

固有ベクトルの変動を考慮した共分散行列 Covariance Matrix rectified by considering Eigenvector Variances

酒井 充[†]
Mitsuru Sakai

米田 政明[†]
Masaaki Yoneda

長谷 博行[‡]
Hiroyuki Hase

丸山 博[†]
Hiroshi Maruyama

1. はじめに

本論文では共分散行列の推定法として標本共分散行列を二重に補正する方法を提案し、それが学習パターンが少ない場合にも有効であることを2次識別関数に適用して検証する。提案する共分散行列の推定法は分布に依存しないが、検証にはクラス内分布が正規分布であり学習サンプル数が等しい場合を扱う。

正規分布にはベイズアプローチによる Keehn の方法 [1] や Geisser の方法 [2],[3] が有効であるが、Keehn の方法では本手法と異なり学習パターンからは決定できないパラメータがある。Geisser の方法については認識実験により比較を行う。

標本共分散行列を修正する方法には修正2次識別関数 [4] や Regularized Discriminant Analysis [5], James と Stein の方法 [6] があるが、修正2次識別関数と RDA では学習パターンからは決定できないパラメータがある。James と Stein の推定量は今回提案する共分散行列のもとになった補正共分散行列とほぼ等しい性能を持つことが分かっている [7],[8]。

補正共分散行列は固有値の偏りの平均的な大きさに基づき標本共分散行列の固有値 (以降、標本固有値 (sample eigenvalue) と呼ぶ) を補正して得られるが、本論文では、さらに標本共分散行列の固有ベクトル (以降、標本固有ベクトル (sample eigenvector) と呼ぶ) の変動に着目することにより、固有値の新しい補正量を示す。そして、この補正を補正共分散行列に適用して得られる共分散行列、すなわち二重に補正された共分散行列を提案する。

以下、2. では、二重補正共分散行列について提案する。

3. では、2次元正規分布に対する二重補正共分散行列の求め方を示す。4. では、二重補正共分散行列を採用した2次識別関数を、2次元正規分布特徴を持つ2クラス認識問題に適用し他の識別関数と認識性能を比較することにより、その有効性を示す。

2. 固有値の新しい補正量と二重補正共分散行列

2.1 母分布のぼけ

2次元正規分布の例を元に説明する。母正規分布を $N(\mu, \Sigma)$ とする。議論を簡単にするため、 $\Sigma = \text{diag}(\lambda_1, \lambda_2)$, ($\lambda_1 \geq \lambda_2$) とし、 Φ は単位行列とする。図1(a)において、 $\tilde{\Phi}_i$, ($1 \leq i \leq 3$) で示した分布は正規分布 $N(\tilde{\mu}, \tilde{\Sigma})$ の散布例であり、その標本平均ベクトルは $\tilde{\mu}_i$, 標本共分散行列は $\tilde{\Sigma}_i$ である。 Φ は Σ の、 $\tilde{\Phi}_i$ は $\tilde{\Sigma}_i$ の固有ベクトル行列である。太い実線の楕円が真の分布を表す。

(b) は (a) の各分布の中心を揃えたものであり、 $\tilde{\Sigma}$ の散布例である。その期待値は真のもの Σ と一致する。

(c) は (b) の各分布の長軸を揃えたものであり、標本固有値の散布例である。破線はこれらの期待値、すなわち、分布の形の平均を表しており、真のものより細長い形になる。標本共分散行列をこの破線と考え、真の形の太い実線に補正することにより補正共分散行列が得られる [8]。

(d) は (c) と同様に各分布の長軸を揃えたものであり、そのとき、母分布は各学習パターン集合ごとに $\tilde{\Phi}_i^t$ だけ回転して観測される。ここで $(\cdot)^t$ は行列の転置である。そのため母共分散行列は $\tilde{\Phi}_i^t \Sigma \tilde{\Phi}_i$ のように観測されることになる。破線はこのように回転して観測された母分布の形の期待値であり、母分布より丸い形になる。

