

# 高次元小標本における サポートベクターマシンの一貫性について

筑波大学大学院・数理物質科学研究科 中山 優吾 (Yugo Nakayama)

Graduate School of Pure and Applied Sciences

University of Tsukuba

筑波大学・数理物質系 矢田 和善 (Kazuyoshi Yata)

Institute of Mathematics

University of Tsukuba

筑波大学・数理物質系 青嶋 誠 (Makoto Aoshima)

Institute of Mathematics

University of Tsukuba

## 1 はじめに

本論文は、高次元小標本データに対する判別分析を考える。母集団が 2 個あると想定し、各母集団  $\pi_i$  ( $i = 1, 2$ ) は平均に  $p$  次元ベクトル  $\mu_i$ 、共分散行列に  $p$  次元正定値対称行列  $\Sigma_i$  ( $> \mathbf{O}$ ) をもつと仮定する。高次元データに対して  $\Sigma_1 = \Sigma_2$  を想定することは現実的ではないので、共分散行列の共通性は仮定しない。ただし、 $\liminf_{p \rightarrow \infty} \{\text{tr}(\Sigma_1)/\text{tr}(\Sigma_2)\} > 0$ ,  $\limsup_{p \rightarrow \infty} \{\text{tr}(\Sigma_1)/\text{tr}(\Sigma_2)\} < \infty$  を仮定する。各母集団  $\pi_i$  から  $n_i$  ( $\geq 2$ ) 個の学習データ  $\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}$  を無作為に抽出する。判別対象の  $p$  次元データを  $\mathbf{x}_0$  とし、 $\mathbf{x}_0 \in \pi_1$  もしくは  $\mathbf{x}_0 \in \pi_2$  を仮定する。 $\mathbf{x}_0 \in \pi_1$  のときに判別対象を誤判別する確率を  $e(1)$  と表記し、 $e(2)$  も同様の表記とする。

高次元データの 2 群判別は、2 群の共分散行列が共通だと仮定すれば、Dudoit et al. (2002) や Bickel and Levina (2004) による標本共分散行列の対角成分だけを使った判別方式がある。しかし、共分散行列の共通性を仮定する問題設定の単純化は、高次元データが本来もつ 2 群の差異に関する情報を損なうことになる。共分散行列に共通性を仮定しない場合、Dudoit et al. (2002) による標本共分散行列の対角成分だけを使った判別方式、Chan and Hall (2009), Aoshima and Yata (2014) 等のユークリッド距離に基づく判別方式、Aoshima and Yata (2011, 2015a) による高次元データの幾何学的表現に基づく判別方

式がある。また, Aoshima and Yata (2015b) は高次元における判別関数のクラスを考え, 非スパース性と適当な正則条件のもとで  $e(i) \rightarrow 0$ ,  $p \rightarrow \infty$ ,  $i = 1, 2$  なる一致性が得られることを証明した。これらの研究は, 数理統計学分野からのアプローチである。

一方, 機械学習分野からのアプローチもある。判別分析は教師あり学習という立場で広く研究されており, 代表的な手法に Vapnik (1999) が考案したサポートベクターマシン (SVM) がある。SVM は高次元データ解析において疎な解を与え, 汎化性能が良いことが知られているものの, SVM の精度保証については理論的な研究が乏しいように思われる。

本論文は, 高次元小標本データが線形分離可能であることに着目し, ハードマージン線形 SVM (HM-LSVM) の漸近的性質を理論的に研究し, これが一致性をもつための正則条件を導出する。さらに, Aoshima and Yata (2014) によるユークリッド距離に基づく判別方式 (DBDA) との比較について, 理論的かつ数値的に性能を検証する。本論文の構成は, 以下の通りである。2 節では, SVM による判別方式を紹介する。3 節では, SVM の漸近的性質を理論的に与え, 一致性をもつための正則条件を導出する。4 節では, DBDA の漸近的性質を紹介し, SVM と DBDA を理論的に比較する。5 節では, 高次元小標本データに対して, SVM と DBDA の性能を数値的に検証する。

