf{y}^t \widehat{\Sigma}^{-1} \mathbf{y} \quad (3.15)$$

ここで、 $[T_1^2 / (M-1)] \cdot [(M-N)/N]$ は自由度 $(N, M-N)$ の F 分布に従う。再び、自由度 (n_1, n_2) の F 分布の期待値、分散に関する関係式(3.11)(3.12)を適用すれば、式(3.9)の右辺の第1項の期待値、および分散も次のように求められる。

$$\begin{aligned} E \{ \mathbf{y}^t \widehat{\Sigma}^{-1} \mathbf{y} \} &= E \{ T_1^2 \} \\ &= \left\{ \frac{N \cdot (M-N)}{M-N-2} \right\} \left\{ \frac{M-1}{M-N} \right\} \\ &= N \cdot \left\{ \frac{M-1}{M-N-2} \right\} \end{aligned} \quad (3.16)$$

$$\begin{aligned} \text{Var} \{ \mathbf{y}^t \widehat{\Sigma}^{-1} \mathbf{y} \} &= \text{Var} \{ T_1^2 \} \\ &= 2 \cdot (M-N)^2 \frac{M-2}{N \cdot (M-N-2)^2 (M-N-4)} \cdot \left\{ \frac{M-1}{M-N} \right\}^2 \cdot N^2 \\ &= 2 \cdot N \cdot \left\{ \frac{M-2}{M-N-4} \right\} \cdot \left\{ \frac{M-1}{M-N-2} \right\}^2 \end{aligned} \quad (3.17)$$

最後に、式(3.9)の右辺の第3項は $(\mathbf{x} - \boldsymbol{\mu})$ と $(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})$ とが独立かつ無相関であり ($\boldsymbol{\mu}$ は定数ベクトルであり、 \mathbf{x} と $\widehat{\boldsymbol{\mu}}$ の計算に用いる標本は互いに独立に抽出される)、共に平均が $\mathbf{0}$ であるので、第3項の期待値は 0 となる。したがって、推定マハラノビス距離の期待値は式(3.9)の第1項と第2項の期待値の和となる(付録3.2)。実際、式(3.13)と(3.16)を加えた値は、証明すべき式(3.7)となり、同時にYoungの求めた理論式(3.3)に等しくなることが容易に確かめられる。

推定マハラノビス距離の分散は式(3.9)の第1～3項の分散の和となる(付録3.2)。ただし、第3項は、分布が $N \left(\mathbf{0}, \frac{\Sigma}{\sqrt{M}} \right)$ の正規分布の場合に対応するHotellingの T^2 と等価となり、分散は次の式(3.18)で与えられる。

$$\begin{aligned} \text{Var} \{ \text{式(3.9)の右辺の第3項} \} &= \text{Var} \{ T_1^2 / \sqrt{M} \} \\ &= 4 \cdot (M-N)^2 \frac{M-2}{N \cdot (M-N-2)^2 (M-N-4)} \cdot \left\{ \frac{M-1}{M-N} \right\}^2 \cdot N^2 \cdot \left\{ \frac{1}{M} \right\} \\ &= 4 \cdot N \cdot \left\{ \frac{M-2}{M-N-4} \right\} \cdot \left\{ \frac{M-1}{M-N-2} \right\}^2 \left\{ \frac{1}{M} \right\} \end{aligned} \quad (3.18)$$

なお、マハラノビス距離全体の分散についても、実際に式(3.14)(3.17)(3.18)を加

第3章 マハラノビス距離の推定量の確率分布

えると証明すべき式(3.8)に等しくなり(付録3.2参照), 同時にYoungの理論式(3.4)に等しくなることが容易に確かめられる.

3.4 考察

式(3.9)中の共分散行列 $\widehat{\Sigma}$ を固有値・固有ベクトル表現を用いて書き表すと, その期待値は次のようになる⁽⁸⁾.

$$E\{\widehat{dm}^2(\mathbf{x})\} = \sum_{i=1}^N \left[\left\{ \sum_{j=1}^N \lambda_j (\widehat{\phi}_i^t \phi_j)^2 \right\} / \widehat{\lambda}_i \right] + \sum_{i=1}^N \left[\{ \widehat{\phi}_i^t (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \}^2 / \widehat{\lambda}_i \right] \quad (3.19)$$

従って, 式(3.19)の右辺の第1項, 第2項の性質は, それぞれ式(3.15), (3.10)の分布を用いて表すことができ, それぞれの期待値, 分散は式(3.16)(3.13)と式(3.17)(3.14)で表すことができる. したがって, 推定マハラノビス距離の期待値に関して, 標本共分散行列の推定誤差に起因する成分と標本共分散行列で重み付けをした標本平均値ベクトルの推定誤差に起因する成分があることが示された.

さらに, その二つの成分について定量的な議論も可能になる. すなわち, 標本共分散行列の推定誤差に起因する成分は, 後者(標本共分散行列で重み付けをした標本平均値ベクトル)に起因する成分の学習標本数倍(=M)だけ, 大きく寄与することが知られる. また, 推定マハラノビス距離の分散については前記二つの成分の他に二つの成分の相互作用の項(式(3.9)右辺の第3項)があることが判明した. なお上記の結果の期待値に関する部分は本文第2章において実験的に確認されている事柄である⁽⁸⁾.

このことは, 推定マハラノビス距離の期待値に関して, たとえばMが10を越えれば第1項に比べて第2項の影響は殆ど無視できることを意味する. 従って, 2次識別関数のようなマハラノビス距離を含む識別関数の設計の際には平均値ベクトルの影響よりも, 標本共分散行列の影響により注意を払うべきであることがわかる. このことは, 文献(4)(11)で指摘された修正2次識別関数を用いた場合に, 標本からの推定量を用いたベイズ識別としての2次識別関数に比べて識別率の向上が見られることを, 文献(11)とは別の角度から例証しているともいえる.

さらに, マハラノビス距離の推定誤差を標本共分散行列の主軸成分(各固有ベクトルへの射影成分)毎に実験的に調べた文献(12)の結果において, 高次の主軸成分(標本共分散行列の小さな固有値に対応する成分)ほどマハラノビス距離の推定誤差が大きくなることが示されていたが, これは主として標本共分散行列の推定誤差に起因し

ていたものとして説明できる。

3.5 むすび

パターン認識において特徴ベクトルで表わされるパターンの集合が多次元正規分布をなす場合には誤り確率最小の決定（ベイズ決定）は2次識別関数になる。この識別関数の値を支配するのは、その中のパターンベクトルに関する2次形式でマハラノビス距離とよばれる量である。本章では、このマハラノビス距離に関して、パターン母集団のパラメータを有限個の学習標本パターンから推定する場合の誤差について、理論的解析を行なった。すなわちマハラノビス距離の推定誤差には、①固有値、固有ベクトルの推定誤差による成分と、②平均値ベクトルの推定誤差による成分とがあることを示し、これら2つの推定誤差の成分の確率分布を明らかにし、それぞれの成分の平均値、分散に関する理論式を導出した。

その結果は、上記②の成分の平均値が①の成分の平均値と学習標本数の逆数との積となること、特徴ベクトルの次元数と学習標本数が接近すると両者の差の逆数のオーダーで推定誤差が増大し、その誤差の主たる原因は標本共分散行列の推定誤差に起因すること、を示している。従って、特徴ベクトルの次元数に比較して十分な学習標本数の確保できないような状況のもとでは、標本からの推定値を用いた2次識別関数よりも、二つのクラスの共分散行列を等しいと仮定し、併合した共分散行列（pooled covariance matrix）を用いた線形識別関数を採用した方が有利となる状況も存在することが示唆される。

今後検討すべき事項としては、標本共分散行列の行列式に関する推定誤差の評価、マハラノビス距離の主軸成分毎の推定誤差の解析的評価、さらにその結果を修正2次識別関数に使用する固有ベクトルの数や置換固有値の決定（現在実験的に決めている）に利用すること、等がある。

文 献

- (1) K. Fukunaga: "Introduction to Statistical Pattern Recognition" 2nd ed., Academic Press, New-York, 1990 (Chapter 3)
- (2) Ness, J. W. V. and Simpson, C.: "On the Effects of Dimension in Discrimi-

第3章 マハラノビス距離の推定量の確率分布

- nant Analysis", TECHNOMETRICS, Vol. 18, No. 2, pp. 175-187(May 1976)
- (3) Jain, A. K. and Chandrasekaran, B.:"Dimensionality and sample size consideration in pattern recognition practice", in Handbook of Statistics, vol. 2, Krishnaiah, P. R. and Kanal, L. N. (Eds), North Holland, pp. 835-855. (1982)
- (4) Kimura, F., Takashina, K., Tsuruoka, S., Miyake, Y.:"Modified Quadratic Discriminant Functions and the Application to Chinese Character Recognition", IEEE Trans. Pattern Anal. Mach. Intell. Vol. PAMI-9 No. 1 pp. 149-153 (Jan. 1987)
- (5) 中川聖一, 榎本康宏:"時系列パターンの統計的認識法を用いた不特定話者の音韻および単語音声の認識", 信学論(D) Vol. J71-D, No. 10, pp. 1977-1983 (1988)
- (6) Young, I. T.:"Further Consideration of Sample and Feature Size", IEEE Trans. Inf. Theory, IT-24, pp. 773-775(Nov. 1978)
- (7) 柳井晴夫, 高木広文共著: 多変量解析ハンドブック p. 272, 現代数学社(1986)
- (8) 竹下鉄夫, 野澤繁之, 掛布英辰, 鳥脇純一郎:"マハラノビス距離の推定およびそれに対するStein補正の効果について", 信学論(D- II) J74-D- II, pp. 1482-1485 (1991)
- (9) Wijsman, R. A.:"Random orthogonal transformation and their use in some classical distribution in multivariate analysis", Ann. Math. Statist., pp. 415-423(1957)
- (10) 広中平祐編: 現代数理科学事典, p. 554, 大阪書籍(1991)
- (11) Kimura, F. and Shridhar, M.:"Handwritten Numerical Recognition Based on Multiple Algorithms", Pattern Recognition, Vol. 24, No. 10, pp. 969-983(1991)
- (12) 竹下鉄夫, 木村文隆, 三宅康二:"マハラノビス距離の推定誤差に関する考察", 信学論(D) Vol. J70-D, No. 3, pp. 567-573(1987)
- (13) 広中平祐編: 現代数理科学事典, p. 564, 大阪書籍(1991)
- (14) 竹村彰通: 多変量推測統計の基礎, p. 32, 共立出版(1991)

付録3.1

$$T_1^2 = \mathbf{y}^t \hat{\Sigma}^{-1} \mathbf{y} \quad (3.20)$$

において、 \mathbf{y} (N 次元ベクトル) は正規分布 $N(\mathbf{0}, \Sigma)$ に従い、 $\hat{\Sigma}$ は次式で与えられる。

$$\hat{\Sigma} = \frac{1}{(M-1)} \sum_{i=1}^M (\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^t \quad (3.21)$$

但し、 \mathbf{x}_i (N 次元ベクトル) は互いに独立で $N(\mu, \Sigma)$ に従う。このとき、

$\left\{ \frac{T_1^2}{M-1} \right\} \cdot \left\{ \frac{M-N}{N} \right\}$ は自由度 $(N, M-N)$ の F 分布に従う

[証明]

\mathbf{x}_i が互いに独立に正規分布 $N(\mu, \Sigma)$ に従うとき、

$$\mathbf{W} = \sum_{i=1}^M \mathbf{x}_i \mathbf{x}_i^t \quad (3.22)$$

は自由度 M のウィシャート分布に従う。これを $W(M, \Sigma)$ で表わす。一方、標本積和行列 \mathbf{S}

$$\mathbf{S} = \sum_{i=1}^M (\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^t \quad (3.23)$$

は自由度 $(M-1)$ のウィシャート分布に従う⁽¹³⁾。

一般に、 \mathbf{W} が $W(n, \Sigma)$ に従い、 \mathbf{y} が $N(\mathbf{0}, \Sigma)$ に従うとき、

$$T^2 = n \mathbf{y}^t \mathbf{W}^{-1} \mathbf{y} \quad (3.24)$$

と、おくと

$$\left\{ \frac{n-p+1}{p} \right\} \left\{ \frac{T^2}{n} \right\} \quad (3.25)$$

は自由度 $(p, n-p+1)$ の F 分布に従う。ただし、

p は \mathbf{y} の次元数である⁽³⁾。

今、自由度 $n \rightarrow (M-1)$ 、次元数 $p \rightarrow N$ 、 $\mathbf{W} \rightarrow \mathbf{S}$ で置き換えれば、式(3.24)(3.25)

は次のようになる。

$$T^2 = (M-1) \mathbf{y}^t \mathbf{S}^{-1} \mathbf{y} \quad (3.26)$$

$$\begin{aligned} \left\{ \frac{n-p+1}{p} \right\} \cdot \left\{ \frac{T^2}{n} \right\} &= \left\{ \frac{M-1-N+1}{N} \right\} \cdot \left\{ \frac{T^2}{M-1} \right\} \\ &= \left\{ \frac{M-N}{N} \right\} \cdot \left\{ \frac{T^2}{M-1} \right\} \end{aligned} \quad (3.27)$$

第3章 マハラノビス距離の推定量の確率分布

式(3.26)に

$$\widehat{\Sigma} = \frac{1}{(M-1)} S \quad (3.28)$$

を代入し、左辺を T_1^2 で表わすと、式(3.26)は次のようになる。

$$T_1^2 = \mathbf{y}^t \widehat{\Sigma}^{-1} \mathbf{y} \quad (3.29)$$

従って、

$\left\{ \frac{T_1^2}{M-1} \right\} \left\{ \frac{M-N}{N} \right\}$ は自由度 $(N, M-N)$ の F 分布に従う。
(証明終わり)

付録3.2

$$\widehat{dm}^2(\mathbf{x}) = (\mathbf{x} - \widehat{\boldsymbol{\mu}})^t \widehat{\Sigma}^{-1} (\mathbf{x} - \widehat{\boldsymbol{\mu}}) \quad (3.30)$$

のとき、

$$E \{ \widehat{dm}^2(\mathbf{x}) \} = E \{ (\mathbf{x} - \boldsymbol{\mu})^t \widehat{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \} + E \{ (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})^t \widehat{\Sigma}^{-1} (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \} \quad (3.31)$$

$$\begin{aligned} \text{Var} \{ \widehat{dm}^2(\mathbf{x}) \} &= \text{Var} \{ (\mathbf{x} - \boldsymbol{\mu})^t \widehat{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \} \\ &\quad + \text{Var} \{ (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})^t \widehat{\Sigma}^{-1} (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \} + \text{Var} \{ 2 \cdot (\mathbf{x} - \boldsymbol{\mu})^t \widehat{\Sigma}^{-1} (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \} \end{aligned} \quad (3.32)$$

[証明]

$$\mathbf{A} = \widehat{\Sigma}^{-1/2} (\mathbf{x} - \boldsymbol{\mu}), \quad \mathbf{B} = \widehat{\Sigma}^{-1/2} (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \quad (3.33)$$

とおくと、式(3.30)は次のようになる。

$$\begin{aligned} \widehat{dm}^2(\mathbf{x}) &= \widehat{dm}^2(\mathbf{A}, \mathbf{B}) \\ &= (\mathbf{A} + \mathbf{B})^t (\mathbf{A} + \mathbf{B}) \end{aligned} \quad (3.34)$$

式(3.31)の計算をする準備として、

$$E \{ \mathbf{A}^t \mathbf{B} \} = E \{ (\mathbf{x} - \boldsymbol{\mu})^t \widehat{\Sigma}^{-1} (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) \} \quad (3.35)$$

を求める。

今、同じ標本ベクトル集合の標本共分散行列 $\widehat{\Sigma}$ と標本平均値ベクトル $\widehat{\boldsymbol{\mu}}$ は互いに独立である⁽¹⁴⁾。さらに、 $\boldsymbol{\mu}$ は母集団の平均値ベクトル(定数)である。 \mathbf{x} は $N(\boldsymbol{\mu}, \Sigma)$ に従う標本であり、 $\widehat{\boldsymbol{\mu}}$ は \mathbf{x} とは独立に $N(\boldsymbol{\mu}, \Sigma)$ の母集団から抽出された M 個の標本から推定した標本平均値ベクトルであるから、 $(\mathbf{x} - \boldsymbol{\mu})$ と $(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})$ とは、互いに独立で平均は共に $\mathbf{0}$ である。よって、その2項と標本共分散行列の逆行列との積の期待値は 0 となる。すなわち、次式が成り立つ。

$$E \{ \mathbf{A}^t \mathbf{B} \} = 0 \quad (3.36)$$

よって、式(3.30)の期待値は次のようになる.

$$\begin{aligned}
 E \{ \widehat{dm}^2(\mathbf{x}) \} &= E [(\mathbf{A} + \mathbf{B})' (\mathbf{A} + \mathbf{B})] \\
 &= E [\mathbf{A}' \mathbf{A} + 2 \cdot \mathbf{A}' \mathbf{B} + \mathbf{B}' \mathbf{B}] \\
 &= E [\mathbf{A}' \mathbf{A}] + E [\mathbf{B}' \mathbf{B}] \quad (3.37)
 \end{aligned}$$

式(3.37)に式(3.33)を代入すれば式(3.31)を導出できる.