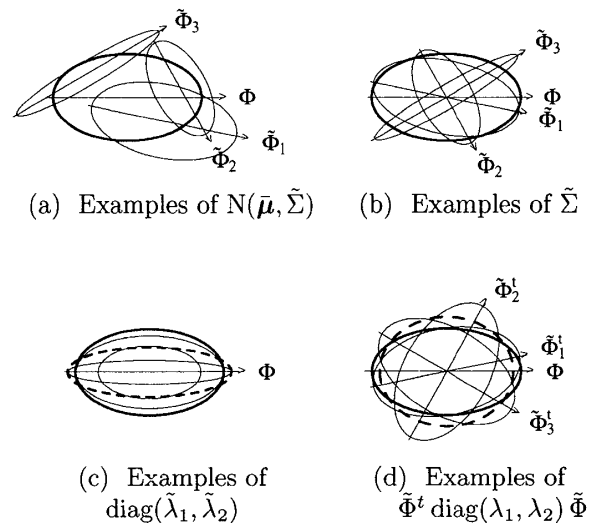


図1: 学習サンプルによる分布, 標本共分散行列, 標本固有値, 母共分散行列のぼけの例

このように標本固有ベクトルが母固有ベクトルから傾くことによって、逆に母分布の形はぼけて見えることになる。筆者らはこのぼけた分布の共分散行列 $\tilde{\Sigma}$ を推定し、認識に使用する。

以下にこのことを式を用いて議論する。

ここでは図1(d)の破線で示されるような分布の共分散行列を求める。そこで母分布を原点移動した正規分布を $f(x|\Sigma)$ と表す。

本論文では標本固有ベクトル $\tilde{\Phi}$ を母固有ベクトルの推定量として用いるものとする。そのため、この標本固有ベクトルの座標系からは母分布は $f(x|\tilde{\Phi}^t \Sigma \tilde{\Phi})$ なる分

[†]富山大学工学部, Faculty of Engineering, Toyama University
[‡]福井大学工学部, Faculty of Engineering, Fukui University

布として観測される[§]. よって, 標本固有ベクトルの座標系から観測される母分布の平均的な分布 $f_b(\mathbf{x}|\check{\Sigma})$ は次式となる.

$$f_b(\mathbf{x}|\check{\Sigma}) = \int f(\mathbf{x}|\check{\Phi}^t \Sigma \check{\Phi}) p(\check{\Phi}) d\check{\Phi} \quad (1)$$

ここで $\check{\Sigma}$ は $f_b(\mathbf{x}|\check{\Sigma})$ の共分散行列であり, $p(\check{\Phi})$ は $\check{\Phi}$ の密度関数を表す. $f_b(\mathbf{x}|\check{\Sigma})$ の共分散行列は次式となる.

$$\begin{aligned} \check{\Sigma} &= \int \mathbf{x} \mathbf{x}^t \int f(\mathbf{x}|\check{\Phi}^t \Sigma \check{\Phi}) p(\check{\Phi}) d\check{\Phi} d\mathbf{x} \\ &= \int \check{\Phi}^t \Sigma \check{\Phi} p(\check{\Phi}) d\check{\Phi} \end{aligned} \quad (2)$$

また $\text{tr}(\check{\Sigma}) = \text{tr}(\check{\Phi}^t \Sigma \check{\Phi})$ となることから, $\check{\Sigma}$ の固有値を $\check{\lambda}_1, \dots, \check{\lambda}_d$, ($\check{\lambda}_1 \geq \dots \geq \check{\lambda}_d$) とすれば, 次式が成り立つ. ここで $\text{tr}(\cdot)$ は行列のトレースである.

$$\sum_{i=1}^d \check{\lambda}_i = \sum_{i=1}^d \lambda_i \quad (3)$$

この共分散行列 $\check{\Sigma}$ による分布が Σ の分布をぼかしたものであることは直感的に明らかである. 証明は省略するが, 次式が成り立つ.

$$\check{\lambda}_1 < \lambda_1, \quad \check{\lambda}_d > \lambda_d \quad (4)$$

すなわち, $\check{\Sigma}$ の最大固有値は母共分散行列のものより小さく, 最小固有値は母共分散行列のものより大きくなり, ぼけることが分かる. そのため, 以降この共分散行列 $\check{\Sigma}$ をぼけ共分散行列と呼ぶ.

一般によく用いられる多変量正規分布の場合には, Σ が対角行列であれば $\check{\Sigma}$ も対角行列となる. すなわち, ぼけの程度は固有値のみに反映されることになる.

