

# Geometric Classifier for Multiclass, High-Dimensional Data

著者別名	青嶋 誠, 矢田 和善
journal or publication title	Sequential analysis
volume	34
number	3
page range	279-294
権利	(C) Taylor & Francis Group, LLC This is an Accepted Manuscript of an article published by Taylor & Francis in Sequential Analysis on Aug 2015 available online: <a href="http://www.tandfonline.com/10.1080/07474946.2015.1063256">http://www.tandfonline.com/10.1080/07474946.2015.1063256</a>
URL	<a href="http://hdl.handle.net/2241/00128944">http://hdl.handle.net/2241/00128944</a>

doi: 10.1080/07474946.2015.1063256

# Geometric Classifier for Multiclass, High-Dimensional Data

**Makoto Aoshima and Kazuyoshi Yata**  
Institute of Mathematics, University of Tsukuba  
Ibaraki, Japan

**Abstract:** In this paper, we consider a geometric classifier which is applicable to multiclass classification for high-dimensional data. We show the consistency property and the asymptotic normality of the geometric classifier under certain mild conditions. We discuss sample size determination so that the geometric classifier can ensure its misclassification rates are less than prespecified thresholds. We give a two-stage procedure to estimate the sample sizes required in such the geometric classifier and propose a *misclassification rate adjusted classifier (MRAC)* based on the geometric classifier. We evaluate the performance of the MRAC theoretically and numerically. Finally, we demonstrate the MRAC in actual data analyses by using a microarray data set.

**Keywords:** Asymptotic normality; Geometric classifier; HDLSS; Sample size determination; Two-stage procedure.

**Subject Classifications:** 62H30; 62H10; 62L10.

## 1. INTRODUCTION

High-dimensional data situations occur in many areas of modern science such as genetic microarrays, medical imaging, text recognition, finance, chemometrics, and so on. A common feature of high-dimensional data is that the data dimension is high, however, the sample size is relatively low. This is the so-called “HDLSS” or “large  $p$ , small  $n$ ” situation where  $p/n \rightarrow \infty$ ; here  $p$  is the data dimension and  $n$  is the sample size. Aoshima and Yata (2011a,b) provided a variety of statistical inference for high-dimensional data such as given-bandwidth confidence region, two-sample test, classification, variable selection, regression, pathway analysis and so on. They considered sample size determination to ensure prespecified high accuracy for high-dimensional, non-Gaussian inference and developed the theory of Stein (1945, 1949)’s two-stage procedure which was originally given for inference on the univariate Gaussian mean. Aoshima and Yata (2015) verified the asymptotic normality of statistics appearing in inference on high-dimensional mean vectors under certain mild conditions. In this paper, we focus on high-dimensional classification and make an attempt to give a multiclass classifier to hold misclassification rates less than prespecified thresholds.

Suppose we have independent and  $p$ -variate populations,  $\pi_i$ ,  $i = 1, \dots, k$ , having an unknown mean vector  $\boldsymbol{\mu}_i$  and unknown covariance matrix  $\boldsymbol{\Sigma}_i (> \mathbf{O})$  for each  $i$ . We assume that  $\limsup_{p \rightarrow \infty} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2/p < \infty$  for all  $i \neq j$ , where  $\|\cdot\|$  denotes the Euclidean norm. Also, we assume that  $\text{tr}(\boldsymbol{\Sigma}_i)/p \in (0, \infty)$  as  $p \rightarrow \infty$  for  $i = 1, \dots, k$ . Here, for a function,  $f(\cdot)$ , “ $f(p) \in (0, \infty)$  as  $p \rightarrow \infty$ ” implies  $\liminf_{p \rightarrow \infty} f(p) > 0$  and  $\limsup_{p \rightarrow \infty} f(p) < \infty$ . We do not assume that  $\boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_k$ . The eigen-decomposition of  $\boldsymbol{\Sigma}_i$  is given by  $\boldsymbol{\Sigma}_i = \mathbf{H}_i \boldsymbol{\Lambda}_i \mathbf{H}_i^T$ , where  $\boldsymbol{\Lambda}_i$  is a diagonal matrix of eigenvalues,  $\lambda_{i1} \geq \dots \geq \lambda_{ip} > 0$ , and  $\mathbf{H}_i$  is an orthogonal matrix of the corresponding eigenvectors. We have independent and identically distributed (i.i.d.) observations,  $\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}$ , from each  $\pi_i$ . Let  $\mathbf{x}_{ij} = \mathbf{H}_i \boldsymbol{\Lambda}_i^{1/2} \mathbf{z}_{ij} + \boldsymbol{\mu}_i$ , where  $\mathbf{z}_{ij}$  is