次に分散について検討する.

$$\begin{aligned}
 \text{Var} \{ \widehat{dm}^2(\mathbf{x}) \} &= E [(\widehat{dm}^2(\mathbf{x}) - E[\widehat{dm}^2(\mathbf{x})])^2] \\
 &= E [\{ (\mathbf{A} + \mathbf{B})' (\mathbf{A} + \mathbf{B}) - E\{\mathbf{A}' \mathbf{A}\} - E\{\mathbf{B}' \mathbf{B}\} \}^2] \\
 &= E [\{ \mathbf{A}' \mathbf{A} + 2 \cdot \mathbf{A}' \mathbf{B} + \mathbf{B}' \mathbf{B} - E\{\mathbf{A}' \mathbf{A}\} - E\{\mathbf{B}' \mathbf{B}\} \}^2] \\
 &= E [(\mathbf{A}' \mathbf{A})^2 - 2 \cdot \mathbf{A}' \mathbf{A} \cdot E\{\mathbf{A}' \mathbf{A}\} + \{E\{\mathbf{A}' \mathbf{A}\}\}^2] \\
 &\quad + E [(\mathbf{B}' \mathbf{B})^2 - 2 \cdot \mathbf{B}' \mathbf{B} \cdot E\{\mathbf{B}' \mathbf{B}\} + \{E\{\mathbf{B}' \mathbf{B}\}\}^2] \\
 &\quad + E [4 \cdot (\mathbf{A}' \mathbf{B})^2] \\
 &= \text{Var} [\mathbf{A}' \mathbf{A}] + \text{Var} [\mathbf{B}' \mathbf{B}] + \text{Var} [2 \cdot \mathbf{A}' \mathbf{B}] \quad (3.38)
 \end{aligned}$$

式(3.38)に式(3.33)を代入すれば式(3.32)を導出できる.

(証明終わり)

第4章 有限学習標本に基づく統計的 パターン分類における認識率の評価

本章前半では、以下のような内容を検討する。

二つのクラスのパターンが多次元正規分布に従う場合、それぞれのパターンの母集団から同数の有限個の標本を抽出し、母集団の共分散行列と平均値ベクトルを推定する。その推定パラメータを用いて、線形識別関数もしくは2次識別関数を構成し、同じ母集団から抽出された他の標本が正しく認識できるか否かをテストし、識別関数の性能を測定する。これを繰り返し、学習標本の有限性に起因する認識率の低下に関する分布を調べ、平均認識率を算出する方法を見いだす。

ついで、後半で特徴パターンの次元数、学習標本数、および平均認識率三者の間の関係をRaudysおよびFukunagaらの近似認識率（理論式）と比較し、それらの評価式の適用限界を明らかにする。さらに認識率の低下分に関する評価を行ない、ベイズ誤り率が高く、特徴ベクトルの次元数が高くなるに従い、認識率の低下分の分布は χ^2 分布に近くなることを見いだす。

4.1 まえがき

パターン集合の確率分布の未知パラメータを学習標本から推定した場合、それを用いて設計した識別関数の認識率はベイズ決定よりは必ず低下する。Raudys, Fukunagaらは互いに等しい共分散行列を共有する二つの多次元正規分布を考え、学習標本数が有限であることに起因する認識率の低下についての近似式を求めた⁽¹⁾⁽²⁾。その近似式は判別を行う識別関数毎に、特徴ベクトルの次元数 N 、学習標本数 M 、およびマハラノビス距離 δ の関数で表現されている。またEfronは線形識別関数について認識率の低下量の分布が χ^2 分布で近似できることを示した⁽³⁾。しかしながら、これらの近似式の精度や分布について、実験によって系統的に検証した報告はない。

先に、筆者らは一つ取って置き法(Leave-one-out method)を用い、学習標本数が認識率に及ぼす影響について実験的検討を行った⁽⁴⁾。しかしながら、認識実験に基づいて認識率を実験的に推定する場合には、設計した識別関数の評価をするためのテスト

第4章 統計的パターン分類における認識率の評価

標本数が少ないことによって、認識実験の結果における認識率のばらつきが極めて大きくなった。したがって、とくにベイズ認識率が高い場合には、平均認識率を精度よく推定しようとするとは非常に大量の標本を用いた実験をする必要があった。さらに、これにテスト標本の有限性に起因する推定のばらつき（推定された認識率の分散）が重なり、学習標本数の有限性にのみ起因する認識率の低下の度合や実験による認識率のばらつきは測定できないという欠点があった。これに対して本章では、実験の一部に理論式による誤り率の評価を組入れることにより、学習標本数の有限性に起因する誤り率の低下分を評価することを試みる。

その結果、ほとんどの場合、平均認識率は上記のRaudysとFukunagaの近似式の間位置すること、Fukunagaの近似式はベイズの誤り率が0.1前後でよい近似を与えること、Raudysの近似はベイズの誤り率が0.01以下であっても学習標本数が少ないところでかなりよい近似となっていることを実験的に明らかにする。

さらに、認識率の低下量の分散と平均の2乗との比がほぼ一定となることから、認識率の低下分の分布がベイズ誤り率が高く、特徴ベクトルの次元数が高くなるに従い、 χ^2 分布に近くなることを実験的に検証する。

以下、4.2で識別関数の定義およびシミュレーション実験の条件を明らかにする。次に4.3で誤り率の理論計算式について述べ、4.4でRaudys, Fukunagaらの提案する評価式の近似精度を実験的に確認した後、認識率の低下分の確率分布についても明らかにする。

4.2 識別関数と認識率との関係

多次元正規分布に従う二つのパターン母集団（クラス1，および2）を考える。以下記法は1.3と同じものを用いる。

4.2.1 共分散行列が等しい場合

多次元正規分布に従う二つのパターン母集団の共分散行列が等しい場合、ベイズ決定は次式の線形識別関数による識別になる。

$$h(\mathbf{x}) = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1} \mathbf{x} + \frac{1}{2} (\boldsymbol{\mu}_1^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2) \\ \text{ただし } \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \quad (4.1)$$

平均認識率は平均誤り率を ε とすれば $(1 - \varepsilon)$ となり、これらの値は次に示すように、二つの分布間のマハラノビス距離 δ （(4.4)式）のみで決まる。

第4章 統計的パターン分類における認識率の評価

$$\varepsilon = \Phi(-\delta/2) \quad (4.2)$$

ただし、 $\Phi(x)$ は標準正規分布関数、すなわち、

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (4.3)$$

であり、 δ^2 は二乗マハラノビス距離 ((4.4)式) である。すなわち、

$$\delta^2 = (\mu_1 - \mu_2)^t \Sigma^{-1} (\mu_1 - \mu_2) \quad (4.4)$$

4.2.2 共分散行列が異なる場合

二つの母集団の共分散行列が異なる場合、ベイズ決定は次式の2次識別関数による識別になる。

$$h(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mu_1)^t \Sigma_1^{-1}(\mathbf{x} - \mu_1) - \frac{1}{2}(\mathbf{x} - \mu_2)^t \Sigma_2^{-1}(\mathbf{x} - \mu_2) + \frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} \quad (4.5)$$

このときの認識率はFukunagaらが求めている⁽⁵⁾。しかしながら、認識率は二つの母集団の平均値ベクトル μ_1 、 μ_2 、共分散行列 Σ_1 、 Σ_2 をパラメータとする関数となるものの、数値積分でしか計算できないため、その値を具体的数値として系統的に求めた例はない。

4.2.3 母集団パラメータの推定と認識率の低下

ここで母集団の確率分布のパラメータ μ_1 、 μ_2 、 Σ_1 、 Σ_2 が未知の場合、有限個の標本(学習標本)によって母集団のパラメータを推定し、そのパラメータを用いて識別関数を設計するのが普通である。このようにして得られた識別関数の性能を実験的に評価する最も適当な方法は、もう一度同じ母集団からテストのための標本を抽出し、その認識率を測定することである。この場合、測定対象となる認識率は母集団のパラメータが既知である場合に達成されるベイズの認識率よりは必ず低い。このベイズ認識率からの低下は、学習標本数が有限であるためパラメータの推定値が真の値と一致しないことに起因するものである。ただし、テスト標本による認識実験の結果の認識率には、テスト標本数の有限性に起因するばらつきも加わる。すなわち、ベイズの誤り率を ε_B とし、学習標本の有限性に基づく誤り率の増加分を $\Delta \varepsilon_d$ 、テスト標本の有限性による変動成分を $\Delta \varepsilon_t$ とおくと、テスト標本に対する認識実験の誤り率 ε_{dt} は、

$$\varepsilon_{dt} = \varepsilon_B + \Delta \varepsilon_d + \Delta \varepsilon_t \quad (4.6)$$

となる。ランダムに抽出されたテスト標本の変動に関する期待値をとると、 $\Delta \varepsilon_t$ の期待値は0と考えられ、

$$E \{ \varepsilon_{d1} \} \quad (\equiv \quad \varepsilon_d) = \varepsilon_B + E \{ \Delta \varepsilon_d \} \quad (4.7)$$

となる.

一方, 誤り率の分散については次の式(4.8)が成立する⁽⁵⁾.

$$\text{Var} \{ \varepsilon_{d1} \} = \text{Var} \{ \Delta \varepsilon_d \} + \text{Var} \{ \Delta \varepsilon_1 \} \quad (4.8)$$

このうち, 右辺の第2項については, 二つのクラスの共分散行列が等しく, かつ二クラスのテスト標本数がそれぞれ等しければ近似的に式(4.9)が成立する⁽⁵⁾.

$$\text{Var} \{ \Delta \varepsilon_1 \} = \varepsilon_d (1 - \varepsilon_d) / n_1 \quad (4.9)$$

ただし, n_1 はテスト標本の総数である.

一方, $\text{Var} \{ \Delta \varepsilon_d \}$ についての評価の例は極めて少ない. ところで, $\Delta \varepsilon_1$ は与えられた識別関数の認識率を有限個のテスト標本の分類結果に基づいて実験的に求めるようにしたために生じる項である. 従ってこの識別関数の認識率を理論的に計算できれば, $\Delta \varepsilon_1$ は生じないため考えなくてよい. このような考え方に基づいて, 本章においては, 次節に述べるように誤り率の理論計算式を認識実験と合わせて用いることにより, 実質的に学習標本の有限性に基づく識別関数のベイズ識別からの劣化の程度のみを評価できるようにすることを意図した. すなわち, 多次元正規分布母集団について, 有限個の学習標本を用いて設計した識別関数が決める多次元空間の決定領域で多次元定積分を行って認識率(誤り率)を求めることにより, 実効的にテスト標本数の有限性に起因する成分 $\Delta \varepsilon_1$ は考えなくてよいようにした. これにより学習標本の有限性に起因する $\text{Var} \{ \Delta \varepsilon_d \}$ を実験的に推定し, 式(4.8), (4.9)の妥当性を実験的に確認した.

4.3 誤り率の理論的評価

本節では次節で用いる誤り率の理論的評価に関する事柄を簡単にまとめておく.

4.3.1 線形識別関数の場合

学習標本から推定した二つのパターン母集団の平均値ベクトル(標本平均)を $\hat{\mu}_1$, $\hat{\mu}_2$ とする. 両クラスの共分散行列は等しいと仮定しているので, 両クラスの学習標本を併合して推定した共分散行列(併合された共分散行列 pooled sample covariance matrix)を $\hat{\Sigma}$ とする. 推定したパラメータによって設計される識別関数は式(4.10)となる.

$$h(\mathbf{x}) = (\hat{\mu}_2 - \hat{\mu}_1)^t \hat{\Sigma}^{-1} \mathbf{x} + \frac{1}{2} (\hat{\mu}_1^t \hat{\Sigma}^{-1} \hat{\mu}_1 - \hat{\mu}_2^t \hat{\Sigma}^{-1} \hat{\mu}_2) \quad (4.10)$$

第4章 統計的パターン分類における認識率の評価

未知のテスト標本 \mathbf{x} については、この $\widehat{h}(\mathbf{x})$ を用いて、

$$\widehat{h}(\mathbf{x}) \begin{cases} < 0 & : \text{クラス1} \\ > 0 & : \text{クラス2} \end{cases} \quad (4.11)$$

のように識別する。いま、クラス1に属する \mathbf{x} は、正規分布 $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ に従う。

式(4.10)は \mathbf{x} に関する1次変換なので、 $\widehat{h}(\mathbf{x})$ も正規分布 $N_1(\eta_1, \sigma^2)$ に従う。ただし、

$$\begin{aligned} \eta_1 &= E\{\widehat{h}(\mathbf{x}) | \omega_1\} \\ &= (\widehat{\boldsymbol{\mu}}_2 - \widehat{\boldsymbol{\mu}}_1)^t \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} (\widehat{\boldsymbol{\mu}}_1^t \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}_1 - \widehat{\boldsymbol{\mu}}_2^t \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}_2) \end{aligned} \quad (4.12)$$

$$\begin{aligned} \sigma^2 &= \text{Var}\{\widehat{h}(\mathbf{x}) | \omega_1\} \\ &= (\widehat{\boldsymbol{\mu}}_2 - \widehat{\boldsymbol{\mu}}_1)^T \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\boldsymbol{\mu}}_2 - \widehat{\boldsymbol{\mu}}_1) \end{aligned} \quad (4.13)$$

同様にして、 \mathbf{x} がクラス2に属するときの $\widehat{h}(\mathbf{x})$ は正規分布 $N_2(\eta_2, \sigma^2)$ に従う。ただし、

$$\begin{aligned} \eta_2 &= E\{\widehat{h}(\mathbf{x}) | \omega_2\} \\ &= (\widehat{\boldsymbol{\mu}}_2 - \widehat{\boldsymbol{\mu}}_1)^t \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}_2 + \frac{1}{2} (\widehat{\boldsymbol{\mu}}_1^t \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}_1 - \widehat{\boldsymbol{\mu}}_2^t \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}_2) \end{aligned} \quad (4.14)$$