## 2 SVM による判別方式

本節では, HM-LSVM による判別方式を紹介する。次の線形モデルを考える。

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b. \quad (1)$$

ここで,  $\mathbf{w}$  は  $p$  次の重みベクトルであり,  $b$  はバイアスパラメータである。  $N = n_1 + n_2$  個の 2 群の学習データを  $(\mathbf{x}_1, \dots, \mathbf{x}_N) = (\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1}, \mathbf{x}_{21}, \dots, \mathbf{x}_{2n_2})$  と並べる。さらに,  $t_j = -1, j = 1, \dots, n_1; t_j = 1, j = n_1 + 1, \dots, N$  とおく。高次元小標本 ( $p > N$ ) データは線形分離可能なので, 学習データに対して

$$y(\mathbf{x}_j) < 0, j = 1, \dots, n_1; \quad y(\mathbf{x}_j) > 0, j = n_1 + 1, \dots, N$$

となる  $(\mathbf{w}, b)$  が存在する。SVM は, 分類境界と学習データの最短距離 (マージン) に着目し, これが最大となるように  $(\mathbf{w}, b)$  を求める。マージンを最大化する解は,

$$t_j(\mathbf{w}^T \mathbf{x}_j + b) \geq 1, \quad j = 1, \dots, N$$

のもとで,

$$\operatorname{argmin}_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (2)$$

なる最適化問題を解くことで求められる。ラグランジュ形式

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{j=1}^N \alpha_j \{t_j (\mathbf{w}^T \mathbf{x}_j + b) - 1\}$$

の最小化を考える。ここで、 $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N)^T$  である。  $L(\mathbf{w}, b, \boldsymbol{\alpha})$  を  $\mathbf{w}$  と  $b$  で微分して、それらを 0 とおくことで、次の 2 条件が得られる。

$$\mathbf{w} = \sum_{j=1}^N \alpha_j t_j \mathbf{x}_j; \quad \sum_{j=1}^N \alpha_j t_j = 0. \quad (3)$$

これらを  $L(\mathbf{w}, b, \boldsymbol{\alpha})$  に代入することで、(2) の双対問題が得られる。すなわち、次の 2 条件

$$\alpha_j \geq 0, \quad j = 1, \dots, N; \quad \sum_{j=1}^N \alpha_j t_j = 0 \quad (4)$$

のもとで、以下の関数を  $\boldsymbol{\alpha}$  について最大化する。

$$L(\boldsymbol{\alpha}) = \sum_{j=1}^N \alpha_j - \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \alpha_j \alpha_k t_j t_k \mathbf{x}_j^T \mathbf{x}_k. \quad (5)$$

$L(\boldsymbol{\alpha})$  を最大化する  $\boldsymbol{\alpha}$  を用いて、線形モデル (1) は次のように書ける。

$$y(\mathbf{x}) = \sum_{j=1}^N \alpha_j t_j \mathbf{x}_j^T \mathbf{x} + b.$$

このとき、 $t_j y(\mathbf{x}_j) = 1$  ( $\alpha_j \neq 0$ ) なるデータ  $\mathbf{x}_j$  をサポートベクターとよぶ。  $S = \{j | \alpha_j \neq 0, j = 1, \dots, N\}$  とおくと、バイアスパラメータは

$$b = \frac{1}{N_S} \sum_{j \in S} \left( t_j - \sum_{k \in S} \alpha_k t_k \mathbf{x}_j^T \mathbf{x}_k \right)$$

で与えられることになる。このようにして求めた  $(b, \boldsymbol{\alpha})$  を用いて、判別関数を

$$y(\mathbf{x}_0) = \sum_{j=1}^N \alpha_j t_j \mathbf{x}_j^T \mathbf{x}_0 + b \quad (6)$$

と定義し、判別方式を  $y(\mathbf{x}_0) < 0$  のとき  $\mathbf{x}_0 \in \pi_1$ ,  $y(\mathbf{x}_0) \geq 0$  のとき  $\mathbf{x}_0 \in \pi_2$  とする。詳細は、Vapnik (1999) 等を参照のこと。

### 3 高次元小標本における SVM の漸近的性質

高次元小標本 ( $p \rightarrow \infty$ ,  $N$  は固定) の枠組みで, 漸近的に目的関数  $L(\alpha)$  を評価する.  
いま,  $p \rightarrow \infty$  で次を仮定する.

$$(A-i) \quad \text{Var}(\|\mathbf{x}_{ik} - \boldsymbol{\mu}_i\|^2) = O\{\text{tr}(\boldsymbol{\Sigma}_i^2)\}, \quad i = 1, 2.$$

$\pi_i$  が正規分布ならば  $\text{Var}(\|\mathbf{x}_{ik} - \boldsymbol{\mu}_i\|^2) = 2\text{tr}(\boldsymbol{\Sigma}_i^2)$  となるので, (A-i) は母集団分布に正規性を課すよりも緩い条件になっていることに注意する. さらに,  $\Delta = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$  とおき,  $p \rightarrow \infty$  で次を仮定する.

$$(A-ii) \quad \frac{\max_{i=1,2} \text{tr}(\boldsymbol{\Sigma}_i^2)}{\Delta^2} = o(1).$$

このとき, 以下の補題が成り立つ.