---

Address correspondence to Makoto Aoshima, Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan; Fax: +81-298-53-6501; E-mail: aoshima@math.tsukuba.ac.jp

considered as a sphered data vector from a distribution with the zero mean vector and the identity covariance matrix. We assume  $n_i \geq 2$ ,  $i = 1, \dots, k$ . We estimate  $\boldsymbol{\mu}_i$  and  $\boldsymbol{\Sigma}_i$  by  $\bar{\boldsymbol{x}}_{in_i} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}/n_i$  and  $\boldsymbol{S}_{in_i} = \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{in_i})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{in_i})^T/(n_i - 1)$ .

As for population  $\pi_i$ ,  $i = 1, \dots, k$ , we make the following assumption:

**(A-i)** Let  $\boldsymbol{y}_{ij}$ ,  $j = 1, \dots, n_i$ , be i.i.d. random  $q_i$ -vectors having  $E(\boldsymbol{y}_{ij}) = \mathbf{0}$  and  $\text{Var}(\boldsymbol{y}_{ij}) = \boldsymbol{I}_{q_i}$  for each  $i$  ( $= 1, \dots, k$ ), where  $q_i \geq p$ . Let  $\boldsymbol{y}_{ij} = (y_{i1j}, \dots, y_{iq_i j})^T$  in which  $\limsup_{p \rightarrow \infty} E(y_{irj}^4) < \infty$  for all  $r$ ,

$$E(y_{irj}^2 y_{isj}^2) = E(y_{irj}^2) E(y_{isj}^2) = 1 \quad \text{and} \quad E(y_{irj} y_{isj} y_{itj} y_{iu j}) = 0$$

for all  $r \neq s, t, u$ . Then, the observations,  $\boldsymbol{x}_{ijs}$ , from each  $\pi_i$  ( $i = 1, \dots, k$ ) are given by

$$\boldsymbol{x}_{ij} = \boldsymbol{\Gamma}_i \boldsymbol{y}_{ij} + \boldsymbol{\mu}_i, \quad j = 1, \dots, n_i, \quad (1.1)$$

where  $\boldsymbol{\Gamma}_i$  is a  $p \times q_i$  matrix such that  $\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i^T = \boldsymbol{\Sigma}_i$ .

Here,  $\boldsymbol{I}_{q_i}$  denotes the identity matrix of dimension  $q_i$ . Note that (1.1) includes the case that  $\boldsymbol{\Gamma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i^{1/2}$  and  $\boldsymbol{y}_{ij} = \boldsymbol{z}_{ij}$ . Also, note that (A-i) is met when  $\pi_i$ s have  $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  for  $i = 1, \dots, k$ . In addition, we assume the following assumptions for  $\boldsymbol{\Sigma}_i$ s as necessary:

**(A-ii)**  $\frac{\text{tr}(\boldsymbol{\Sigma}_i^4)}{\text{tr}(\boldsymbol{\Sigma}_i^2)^2} \rightarrow 0$  and  $\frac{\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_l)}{\text{tr}(\boldsymbol{\Sigma}_j^2)} \in (0, \infty)$  as  $p \rightarrow \infty$  for  $i, j, l = 1, \dots, k$ .

Note that “ $\text{tr}(\boldsymbol{\Sigma}_i^4)/\text{tr}(\boldsymbol{\Sigma}_i^2)^2 \rightarrow 0$  as  $p \rightarrow \infty$ ” is equivalent to the condition that “ $\lambda_{i1}/\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2} \rightarrow 0$  as  $p \rightarrow \infty$ ”. Also, the sphericity condition such as “ $\text{tr}(\boldsymbol{\Sigma}_i^2)/\text{tr}(\boldsymbol{\Sigma}_i)^2 \rightarrow 0$  as  $p \rightarrow \infty$  for  $i = 1, \dots, k$ ” holds under (A-ii).

