

高次元における Distance Weighted Discrimination について

筑波大学大学院・数理物質科学研究科 江頭 健斗 (Kento Egashira)

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

本論文では、高次元小標本データに対する2クラスの判別分析を考える。独立な d 次元の母集団が2個あると考え、各母集団 π_i ($i = 1, 2$) は平均に未知の d 次元ベクトル $\boldsymbol{\mu}_i$ 、共分散行列に未知の d 次元正定値対称行列 $\boldsymbol{\Sigma}_i$ ($> \mathbf{O}$) をもつと仮定する。ここで、高次元データに対して $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ を考えることは現実的でないため、共分散行列の共通性は仮定しない。各 $i = 1, 2$ について、母集団 π_i から、 n_i 個の訓練データ $\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{in_i}$ を無作為に抽出する。但し、 $N = n_1 + n_2$ とする。簡単のため、訓練データを、 $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) = (\boldsymbol{x}_{11}, \dots, \boldsymbol{x}_{1n_1}, \boldsymbol{x}_{21}, \dots, \boldsymbol{x}_{2n_2})$ と表記する。また、訓練データのラベル y_i を、 $i = 1, \dots, n_1$ に対しては、 -1 、 $i = n_1 + 1, \dots, N$ に対しては、 $+1$ と定義する。判別するテストデータを、 \boldsymbol{x}_0 ($\in \pi_1$ もしくは $\in \pi_2$) とし、 $\boldsymbol{x}_0 \in \pi_i$ を誤判別する確率を $e(i)$ と表記する。ここで、 \boldsymbol{x}_0 は、訓練データと独立に抽出されたものとする。

高次元の2群における判別分析について、2群の共分散行列が等しいと仮定すれば、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などがある。しかし、2群の共分散行列が等しいと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に目を瞑ることになる。2群の共分散行列

が等しいことを仮定しない2次判別分析の場合, 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 (2011,2015) による高次元小標本の幾何学的表現に基づく判別方式がある. さらに, Aoshima and Yata (2019) は, 非スパース性と適当な正則条件のもと, 一般的な高次元判別方式に $e(i) \rightarrow 0, d \rightarrow \infty, i = 1, 2$ なる一致性が得られることを証明した.

工学における機械学習の方面から, 高次元データ解析の方法論に関する研究が膨大にある. 例えば, 判別分析は, 機械学習の領域では教師あり学習という立場で広く研究され, 代表的な手法に Vapnik (2000) が考案したサポートベクトルマシン (SVM) がある. Nakayama et al. (2017, 2019) では, 高次元小標本データにおける SVM の漸近的性質を導出し, 一致性を与えるための正則条件も導出した. 一方で, Marron et al. (2007) は, SVM が高次元データに対してオーバーフィッティングを起こすことを指摘し, その問題を解決するために Distance weighted discrimination (DWD) と呼ばれる新たな判別手法を提案した. しかしながら, 理論的に DWD の精度保証を与えるような研究が乏しく, 理論研究に改善の余地がある. 本論文では, 高次元小標本データにおける DWD の漸近的性質を導出し, 一致性を与えるための正則条件も導出する.

2節では, DWD の判別方式を説明する. 3節では, DWD における判別関数の漸近的性質を導出し, 一致性を与えるための正則条件も導出する. 4節では, 数値シミュレーションにより, 高次元小標本における DWD の精度を検証する.

2 DWD による判別方式

本節では, DWD の判別方式を紹介する. $\mathbf{w} \in \mathbb{R}^d$ を重みベクトルとし, $b \in \mathbb{R}$ を適当に選択すれば, 分離平面とデータ点 \mathbf{x}_i のマージンは, $r_i = y_i(\mathbf{w}^T \mathbf{x}_i + b)$ と表すことができる. 訓練データが, 超平面で分離可能であるとき, DWD は, $1/r_i$ の和を最小にするように, 分離平面を求めていく. ここで, 高次元小標本データは線形分離可能であることに注意する. 従って, DWD の定式化は, 以下の最適化問題として表すことができる.