補題 1. (A-i) と (A-ii) を仮定する. (4) のもと,  $p \rightarrow \infty$  のとき次が成り立つ.

$$\begin{aligned} L(\alpha) &= \sum_{j=1}^N \alpha_j - \frac{\Delta}{8} \left( \sum_{j=1}^N \alpha_j \right)^2 \{1 + o_P(1)\} \\ &\quad - \frac{1}{2} \left( \text{tr}(\boldsymbol{\Sigma}_1) \sum_{j=1}^{n_1} \alpha_j^2 + \text{tr}(\boldsymbol{\Sigma}_2) \sum_{j=n_1+1}^N \alpha_j^2 \right) \{1 + o_P(1)\}. \end{aligned}$$

補題 1 の第 3 項について, (4) と  $\sum_{j=1}^N \alpha_j = C (> 0)$  のもとで以下を得る.

$$\max_{\alpha} \left\{ -\frac{1}{2} \left( \text{tr}(\boldsymbol{\Sigma}_1) \sum_{j=1}^{n_1} \alpha_j^2 + \text{tr}(\boldsymbol{\Sigma}_2) \sum_{j=n_1+1}^N \alpha_j^2 \right) \right\} = -\frac{C^2}{8} \delta. \quad (7)$$

ここで,  $\delta = \{\text{tr}(\boldsymbol{\Sigma}_1)/n_1 + \text{tr}(\boldsymbol{\Sigma}_2)/n_2\}$  である.  $\Delta_* = \Delta + \delta$  とおく. (7) と補題 1 から,  $L(\alpha)$  の主要項は

$$C - \frac{C^2}{8} \Delta - \frac{C^2}{8} \delta = -\frac{\Delta_*}{8} \left( C - \frac{4}{\Delta_*} \right)^2 + \frac{2}{\Delta_*} \quad (8)$$

と書け,  $C = \sum_{j=1}^N \alpha_j = 4/\Delta_*$  のとき最大値は  $2/\Delta_*$  になる. 次を仮定する.

$$(A-iii) \quad \limsup_{p \rightarrow \infty} \frac{\delta}{\Delta} < \infty \quad \text{かつ} \quad \liminf_{p \rightarrow \infty} \frac{\delta}{\Delta} > 0.$$

そのとき, 以下の補題が成り立つ.

補題 2. (A-i)-(A-iii) を仮定する.  $p \rightarrow \infty$  のとき次が成り立つ.

$$\alpha_j = \frac{2}{\Delta_* n_1} \{1 + o_P(1)\}, j = 1, \dots, n_1; \quad \alpha_j = \frac{2}{\Delta_* n_2} \{1 + o_P(1)\}, j = n_1 + 1, \dots, N.$$

さらに,  $p \rightarrow \infty$  のとき判別関数 (6) について次が成り立つ.

$$y(\mathbf{x}_0) = \frac{2(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T}{\Delta_*} \mathbf{x}_0 + \frac{\|\boldsymbol{\mu}_1\|^2 - \|\boldsymbol{\mu}_2\|^2}{\Delta_*} + \frac{\text{tr}(\boldsymbol{\Sigma}_1)/n_1 - \text{tr}(\boldsymbol{\Sigma}_2)/n_2}{\Delta_*} + o_P(1).$$

補題 2 を用いると, SVM に関して高次元小標本における一致性が主張できる. 次を仮定する.

$$(A\text{-iv}) \quad \limsup_{p \rightarrow \infty} \frac{|\text{tr}(\boldsymbol{\Sigma}_1)/n_1 - \text{tr}(\boldsymbol{\Sigma}_2)/n_2|}{\Delta} < 1.$$

0

定理 1. (A-i)-(A-iv) を仮定する.  $p \rightarrow \infty$  のとき判別方式 (6) について次が成り立つ.

$$e(i) \rightarrow 0, \quad i = 1, 2.$$

## 4 ユークリッド距離に基づく判別方式 (DBDA)

Aoshima and Yata (2014) は, ユークリッド距離に基づく次のような判別関数を考えた.