一次元正規分布 $N_1(\eta_1, \sigma^2)$, $N_2(\eta_2, \sigma^2)$ の確率密度関数を $\phi_1(t)$, $\phi_2(t)$ で表わせば、クラス1 (クラス2) のパターンをクラス2 (クラス1) と誤識別する確率 ε_1 (ε_2) は、それぞれ次式で与えられる。

$$\varepsilon_1 = \int_0^\infty \phi_1(t) dt \quad (4.15)$$

$$\varepsilon_2 = \int_{-\infty}^0 \phi_2(t) dt \quad (4.16)$$

式(4.15), (4.16)を式(1.4)に代入すると、有限個の学習標本から設計された識別関数 $\widehat{h}(\mathbf{x})$ の識別能力が評価できる。ただし、式(4.15)(4.16)の計算は数値積分による。

4.3.2 2次識別関数の場合

学習標本から推定した二つのパターン母集団の平均値ベクトル (標本平均) を $\widehat{\boldsymbol{\mu}}_1$, $\widehat{\boldsymbol{\mu}}_2$, 共分散行列 (標本共分散行列) を $\widehat{\boldsymbol{\Sigma}}_1$, $\widehat{\boldsymbol{\Sigma}}_2$ とする。推定したパラメータによって設計される2次識別関数は、これらを式(4.5)に代入して、

$$\widehat{h}(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \widehat{\boldsymbol{\mu}}_1)^t \widehat{\boldsymbol{\Sigma}}_1^{-1} (\mathbf{x} - \widehat{\boldsymbol{\mu}}_1) - \frac{1}{2} (\mathbf{x} - \widehat{\boldsymbol{\mu}}_2)^t \widehat{\boldsymbol{\Sigma}}_2^{-1} (\mathbf{x} - \widehat{\boldsymbol{\mu}}_2) + \frac{1}{2} \ln \frac{|\widehat{\boldsymbol{\Sigma}}_1|}{|\widehat{\boldsymbol{\Sigma}}_2|} \quad (4.17)$$

第4章 統計的パターン分類における認識率の評価

式(4.10)の $\hat{h}(\mathbf{x})$ の代わりに上式(4.17)を用い、式(4.11)と同様にして分類を行うとすると、クラス1のパターンをクラス2と誤識別する確率 ε_1 は次式で与えられる。

$$\varepsilon_1 = \int_{\hat{h}(\mathbf{x}) > 0} p_1(\mathbf{x}) d\mathbf{x} \quad (4.18)$$

ただし、 $p_1(\mathbf{x})$ は \mathbf{x} がクラス1に属するときの確率密度関数。1.3節の結果を用いると式(4.18)は次のように変換でき、1次元定積分に帰着できる。

$$\begin{aligned} \varepsilon_1 &= \int_{\hat{g}(\mathbf{y}) > 0} r(\mathbf{y}) d\mathbf{y} \\ &= \frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1}{j\omega} F(\omega) d\omega \end{aligned} \quad (4.19)$$

ただし、

$$F(\omega) = \left\{ \prod_{i=1}^N \frac{1}{\sqrt{1-j\omega/\lambda_i}} e^{-\frac{1}{2} \left(\frac{\omega^2 \nu_i^2}{1-j\omega/\lambda_i} \right)} \right\} \cdot e^{j\omega c}, \quad j=\sqrt{-1} \quad (4.20)$$

c, ν_i, λ_i はここで用いる変数変換で定まる定数であるが詳細は1.3節参照。式(4.19)の計算は数値積分による。同様にして ε_2 も求められる。

算出された ε_1 と ε_2 と式(1.4)により、有限個の学習標本から設計された個々の2次識別関数の識別能力が評価できる。

4.3.3 誤り率の増加分の評価

この節では本章で検討の対象とする誤り率の増加分の評価式をまとめておく。

(1) 線形識別関数の場合 前節の結果によって個々の具体的な識別関数に対する認識率を数値的に評価できるが、学習標本の統計的変動に基づく $\hat{h}(\mathbf{x})$ の変動は求められないため式(4.7)の値も評価できない。これに関しては、Raudysおよび Fukunagaがそれぞれ、 $E\{\Delta\varepsilon_d\}$ の近似式として次式(4.21)(4.22)を与えている⁽¹⁾⁽²⁾。

$$\begin{aligned} & \text{(Raudysの近似式)} \\ \Delta\varepsilon_d &= \frac{1}{2M} \cdot \frac{\phi(\delta/2)}{\delta} \left\{ 1 + \frac{\delta^2}{2} + N + \frac{\frac{\delta^4}{8} + \frac{N\delta^2}{4}}{1 - \frac{N}{2M}} \right\} \end{aligned} \quad (4.21)$$

$$\begin{aligned} & \text{(Fukunagaの近似式)} \\ \Delta\varepsilon_d &= \frac{1}{2M} \cdot \frac{\phi(\delta/2)}{\delta} \{ N (1 + \frac{\delta^2}{4}) - 1 \} \end{aligned} \quad (4.22)$$

第4章 統計的パターン分類における認識率の評価

ただし、 δ は式(4.4)で定義されるマハラノビス距離であり、 $\phi(x)$ は標準正規分布の密度関数(次式)である。

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (4.23)$$

なお、Raudysは識別関数の分布が正規分布を成すと仮定して近似式を導出している⁽¹⁾が厳密にはこれは成り立たない。一方 Fukunagaは学習標本から設計した識別関数とベイズ識別を達成する識別関数の差を母集団のパラメータによってTaylor展開し、3次以上の項を切り捨てて近似式を導出している⁽²⁾。このため二つの近似式とも近似精度に限界があることは明らかであるが、その定量的な評価は文献(1),(2)には与えられていない。また、その他にもこれらの近似式が実際にどの程度よい近似を与えるかを数値的に評価した報告は見当たらない。

さらに、Efronは線形識別関数の場合、 $\Delta \varepsilon_d / A$ が近似的に自由度Nの χ^2 分布に従うことを示した⁽³⁾。

$$\text{ただし, } A = \frac{1}{2M} \cdot \frac{\phi(\delta/2)}{\delta} \left(1 + \frac{\delta^2}{4} \right) \quad (4.24)$$

(2) 2次識別関数の場合 識別関数の設計者が母集団の共分散行列が等しいことを知らない場合には、有限個の学習標本から推定した二つの母集団の平均ベクトルと共分散行列を個別に用いて2次識別関数を設計することになる。この場合の $E\{\Delta \varepsilon_d\}$ の近似値として Raudys および Fukunaga はそれぞれ、次の式(4.25)(4.26)を与えている⁽¹⁾⁽²⁾。

(Raudysの近似式)

$$E\{\Delta \varepsilon_d\} = \frac{1}{2M} \cdot \frac{\phi(\delta/2)}{\delta} \left[1 + \frac{\delta^2}{2} + N + \frac{2 \left\{ \frac{\delta^4}{8} + \frac{N(N+\delta^2)}{4} \right\}}{1 - \frac{N}{M}} \right] \quad (4.25)$$

(Fukunagaの近似式)

$$E\{\Delta \varepsilon_d\} = \frac{1}{4M} \cdot \frac{\phi(\delta/2)}{\delta} \left\{ N^2 + N \left(1 + \frac{\delta^2}{4} \right) + \frac{\delta^4}{16} - \frac{\delta^2}{2} - 1 \right\} \quad (4.26)$$

式(4.25)(4.26)中にマハラノビス距離 δ が含まれていることから明らかなように、これらの近似式は二つのクラスの真の共分散行列が等しい場合($\Sigma_1 = \Sigma_2$)にしか適用できない。

4. 4 実験

4. 4. 1 実験方法

本節では、式(4. 21)(4. 22)(4. 25)(4. 26)の $\Delta \varepsilon_d$ の評価式の精度を実験的に評価する。その手順は以下の通りである。

(1)次元数 N で、二つの母集団のパラメータがそれぞれ平均値ベクトル μ_1, μ_2 , 共分散行列 Σ_1, Σ_2 の正規母集団を想定する。この母集団からの標本を、各クラスにつき M 個ずつ乱数を用いて発生させ、大きさ M の学習標本を1セット作成する。

(2)この学習標本を用いて識別関数の設計を行う。

(3)3節で述べた数値積分(式(4. 15)(4. 16)あるいは(4. 19))により、このセットに対応する識別関数の認識率を算出する。これとベイズ誤り率(発生させた母集団に対する)の理論値から $\Delta \varepsilon_d$ のサンプル値を得る。

(4)さらにこれを複数回実験し、セット毎の認識率の平均、分散をもって各条件における認識率並びにその分散とした。繰り返し回数は分散のばらつきが十分小さくなるように、最低でも200回以上とした。識別関数としては、線形識別関数と2次識別関数の2種類について実験を行った。

パターンの母集団のパラメータは以下のように設定した。

クラス1 : $\mu_1 = \mathbf{0}, \Sigma_1 = \mathbf{I}$ (単位行列)

クラス2 : $\mu_2 = (a, a, \dots, a)^t, \Sigma_2 = \mathbf{I}, (a \text{ は適当な定数})$

ここで、 a の値を変えて、ベイズの誤り率 $\varepsilon_B = 0.0001 \sim 0.1$ の間の値を実現した。次元数 $N = 5 \sim 100$ 次元で、学習標本数 $M = 1.5N \sim 20N$ の範囲で実験を行った。

4. 4. 2 実験結果

(1)線形識別関数の場合 式(4. 21)と(4. 22)の差をとれば、(次元数 $N <$ 学習標本数 M) の範囲では常に式(4. 21)が式(4. 22)より大きくなるのが容易に確かめられる。 N が $2M$ に接近するに従い式(4. 21)の第4項が支配的となり、式(4. 21)と式(4. 22)の差はますます増大する。

例として、次元数 $N = 10$ でベイズの誤り率 $\varepsilon_B = 0.01, 0.1$ の場合の実験結果を、RaudysとFukunagaの近似式による値とともに図4. 1, 4. 2に示す。

図より明らかなように、Raudysの近似式の方が実験結果よりも誤り率を大きく推定しており、Fukunagaの近似式は実験値より小さい誤り率を与えている。Fukunagaの近

第4章 統計的パターン分類における認識率の評価

似式は、学習標本数 M が次元数 N の4倍以上ある範囲ではかなり良い近似値を与えているが、標本数の少ないところで誤り率の推定が小さすぎることがわかる。この傾向は次元数 N の増加とともに、またベイズの誤り率 ε_B の低下とともに強くなり、 $N=20$ 以上では $\varepsilon_B=0.01$ 以下の場合、学習標本数 M が $2N$ 以下の範囲ではFukunagaの近似値が実験値の $1/2$ 以下になることさえある。一方、 $\varepsilon_B=0.1$ の場合にはかなり近似精度が良い⁽⁶⁾。

図4.3はベイズの誤り率 ε_B を横軸にとり、標本数 M が次元数 N の2倍および5倍の場合において、 $\Delta \varepsilon_d$ (近似値) / $\Delta \varepsilon_d$ (実験値) を求めたものである。例えば、 $N=10$ に対して、Raudysの近似は $M=2N$ の場合に、Fukunagaの近似は $M=5N$ の場合に、それぞれ良い近似を与えていることがわかる。

(2) 2次識別関数の場合 次元数 $N <$ 学習標本数 M の範囲では常に式(4.25)が式(4.26)より大きくなる。式(4.25)と式(4.26)の差は N と M が接近するに従い式(4.25)の第4項が支配的となり、ますます増大する。Fukunagaの評価式は線形識別関数の場合と比較して、中括弧内の第1項と第3・4項が追加されている。第1項の N^2 は線形識別関数に比較して無条件に $E\{\Delta \varepsilon_d\}$ の増加要因になっている。第3・4項は

$$\delta^4/16 - \delta^2/2 = \delta^2(\delta^2 - 8)/16 \quad (4.27)$$

と変形する事により、 $\delta^2 > 8$ の場合 (式(4.2)よりベイズの誤り率 $\varepsilon_B < 0.08$ に相当する)、 $E\{\Delta \varepsilon_d\}$ は線形識別関数の場合より増加することが知られる。

次元数 $N=10$ でベイズの誤り率 $\varepsilon_B=0.01, 0.1$ の場合の例を図4.4, 4.5に示す。

2次識別関数の場合も4.2.1の線形識別関数の場合と同じ現象が見られる。すなわち、図より明らかなように、Raudysの近似式は実験結果よりも誤り率を大きく推定し、Fukunagaの近似式は小さく推定する。Fukunagaの近似式は ε が0.05以上の範囲でかつ、標本数 M が次元数 N の5倍以上ある範囲ではかなり良い近似値を与えるが、標本数の少ないところでは推定値は小さい方へ片寄る。この傾向は次元数 N の増加とともに、またベイズの誤り率 ε_B の低下とともに強まる。例えば、 $N=20$ 以上で $\varepsilon_B=0.01$ 以下の場合、標本数 M が $2N$ 以下の範囲ではFukunagaの近似値が実験値の $1/10$ 以下になることもある⁽⁶⁾。

図4.6は図4.3と同様に、 $\Delta \varepsilon_d$ (近似値) / $\Delta \varepsilon_d$ (実験値) の値を求めたものである。 $N=10$ ではベイズの誤り率 ε_B が小さい場合、Raudysの近似の方がFukunagaの近似よりも優れていることがわかる。

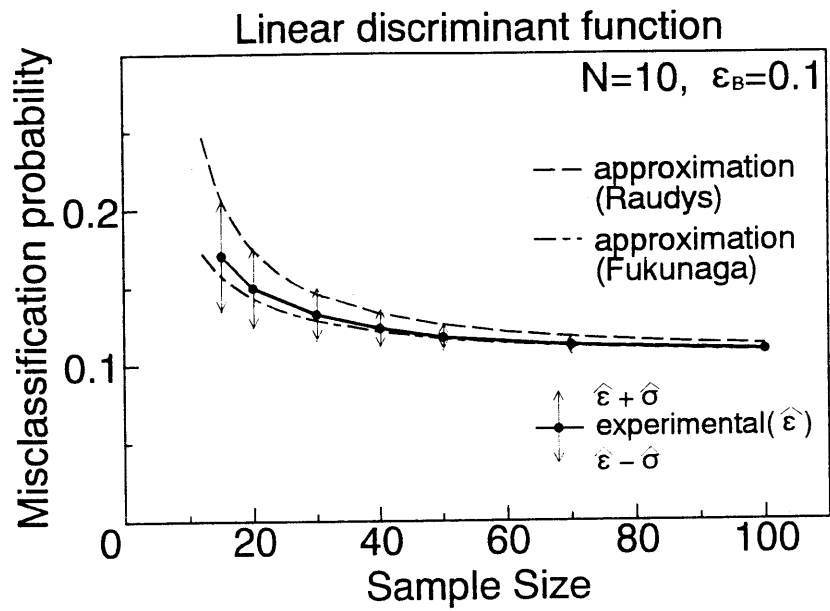


図4.1 標本数と誤り率の関係－線形識別関数－(1)

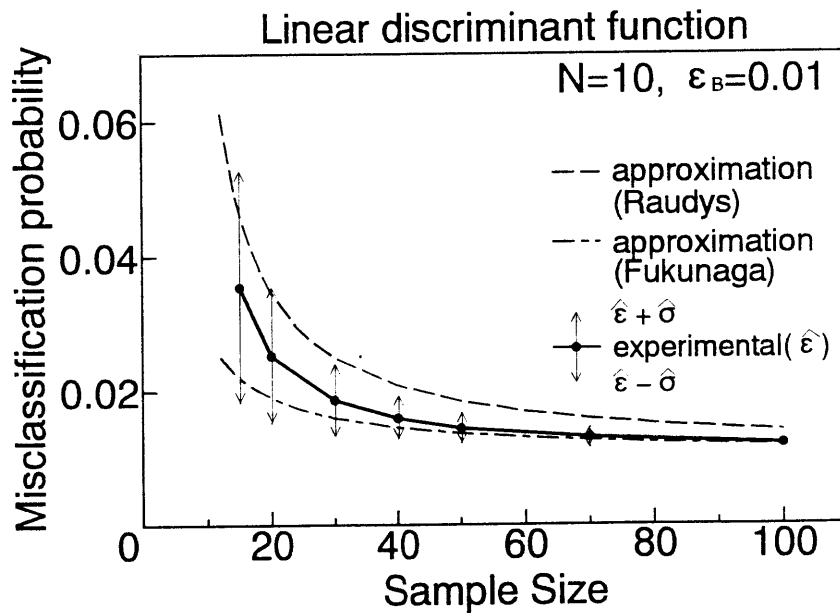


図4.2 標本数と誤り率の関係－線形識別関数－(2)