$$\min_{\mathbf{w}, b} \sum_{i=1}^N \frac{1}{r_i} \text{ subject to } r_i = y_i(\mathbf{w}^T \mathbf{x}_i + b) > 0, i = 1, \dots, N, \|\mathbf{w}\| = 1$$

但し, $\|\cdot\|$ は, ユークリッド距離とする. 上記の最適化問題は, 次のように書き換えることができる.

$$\max_{\alpha, \lambda > 0} \min_{\mathbf{w}, b, \mathbf{r}} \left\{ \sum_{i=1}^N \frac{1}{r_i} + \sum_{i=1}^N \alpha_i (r_i - y_i (\mathbf{w}^T \mathbf{x}_i + b)) + \lambda (\mathbf{w}^T \mathbf{w} - 1) \right\}$$

但し, $\alpha = (\alpha_1, \dots, \alpha_N)^T$ である. 次に,

$$A = \sum_{i=1}^N \frac{1}{r_i} + \sum_{i=1}^N \alpha_i (r_i - y_i (\mathbf{w}^T \mathbf{x}_i + b)) + \lambda (\mathbf{w}^T \mathbf{w} - 1)$$

とおくと,

$$\begin{aligned} \frac{\partial A}{\partial r_i} = 0 &\Leftrightarrow r_i^2 = \frac{1}{\alpha_i}, \\ \frac{\partial A}{\partial \mathbf{w}} = 0 &\Leftrightarrow \mathbf{w} = \frac{1}{2\lambda} \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i, \\ \frac{\partial A}{\partial b} = 0 &\Leftrightarrow \sum_{i=1}^N \alpha_i y_i = 0 \end{aligned} \quad (1)$$

を得る. そのとき,

$$\max_{\alpha, \lambda > 0} \left\{ 2 \sum_{i=1}^N \sqrt{\alpha_i} - \frac{1}{4\lambda} \left\| \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \right\|^2 - \lambda \right\} \quad (2)$$

なる最適化問題を考えれば良い. また,

$$A^* = 2 \sum_{i=1}^N \sqrt{\alpha_i} - \frac{1}{4\lambda} \left\| \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \right\|^2 - \lambda$$

とおき,

$$\frac{\partial A^*}{\partial \lambda} = 0 \Leftrightarrow \lambda = \frac{1}{2} \left\| \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \right\| \quad (3)$$

に注意すると, 最適化問題 (2) を次のように書き換えることができる.

$$\max_{\alpha} \left\{ 2 \sum_{j=1}^N \sqrt{\alpha_j} - \sqrt{\sum_{j=1}^N \sum_{l=1}^N \alpha_j \alpha_l y_j y_l \mathbf{x}_j^T \mathbf{x}_l} \right\}$$

$$\text{subject to } \alpha_j \geq 0, j = 1, \dots, N, \sum_{j=1}^N \alpha_j y_j = 0 \quad (4)$$

ここで,

$$L(\boldsymbol{\alpha}) = 2 \sum_{j=1}^N \sqrt{\alpha_j} - \sqrt{\sum_{j=1}^N \sum_{l=1}^N \alpha_j \alpha_l y_j y_l \mathbf{x}_j^T \mathbf{x}_l}$$

とおき,

$$\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_N)^T = \operatorname{argmax}_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}) \text{ subject to (4)}$$

とする。このとき、(1) と (3) より,

$$\hat{\mathbf{w}} = \frac{1}{\sqrt{\sum_{j,l=1}^N \hat{\alpha}_j \hat{\alpha}_l y_j y_l \mathbf{x}_j^T \mathbf{x}_l}} \sum_{j=1}^N \hat{\alpha}_j y_j \mathbf{x}_j$$

である。切片項 b は、適当な j を用いて,

$$\hat{b}_j = \frac{y_j}{\sqrt{\hat{\alpha}_j}} - \hat{\mathbf{w}}^T \mathbf{x}_j$$