$$\omega(\mathbf{x}_0) = \left( \mathbf{x}_0 - \frac{\bar{\mathbf{x}}_{1n_1} + \bar{\mathbf{x}}_{2n_2}}{2} \right)^T (\bar{\mathbf{x}}_{2n_2} - \bar{\mathbf{x}}_{1n_1}) - \frac{\text{tr}(\mathbf{S}_{1n_1})}{2n_1} + \frac{\text{tr}(\mathbf{S}_{2n_2})}{2n_2}. \quad (9)$$

ここで,  $\bar{\mathbf{x}}_{in_i} = \sum_{j=1}^{n_i} \mathbf{x}_{ij}/n_j$ ,  $\mathbf{S}_{in_i} = \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_{in_i})(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{in_i})^T / (n_i - 1)$  である. 判別方式は,  $\omega(\mathbf{x}_0) < 0$  のとき  $\mathbf{x}_0 \in \pi_1$ ,  $\omega(\mathbf{x}_0) \geq 0$  のとき  $\mathbf{x}_0 \in \pi_2$  である. Aoshima and Yata (2014) は次の定理を与えた.

定理 2. (A-ii) のもとで,  $p \rightarrow \infty$  のとき判別方式 (9) について次が成り立つ.

$$e(i) \rightarrow 0, \quad i = 1, 2.$$

上記の結果は, 母集団  $\pi_i$  ( $i = 1, 2$ ) の分布に依存しない. つまり判別方式 (9) は, 母集団分布に正規性等を仮定せずに, (A-ii) の仮定だけで一致性を主張している. それに対して, 判別方式 (6) は, 定理 1 で示した通り (A-i), (A-iii), (A-iv) なる正則条件も必要になる. さらに, もしも

$$\frac{\text{tr}(\boldsymbol{\Sigma}_1)/n_1 - \text{tr}(\boldsymbol{\Sigma}_2)/n_2}{\Delta_*}$$

なるバイアス項が大きいとき、判別方式 (6) は誤判別確率についてアンバランスな結果を与えることが予想される。バイアス項が非常に大きく (A-iv) を満たさない状況においては、補題 2 から  $e(i) \rightarrow 1$  となることも危惧される。それゆえ、高次元小標本の枠組みにおいては、判別方式 (6) ではなく、判別方式 (9) を用いることが理論的に推奨される。

## 5 シミュレーション

本節では、高次元小標本のもとで、判別方式 (6) (HM-LSVM) と判別方式 (9) (DBDA) の精度を数値的に検証する。  $p = 2^s$ ,  $s = 5, \dots, 11$  とおく。  $\pi_i$  ( $i = 1, 2$ ) の母集団分布には、  $p$  次元正規分布  $N_p(\mu_i, \Sigma_i)$  を考える。ここで、  $\Sigma_1 = B(0.3^{|i-j|^{1/3}})B$ ,  $\Sigma_2 = 1.5B(0.3^{|i-j|^{1/3}})B$  とおき、

$$B = \text{diag}(\sqrt{0.5 + 1/(p+1)}, \sqrt{0.5 + 2/(p+1)}, \dots, \sqrt{0.5 + p/(p+1)})$$

とする。このとき、  $\text{tr}(\Sigma_1) = p$ ,  $\text{tr}(\Sigma_2) = 1.5p$  となる。  $\mu_1 = \mathbf{0}$  とし、次の 3 つの場合を考える。

- (a)  $\mu_2 = (1, \dots, 1)^T/2$ ,  $(n_1, n_2) = (6, 9)$ .
- (b)  $\mu_2 = (1, \dots, 1)^T/2$ ,  $(n_1, n_2) = (9, 6)$ .
- (c) 最初の  $\lceil p^{2/3} \rceil$  個の成分が 1, それ以外が 0 である  $\mu_2 = (1, \dots, 1, 0, \dots, 0)^T$ ,  $(n_1, n_2) = (9, 6)$ 。ここで、  $\lceil x \rceil$  は  $x$  以上の最小の整数を表す。

(a) と (b) は  $\Delta = p/4$ , (c) は  $\Delta = \lceil p^{2/3} \rceil$  となる。(A-iii) と (A-iv) は (a) と (b) のもとで成り立つが、(c) のもとでは成り立たない。さらに、(a) のもとでは  $\{\text{tr}(\Sigma_1)/n_1 - \text{tr}(\Sigma_2)/n_2\}/\Delta_* = 0$  となるが、(b) と (c) のもとでは  $\{\text{tr}(\Sigma_1)/n_1 - \text{tr}(\Sigma_2)/n_2\}/\Delta_* \neq 0$  であることに注意する。一方で、これらの設定において、(A-i) と (A-ii) を満たすことに注意する。