第4章 統計的パターン分類における認識率の評価

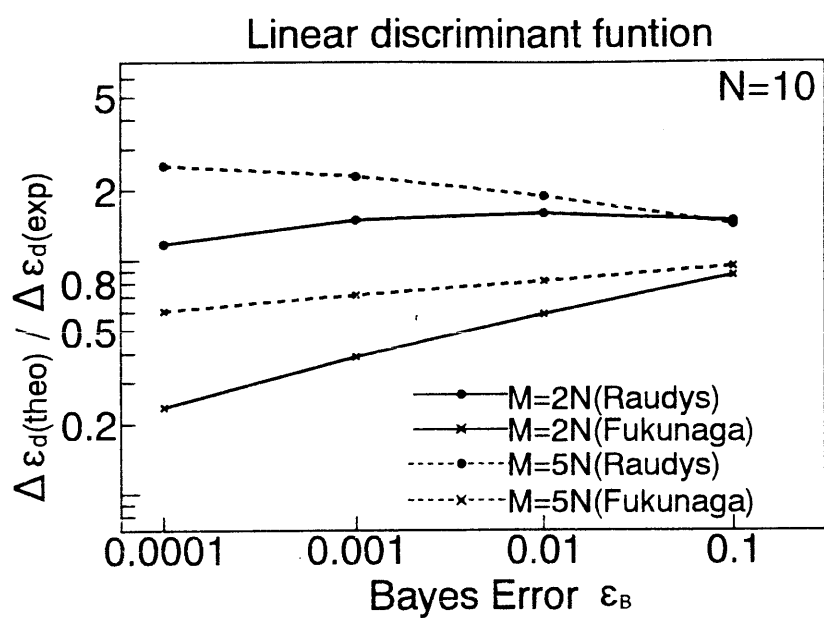


図4.3 近似精度－線形識別関数－

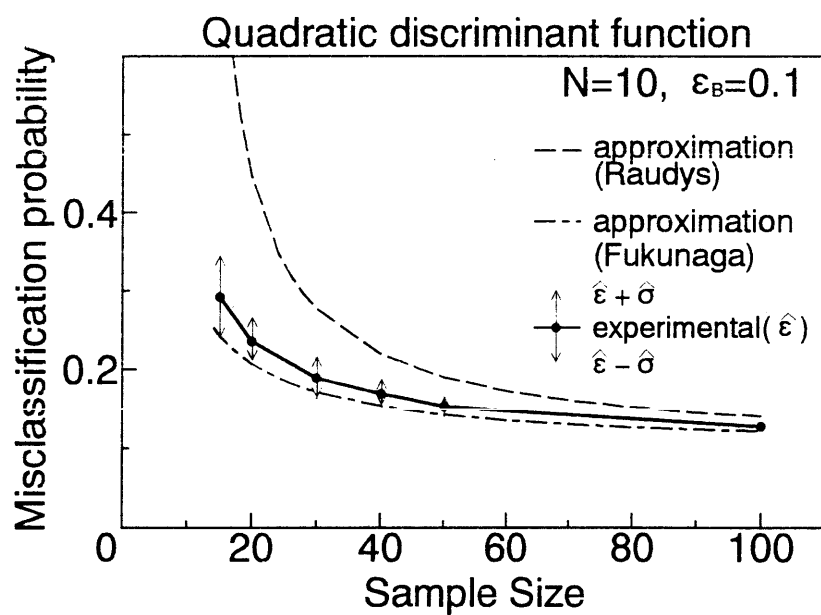


図4.4 標本数と誤り率の関係－2次識別関数－(1)

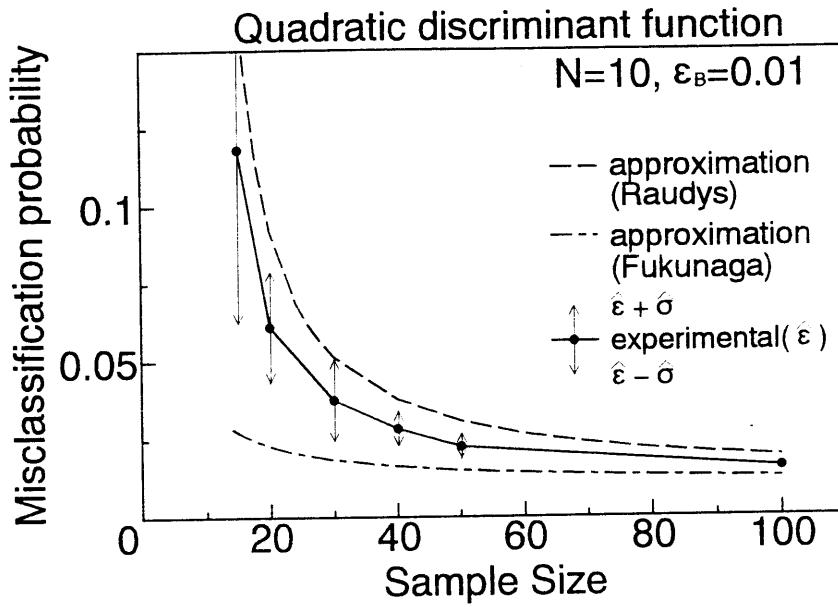


図4.5 標本数と誤り率の関係－2次識別関数－(2)

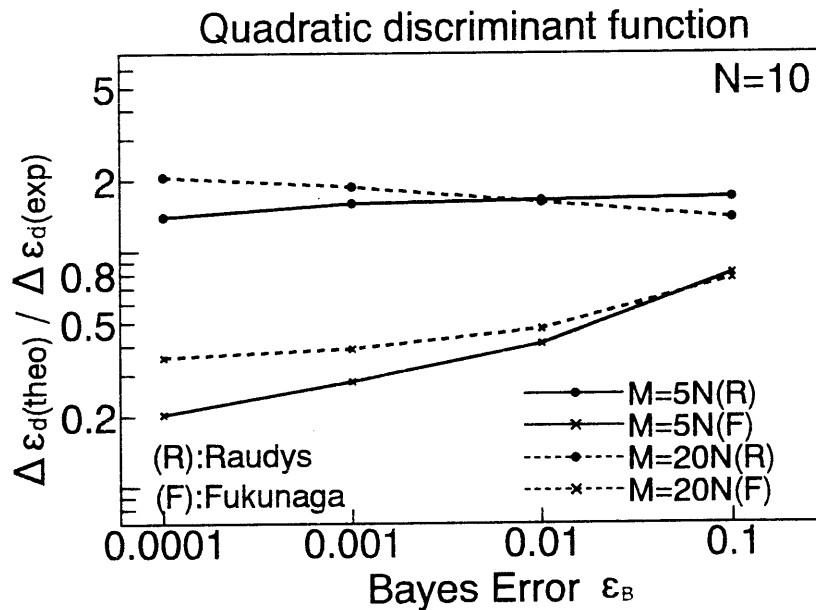


図4.6 近似精度－2次識別関数－

第4章 統計的パターン分類における認識率の評価

4.4.3 識別実験の誤り率の平均値と分散

先に、筆者らが行ったLeave-one-out法による認識率の実験的評価⁽⁴⁾と今回の実験の結果を比較すると、平均誤り率（平均認識率）については、二つの実験とも完全に一致した値が得られた。すなわちLeave-one-out法による認識実験の認識率は平均すれば真の認識率の良い推定を与えている。また式(4.7)に関連して、この実験の場合の $E\{\Delta \varepsilon_i\}$ に対応する項はほぼ0とみなされる。

次に、認識実験における誤り率のばらつきについて検討する。図4.7に線形識別関数で次元数 $N=10$ 、ベイズの誤り率 $\varepsilon_B=0.1$ の場合の学習標本数と誤り率の分散の関係を示す。図中の黒丸(●)はLeave-one-out法によって識別実験を行った場合⁽⁴⁾の誤り率の分散 $\text{Var}\{\varepsilon_{di}\}$ を示す。また×印は、誤り率の計算には理論計算式を用いた本文の実験から得られた誤り率の分散 $\text{Var}\{\Delta \varepsilon_d\}$ を示す。さらに式(4.7)による ε_d の値として本文の実験で求めた誤り率を用い、これを式(4.9)に入れて求めた値を $\text{Var}\{\Delta \varepsilon_i\}$ とみなして、この $\text{Var}\{\Delta \varepsilon_i\}$ を×印の実験値 $\text{Var}\{\Delta \varepsilon_d\}$ に加えたものを折線で結び、図4.7に破線で示す。破線は●印の実験値 $\text{Var}\{\varepsilon_{di}\}$ によく一致し、式(4.8)(4.9)が本文の実験結果とよく合うことが確認できた。

ここでは、線形識別関数を例にとり説明したが同様なことが2次識別関数についても確認できた。

4.4.4 誤り率の増加分の分布（線形識別関数）

前記Efron⁽³⁾の結果を実験的に確かめる。まず、 $\Delta \varepsilon_d/A$ (A は式(4.24))の実験値の分布に対する χ^2 分布の適合度を調べた。図4.8は次元数 $N=10$ 、学習標本数 $M=50$ 、ベイズの誤り率 $\varepsilon=0.1$ 、実験回数2000の場合の $\Delta \varepsilon_d/A$ のヒストグラムである。図中破線は χ^2 分布(自由度= N)の確率密度関数を示す。ただし、ヒストグラムを作成する際、実験から得られた値 $E\{\Delta \varepsilon_d\}=0.0186$ より、 $E\{\Delta \varepsilon_d\}/A'=N$ を満たすようにして A' を選んだ。その上で区間数10にて χ^2 適合度検定を行ったところ、有意水準10%で帰無仮説《 $\Delta \varepsilon_d/A$ が χ^2 分布にしたがう》が棄却できなかった。

また、 $\Delta \varepsilon_d/A$ が χ^2 分布に従えば、次式が成立する。

$$\text{Var}\{\Delta \varepsilon_d\}/[E\{\Delta \varepsilon_d\}]^2 = 2/N \quad (4.28)$$

各種の条件のもとでの式(4.28)の実験値を表に示す。

表中、最も右側の欄に、 $\Delta \varepsilon_d/A$ が χ^2 分布に従う場合の理論値を示す。この表より、次元数 N が大きくなる程、ベイズの誤り率 ε_B が大きくなるにつれ、学習標本数 M

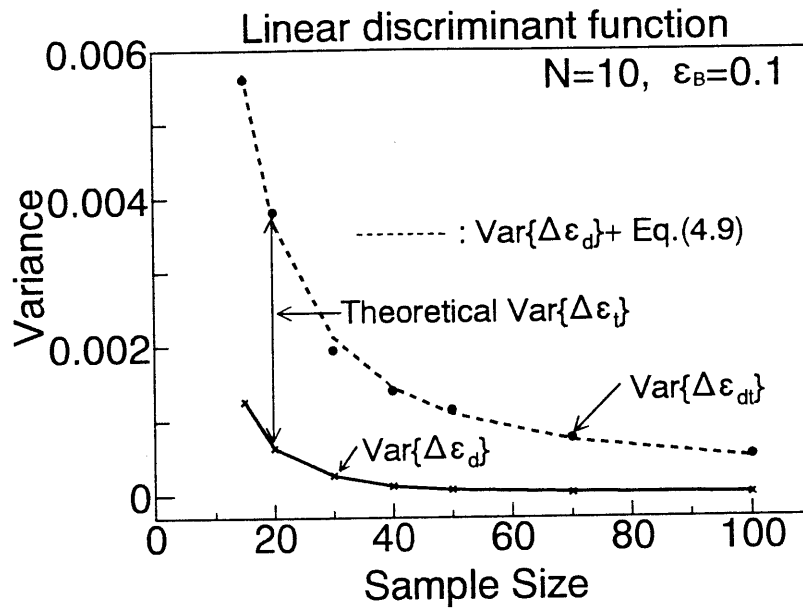


図4.7 誤り率の分散

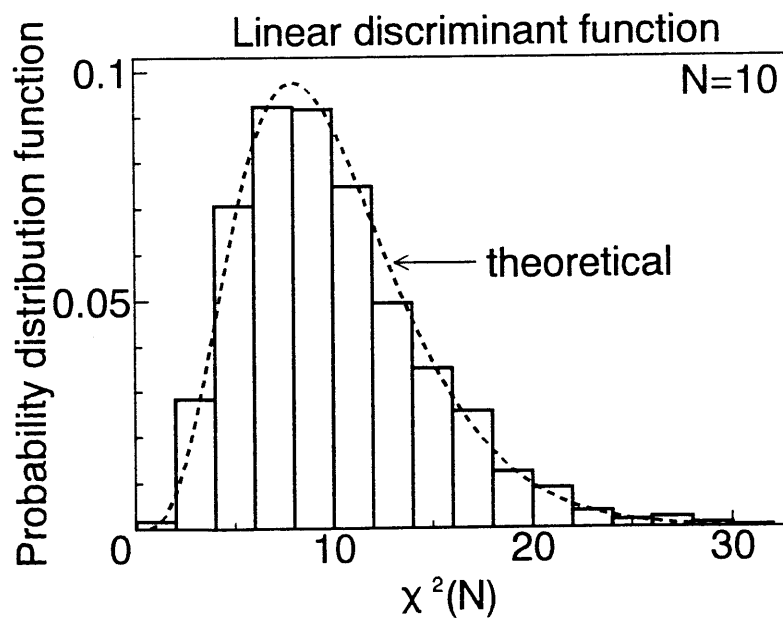


図4.8 $\Delta\epsilon_d$ の分布

第4章 統計的パターン分類における認識率の評価

表 $\text{Var}(\Delta \varepsilon_d) / \{E(\Delta \varepsilon_d)\}^2$ (線形識別関数)

$N \backslash \varepsilon_B$	0.0001	0.001	0.01	0.1	2/Nの場合
5	7.84	2.72	1.05	0.52	0.4
	0.98	0.71	0.53	0.43	
20	0.48	0.30	0.18	0.12	0.1
	0.21	0.17	0.13	0.11	
100	0.084	0.056	0.036	0.024	0.02
	0.038	0.031	0.025	0.020	

(上段は $M=2N$, 下段は $M=5N$)