として与えられる。したがって、 b を以下のようにその平均を用いて与える。

$$\hat{b} = \frac{1}{N} \sum_{j=1}^N \left(\frac{y_j}{\sqrt{\hat{\alpha}_j}} - \hat{\mathbf{w}}^T \mathbf{x}_j \right)$$

上記で求めた $(\hat{b}, \hat{\mathbf{w}})$ を用いて、判別関数を

$$y(\mathbf{x}_0) = \hat{\mathbf{w}}^T \mathbf{x}_0 + \hat{b}$$

とし、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、Marron et al. (2007) や Qiao et al. (2010) を参照のこと。

3 高次元データにおける DWD の漸近的性質

高次元小標本 ($d \rightarrow \infty$, N は固定) の枠組みで、漸近的に目的関数 $L(\boldsymbol{\alpha})$ を評価する。 $\Delta = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$ とし、 $d \rightarrow \infty$ の高次元のもとで、以下を仮定する。

$$(A-i) \quad \frac{\operatorname{Var}(\|\mathbf{x}_{ij} - \boldsymbol{\mu}_i\|^2)}{\Delta^2} \rightarrow 0, i = 1, 2$$

$$(A-ii) \quad \frac{\operatorname{tr}(\boldsymbol{\Sigma}_i^2)}{\Delta^2} \rightarrow 0, i = 1, 2$$

母集団に正規分布を仮定したとき、 $\text{Var}(\|\mathbf{x}_{ij} - \boldsymbol{\mu}_i\|^2) = 2\text{tr}(\boldsymbol{\Sigma}_i^2)$ が成立し、(A-i) と (A-ii) が同値であることに注意する。

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

$$L(\boldsymbol{\alpha}) = 2 \sum_{j=1}^N \sqrt{\alpha_j} - \sqrt{\frac{\Delta}{4} \left(\sum_{j=1}^N \alpha_j \right)^2 \{1 + o_p(1)\} + \text{tr}(\boldsymbol{\Sigma}_1) \sum_{j=1}^{n_1} \alpha_j^2 + \text{tr}(\boldsymbol{\Sigma}_2) \sum_{j=n_1+1}^N \alpha_j^2}$$

次に、 $L(\boldsymbol{\alpha})$ の最大化を考える。 $\Delta_* = \Delta + \text{tr}(\boldsymbol{\Sigma}_1)/n_1 + \text{tr}(\boldsymbol{\Sigma}_2)/n_2$ とする。 ある正の定数 B について、制約条件 $\sum_{j=1}^N \alpha_j = B$ を考える。 このとき、(4) のもと、以下が成り立つ。

$$\min_{\boldsymbol{\alpha}} \left(\text{tr}(\boldsymbol{\Sigma}_1) \sum_{j=1}^{n_1} \alpha_j^2 + \text{tr}(\boldsymbol{\Sigma}_2) \sum_{j=n_1+1}^N \alpha_j^2 \right) = \frac{B^2}{4} \left(\frac{\text{tr}(\boldsymbol{\Sigma}_1)}{n_1} + \frac{\text{tr}(\boldsymbol{\Sigma}_2)}{n_2} \right) \quad (5)$$

このとき、補題 1 より、Nakayama et al. (2017) と同様にして、 $B = 2(\sqrt{n_1} + \sqrt{n_2})^2/\Delta_*$ とすることで、 $L(\boldsymbol{\alpha})$ を漸近的に最大化することができ、以下を得る。

補題 2. (A-i) と (A-ii) を仮定する。 このとき、 $d \rightarrow \infty$ のもとで、以下が成立する。

$$\begin{aligned} \hat{\alpha}_j &= \frac{(\sqrt{n_1} + \sqrt{n_2})^2}{n_1 \Delta_*} \{1 + o_p(1)\} \quad j = 1, \dots, n_1, \\ \hat{\alpha}_j &= \frac{(\sqrt{n_1} + \sqrt{n_2})^2}{n_2 \Delta_*} \{1 + o_p(1)\} \quad j = n_1 + 1, \dots, N \end{aligned}$$