判別方式 (6) と (9) について、上記設定において  $\mathbf{x}_0 \in \pi_1$  と  $\mathbf{x}_0 \in \pi_2$  が正しく判別されるかを確認した。実験を 2000 回繰り返し、  $\mathbf{x}_0 \in \pi_1$  と  $\mathbf{x}_0 \in \pi_2$  のそれぞれの場合における誤判別の割合で  $e(1)$  と  $e(2)$  を推定し、それらを図 1 から 3 に示した。標準誤差は 0.011 以下である。

判別方式 (9) は、全体的に良い結果を与えた。これは、判別方式 (9) が (A-iii) や (A-iv) に依存せずに一致性をもつことが理由と考えられる。さらに、判別方式 (9) は  $\{\text{tr}(\Sigma_1)/n_1 - \text{tr}(\Sigma_2)/n_2\}/\Delta_*$  なるバイアス項にも依存しない。一方で、判別方式 (6) は、  $\text{tr}(\Sigma_1)/n_1 = \text{tr}(\Sigma_2)/n_2$  となる (a) の場合を除いて、判別方式 (9) よりも判別精度が低い。特に、(A-iii)

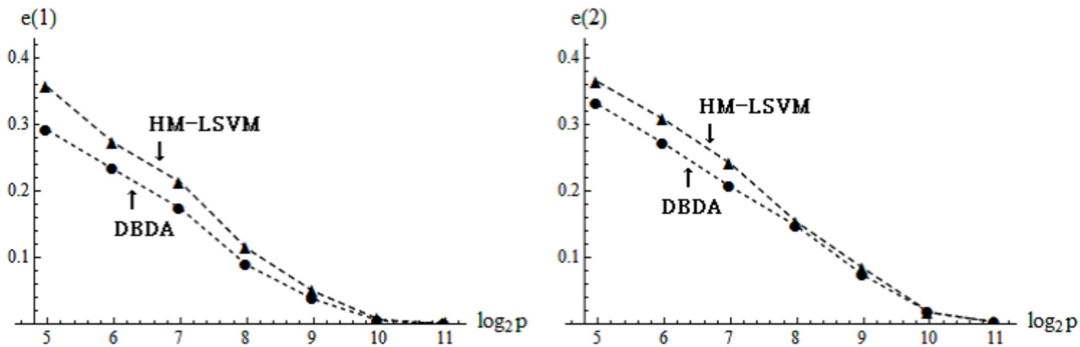


図1  $\mu_2 = (1, \dots, 1)^T/2$ ,  $(n_1, n_2) = (6, 9)$  のときのシミュレーション結果. 左図は  $e(1)$ , 右図は  $e(2)$  をプロットしている.

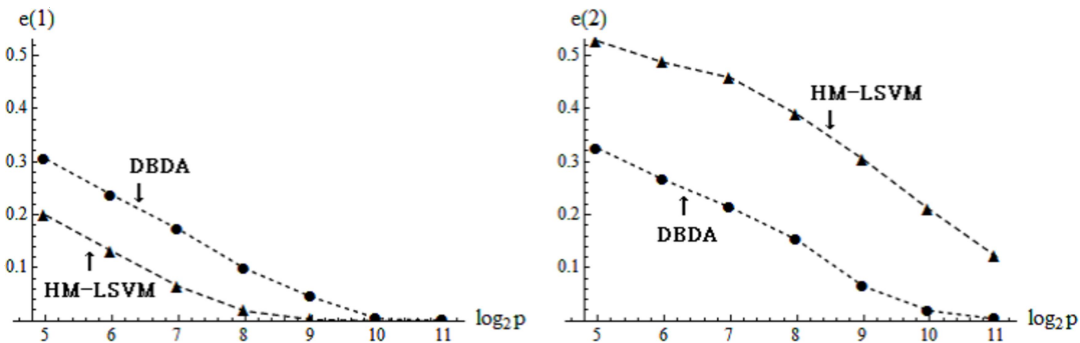


図2  $\mu_2 = (1, \dots, 1)^T/2$ ,  $(n_1, n_2) = (9, 6)$  のときのシミュレーション結果. 左図は  $e(1)$ , 右図は  $e(2)$  をプロットしている.