**Remark 1.1.** If all  $\lambda_{ij}$ s are bounded such as  $\lambda_{ij} \in (0, \infty)$  as  $p \rightarrow \infty$ , (A-ii) trivially holds. For a spiked model such as  $\lambda_{ij} = a_{ij} p^{\alpha_{ij}}$  ( $j = 1, \dots, t_i$ ) and  $\lambda_{ij} = c_{ij}$  ( $j = t_i + 1, \dots, p$ ) with positive constants,  $a_{ij}$ s,  $c_{ij}$ s and  $\alpha_{ij}$ s, and positive integers  $t_i$ s, (A-ii) holds under the condition that  $\alpha_{ij} < 1/2$  for  $j = 1, \dots, t_i$  ( $< \infty$ );  $i = 1, \dots, k$ .

Let  $\boldsymbol{x}_0$  be an observation vector of an individual belonging to one of the  $k$  populations. When  $k = 2$ , a typical classification rule is that one classifies the individual into  $\pi_1$  if

$$(\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_{1n_1})^T \boldsymbol{S}_{1n_1}^{-1} (\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_{1n_1}) - \log \left\{ \frac{\det(\boldsymbol{S}_{2n_2})}{\det(\boldsymbol{S}_{1n_1})} \right\} < (\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_{2n_2})^T \boldsymbol{S}_{2n_2}^{-1} (\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_{2n_2}),$$

and into  $\pi_2$  otherwise. However, the inverse matrix of  $\boldsymbol{S}_{in_i}$  does not exist in the HDLSS context ( $p > n_i$ ). Dudoit et al. (2002) considered substituting the inverse matrix defined by only diagonal elements of  $\boldsymbol{S}_{in_i}$ . Chan and Hall (2009) and Aoshima and Yata (2014a) considered distance-based classifiers. Particularly, Aoshima and Yata (2014a) gave a distance-based classifier for multiclass, non-Gaussian high dimensional data and considered sample size determination to hold misclassification rates less than prespecified thresholds. When  $k = 2$ , the distance-based classifier is simplified as follows: One classifies the individual into  $\pi_1$  if

$$\left( \boldsymbol{x}_0 - \frac{\bar{\boldsymbol{x}}_{1n_1} + \bar{\boldsymbol{x}}_{2n_2}}{2} \right)^T (\bar{\boldsymbol{x}}_{2n_2} - \bar{\boldsymbol{x}}_{1n_1}) - \frac{\text{tr}(\boldsymbol{S}_{1n_1})}{2n_1} + \frac{\text{tr}(\boldsymbol{S}_{2n_2})}{2n_2} < 0 \quad (1.2)$$

and into  $\pi_2$  otherwise. Here,  $-\text{tr}(\boldsymbol{S}_{1n_1})/(2n_1) + \text{tr}(\boldsymbol{S}_{2n_2})/(2n_2)$  is a bias-correction term. Aoshima and Yata (2014a) showed that the classifier holds a consistency property in which misclassification rates go to zero as  $p \rightarrow \infty$  even when (A-i) is not met. In that sense, the classifier is quite robust and applicable

to actual high-dimensional data. On the other hand, Aoshima and Yata (2011a) considered substituting  $\{\text{tr}(\mathbf{S}_{in_i})/p\}\mathbf{I}_p$  for  $\mathbf{S}_{in_i}$  in order to use a geometric representation of HDLSS data from each  $\pi_i$  and gave a two-class quadratic classifier called the *geometric classifier* as follows: One classifies the individual into  $\pi_1$  if

$$\frac{p\|\mathbf{x}_0 - \bar{\mathbf{x}}_{1n_1}\|^2}{\text{tr}(\mathbf{S}_{1n_1})} - \frac{p\|\mathbf{x}_0 - \bar{\mathbf{x}}_{2n_2}\|^2}{\text{tr}(\mathbf{S}_{2n_2})} - p \log \left\{ \frac{\text{tr}(\mathbf{S}_{2n_2})}{\text{tr}(\mathbf{S}_{1n_1})} \right\} - \frac{p}{n_1} + \frac{p}{n_2} < 0 \quad (1.3)$$

and into  $\pi_2$  otherwise. Here,  $-p/n_1 + p/n_2$  is a bias-correction term. Aoshima and Yata (2014a,b) showed that the classifier holds the consistency property even when  $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ . Recently, Aoshima and Yata (2014b) provided a general theory of quadratic classifiers for high-dimensional data in non-sparse settings.