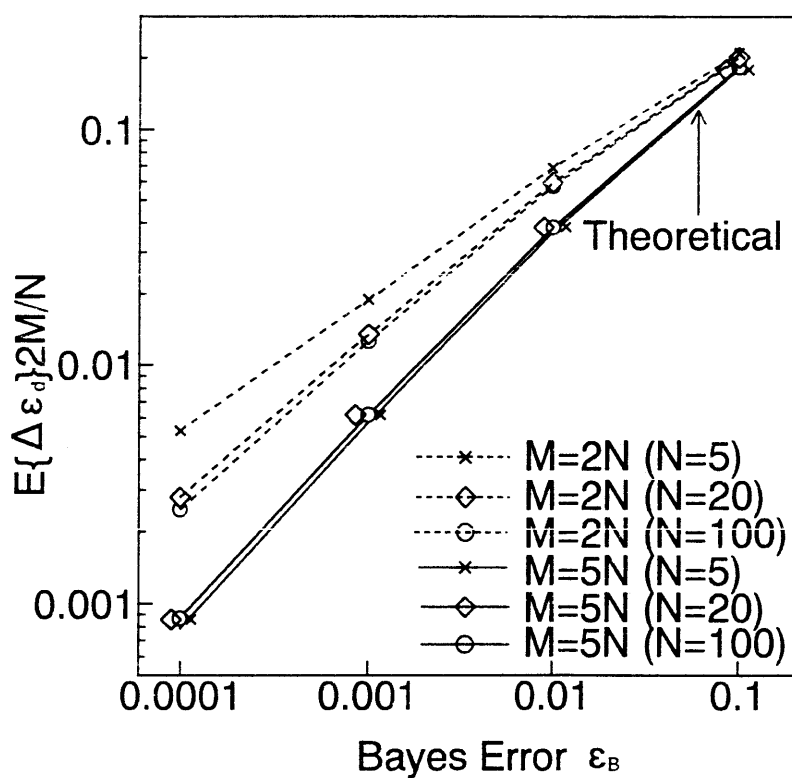


図4.9 $E(\Delta \varepsilon_d)$ と χ^2 分布との関係

が大きい程、この値は χ^2 分布の場合の理論値に近づくことがわかる。

次に、 $\Delta \varepsilon_d$ の期待値について、ベイズの誤り率 ε_B 毎に次式(4.29)を求める。

$$E\{\Delta \varepsilon_d\} = 2M/N \quad (4.29)$$

すなわち、 $\Delta \varepsilon_d/A$ が自由度 N の χ^2 分布に従えば、

$$E\{\Delta \varepsilon_d\} / A = N \quad (4.30)$$

より、式(4.24)を用い、

$$\frac{E\{\Delta \varepsilon_d\} \cdot 2M}{N} = \frac{\phi(\delta/2)}{\delta} \left(1 + \frac{\delta^2}{4}\right) \quad (4.31)$$

となり、式(4.31)の右辺は、 δ （換言すればベイズ誤り率 ε_B ）のみに依存し、次元数 N や学習標本数 M に依存しない値となる。これを、 ε_B 毎にプロットしたものを図4.9に示す。 $M=5N$ の場合には、 $N=5, 20, 100$ とも同一のグラフとなる。図4.9中のTheoreticalとは、式(4.31)に基づくものである。

この図4.9より、ベイズ誤り率 ε_B が大きくなる程、次元数 N が大きくなるにつれ、学習標本数 M が大きい程、式(4.31)の理論的な近似値に近づくことがわかる。

4.5 むすび

二つのクラスが多次元正規分布に従う場合、学習標本から推定したパラメータに基づいて設計した識別関数における認識率の低下量 $\Delta \varepsilon_d$ について、テスト標本の影響があらわれないようなモンテカルロシミュレーションで実験的評価を行った。すなわち、認識率の低下量 $\Delta \varepsilon_d$ の期待値を与えるRaudysとFukunagaの近似理論式を実験的に評価した。また、Efronによる $\Delta \varepsilon_d$ の分布に関する近似を実験的に検討した。その結果、互いに等しい共分散行列をもつ正規分布に従うパターンの2クラス分類の場合に、次のようなことが明らかになった。

- (1)ベイズ決定に基づく誤り率が0.1程度の場合には常にFukunagaの評価式がよい推定を与える。
- (2)線形識別関数においては、学習標本数 M が N のおよそ2倍以下の場合Raudysの評価式が、また、およそ M が N の5倍以上あればFukunagaの評価式はよい推定を与える。
- (3)2次識別関数においてはベイズ決定に基づく誤り率が約0.05以下では、Raudysの評価式がよい推定を与える。このとき、Fukunagaの評価式は実験値に比べ楽観的に過ぎる値を示す。

第4章 統計的パターン分類における認識率の評価

(4)線形識別関数においては、学習標本数の有限性に起因する認識率の低下は、ベイズ誤り率が高く、特徴ベクトルの次元数が高くなるに従い、自由度 N の χ^2 分布に近づく。

今後の課題として、2次識別関数における認識率の低下分のなす分布の性質を明らかにすること、二つの母集団の共分散行列が異なる場合についての学習標本数の有限性に起因する認識率の低下の程度について調べるなどがある。

文 献

- (1)Raudys, S. J. and Jain, A. K. : "Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners", IEEE Trans. Pattern Anal. and Machine Intell., Vol.13, No.3, pp.252-264 (March 1991)
- (2)Fukunaga, K. and Hayes, R. R. : "Effects of sample size in classifier design", IEEE Trans. Pattern Anal. and Machine Intell., Vol.2, No.8, pp.873-885 (Aug. 1989)
- (3)Efron, B. : "The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis", J. Amer. Statist. Assoc., Vol. 70, No.352 pp.892-898, (1975)
- (4)竹下, 野澤, 鳥脇 : "多次元正規分布において標本数が認識率に及ぼす影響について", PRU92-27, 信学技報Vol. 92, No. 227, pp.17-24 (1992)
- (5)Fukunaga, K. : "Introduction to Statistical Pattern Recognition" 2nd ed., Academic Press, New-York (1990) (Chapter 2, 3, and 5)
- (6)Takeshita, T., Toriwaki, J. : "Experimental study of performance of pattern classifiers and the size of design samples", Pattern Recognition letters, Vol. 16, No. 3, pp.307-312(1995)

第5章 有限学習標本に基づく 統計的パターン分類器の改良

第4章において、二つの母集団の共分散行列が等しい場合について有限学習標本に基づく統計的パターン分類器のベイズ誤り率からの性能低下の程度を定量的に評価した。線形識別関数および2次識別関数の分類性能を評価する際には、特に、母集団の真の共分散行列が等しいと仮定しても一般性を失うことはなかった。しかしながら、識別関数の改良を行う際には、二つの母集団の分離の程度が識別関数の改良に密接に関係するため、二つの母集団の真のパラメータをどのように設定するかという問題も生じてくる。普通ベイズ誤り率を固定しても、母集団の真のパラメータは一意に定まらない。このため、本章では二つの母集団の間の分離の程度を測定できる Bhattacharyya 距離を用い、ベイズ誤り率を一定に保ちながら、二つの母集団の真のパラメータを設定できるようにする。このようにして二つの母集団の共分散行列が異なる場合も含み、学習標本の有限性に起因する識別関数の性能低下を改良する試みについてシミュレーションを行う。その結果、2次識別関数が学習標本数に敏感な識別関数であるということ、および、Friedman の提案する Regularized Discriminant Analysis が学習標本数の少ない状況において、識別性能の低下を防止することに効果があることを確認する。さらに、具体的な数字認識実験を通じて、識別関数を改良できる見通しを与える。

5.1 まえがき

第3章において、2次識別関数の値を実質的に支配するマハラノビス距離の推定量の確率分布を明らかにし、母集団のパラメータが正確に推定できないことに起因するマハラノビス距離の推定誤差について、標本共分散行列の推定誤差が主要な要素であることを明らかにした。ところで、この標本共分散行列の推定誤差に起因して識別関数の能力が低下することを防止する方法が、これまでもいくつか提案されてきた。その中で、Friedman は学習標本の数が特徴パターンの次元数に比較して十分確保でき

第5章 統計的パターン分類器の改良

ない状況における識別手順としてRegularized Discriminant Analysis（以下RDA法と呼ぶ）を提案している⁽¹⁾。

Friedmanによれば、RDA法は特徴ベクトルの次元数よりも学習標本数が少ない場合にも有効であるとしており、RayensとGreene⁽²⁾やAeberhardら⁽³⁾もシミュレーション実験や具体的な分類実験を通じてその有効性をある程度は確認している。しかしながら、RDA法は識別関数の設計に際して二つのパラメータに自由度が与えられており、その最適な値の選択法は得られていない。また、特徴ベクトルの次元数に比較して学習標本数が少ないという条件のもとでの実験では、平均認識率を分類実験で評価する場合の分散を大きくする要因を内包している。このためベイズ決定に基づく認識率が高い場合には、識別関数の性能を評価するためのテスト標本を極めて多数準備する必要がある、認識実験のみによる評価を困難にしている。また、RDA法の効果は母集団のパラメータに強く依存する性格を有していると考えられるが、それについて系統的に検証した報告はない。

第4章において、二つのクラスの共分散行列が共に単位行列に等しい場合について、学習標本から推定したパラメータを用いて設計した識別関数のベイズ認識率からの低下分を数値積分を用いて算出する方法を見出し、有限個のテスト標本を用いて識別関数の性能評価をするかわりに、母集団のパラメータを用いて数値積分を行い、理論的な平均認識率を評価した。この方法は、識別関数に対して、テスト標本の数に依存しない性能評価をすることができ、ベイズ認識率が高い場合に特に有効な手法である。

そこで、本章では、正規分布に従うパターンにおいて二つのクラスの共分散行列が異なる場合についても同じ手法を適用し、分類実験のみに依存することのないやり方でRDA法に基づく識別関数の性能評価を試みた。

ここで考察の対象となる二つのクラスの母集団のパラメータを設定する際、RDA法の評価が広い範囲でできるようBhattacharyya距離⁽⁴⁾を用い、二つのクラスの分離度を測定した。正規母集団に対するBhattacharyya距離は二つの成分、すなわち、平均値ベクトルの相違に基づく成分、および共分散行列の相違に基づく成分から構成されている。本研究ではこの二つの成分の構成比を変えることで、ベイズ認識率が等しいいくつかの母集団パラメータを設定した。さらに、具体的な認識性能への影響を調べるため数字認識実験を試みた。

その結果、2次識別関数が学習標本数に極めて敏感なパターン分類器であることを

確認した。また、RDA法の有効性を確認できたものの、ベイズ認識率の高い状況ではRDA法の方のパラメータがあまり有効に作用しないことを見いだした。このことから、RDA法が有している二つのパラメータの自由度の内、より有効なパラメータを指摘でき、より優れた統計的パターン分類器の設計ができる見通しを得た。

以下、5.2でRDA法に基づく識別関数およびBhattacharyya距離について述べ、5.3で学習標本から推定したRDA法に基づく識別関数の性能を明らかにするためのシミュレーション実験の方法およびその結果、5.4で数字認識実験の結果について述べる。

5.2 母集団パラメータと識別関数

多次元正規分布に従う二つのパターン母集団（クラス1、および2）の平均値ベクトルをそれぞれ μ_1 、 μ_2 、共分散行列を Σ_1 、 Σ_2 とする。

二つの母集団の共分散行列が異なる場合、ベイズ決定は次式の二次識別関数による識別になる。

$$h(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mu_1)^t \Sigma_1^{-1}(\mathbf{x} - \mu_1) - \frac{1}{2}(\mathbf{x} - \mu_2)^t \Sigma_2^{-1}(\mathbf{x} - \mu_2) + \frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} \quad (5.1)$$

学習標本から推定した二つのパターン母集団の標本平均値ベクトルを $\hat{\mu}_1$ 、 $\hat{\mu}_2$ 、共分散行列（標本共分散行列）を $\hat{\Sigma}_1$ 、 $\hat{\Sigma}_2$ とする。推定したパラメータによって設計される二次識別関数は、これらを式(5.1)に代入して、

$$\hat{h}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \hat{\mu}_1)^t \hat{\Sigma}_1^{-1}(\mathbf{x} - \hat{\mu}_1) - \frac{1}{2}(\mathbf{x} - \hat{\mu}_2)^t \hat{\Sigma}_2^{-1}(\mathbf{x} - \hat{\mu}_2) + \frac{1}{2} \ln \frac{|\hat{\Sigma}_1|}{|\hat{\Sigma}_2|} \quad (5.2)$$

5.2.1 RDA法に基づく識別関数

RDA法は、学習標本が十分確保できない状況における共分散行列のロバスト性(robustness)を増すための識別関数の設計法に関する一つの改良である。

学習標本を N 次元のベクトル \mathbf{X}_{ji} （添え字 i はクラス、 j は標本番号）で表す。総クラス数は g とし、第 i クラスの学習標本数は m_i である。ここで、総学習標本数は m とする。すなわち、

$$m = \sum_{i=1}^g m_i \quad (5.3)$$

クラス毎の標本平均値ベクトルは式(5.4)となる。

第5章 統計的パターン分類器の改良

$$\hat{\mu}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \mathbf{X}_{ji} \quad (5.4)$$

2次識別関数の設計に必要なクラス毎の標本共分散行列は式(5.5)で得られる.

$$\hat{\Sigma}_i = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (\mathbf{X}_{ji} - \hat{\mu}_i)(\mathbf{X}_{ji} - \hat{\mu}_i)^t \quad (5.5)$$

一方, すべてのクラスの共分散行列が等しいと仮定して得られるプールされた共分散行列 (pooled covariance matrix) は

$$\hat{\Sigma}_{pool} = \sum_{i=1}^g \sum_{j=1}^{m_i} \frac{(\mathbf{X}_{ji} - \hat{\mu}_i)(\mathbf{X}_{ji} - \hat{\mu}_i)^t}{m - g} \quad (5.6)$$

となる. ベイズ推定に基づく2次識別関数ではクラス毎の標本共分散行列として式(5.5)を用いるが, RDA法ではそのかわりに, 式(5.7)に示すように式(5.5)と(5.6)を組み合わせて使う.

$$\hat{\Sigma}_i(\lambda) = \frac{(1-\lambda)(m_i-1)\hat{\Sigma}_i + \lambda(m-g)\hat{\Sigma}_{pool}}{(1-\lambda)(m_i-1) + \lambda(m-g)} \quad (5.7)$$

但し, $0 \leq \lambda \leq 1$

$\lambda = 0$ のとき, 式(5.7)は式(5.5)に等しく, $\lambda = 1$ のとき式(5.7)は式(5.6)と等しい. すなわち λ はクラス毎の共分散行列の相似度を示すパラメータと考えることもできる. $\lambda = 0$ では, クラス毎の共分散行列は互いに異なると仮定して識別関数の設計にクラス毎の標本共分散行列を用いる. これに対して $\lambda = 1$ ではすべてのクラスの共分散行列が等しいと仮定して識別関数を設計する. このとき, 識別関数は線形識別関数となる. λ が0と1の間では個別の標本共分散行列とプールした併合共分散行列を混合してクラス毎の標本共分散行列としている. 識別関数式(5.2)の標本共分散行列のかわりに形式的に式(5.7)の行列をクラス毎の標本共分散行列として, 2次識別関数を設計する.