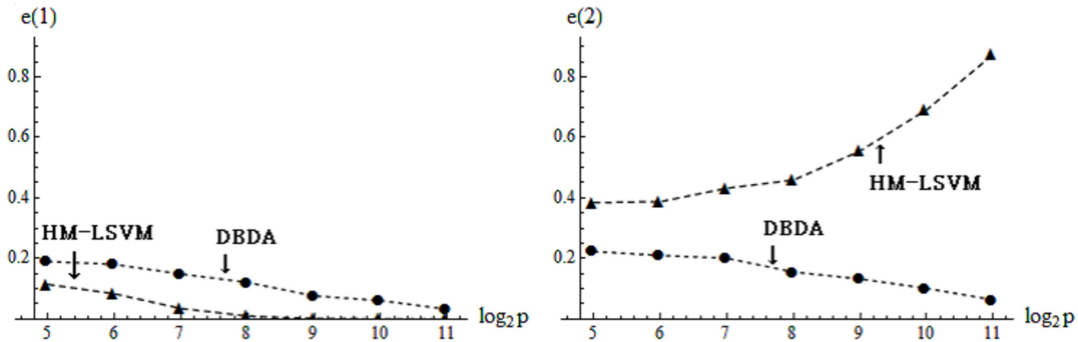


図3  $\mu_2 = (1, \dots, 1, 0, \dots, 0)^T$ ,  $(n_1, n_2) = (9, 6)$  のときのシミュレーション結果. 左図は  $e(1)$ , 右図は  $e(2)$  をプロットしている.

と (A-iv) を満たさない (c) の場合には、判別精度は極端に低くなり、高次元で  $e(2)$  は 1 に近づくことが確認できる。これは、判別方式 (6) が、(A-iv) を満たさない状況で高次元のとき  $e(i) \rightarrow 1$  となり得るからである。さらに、(b) の場合も、 $\{\text{tr}(\Sigma_1)/n_1 - \text{tr}(\Sigma_2)/n_2\}/\Delta_* \neq 0$  となるためにアンバランスな結果を与えている。このシミュレーション結果からも、高次元小標本の枠組みにおいて、判別方式 (6) よりも判別方式 (9) を用いることが推奨される。

## 6 付録

以降、 $\mu = \mu_1 - \mu_2$ ,  $\mu_* = (\mu_1 + \mu_2)/2$  と表記する。

補題 1 の証明。シュワルツの不等式を用いると、仮定 (A-ii) のもとで次が成り立つ。

$$\mu^T \Sigma_i \mu / \Delta^2 \leq \text{tr}(\Sigma_i^2)^{1/2} / \Delta = o(1), \quad i = 1, 2.$$

このとき、 $\text{tr}(\Sigma_1 \Sigma_2) \leq \max_{i=1,2} \text{tr}(\Sigma_i^2)$  に注意し、チェビシェフの不等式を用いると、任意の  $\tau > 0$  について、仮定 (A-ii) のもとで次が成り立つ。

$$\begin{aligned} & P(|(\mathbf{x}_j - \mu_*)^T (\mathbf{x}_k - \mu_*) - \Delta/4| \geq \tau \Delta) \\ & \leq (\tau \Delta)^{-2} E[\{(\mathbf{x}_j - \mu_*)^T (\mathbf{x}_k - \mu_*) - \Delta/4\}^2] \\ & = O\{\text{tr}(\Sigma_1^2) + \mu^T \Sigma_1 \mu\} / \Delta^2 = o(1), \quad 1 \leq j < k \leq n_1; \\ & P(|(\mathbf{x}_j - \mu_*)^T (\mathbf{x}_k - \mu_*) - \Delta/4| \geq \tau \Delta) \\ & = O\{\text{tr}(\Sigma_2^2) + \mu^T \Sigma_2 \mu\} / \Delta^2 = o(1), \quad n_1 + 1 \leq j < k \leq N; \\ & P(|(\mathbf{x}_j - \mu_*)^T (\mathbf{x}_k - \mu_*) + \Delta/4| \geq \tau \Delta) \\ & = O\{\text{tr}(\Sigma_1 \Sigma_2) + \mu^T (\Sigma_1 + \Sigma_2) \mu\} / \Delta^2 = o(1), \quad 1 \leq j \leq n_1, n_1 + 1 \leq k \leq N. \end{aligned} \quad (10)$$