2.2 二重補正共分散行列

標本共分散行列からぼけ共分散行列を推定する方法は以下に示すように2段階に分けられる. まず, 文献 [7] に従って標本固有値の偏りを補正して母固有値の推定量とし, それをもとに母共分散行列の推定量を求める. 次に, 母共分散行列の推定量を式 (2) に代入してぼけ共分散行列の推定量を求める. 母分布として多変量正規分布が仮定できれば母固有値の推定量を式 (2) に適用してぼかした固有値を求めればよい. 以降, このようにして推定した共分散行列を二重補正共分散行列と呼ぶ.

3. 2次元正規分布の二重補正共分散行列

2次元正規分布の標本固有値の期待値及びぼけ共分散行列の固有値を示す. 4. の認識実験では, これらの結果を用いて二重補正共分散行列を正確に計算している.

3.1 標本固有値の期待値

2次元正規分布の共分散行列は一般性を失わずに $\Sigma = \text{diag}(\lambda_1, \lambda_2)$, ($\lambda_1 \geq \lambda_2$) とする. 標本共分散行列は学

[§]簡単のため, $\check{\Phi}$ を基底とするベクトル変数に対しても \mathbf{x} を用いる.

習パターン $\mathbf{x}_1, \dots, \mathbf{x}_N$ から $\check{\Sigma} = \sum_{i=1}^N (\mathbf{x}_i - \bar{\boldsymbol{\mu}})(\mathbf{x}_i - \bar{\boldsymbol{\mu}})^t / (N-1)$ で計算される. $\check{\Sigma}$ の固有値を $\check{\lambda}_1, \check{\lambda}_2$, ($\check{\lambda}_1 \geq \check{\lambda}_2$) と表し, $\check{\lambda}_1$ と $\check{\lambda}_2$ の平均をそれぞれ $\bar{\check{\lambda}}_1, \bar{\check{\lambda}}_2$ と表す.

図2には, $\lambda_1 + \lambda_2 = 2$ の場合における, いくつかの学習パターン数 N に対する標本固有値 $\check{\lambda}_2$ の期待値 $\bar{\check{\lambda}}_2$ を示してある [7].

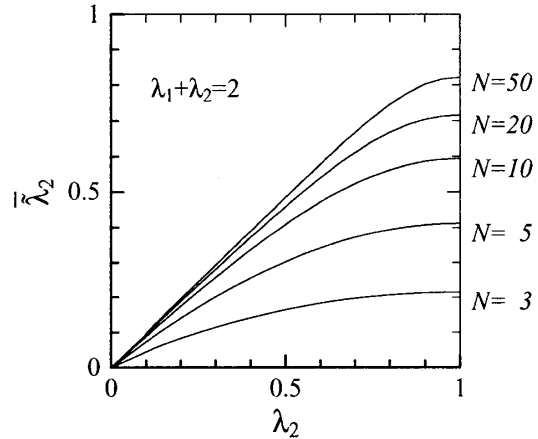


図2: λ_2 と $\bar{\check{\lambda}}_2$ の関係 ($\lambda_1 + \lambda_2 = 2$ の場合)

このとき $\bar{\check{\lambda}}_1$ は次の関係式から求めることができる.

$$\bar{\check{\lambda}}_1 + \bar{\check{\lambda}}_2 = \lambda_1 + \lambda_2 \quad (5)$$

3.2 標本固有値から母固有値の推定

学習パターン $\mathbf{x}_1, \dots, \mathbf{x}_N$ から得られる標本固有値 $\check{\lambda}_1, \check{\lambda}_2$ から母固有値の推定量 λ^*_1, λ^*_2 を求める方法を示す. ここで $\bar{\check{\lambda}}_2(\lambda_1, \lambda_2)$ は λ_1, λ_2 に対する $\bar{\check{\lambda}}_2$ の値を表す.

もし $\bar{\check{\lambda}}_2 < \bar{\check{\lambda}}_2((\check{\lambda}_1 + \check{\lambda}_2)/2, (\check{\lambda}_1 + \check{\lambda}_2)/2)$ であれば, 次式を同時に満たす λ^*_1, λ^*_2 の値を求める.

$$\begin{aligned} \bar{\check{\lambda}}_2 &= \bar{\check{\lambda}}_2(\lambda^*_1, \lambda^*_2) \\ \lambda^*_1 + \lambda^*_2 &= \check{\lambda}_1 + \check{\lambda}_2 \end{aligned} \quad (6)$$

認識実験では $\lambda^*_1 = \check{\lambda}_1 + \check{\lambda}_2 - \lambda^*_2$ とおき, Newton-Raphson 法により求めた.