RDA法においては, クラス毎の共分散行列の相似度を反映するパラメータ λ とは別に, ridge regression と呼ばれる技法に対応するもう一つのパラメータ γ がある. これら二つのパラメータを組み合わせて, RDA法による識別関数の設計の際には式(5.8)の標本共分散行列を用いる.

$$\hat{\Sigma}_i(\lambda, \gamma) = (1-\gamma)\hat{\Sigma}_i(\lambda) + \gamma c_i \mathbf{I}_N \quad (5.8)$$

但し, $c_i = \{\text{tr} \hat{\Sigma}_i(\lambda)\} / N$

ここで, \mathbf{I}_N は $N \times N$ の単位行列 (N はパターンの次元数) であり, c_i はクラス i の

標本共分散行列の平均固有値である。パラメータ γ の取り得る範囲は $[0, 1]$ である。

γ に基づく補正は、標本共分散行列の推定誤差を平均二乗誤差で評価するとき、標本共分散行列の固有値が大きな値はより小さな値で置換し、小さな固有値はより大きな値で置換すると平均二乗誤差がより小さくなるという事実⁽⁵⁾⁻⁽⁷⁾に依っている。 $\gamma = 0$ では、補正をしないことに等しく、 $\gamma = 1$ では特徴ベクトルのすべての成分が同じ分散を有する分布とみなすことと等価である。このようにしてRDA法は標本共分散行列に対して二つのパラメータ (λ, γ) に基づく補正を実施し、より高い識別性能を有する識別関数を設計しようとするものである。

パラメータ (λ, γ) の組み合わせに応じて、 $(\lambda = 0, \gamma = 0)$ は従来の2次識別関数に対応し、 $(\lambda = 1, \gamma = 0)$ は線形識別関数に対応する。そして $(\lambda = 1, \gamma = 1)$ の場合は平均パターンからのユークリッド距離に基づく識別関数と等価である。

パラメータ (λ, γ) の最適値の決定法については、Friedmanは、最適値は母集団のパラメータに依存するので標本を識別関数の設計用とテスト用に分割しておき、二つのパラメータ (λ, γ) が異なる多数の識別関数を具体的に設計し、それをテスト用標本で評価し、最も良い性能を示す識別関数に対応するパラメータの組を見いだす、ということしか提案していない。そこで、本章ではRDA法の能力がパラメータ (λ, γ) とともにどのように変化するかを数値的に評価してみる。

RDA法はパラメータ λ が示すように母集団の共分散行列の相似度に応じて補正するため、母集団の性質に応じてパラメータ (λ, γ) の最適値も変動することが予想される。そのため、母集団の性質—主としてクラス間の距離、分離の度合いに関する性質—を表す尺度を用意しておく都合がよい。本章ではこのためにBhattacharyya距離を用いる。

5.2.2 Bhattacharyya 距離

Bhattacharyya距離は二つの正規母集団のクラス間距離を測定する尺度のひとつである。多次元正規分布に従う二つの母集団（クラス）の平均値ベクトルをそれぞれ μ_1, μ_2 、共分散行列を Σ_1, Σ_2 とする。この2クラス間のBhattacharyya距離を式(5.9)に示す⁽⁴⁾。

$$d_B = \frac{1}{8}(\mu_2 - \mu_1)^t \left[\frac{\Sigma_2 + \Sigma_1}{2} \right]^{-1} (\mu_2 - \mu_1) + \frac{1}{2} \ln \frac{\left| \frac{\Sigma_2 + \Sigma_1}{2} \right|}{\sqrt{|\Sigma_2| |\Sigma_1|}} \quad (5.9)$$

第5章 統計的パターン分類器の改良

式(5.9)より、 $\mu_1 = \mu_2$ のとき第1項は0となる。すなわち第1項は二つのクラスの平均値ベクトル間の距離を表わす項である。第2項は $\Sigma_1 = \Sigma_2$ のとき0となり、二つのクラスの共分散行列の相違の程度を示す項である。二つのクラスのそれぞれの平均値ベクトルと共分散行列を決定すれば、Bhattacharyya距離が求められると同時に、数値積分によりベイズ誤り率も求めることができる。

二つのクラスの共分散行列の相違度を広い範囲に亘って確認するという観点から、Bhattacharyya距離の第1成分と第2成分の比に着目する。式(5.9)の右辺の第1項を d_{B1} 、第2項を d_{B2} とし、 R を

$$R = d_{B1} / (d_{B1} + d_{B2}) \quad (5.10)$$

と定義する。以下の実験では、同一のベイズ誤り率を有する母集団パラメータの組み合わせから、 R が $[0, 1]$ の範囲をとるよう広い範囲の組み合わせを実現する。 $R = 0$ のときには、Bhattacharyya距離の第1成分が0であり、二つのクラスの平均値ベクトルは等しく、二つのクラスを分離するのは共分散行列の相違のみである。逆に $R = 1$ のときには、Bhattacharyya距離の第2成分が0であり、二つのクラスの共分散行列が等しい場合に当たる。現実の個々のパターン認識の問題においては、母集団分布のパラメータが既知でないならば R がいくつであるか事前に知ることはできない。

5.3 シミュレーション実験

5.3.1 母集団パラメータの決定

クラス1のパラメータは $\mu_1 = 0$ 、 $\Sigma_1 = I$ （単位行列）とし、クラス2のパラメータは $\mu_2 \neq 0$ 、 $\Sigma_2 = \Lambda$ （対角行列）とした。任意の二つの多次元正規分布は一つの線形変換によって同時に、一方の共分散行列を単位行列、他方を対角行列に変換でき、さらに平行移動によって単位行列に対応するクラスの平均値を原点に移すことができるので、前記のような仮定によっても一般性を失わない。

ここでは、 μ_2 、 Σ_2 は次のように設定した。まず、 μ_2 としては式(5.11)に示すように、すべての成分が等しいベクトル（成分の値を a とする）を仮定した。すなわち、

$$\mu_2 = (a, a, \dots, a)^t. \quad (5.11)$$

次に、 $\Sigma_2 = \Lambda$ とし、 Λ は次の式(5.12)、(5.13)のように設定した。

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_N) \quad (5.12)$$

$$\lambda_i = \sigma^2 \cdot e^{-\alpha(i-1)}, \quad i=1, 2, \dots, N \quad (5.13)$$

二つのクラスの母集団のパラメータ $\mu_1, \mu_2, \Sigma_1, \Sigma_2$ を定めることによりベイズ誤り率が決定する.

二つのクラス間の Bhattacharyya 距離も R も母集団のパラメータを定めれば一意に決まる. $R=0$ とするためには,

$$\mu_1 = \mu_2 = 0 \quad (5.14)$$

とし, 式(5.11)において

$$a = 0 \quad (5.15)$$

とすればよい. また $R=1$ とするためには,

$$\Sigma_1 = \Sigma_2 = I \quad (5.16)$$

とし, 式(5.13)において,

$$\sigma^2 = 1, \quad \alpha = 0 \quad (5.17)$$

とすればよい. 以下では $R=0, 1/3, 2/3, 1$ の各場合についてシミュレーション実験を行った. なお, $\sigma^2=1$ とすると, 指定されたベイズ誤り率に対して, 可変パラメータは式(5.11)の a と式(5.13)の α のみとなり, その上で R を指定することで母集団のパラメータは一意に定まる.

5.3.2 実験方法

実験はモンテカルロシミュレーションにより行った. その手順は以下の通りである.

まず, ベイズの誤り率 ε_B を適当に設定して, 以下の手順を実行する.

- (1) 5.3.1 で求めた母集団パラメータに従う標本を各クラスにつき M 個ずつ乱数を用いて発生させ, 学習標本を 1 セット作成する.
- (2) この学習標本を用いて, クラス毎に, 式(5.4)より平均値ベクトルを求め, 式(5.5)~(5.8)により標本共分散行列の推定を行う. これらの推定値を用いて RDA 法による識別関数を設計する.
- (3) 上記の識別関数の誤り率の理論値を多次元積分によって求める. 計算法については 1.3.2 参照.
- (4) 上記(1)~(3)の処理を L 回繰り返す. こうして得られた L 個の誤り率の平均, および分散をもって各条件における誤り率およびその分散とした. 繰り返し回数 L は分散

第5章 統計的パターン分類器の改良

のばらつきが十分小さくなるように、最低でも100回以上とした。

ベイズの誤り率 ε_B は0.1および0.01で実験した。これは、応用上有用な識別関数では誤り率はかなり小さくなくてはならないからである。

その他のパラメータについては以下の値を用いた。

- ・ パターンの次元数 $N = 10$,
- ・ 学習標本数 $M = 15 \sim 200$,
- ・ パラメータ $R = 0, 1/3, 2/3, 1$
- ・ パラメータ (λ, γ) ; 共に $[0, 1]$ で0.1間隔で変える。

5.3.3 実験結果

実験結果の一部を図5.1～5.5に示す。図ではRDA法の二つのパラメータ (λ, γ) をx軸とy軸にとり、z軸（高さ方向）が平均認識率である。グラフの格子点がそれぞれ識別関数の認識性能を表わしており、認識率が高いほど性能が優れている。

実験結果より次のようなことが明らかになった。

- ①従来の2次識別関数の性能を示す $\lambda = 0, \gamma = 0$ の点では認識率が極端に低下している（図5.1）。これは2次識別関数の学習標本数依存性が極めて強いことを示唆している。
- ② $R = 1$ の場合を除き、平均認識率が最高になるのは、 $\lambda = 0$ の付近である。
- ③ $R = 1$ の場合には2クラスの共分散行列が同一であることから理論的には線形識別関数を使って認識することが最適となる。図5.4の実験結果もこれを裏付けており、 $\lambda = 1$ の場合に認識率が最高になっている。
- ④ $R = 0$ の場合、二つのクラスの平均値ベクトルに差がないことから、線形識別関数では原理的にまったく認識ができず、認識率は0.5である（図5.1の $\lambda = 1$ の場合）。この場合には $\lambda = 0$ のときに最大の認識率(0.99)を達成できるはずであるが、①で指摘したようにこの点での認識率は高くない。しかし、それ以外では λ が小さいほど認識率が高くなることは、RDA法が一定の効果をもつことを示している。

次に、RDA法の二つのパラメータのうち、どちらがより優れた性能を付与するパラメータか判断するため、 γ と認識率の関係を図5.6に、 λ と認識率の関係を図5.7に示す。グラフの中心は平均認識率を示し、上下の横棒は標準偏差を示す。実験条件は前記図5.1の場合と同じである。図5.6のグラフより、極端に識別関数の性能

$N=10, M=15, \varepsilon_B=0.01$

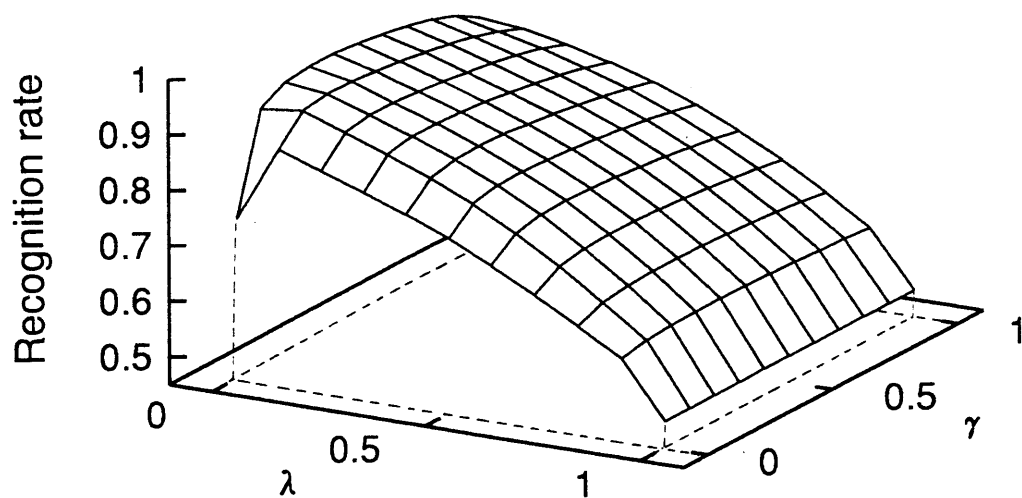


図 5. 1 平均認識率 ($R = 0$)

$N=10, M=15, \varepsilon_B=0.01$

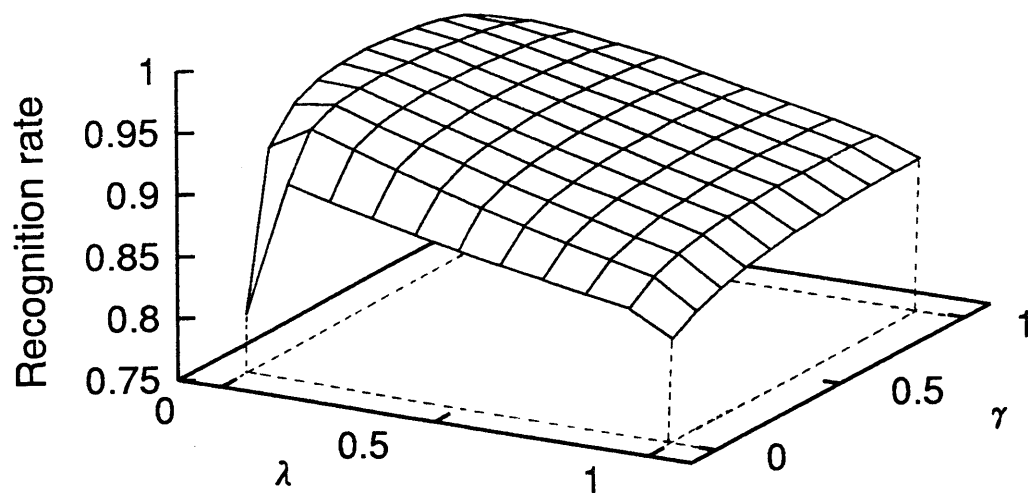


図 5. 2 平均認識率 ($R = 1/3$)

$N=10, M=15, \epsilon_B=0.01$

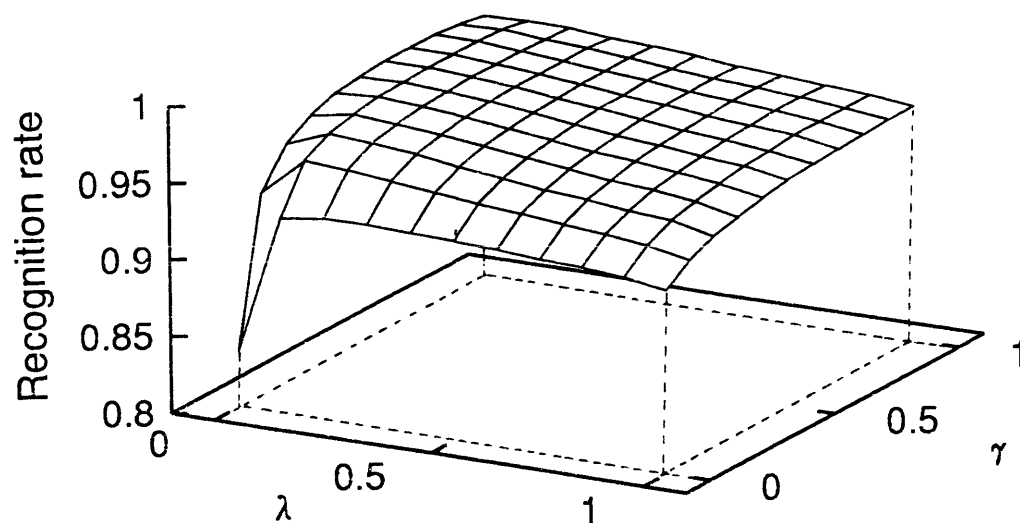


図 5. 3 平均認識率 ($R = 2 / 3$)

$N=10, M=15, \epsilon_B=0.01$

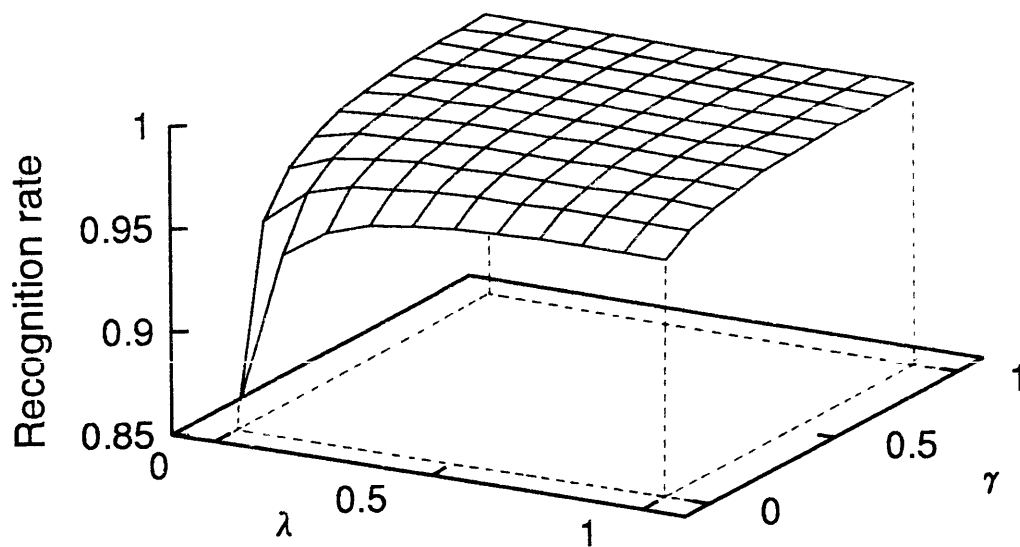


図 5. 4 平均認識率 ($R = 1$)