さらに、(A-i) のもと、

$$\text{Var}\{(\mathbf{x}_{ij} - \mu_i)^T (\mathbf{x}_{ij} - \mu_i)\} = O\{\text{tr}(\Sigma_i^2)\}, \quad j = 1, \dots, n_i; \quad i = 1, 2$$

となる。このとき、任意の  $\tau > 0$  について、仮定 (A-i) と (A-ii) のもとで次が成り立つ。

$$\begin{aligned} & P(|(\mathbf{x}_j - \mu_*)^T (\mathbf{x}_j - \mu_*) - \Delta/4 - \text{tr}(\Sigma_1)| \geq \tau \Delta) \\ & = O\{\text{tr}(\Sigma_1^2) + \mu^T \Sigma_1 \mu\} / \Delta^2 = o(1), \quad j = 1, \dots, n_1; \\ & P(|(\mathbf{x}_j - \mu_*)^T (\mathbf{x}_j - \mu_*) - \Delta/4 - \text{tr}(\Sigma_2)| \geq \tau \Delta) = o(1), \quad j = n_1 + 1, \dots, N. \end{aligned} \quad (11)$$

ここで、(4) のもとで、 $L(\alpha)$  は次のように書ける。

$$L(\alpha) = \sum_{j=1}^N \alpha_j - \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \alpha_j \alpha_k t_j t_k (\mathbf{x}_j - \mu_*)^T (\mathbf{x}_k - \mu_*).$$



このとき, (10) と (11) から, 仮定 (A-i) と (A-ii) のもとで次が成り立つ.

$$L(\alpha) = \sum_{j=1}^N \alpha_j - \frac{\Delta}{8} \left( \sum_{j=1}^N \alpha_j \right)^2 \{1 + o_P(1)\} \\ - \frac{1}{2} \left( \text{tr}(\Sigma_1) \sum_{j=1}^{n_1} \alpha_j^2 + \text{tr}(\Sigma_2) \sum_{j=n_1+1}^N \alpha_j^2 \right) \{1 + o_P(1)\}.$$

以上から, 補題 1 を得る. □

補題 2 の証明. まず, 1 つ目の結果を示す. (7) と (8) から, (4) のもと,

$$\alpha_j = \frac{2}{\Delta_* n_1}, \quad j = 1, \dots, n_1; \quad \alpha_j = \frac{2}{\Delta_* n_2}, \quad j = n_1 + 1, \dots, N$$

は

$$\sum_{j=1}^N \alpha_j - \frac{\Delta}{8} \left( \sum_{j=1}^N \alpha_j \right)^2 - \frac{1}{2} \left( \text{tr}(\Sigma_1) \sum_{j=1}^{n_1} \alpha_j^2 + \text{tr}(\Sigma_2) \sum_{j=n_1+1}^N \alpha_j^2 \right)$$

の最大値を与える. よって, 補題 1 より, (A-i)-(A-iii) のもとで次を得る.

$$\alpha_j = \frac{2}{\Delta_* n_1} \{1 + o_P(1)\}, \quad j = 1, \dots, n_1; \quad \alpha_j = \frac{2}{\Delta_* n_2} \{1 + o_P(1)\}, \quad j = n_1 + 1, \dots, N. \quad (12)$$

次に, 2 つ目の結果を示す.  $S = \{1, \dots, N\}$  のとき, (4) のもと, (6) は次のように書ける.

$$y(x_0) = \sum_{j=1}^N \alpha_j t_j (\mathbf{x}_j - \boldsymbol{\mu}_*)^T (\mathbf{x}_0 - \boldsymbol{\mu}_*) + \sum_{j=1}^N \alpha_j t_j (\mathbf{x}_j - \boldsymbol{\mu}_*)^T \boldsymbol{\mu}_* + b \\ = \sum_{j=1}^N \alpha_j t_j (\mathbf{x}_j - \boldsymbol{\mu}_*)^T (\mathbf{x}_0 - \boldsymbol{\mu}_*) + \frac{-n_1 + n_2}{N} \\ - \frac{1}{N} \sum_{j=1}^N \sum_{k=1}^N \alpha_k t_k (\mathbf{x}_j - \boldsymbol{\mu}_*)^T (\mathbf{x}_k - \boldsymbol{\mu}_*).$$

このとき, (10), (11) と (12) から, (A-i)-(A-iii) のもとで次を得る.