さらに、 $\mathbf{x}_0 \in \pi_i$, $i = 1, 2$ ($i' \neq i$) について、 $d \rightarrow \infty$ のもとで、

$$\frac{\Delta_*^{1/2}}{\Delta} \hat{y}(\mathbf{x}_0) = \frac{(-1)^i n_{i'}}{N} + \delta + o_p(1) \quad (6)$$

が成立する。 但し、

$$\delta = -\frac{(n_1^{3/2} - n_2^{3/2})\Delta_*}{N(\sqrt{n_1} + \sqrt{n_2})\Delta} + \frac{\text{tr}(\boldsymbol{\Sigma}_1) - \text{tr}(\boldsymbol{\Sigma}_2)}{N\Delta}$$

である。

次に、 $d \rightarrow \infty$ のもとで、以下の仮定を考える。

$$(A-iii) \quad \limsup_{d \rightarrow \infty} |\delta| < \frac{\min_{i=1,2} n_i}{N}$$

このとき、次の結果を得ることができる。

定理 1. (A-i) から (A-iii) を仮定する. DWD について, $d \rightarrow \infty$ の高次元のもとで, 以下が成立する.

$$e(1) \rightarrow 0, \quad e(2) \rightarrow 0$$

一方で, (A-iii) が仮定できないとき, 次の強不一致性が成り立つ.

系 1. (A-i) と (A-ii) を仮定する. DWD について, $d \rightarrow \infty$ のもとで, 以下が成立する.

$$\begin{aligned} \liminf_{d \rightarrow \infty} \delta > \frac{n_2}{N} \quad \text{ならば,} \quad e(1) \rightarrow 1, \quad e(2) \rightarrow 0, \\ \limsup_{d \rightarrow \infty} \delta < -\frac{n_1}{N} \quad \text{ならば,} \quad e(1) \rightarrow 0, \quad e(2) \rightarrow 1 \end{aligned}$$

系 1 より, DWD は, n_1 と n_2 や $\text{tr}(\Sigma_1)$ と $\text{tr}(\Sigma_2)$ の不均等さによって, 強不一致性を与えることがわかる.

4 数値シミュレーション

本節では, 高次元小標本のもとで, DWD の精度を数値的に検証する. 母集団 π_i ($i = 1, 2$) について, 訓練データ $\mathbf{x}_{ij} - \boldsymbol{\mu}_i, j = 1, \dots, n_i$ を

(i) $N_d(\mathbf{0}, \Sigma_i)$;

(ii) 平均 $\mathbf{0}$, 共分散行列 Σ_i , 自由度 10 の d 次元 t 分布 $t_d(\Sigma_i, 10)$,

のいずれかから抽出するものとする. 最初の $\lceil d^{2/3} \rceil$ 個の成分が 1, それ以外が 0 である $\boldsymbol{\mu}_1 = (1, \dots, 1, 0, \dots, 0)^T$ とし, $\boldsymbol{\mu}_2 = \mathbf{0}$ とする. ただし, $\lceil x \rceil$ は x 以上の最小の整数を表す. 共分散行列とサンプル数について, 次の 4 つの場合を考える.