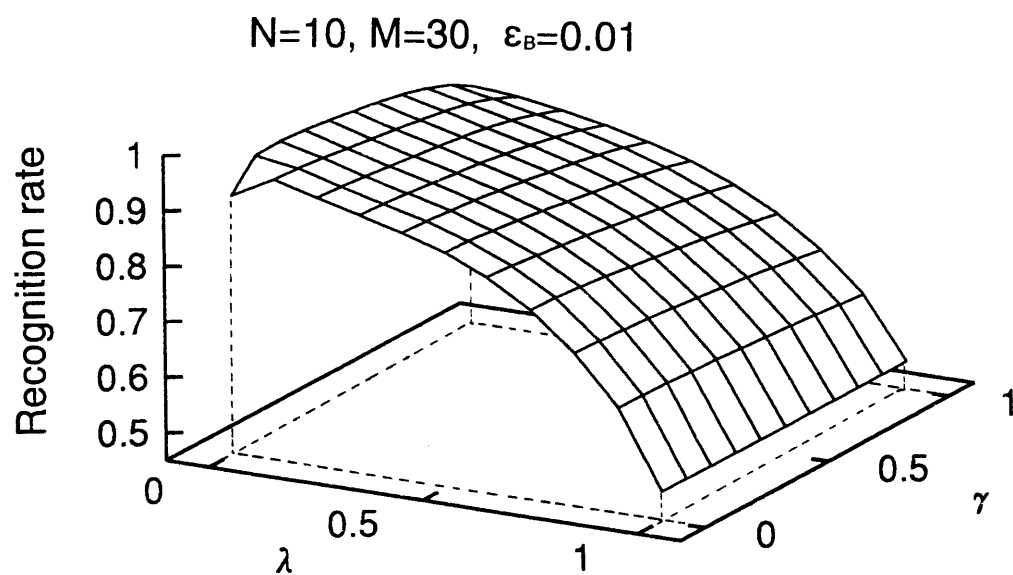


図 5. 5 平均認識率 ($R = 0$)

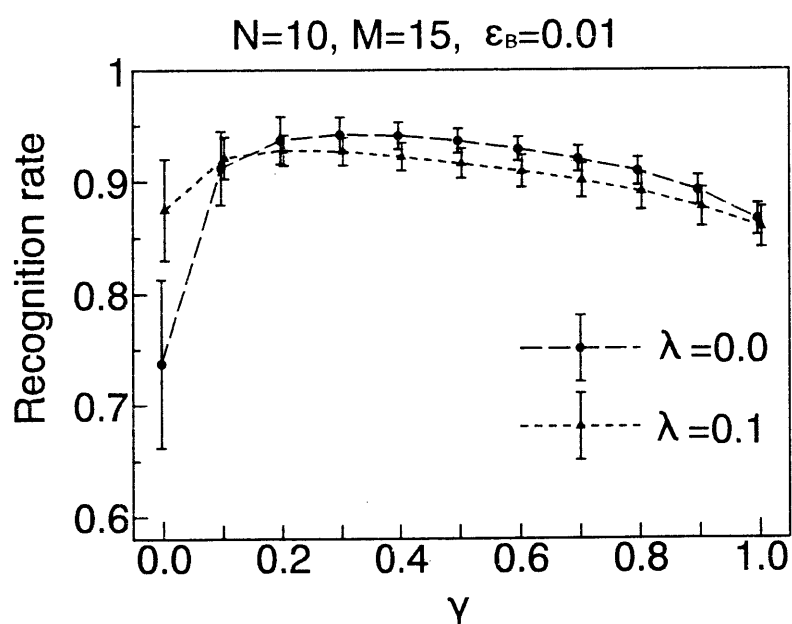


図 5. 6 γ の変化と平均認識率 ($R = 0$)

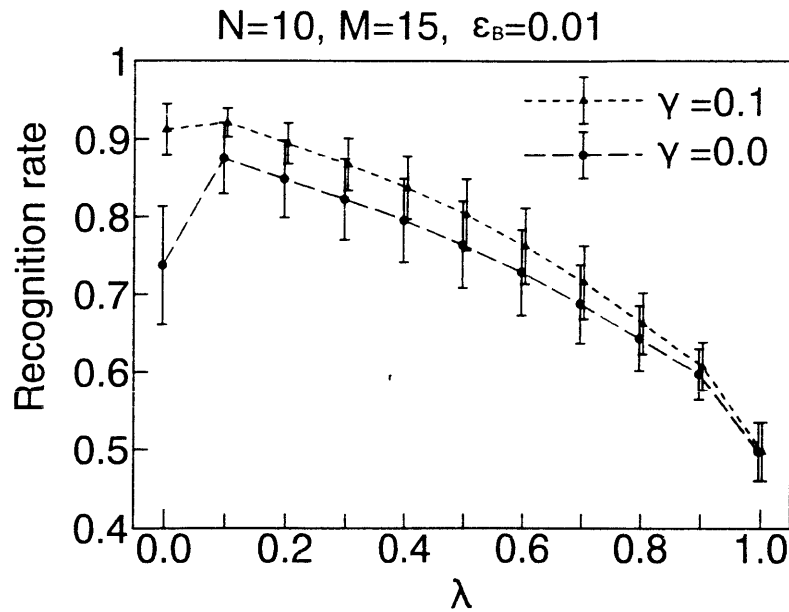


図5.7 λ の変化と平均認識率 ($R = 0$)

が劣化する $\gamma = 0$ の場合を除き, $\gamma \geq 0.2$ の範囲では $\lambda = 0$ の方が $\lambda = 0.1$ に比べて高い認識性能を示している. また, 高い認識率を示すところでは認識率の標準偏差も小さい. $\lambda = 0$ の場合のグラフは $\gamma \geq 0.2$ の範囲では γ の変化に対して認識率の変動が比較的少ない. 他の λ についてもおおむね, 同様の傾向がみられる. 従って, λ の値の方が認識率に対する寄与が大きい.

図5.7のグラフの例からも知られるように, λ の変化につれて認識率も大きく変動しており, (λ, γ) に対して予め最適な条件を見つけ出すことは必ずしも容易でないことが示唆されている.

5.4 数字認識実験

RDA法の具体的な効果を調べるため, 手書き数字認識への応用実験を行った. ここでは, より現実的な条件で実験するためシミュレーションと異なり, 2クラスの認識実験ではなく, 0から9までの数字, 10クラスを対象とする認識実験を行った.

5.4.1 実験方法

認識実験には電子技術総合研究所作成の手書き文字データベースETL-1⁽⁸⁾を用

いた。ETL-1は、1445人の筆記者により書かれた数字、アルファベット、カタカナが格納されているが、この中の数字のみを対象として認識実験を行った。

認識に使用した特徴抽出方法は、テストのための標本を数多く確保できるよう、比較的特徴パターンの次元数が小さくても認識性能が得られやすい「加重方向指数ヒストグラム法」⁽⁹⁾によった。実験は分割法により、学習標本とテスト標本とが重複しないよう選んだ。実験結果のばらつきを考慮し、学習標本の選び方を変え、複数回実験を繰り返した。実験に用いた諸パラメータを以下に記す。

- ・クラス数 $g = 10$ (数字0～9)
- ・特徴パターンの次元数 $N = 64$
- ・学習標本数 $M = 1.5N \sim 5N / \text{クラス}$
- ・テスト標本数 $= 1,000 \text{ 字} / \text{クラス}$
- ・パラメータ (λ, γ) ; 共に $[0, 1]$ で0.1間隔で変える。
- ・繰り返し回数 ; 10回

式(5.2)に基づく2次識別関数では2クラスの識別しかできないため、式(5.2)の2次識別関数をクラス毎に分離し、特徴パターン \mathbf{x} に対するクラス毎の距離 $\hat{g}_i(\mathbf{x})$ を次式により求めた。

$$\hat{g}_i(\mathbf{x}) = (\mathbf{x} - \hat{\boldsymbol{\mu}}_i)^t \hat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_i) - \ln |\hat{\boldsymbol{\Sigma}}_i| \quad (5.18)$$

式(5.18)に基づく距離値が最も小さいクラスの数字を判別結果として判定し、棄却処理はしていない。1回の実験につき、1,000字×10クラス=10,000字の数字認識を行い、認識率を求めた。これを10回繰り返し、その平均と分散を求め、各条件における認識率とその分散とした。

5.4.2 実験結果

前記組み合わせの中で、最高の認識率が得られた条件は、

- ・ $M = 5N (= 320)$,
- ・ $(\lambda, \gamma) = (0.0, 0.1)$,

のときであり、平均認識率は93.9%，その標準偏差は0.8%であった。この条件での認識率をクラス毎に見ると、最も認識率の良かった数字は「2」であり、平均認識率は96.7%，逆に最も認識率の悪かった数字は「6」であり、平均認識率は

第5章 統計的パターン分類器の改良

83.0%であった。以下図5.8～5.11に実験結果の一部を示す。図5.8, 5.9はRDA法の二つのパラメータ(λ , γ)をx軸, y軸にとり, z軸(高さ方法)が平均認識率を示す。

実験結果より, 次のことが知られる。

- ①シミュレーション実験の場合と同じように, 従来の2次識別関数, $\lambda = 0$, $\gamma = 0$ のとき, 認識率が著しく低下している。
- ②シミュレーション実験と異なり, λ の増加に伴う認識率の顕著な低下の傾向は認められない。 $\gamma = 1$ の点における認識率の低下が著しい。このことは, 実際の特徴パターンにおいては, 各要素毎に分散が大きくばらついており, シミュレーション実験の際に設定した, 式(5.13)とは異なることを示唆している。
- ③RDA法の二つのパラメータのうち λ に関して, 学習標本の増加に伴い, 最適値は0に漸近する。また, 実験した範囲内では, γ の最適値が0になることはなく, 必ず0より大きな数値であった。

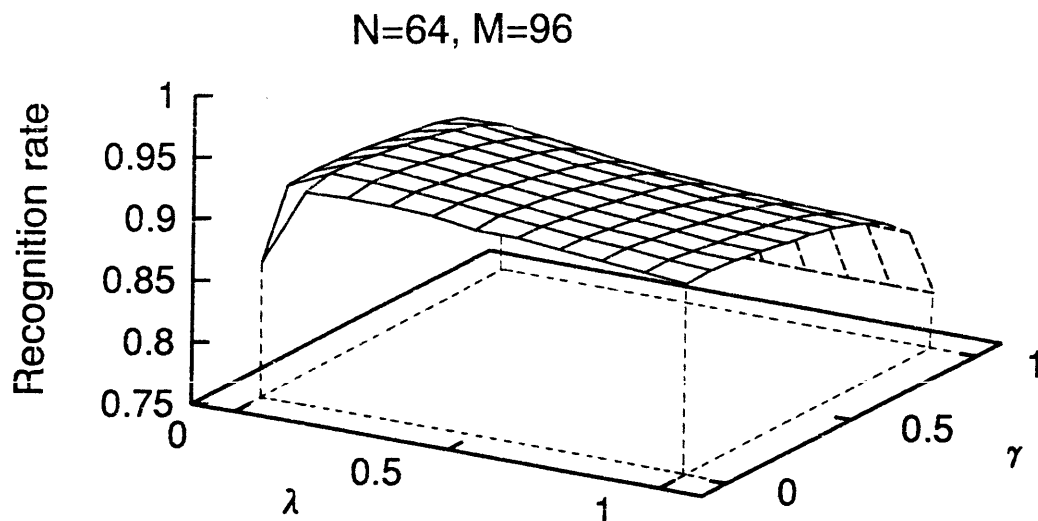


図5.8 平均認識率

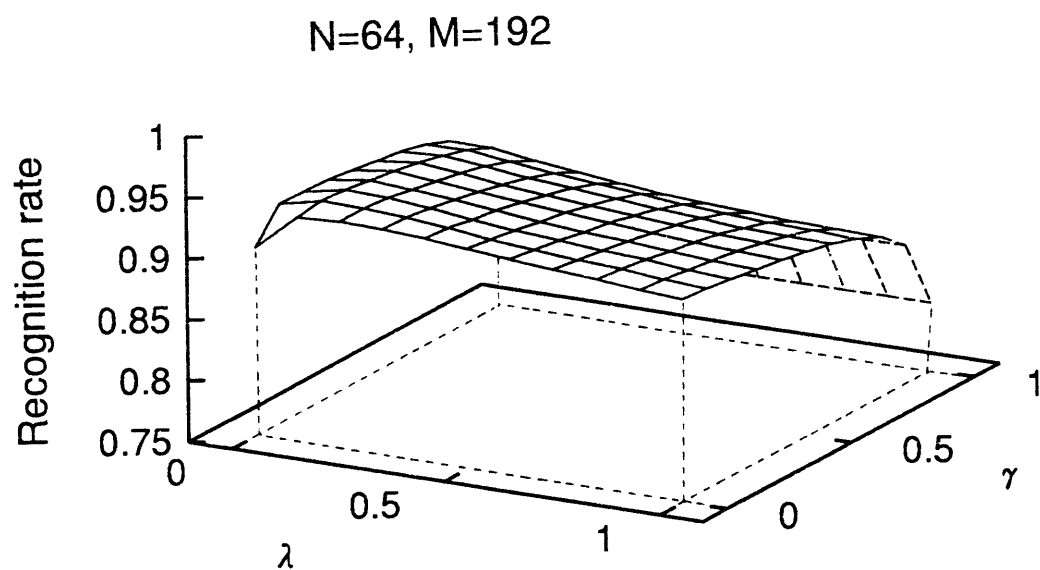


図 5. 9 平均認識率

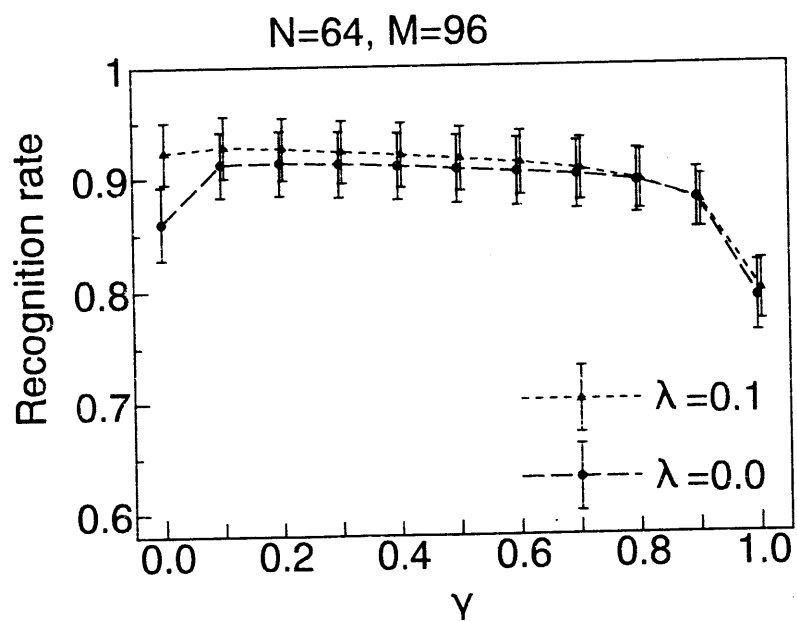


図 5. 1 0 γ の変化と平均認識率

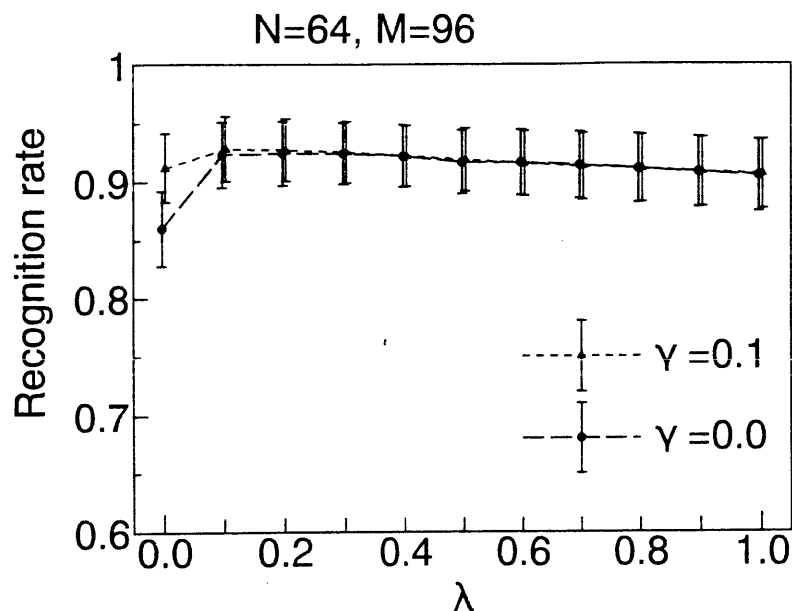


図 5.11 λ の変化と平均認識率

5.5 むすび

学習標本数が十分に確保できない場合の標本共分散行列のロバスト性を増すため Friedman が提案した RDA 法について、有限個の学習標本から識別関数を設計し、テスト標本数の影響を受けないよう、理論的な誤り率を数値積分により求めて、その効果を調べた。また、RDA 法の二つのパラメータの値の効果をシミュレーションによって調べ、比較した。RDA 法の振る舞いが母集団の真のパラメータに強く依存することから、二つの母集団のパラメータをベイズ誤り率と Bhattacharyya 距離の第 1 成分と第 2 成分の比に着目し、その比率を一定にしてコンピュータシミュレーションを実施した。さらに、具体的な RDA 法の効果を調べるため数字認識実験を行った。その結果、次のようなことが明らかになった。