もしそうでなければ次式と置く.

$$\lambda^*_1 = \lambda^*_2 = (\check{\lambda}_1 + \check{\lambda}_2)/2 \quad (7)$$

よって, 補正共分散行列 Σ^* は次式により求められる.

$$\Sigma^* = \check{\Phi} \text{diag}(\lambda^*_1, \lambda^*_2) \check{\Phi}^t \quad (8)$$

3.3 ぼけ共分散行列の固有値

ぼけ共分散行列は式 (2) から求められる.

$\lambda_1 + \lambda_2 = 2$ の場合におけるぼけ固有値 $\check{\lambda}_2$ の値は, いくつかの学習パターン数 N に対して, 図3のように求まる.

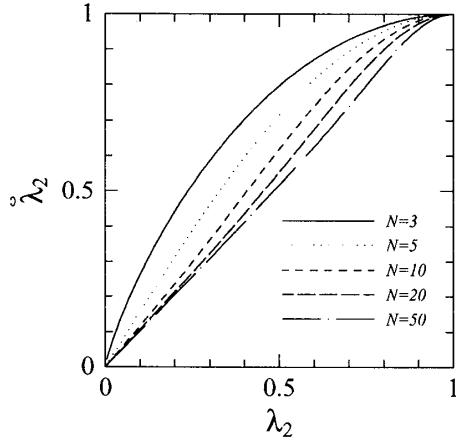


図 3: λ_2 と $\check{\lambda}_2$ の関係 ($\lambda_1 + \lambda_2 = 2$ の場合)

3.4 二重補正共分散行列

標本共分散行列 $\hat{\Sigma} = \hat{\Phi} \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2) \hat{\Phi}^t$ の固有値に対して 3.2 で示した方法により母固有値の推定量 λ_1^*, λ_2^* が求められ、それらの推定量を 3.3 で求めた図 3 に適用し、ばけ固有値の推定量 $\check{\lambda}_1^*, \check{\lambda}_2^*$ が求まる。そこで二重補正共分散行列 $\check{\Sigma}^*$ は次式で与えられる。

$$\check{\Sigma}^* = \check{\Phi} \text{diag}(\check{\lambda}_1^*, \check{\lambda}_2^*) \check{\Phi}^t \quad (9)$$

4. 認識実験

二重補正共分散行列を採用した 2 次識別関数を 2 次元正規分布特徴を持つ 2 クラス認識問題に適用し、いくつかの識別関数と比較検討する。クラス c_j , ($j = 1, 2$) の分布は平均ベクトルと共分散行列がそれぞれ μ_j, Σ_j の 2 変量正規分布に従うものとする。ここでは各クラスの生起確率 P_j は等しく、すなわち $P_1 = P_2 = 1/2$ とし、学習パターン数も等しく共に N 個とする。

4.1 比較に用いる識別関数

比較に用いる識別関数を以下に示す。

• QDF

よく知られているように 2 次識別関数 QDF は次式で表される。

$$\begin{aligned} H_j(x) &= (x - \mu_j)^t \Sigma_j^{-1} (x - \mu_j) + \log |\Sigma_j| \\ &\quad - 2 \log(P_j), \quad j = 1, 2 \\ QDF(x) &= H_1(x) - H_2(x) \end{aligned} \quad (10)$$

これは母平均、母共分散行列を使用しているため、最良の識別関数となる。グラフの最小値がこの誤り率を示す。

• UQDF

式 (10) の μ_j と Σ_j の代わりに、標本平均ベクトル $\bar{\mu}_j$ と標本共分散行列 $\hat{\Sigma}_j$ をそれぞれ用いる 2 次識別関数である。

• RQDF

文献 [7] で提案した 2 次識別関数であり、式 (10) の μ_j と Σ_j の代わりに、 $\bar{\mu}_j$ と式 (8) で求まる補正共分散行列