(a) 母集団 (i), $\Sigma_1 = 1.2\Phi_1, \Sigma_2 = \Sigma_1, (n_1, n_2) = (15, 5)$

(b) 母集団 (ii), $\Sigma_1 = 1.2\Phi_1, \Sigma_2 = \Sigma_1, (n_1, n_2) = (15, 5)$

(c) 母集団 (i), $\Sigma_1 = \Phi_1, \Sigma_2 = 2\Sigma_1, (n_1, n_2) = (10, 10)$

(d) 母集団 (ii), $\Sigma_1 = \Phi_1, \Sigma_2 = 2\Sigma_1, (n_1, n_2) = (10, 10)$

但し, $\Phi_1 = \mathbf{B}(0.3^{|i-j|^{1/3}})\mathbf{B}$, $\mathbf{B} = \text{diag}(\{0.5 + 1/(d+1)\}^{1/2}, \dots, \{0.5 + d/(d+1)\}^{1/2})$ とする. 次元を $d = 2^s, s = 5, \dots, 10$ と設定する. 上記の設定はいずれも (A-iii) を満たさないことに注意する. DWD について, 各設定で, $\mathbf{x}_0 \in \pi_1$ 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 $e(1), e(2)$ と全体の誤判別の割合 $e = (e(1) + e(2))/2$ を纏めたものが図 1 である. 標準誤差は 0.011 以下であることを注意する.