- ① 母集団の性質によらず、2 次識別関数は極めて学習標本数に敏感な識別関数である。その主たる要因は標本共分散行列の推定精度が悪いことであり、学習標本の欠如に対して RDA 法はある程度有効な解決手段を提供している。
- ② RDA 法は、二つのパラメータ (λ , γ) を選択する必要があるが、二つの母集団

の共分散行列が等しく線形識別関数を用いるとき以外では、 λ は0付近でよい。

③RDA法のパラメータ γ の最適値は学習標本数の逆数にほぼ比例する。

④ (λ, γ) の最適値を予め設定しておくことは必ずしも容易ではない。

⑤RDA法を実際の多クラスの数字認識に応用しても、認識率の向上に効果がある。

⑥具体的なパターン認識に応用するときクラス数が多いことも影響し、RDA法の二つのパラメータ (λ, γ) と平均認識率は複雑な振る舞いを示す。

今後の課題としては、母集団のパラメータに依存しない一般的な条件下で、識別性能を改善するパラメータ γ と学習標本数との関係を見いだすこと、学習標本数が極端に少なく標本共分散行列が正則行列にならない場合における統計的識別関数の構成法、2クラスの分類にとどまらないクラス数が増加したときの影響などを調べることがある。

文 献

- (1)Friedman J. H. : "Regularized Discriminant Analysis", Journal of American Statistical Association, 84, No. 405, pp. 165-175(1989)
- (2)Rayens W. and Greene T. : "Covariance pooling and stabilization for classification", Computational Statistics & Data Analysis, 11, pp. 17-42(1991)
- (3)Aeberhard S., Coomans D. and Oliver D. V. : "Comparative Analysis of Statistical Pattern Recognition Methods in High Dimensional Settings", 27, No. 8, pp. 1065-1077(1994)
- (4)Bhattacharyya, A. : "On a measure of divergence between two statistical populations defined by their probability distributions", Bull. Calcutta Math. Soc., 35, pp. 99-110(1943)
- (5)James, W. and Stein, C. : "Estimation with quadratic loss", Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Jerzy Neyman, Ed. Vol. I, pp. 361-379(1961)
- (6)Dey, D. K. and Srinivasan, C. : "Estimation of a covariance matrix under Stein's loss", Ann. Statist., Vol. 13, pp. 1581-91(1985)

第5章 統計的パターン分類器の改良

(7) Takeshita, T., Toriwaki, J. : "Experimental study of performance of pattern classifiers and the size of design samples", Pattern Recognition letters, Vol. 16, No. 3, pp. 307-312(1995)

(8) 山田博三, 森俊二 : "手書き文字データベースの解析 (I)", 電総研彙報, Vol. 39, No. 8, pp. 580-599(1975)

(9) Kimura, F., Takashina, K., Tsuruoka, S., Miyake, Y. : "Modified quadratic discriminant functions and the application to Chinese character recognition", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-9, pp. 149-153 (1987)

第6章 むすび

本章では、本論文の内容とその成果について要約するとともに、学習標本数に制限があるパターン認識システムの設計・構築の際、本研究の成果がどのように適用可能かを論ずる。最後に今後の展望と課題について述べる。

6.1 本研究のまとめ

本論文では、有限学習標本に基づく統計的パターン分類器の設計と評価について論じた。

第1章では、研究の意義、および、数学的基礎となる多次元正規分布と統計的識別関数について述べた。

第2章では、2次識別関数の主要な要素を占めるマハラノビス距離の推定誤差について、Steinが示した単純標本平均に関するパラドックス「3次元以上のパターンでは、単純平均が平均二乗誤差最小という基準のもとでは最良の推定を与えない」という現象についてシミュレーション実験により確認した。またSteinの提案した平均値ベクトルについての補正を施すことにより、平均値ベクトルの推定精度が向上するのみでなくマハラノビス距離の推定精度も向上することを見いだした。

第3章では、まず、マハラノビス距離の推定値の確率分布について検討し、マハラノビス距離の推定誤差に影響を与える要因として、平均値ベクトルの推定誤差と共分散行列の推定誤差があることを見いだした。次に、マハラノビス距離の推定値の確率分布を明らかにし、それを二つの成分に分解できることを示した後、各々の成分の平均値と分散に関する理論式を導出し、マハラノビス距離の推定誤差に共分散行列の推定誤差が大きな影響を与えていることを明らかにした。そして、これらの結果に基づいて識別関数の改良についての示唆を得た。

第4章では、まず有限学習標本に基づいて設計された線形識別関数や2次識別関数の振る舞いに関して、テスト標本を用いず、理論的に平均誤り率を算出する方法を見だし、ベイズ誤り率からの性能低下を定量的に評価した。ついで、特徴パターンの次元数、学習標本数、および平均誤り率三者の間の関係をRaudysおよびFukunagaらの近似誤り率（理論式）と比較し、それらの近似理論式の適用限界を明らかにした。さ

第6章 むすび

らに、ベイズ誤り率と標本に基づく識別関数が示す平均誤り率との差の分布に関する評価を行い、ベイズ誤り率が低くなる（二つの母集団がより離れる）ほど、また特徴パターンの次元数が大きくなるに従い、認識性能の低下分の分布は χ^2 分布に近くなることを見いだした。

第5章では、まず準備として、二つの母集団の組み合わせについて、より一般的に取り扱うことができるよう、二つの母集団間の距離をBhattacharyya距離により計測した。Bhattacharyya距離は、平均値ベクトルの差に基づく成分と共分散行列の差に基づく成分からなっているので、二つの成分の比を一定に保ちながらBhattacharyya距離を増減し、所定のベイズ誤り率となるよう母集団パラメータを設定した。これにより同一のベイズ誤り率を有しながら、広い範囲の母集団の組み合わせに対応できるようになり、識別関数の能力を特定の分布に偏らないで、幅広く評価できるようになった。

次に、この方法に基づくシミュレーションによって、学習標本の有限性に基づく識別性能の低下に対してロバスト性を増す目的で考えられたRDA法は、2つの自由度を持つパラメータがあるが、そのうちの一つは性能の改善に余り寄与していないことを見いだした。しかしながら、2次識別関数は学習標本数の減少に対して敏感に反応し、認識性能が極端に低下することが明らかになった。そして、これに対処する方法としてRDA法は優れた特性を有していることを見いだした。

6.2 今後の展望と課題

本文の第4章と第5章では、識別関数の性能を評価する理論式から数値積分により、学習標本1セット毎の誤り率を求めている。さらにこれを繰り返すことで誤り率の平均値や分散を求めているためかなりの計算量となっている。識別関数の確率分布は常に二つの距離尺度の差に基づく分布となっていることから、マハラノビス距離の推定値の確率分布を求める手法がより一般化できれば、誤り率そのものの確率分布を求められる可能性がある。誤り率自身の確率分布を明らかにできれば、多少複雑な分布の形をとったとしても現在行っているシミュレーション実験によらなくても誤り率の平均値や分散を求めることができるようになり、計算コストを大幅に減らすことができるようになる。この点については今後の課題としたい。

統計的パターン分類器を用いる場合、一般のパターン認識においては二つの母集団

の共分散行列が等しいか否かということは知られていない。このため、共分散行列についての特別の前提知識を有しないときは2次識別関数を用いるべきである。ところが、2次識別関数は学習標本の不足に敏感な分類器であるため、線形識別関数を用いた方がよい場合もあると思われる。RDA法はそのような問題に対する一つの解決を与えるために提案されたものとも見られるが、第5章で明らかになったようにRDA法のパラメータを決定する簡単な方法は見いだされていない。むしろ、特徴パターンの次元数と学習標本数と母集団のパラメータの三者の関係から、線形識別関数か2次識別関数のどちらを用いるかを決定する方が現実的であると思われる。これは、RDA法の一つのパラメータ λ を0か1にすることに相当する。前記三者の関係からパラメータ λ を0か1に決めることができれば、その方がRDA法のように小刻みにパラメータを調節するより現実的な対処の方法であろう。この後、他のパラメータ γ の最適値と学習標本数と特徴パターンの次元数の間の関係を明らかにすることは今後の課題である。

特徴パターンの次元数が大きく、それに対して学習標本が十分に確保できない状況におけるパターン分類器のロバスト性については、統計的パターン分類器に比較して、ニューラルネットワークシステムの方がより優れているという報告もある⁽¹⁾。ニューラルネットワークシステムについてはネットワーク構成、ネットワークの初期状態の設定や学習の際のパラメータ等、理論的に未解明の部分が多く、どのようなアプローチを用いるかという基本的な部分からの研究が必要であるが、より大きな研究の課題ととらえ、今後機会をとらえ挑戦したい。

最後に、特徴パターンの次元数と学習標本数と平均認識率の間の関係を明確にしたいと考えた最初の動機は、特徴抽出方法の異なるパターン認識システムの性能比較ができないか考えたことにある。音声認識や文字認識を行うシステムは数多く製品化されているが、それらの製品の性能比較はやさしいことではない。各認識システムは対象パターンに固有な処理を行う観測系や前処理系、決定に有効な特徴を抽出・選択する特徴抽出系を経て、多次元の特徴パターンに変換される。1.2節の研究の意義で述べたように文字や音声を認識するための特徴抽出系は、そのみで比較検討が行われることはまれである。特徴抽出系の比較検討を妨げている主要な要因の一つに認識標本の共通性の問題がある。これについては、音声認識や文字認識の分野ではデータベースが整備され、問題点は次第に解消されつつある。しかしながら、決定・分類系の

第6章 むすび

比較は特徴パターンの次元数が異なることが障害となっていなかった。第4章で明らかにできた特徴パターンの次元数と学習標本数と平均認識率の関係を用い、具体的な認識システムにおける特徴抽出系の性能比較を行うことも今後の一つの目標である。幸い、筆者が属する文字認識用プログラムパッケージの開発グループには文字認識に関して、何種類かの特徴抽出アルゴリズムが集められ、動作確認がなされている。特徴パターンの分布が実際に正規分布を仮定してよいかどうか、十分注意を払う必要があるが、国の内外における主要な文字認識用特徴抽出アルゴリズムの性能比較を行いたいと考えている。

文 献

(1) Raudys, S. : " Why do multilayer perceptrons have favorable small sample properties?", Proc. of Pattern Recognition in Practice IV, pp. 287-298, (1994)

謝辞

本研究は筆者が豊田工業高等専門学校に赴任して以来、約10年間の研究についてまとめたものである。本研究を進めるにあたり、名古屋大学工学部情報工学科教授鳥脇純一郎先生には懇切丁寧なご指導ご助言を賜りました。先生の叱咤激励なしには、本研究をこのような形にまとめることは不可能なことでありました。深甚なる感謝の意を表する次第です。

論文の審査にあたり、名古屋大学工学部電子情報学科教授板倉文忠先生ならびに情報工学科教授大西昇先生には多忙な中、貴重な時間を割いて、多くのご指摘、ご助言を頂きました。ここに記して深く感謝の意を表します。

豊田工業高等専門学校前校長堀井憲爾先生には研究の姿勢やあり方を教えていただくと同時に、豊田工業高等専門学校における筆者の研究遂行に必要な種々のご配慮を賜った。歴代の豊田工業高等専門学校校長、市川真人先生、岩田幸二先生には筆者が企業人から教員へと、それまでの開発設計から教育研究の生活への環境の変化に対応できるよう種々のご配慮を賜ると同時に温かい激励を戴いた。深く感謝の意を表する次第です。

筆者の恩師である豊田工業高等専門学校情報工学科教授野澤繁之先生には学生時代は言うに及ばず、筆者が豊田工業高等専門学校電気工学科に着任して以来、教育に関する手ほどきからその醍醐味、コンピュータの素晴らしさと研究の哲学に至るまで、公私両面にわたって、幅広くご指導ご助言を賜りました。ここに記して深く感謝の意を表する次第です。

三重大学工学部情報工学科教授三宅康二先生、同助教授木村文隆先生、同電気電子工学科助教授鶴岡信治先生には、筆者が文部省内地研究員として同学部に滞在した昭和60年5月から10カ月の長きにわたって、筆者の実質的なパターン認識研究のスタートの場を与えていただいた。同時に手書き文字認識システムに関する知識を数多くお教えいただいた。さらに、中部大学経営情報学部経営情報学科教授吉村ミツ先生には研究の進め方についてご指導ご助言を賜った。豊田工業高等専門学校情報工学科教授岡部直木先生には先生の名古屋大学大型計算機センター在職中から筆者の計算機利用に関するお世話をしていただいた。ここに記して深く感謝の意を表します。

豊田工業高等専門学校教官・職員の方々には筆者が研究をすすめる上でいろいろなご配慮、ご指導、ご助言をいただいた。同時に豊田工業高等専門学校の電気工学科・情報工学科の卒業研究の学生諸君にはコンピュータシミュレーション、ならびにその実験結果の整理をしてもらった。ここに記し感謝の意を表する次第である。

シミュレーションには名古屋大学大型計算センターならびに豊田工業高等専門学校学内共同利用データステーションを利用した。また、シミュレーションの一部には名古屋大学大型計算センター数値計算ライブラリーNUMPAC、および、同手書き文字認識用プログラムパッケージSMARTを利用させていただいた。

最後に、内藤科学技術振興財団には第2章のコンピュータシミュレーションのため、堀情報科学財団には第4章ならびに第5章のコンピュータシミュレーションのための計算機使用料を中心に研究費の一部を助成していただいた。ここに記し感謝の意を表する次第である。