In this paper, we develop the geometric classifier by (1.3) to multiclass classification when  $k (\geq 2)$ . In Section 2, we show the consistency property and the asymptotic normality of the geometric classifier for multiclass high-dimensional data. In Section 3, we discuss sample size determination so that the geometric classifier can ensure its misclassification rates are less than prespecified thresholds. We give a two-stage procedure to estimate the sample sizes required in such the geometric classifier and propose a *misclassification rate adjusted classifier (MRAC)* based on the geometric classifier. In Section 4, we evaluate the performance of the MRAC numerically as well. Finally, in Section 5, we demonstrate the MRAC in actual data analyses by using a microarray data set.

## 2. ASYMPTOTIC PROPERTIES OF THE GEOMETRIC CLASSIFIER

Let

$$W_i(\mathbf{x}_0|n_i) = \frac{p\|\mathbf{x}_0 - \bar{\mathbf{x}}_{in_i}\|^2}{\text{tr}(\mathbf{S}_{in_i})} - \frac{p}{n_i} + p \log \{\text{tr}(\mathbf{S}_{in_i})\} \quad (2.1)$$

for  $i = 1, \dots, k$ . We consider the geometric classifier when  $k (\geq 2)$  as follows: One classifies the individual into  $\pi_i$  if

$$\max \left\{ \underset{j=1, \dots, k}{\text{argmin}} W_j(\mathbf{x}_0|n_j) \right\} = i. \quad (2.2)$$

When  $\underset{j=1, \dots, k}{\text{argmin}} W_j(\mathbf{x}_0|n_j) = \{i_1, \dots, i_l\}$  with integers  $l \in [2, k]$  and  $i_1 < \dots < i_l$ , we have  $\max \{\underset{j=1, \dots, k}{\text{argmin}} W_j(\mathbf{x}_0|n_j)\} = i_l$ . Note that the difference,  $W_1(\mathbf{x}_0|n_1) - W_2(\mathbf{x}_0|n_2)$ , is equivalent to (1.3).

### 2.1. Consistency Property

Let  $\Delta_{ij(1)} = \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2$  and  $\Delta_{ij(2)} = \text{tr}(\boldsymbol{\Sigma}_i) - \text{tr}(\boldsymbol{\Sigma}_j) + \text{tr}(\boldsymbol{\Sigma}_j) \log \{\text{tr}(\boldsymbol{\Sigma}_j)/\text{tr}(\boldsymbol{\Sigma}_i)\}$  for all  $i \neq j$ . Note that  $\Delta_{ij(2)} \geq 0$  ( $i \neq j$ ) with equality if and only if  $\text{tr}(\boldsymbol{\Sigma}_i) = \text{tr}(\boldsymbol{\Sigma}_j)$ . Let

$$\Delta_{ij} = \frac{p}{\text{tr}(\boldsymbol{\Sigma}_j)} (\Delta_{ij(1)} + \Delta_{ij(2)})$$

for all  $i \neq j$ . We assume the followings as  $p \rightarrow \infty$  either when  $n_i$  is fixed or  $n_i \rightarrow \infty$  for  $i = 1, \dots, k$ :

$$\text{(A-iii)} \quad \frac{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_i (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{\Delta_{ij}^2} = o(1) \quad \text{and} \quad \frac{\text{tr}(\boldsymbol{\Sigma}_i^2) \text{tr}(\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_j)^2}{\text{tr}(\boldsymbol{\Sigma}_i)^2 \Delta_{ij}^2} = o(1) \quad \text{for all } i \neq j;$$

$$\text{(A-iv)} \quad \frac{\max \{\text{tr}(\boldsymbol{\Sigma}_i^2), \text{tr}(\boldsymbol{\Sigma}_j^2)\}}{\min \{n_i, n_j\} \Delta_{ij}^2} = o(1) \quad \text{for all } i \neq j.$$

We denote the error rate of misclassifying an individual from  $\pi_i$  (into another class) by  $e(i)$ . Then, we have the following result.