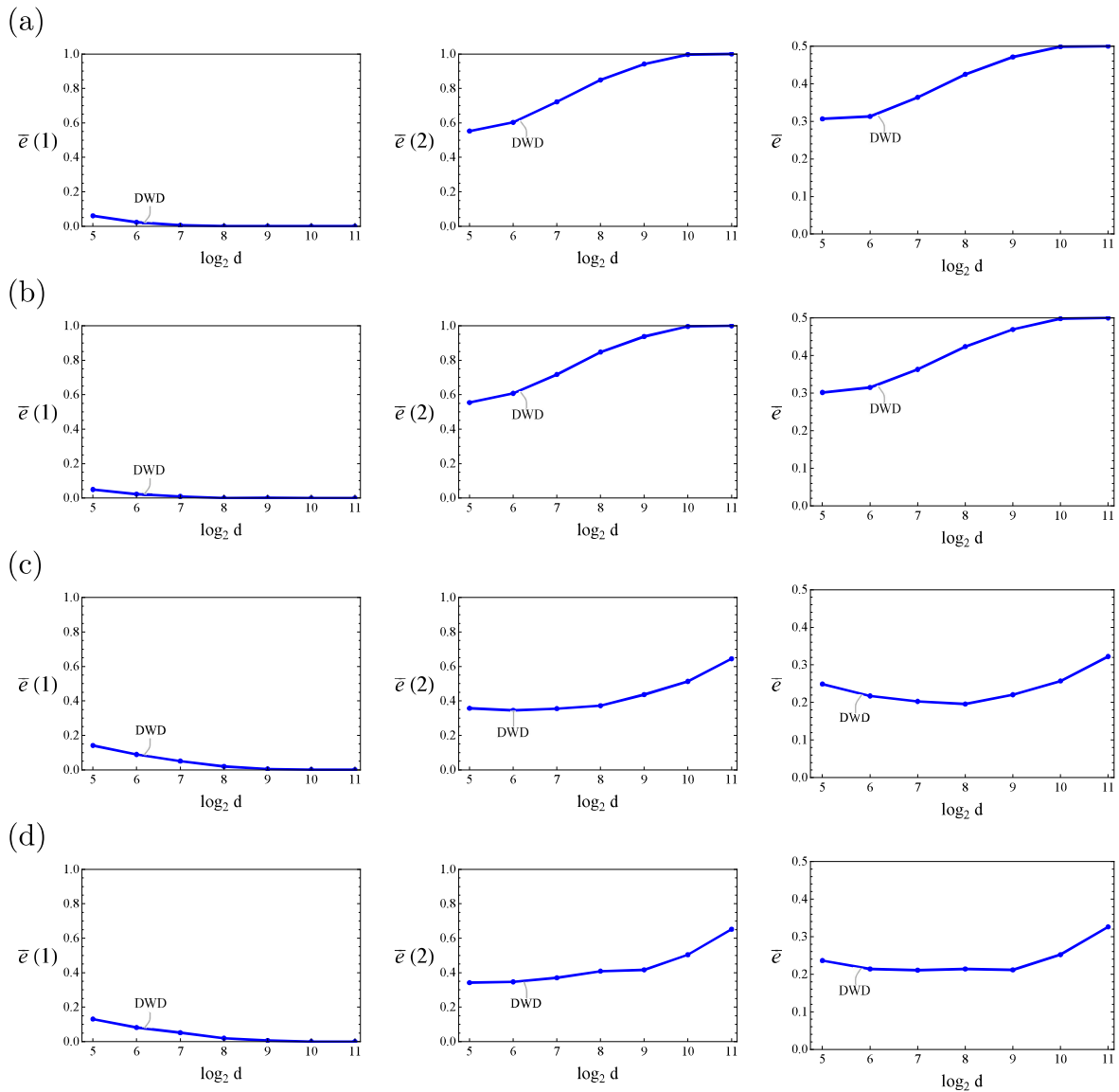


図1 (a), (b), (c), (d) のシミュレーション結果. 左からそれぞれ $\bar{e}(1)$, $\bar{e}(2)$, \bar{e} をプロットしている.

図1のように, (A-iii) を満たさない場合は, アンバランスな結果を与えることが数値的に確認された. 系1や本シミュレーションから, 高次元小標本の枠組みにおいて, DWDをそのまま用いることは推奨されない.

5 付録

以降, $\boldsymbol{\mu}_{12} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2$ と表記する.

補題 1 の証明. (A-i) と (A-ii) を仮定する. $d \rightarrow \infty$ のもと,

$$\frac{\boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}}{\Delta^2} \leq \frac{\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}}{\Delta} = o(1), \quad i = 1, 2 \quad (7)$$

であり, チェビシエフの不等式より, 任意の $\tau > 0$ に対して, 以下が成立する.

$$\begin{aligned} & P(|(\boldsymbol{x}_j - \boldsymbol{\mu})^T (\boldsymbol{x}_l - \boldsymbol{\mu}) - \Delta/4| \geq \tau \Delta) \\ & \leq \frac{E\{[(\boldsymbol{x}_j - \boldsymbol{\mu})^T (\boldsymbol{x}_l - \boldsymbol{\mu}) - \Delta/4]^2\}}{(\tau \Delta)^2} \\ & = O\{\text{tr}(\boldsymbol{\Sigma}_1^2) + \boldsymbol{\mu}_{12}^T \boldsymbol{\Sigma}_1 \boldsymbol{\mu}_{12}\} / \Delta^2 = o(1), \quad 1 \leq j < l \leq n_1, \\ & P(|(\boldsymbol{x}_j - \boldsymbol{\mu})^T (\boldsymbol{x}_l - \boldsymbol{\mu}) - \Delta/4| \geq \tau \Delta) \\ & = O\{\text{tr}(\boldsymbol{\Sigma}_2^2) + \boldsymbol{\mu}_{12}^T \boldsymbol{\Sigma}_2 \boldsymbol{\mu}_{12}\} / \Delta^2 = o(1), \quad n_1 + 1 \leq j < l \leq N, \\ & P(|(\boldsymbol{x}_j - \boldsymbol{\mu})^T (\boldsymbol{x}_l - \boldsymbol{\mu}) + \Delta/4| \geq \tau \Delta) \\ & = O\{\text{tr}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2) + \boldsymbol{\mu}_{12}^T (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \boldsymbol{\mu}_{12}\} / \Delta^2 = o(1), \quad 1 \leq j \leq n_1, n_1 + 1 \leq l \leq N \quad (8) \end{aligned}$$

但し, $\text{tr}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2) \leq \{\text{tr}(\boldsymbol{\Sigma}_1^2) \text{tr}(\boldsymbol{\Sigma}_2^2)\}^{1/2}$ であることに注意する. (7) より, 任意の $\tau > 0$ に対して, 以下が成立する.

$$\begin{aligned} & P(\|\boldsymbol{x}_j - \boldsymbol{\mu}\|^2 - \Delta/4 - \text{tr}(\boldsymbol{\Sigma}_1) \geq \tau \Delta) \\ & = O\{\text{Var}(\|\boldsymbol{x}_{1j} - \boldsymbol{\mu}_1\|^2) + \boldsymbol{\mu}_{12}^T \boldsymbol{\Sigma}_1 \boldsymbol{\mu}_{12}\} / \Delta^2 = o(1), \quad j = 1, \dots, n_1, \\ & P(\|\boldsymbol{x}_j - \boldsymbol{\mu}\|^2 - \Delta/4 - \text{tr}(\boldsymbol{\Sigma}_2) \geq \tau \Delta) = o(1), \quad j = n_1 + 1, \dots, N \quad (9) \end{aligned}$$