$\check{\Sigma}_j^*$ をそれぞれ用いる。

• WRQDF

本論文で提案している二重補正共分散行列を用いる 2 次識別関数であり、式 (10) の μ_j と Σ_j の代わりに、 $\bar{\mu}_j$ と式 (9) で求まる補正共分散行列 $\check{\Sigma}_j^*$ をそれぞれ用いる。

• GeisserDF

Geisser が無情報事前分布を用いるベイズ推定により導出した次式 [2],[3] を用いる。 $\hat{\Sigma}_j$ は最ゆう推定量の方の標本共分散行列である。

$$\begin{aligned} GeisserDF_j(x) &= N \log \left(1 + \frac{1}{N-1} (x - \bar{\mu}_j)^t \hat{\Sigma}_j^{-1} \right. \\ &\quad \left. \cdot (x - \bar{\mu}_j) \right) + \log |\hat{\Sigma}_j| \end{aligned} \quad (11)$$

4.2 認識対象

認識実験には以下の典型的な 4 種類のデータ集合を使用する。クラス c_1 は全データ集合で共通である。

$$c_1: \mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_1 = \text{diag}(1, 1)$$

データ集合 1 中心が異なる単位正規分布

$$c_2: \mu_2 = \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \Sigma_2 = \text{diag}(1, 1)$$

データ集合 2 同心円で大きさの異なる正規分布

$$c_2: \mu_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_2 = \text{diag}(9, 9)$$

データ集合 3 中心も形も異なる正規分布

$$\begin{aligned} c_2: \mu_2 &= \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \\ &\cdot \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \theta = \pi/4 \end{aligned}$$

データ集合 4 中心も形も異なる正規分布

$$\begin{aligned} c_2: \mu_2 &= \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \\ &\cdot \begin{pmatrix} 9 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \theta = \pi/4 \end{aligned}$$

4.3 実験方法

各クラスごとにそれぞれ N 個発生させて得られる学習パターンから分類器を設計し、学習パターンとは独立に発生させた各クラス 100 個の評価用パターンにより誤り率を求めた。それを 50000 回繰り返して行い、その誤り率の平均、すなわち平均誤り率 (mean error rate) を求めた。 $N = 3, 4, 5, 7, 10, 15, 20$ について実験を行った。

4.4 結果及び考察

データ集合 1 ~ 4 に対する認識実験の結果を図 4 ~ 7 に示す。各図は学習パターン数 N に対する平均誤り率を示している。全てのデータ集合 1 ~ 4 に対して、本論文で提案した WRQDF が、各 N において文献 [7] で提案した RQDF より明らかに良い性能を示している。このことは標本固有ベクトルの変動による母共分散行列のばけを考慮する考え方が正しいことを示唆している。