$$\frac{-n_1 + n_2}{N} - \frac{1}{N} \sum_{j=1}^N \sum_{k=1}^N \alpha_k t_k (\mathbf{x}_j - \boldsymbol{\mu}_*)^T (\mathbf{x}_k - \boldsymbol{\mu}_*) \\ = \frac{-n_1 + n_2}{N} + \frac{(n_1 - n_2)\Delta}{\Delta_* N} + 2 \frac{\text{tr}(\Sigma_1) - \text{tr}(\Sigma_2)}{\Delta_* N} + o_P(1) \\ = \frac{\text{tr}(\Sigma_1)/n_1 - \text{tr}(\Sigma_2)/n_2}{\Delta_*} + o_P(1). \quad (13)$$

(10) と同様にして,  $\mathbf{x}_0 \in \pi_1$  もしくは  $\mathbf{x}_0 \in \pi_2$  に対して, (A-ii) のもと,  $\text{Var}\{(\mathbf{x}_{ij} - \boldsymbol{\mu}_i)^T(\mathbf{x}_0 - \boldsymbol{\mu}_*)/\Delta\} = o(1)$  がすべての  $i, j$  で成り立つ. よって, (12) から, (A-i)-(A-iii) のもとで次を得る.

$$\sum_{j=1}^N \alpha_j t_j (\mathbf{x}_j - \boldsymbol{\mu}_*)^T (\mathbf{x}_0 - \boldsymbol{\mu}_*) = \frac{-2\boldsymbol{\mu}^T(\mathbf{x}_0 - \boldsymbol{\mu}_*)}{\Delta_*} + o_P(1). \quad (14)$$

(13) と (14) から, 2つ目の結果が得られる.  $\square$

定理 1 の証明.  $\mathbf{x}_0 \in \pi_i$  に対して,

$$\begin{aligned} E\{\boldsymbol{\mu}^T(\mathbf{x}_0 - \boldsymbol{\mu}_*)/\Delta\} &= (-1)^{i+1}/2, \\ \text{Var}\{\boldsymbol{\mu}^T(\mathbf{x}_0 - \boldsymbol{\mu}_*)/\Delta\} &= O(\boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu} / \Delta^2) = O(\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2} / \Delta) \end{aligned}$$

となる. 補題 2 から, (A-i)-(A-iii) のもとで  $\mathbf{x}_0 \in \pi_i$  に対して次を得る.

$$y(\mathbf{x}_0) = (-1)^i \frac{\Delta}{\Delta_*} + \frac{\text{tr}(\boldsymbol{\Sigma}_1)/n_1 - \text{tr}(\boldsymbol{\Sigma}_2)/n_2}{\Delta_*} + o_P(1).$$

それゆえ, (A-iv) から, 誤判別確率について結果を得る.  $\square$

謝辞 本研究は, 科学研究費補助金 基盤研究 (A) 15H01678 研究代表者: 青嶋 誠「大規模複雑データの理論と方法論の総合的研究」, 学術研究助成基金助成金 挑戦的萌芽研究 26540010 研究代表者: 青嶋 誠「ビッグデータの統計学: 理論の開拓と 3V への挑戦」, および, 若手研究 (B) 26800078 研究代表者: 矢田和善「高次元漸近理論の統一的研究」から研究助成を受けています. 本論文の作成にあたって江面友希氏に感謝します.

## 参考文献

- [1] Aoshima, M. and Yata, K. (2011). Two-stage procedures for high-dimensional data. *Sequential Anal. (Editor's special invited paper)*, **30**, 356-399.
- [2] Aoshima, M. and Yata, K. (2014). A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data. *Ann. Inst. Statist. Math.*, **66**, 983-1010.
- [3] Aoshima, M. and Yata, K. (2015a). Geometric classifier for multiclass, high-dimensional data. *Sequential Anal.*, **34**, 279-294.

- [4] Aoshima, M. and Yata, K. (2015b). High-dimensional quadratic classifiers in non-sparse settings. arXiv:1503.04549.
- [5] Bickel, P.J. and Levina, E. (2004). Some theory for Fisher's linear discriminant function, 'naive Bayes', and some alternatives when there are many more variables than observations. *Bernoulli*, **10**, 989-1010.
- [6] Chan, Y.-B. and Hall, P. (2009). Scale adjustments for classifiers in high-dimensional, low sample size settings. *Biometrika*, **96**, 469-478.
- [7] Dudoit, S., Fridlyand, J. and Speed, T.P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Amer. Statist. Assoc.*, **97**, 77-87.
- [8] Vapnik, V.N. (1999). *The Nature of Statistical Learning Theory (second ed.)*. New York: Springer-Verlag.