ここで, (4) のもと, 以下を得ることができる.

$$L(\boldsymbol{\alpha}) = 2 \sum_{j=1}^N \sqrt{\alpha_j} - \sqrt{\sum_{j=1}^N \sum_{l=1}^N \alpha_j \alpha_l y_j y_l (\boldsymbol{x}_j - \boldsymbol{\mu})^T (\boldsymbol{x}_l - \boldsymbol{\mu})}$$

従って, (4) のもと, すべての j について $\alpha_j \geq 0$ であることに注意すれば, (8) と (9) より,

$$L(\boldsymbol{\alpha}) = 2 \sum_{j=1}^N \sqrt{\alpha_j} - \sqrt{\frac{\Delta}{4} \left(\sum_{j=1}^N \alpha_j \right)^2 \{1 + o_p(1)\} + \left(\text{tr}(\boldsymbol{\Sigma}_1) \sum_{j=1}^{n_1} \alpha_j^2 + \text{tr}(\boldsymbol{\Sigma}_2) \sum_{j=n_1+1}^N \alpha_j^2 \right)}$$

が得られ, 補題 1 を示すことができる. \square

補題 2 の証明. Nakayama et al. (2017) と同様にして, 補題 1 と (5) より, 結果を得る. □

定理 1 と系 1 の証明. (6) より, 結果を得る. □

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 挑戦的研究 (萌芽) 19K22837 研究代表者: 青嶋 誠「超高次元データによる個別化モデリングへの挑戦」, および, 科学研究費補助金 基盤研究 (C) 18K03409 研究代表者: 矢田 和善「高次元データにおける高次漸近理論の開拓とその応用」から研究助成を受けています.

参考文献

- [1] Aoshima, M., Yata, K. (2011). Two-stage procedures for high-dimensional data. *Sequential Analysis (Editor's special invited paper)*, 30, 356–399.
- [2] Aoshima, M., Yata, K. (2014). A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data. *Annals of the Institute of Statistical Mathematics*, 66, 983–1010.
- [3] Aoshima, M., Yata, K. (2015). Geometric classifier for multiclass, high-dimensional data. *Sequential Analysis*, 34, 279–294.
- [4] Aoshima, M., Yata, K. (2019). High-dimensional quadratic classifiers in non-sparse settings. *Methodology and Computing in Applied Probability* 21, 663–682.
- [5] Bickel, P.J., 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., Hall, P. (2009). Scale adjustments for classifiers in high-dimensional, low sample size settings. *Biometrika*, 96, 469–478.
- [7] Dudoit, S., Fridlyand, J., Speed, T.P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data, *Journal of the American Statistical Association*, 97, 77–87.
- [8] Marron, J.S., Todd, M.J., Ahn, J. (2007). Distance-weighted discrimination. *Journal of the American Statistical Association*, 102, 1267–1271.

- [9] Nakayama, Y., Yata, K., Aoshima, M. (2017). Support vector machine and its bias correction in high-dimension, low-sample-size settings. *Journal of Statistical Planning and Inference*, 191, 88–100.
- [10] Nakayama, Y., Yata, K., Aoshima, M. (2019). Bias-corrected support vector machine with Gaussian kernel in high-dimension, low-sample-size settings. *Annals of the Institute of Statistical Mathematics*, in press.
- [11] Qiao, X., Zhang, H. H., Liu, Y., Todd, M.J., Marron, J.S. (2010). Weighted distance weighted discrimination and its asymptotic properties. *Journal of the American Statistical Association*, 105, 401–414.
- [12] Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory (second ed.)*. New York: Springer.