文献 [7] では RQDF が通常使用される UQDF より優れていたが、WRQDF はさらに優れていることは明らかである。

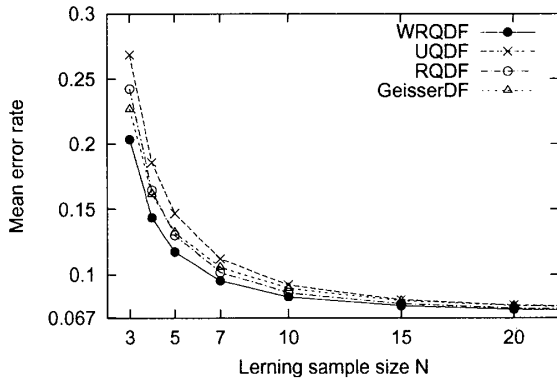


図 4: データ集合 1 における平均誤り率

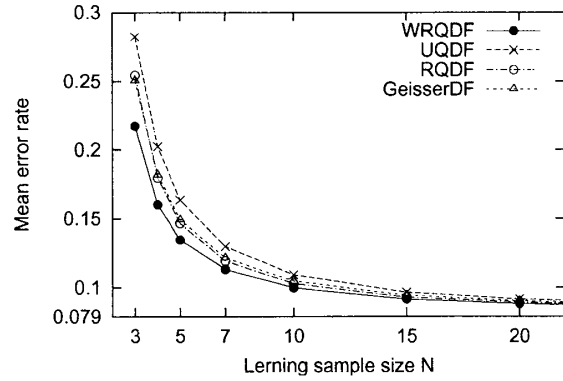


図 6: データ集合 3 における平均誤り率

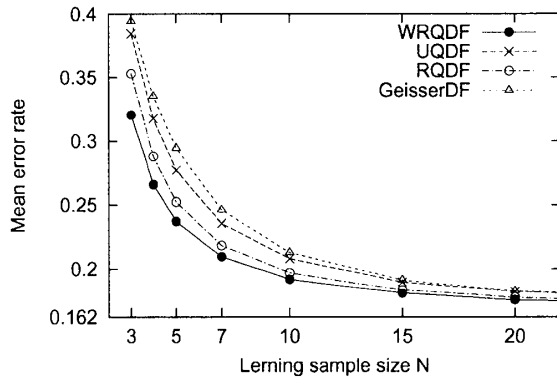


図 5: データ集合 2 における平均誤り率

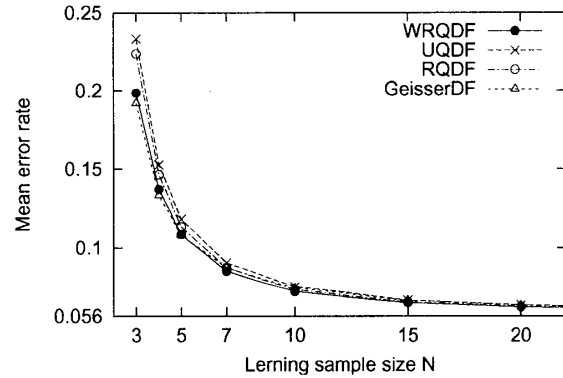


図 7: データ集合 4 における平均誤り率

ベイズ推定による式 (11) を用いる GeisserDF に対しては、データ集合 1～3 の場合には、提案手法の WRQDF の方が認識性能が高かった。逆に、データ集合 4 に対しては GeisserDF の方が高かった。よって、WRQDF は Geisser のベイズ推定による方式と同程度以上の性能があることが示唆される。

典型的な 4 種類のデータセットに対しては概ね上記の傾向が見られたが、更に詳細な検討が必要であると思われる。

5. おわりに

学習パターンが少ない場合には標本共分散行列はあまり適切な推定量ではなく、これまで小さな固有値を大きくするような認識方法が効果を上げてきた。我々は、以前に指摘した標本固有値の偏りに加えて、標本固有ベクトルの変動に起因する母分布のばけもその原因の一つであると考察し、これをもとに標本共分散行列を 2 段階で補正する二重補正共分散行列を提案した。そして、2 次元正規分布の分布モデルを用いた認識実験ではベイズ推定で得られる手法と同程度以上に優れているという結果が得られた。このことから、標本固有値の偏りと標本固有ベクトルの影響に関する我々の考察の正しさとそれに基づく提案手法の有効性が示された。

本論文では 2 次元正規分布の場合に有効性を検証したが、多次元の場合においても検証したいと考えている。

参考文献

- [1] D. Keehn, "A note on learning for gaussian properties," *IEEE Trans. Inform. Theory*, vol. IT-11, no. 1, pp. 126–132, January 1965.
- [2] S. Geisser, "Posterior odds for multivariate normal parameters," *Journal of the Royal Statistical Society, Series B*, vol. 26, no. 1, pp. 69–76, 1964.
- [3] 岩村雅一, 大町真一郎, 阿曾弘具, "パターン認識における予測分布の偏りに関する一考察," *信学論 (D-II)*, vol. J87-D-II, no. 2, pp. 407–415, February 2004.
- [4] F. Kimura, K. Takashina, S. Tsuruoka, and Y. Miyake, "Modified quadratic discriminant functions and the application to Chinese character recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 1, pp. 149–153, January 1987.
- [5] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175, 1989.
- [6] W. James and C. Stein, "Estimation with quadratic loss," *Proc. 4th Berkeley Symp. on Math. Statist. and Prob.*, pp. 361–379, 1961.
- [7] 酒井充, 米田政明, 長谷博行, 丸山博, "新しいロバストな 2 次識別関数," *信学論 (A)*, vol. J81-A, no. 11, pp. 1619–1623, November 1998.
- [8] 酒井充, 米田政明, 長谷博行, 丸山博, 直江美知子, "固有値の偏り補正に基づく 2 次識別関数," *信学論 (D-II)*, vol. J82-D-II, no. 4, pp. 631–640, April 1999.