**Theorem 2.1.** *Under (A-i), (A-iii) and (A-iv), it holds that as  $p \rightarrow \infty$*

$$\frac{W_j(\mathbf{x}_0|n_i) - W_i(\mathbf{x}_0|n_j)}{\Delta_{ij}} = 1 + o_P(1) \text{ when } \mathbf{x}_0 \in \pi_i \text{ for all } i \neq j;$$

$$e(i) \rightarrow 0 \text{ for } i = 1, \dots, k.$$

**Remark 2.1.** When  $k = 2$ , Aoshima and Yata (2014a) gave partial results of Theorem 2.1 under different conditions.

**Remark 2.2.** If  $\max_{i' \neq i, \dots, k} \{\text{tr}(\boldsymbol{\Sigma}_{i'}^2)\} / \Delta_{ij}^2 \rightarrow 0$  as  $p \rightarrow \infty$  for all  $i \neq j$ , (A-iii) and (A-iv) naturally hold. Then, one can claim Theorem 2.1 even when  $n_i$  is fixed for  $i = 1, \dots, k$ .

## 2.2. Asymptotic Normality

Let

$$\widetilde{W}(\mathbf{x}_0|n_i, n_j) = \frac{p\|\mathbf{x}_0 - \bar{\mathbf{x}}_{in_i}\|^2}{\text{tr}(\boldsymbol{\Sigma}_i)} - \frac{p\|\mathbf{x}_0 - \bar{\mathbf{x}}_{jn_j}\|^2}{\text{tr}(\boldsymbol{\Sigma}_j)} - p \log \left\{ \frac{\text{tr}(\boldsymbol{\Sigma}_j)}{\text{tr}(\boldsymbol{\Sigma}_i)} \right\} - \frac{p \text{tr}(\mathbf{S}_{in_i})}{\text{tr}(\boldsymbol{\Sigma}_i)n_i} + \frac{p \text{tr}(\mathbf{S}_{jn_j})}{\text{tr}(\boldsymbol{\Sigma}_j)n_j}$$

for all  $i \neq j$ . Note that  $W_i(\mathbf{x}_0|n_i) - W_j(\mathbf{x}_0|n_j)$  is equivalent to  $\widetilde{W}(\mathbf{x}_0|n_i, n_j)$  with  $\boldsymbol{\Sigma}_i = \mathbf{S}_{in_i}$  and  $\boldsymbol{\Sigma}_j = \mathbf{S}_{jn_j}$  for all  $i \neq j$ . We have that  $E\{\widetilde{W}(\mathbf{x}_0|n_i, n_j)\} = -\Delta_{ij}$  when  $\mathbf{x}_0 \in \pi_i$  for all  $i \neq j$ . Under (A-i), it holds that

$$\begin{aligned} \frac{\text{tr}(\boldsymbol{\Sigma}_j)^2}{4p^2} \text{Var}\{\widetilde{W}(\mathbf{x}_0|n_i, n_j)\} &= \frac{\text{tr}(\boldsymbol{\Sigma}_j)^2}{\text{tr}(\boldsymbol{\Sigma}_i)^2} \left( \frac{\text{tr}(\boldsymbol{\Sigma}_i^2)}{n_i} + \frac{\text{tr}(\boldsymbol{\Sigma}_i^2)}{2n_i(n_i - 1)} \right) + \frac{\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)}{n_j} + \frac{\text{tr}(\boldsymbol{\Sigma}_j^2)}{2n_j(n_j - 1)} \\ &\quad + (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T (\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j/n_j) (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + O\left( \frac{\text{tr}(\boldsymbol{\Sigma}_i^2) \text{tr}(\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_j)^2}{\text{tr}(\boldsymbol{\Sigma}_i)^2} \right) \end{aligned}$$

when  $\mathbf{x}_0 \in \pi_i$  for all  $i \neq j$ . Let

$$\delta_{ij} = \frac{2p}{\text{tr}(\boldsymbol{\Sigma}_j)} \left\{ \frac{\text{tr}(\boldsymbol{\Sigma}_j)^2}{\text{tr}(\boldsymbol{\Sigma}_i)^2} \left( \frac{\text{tr}(\boldsymbol{\Sigma}_i^2)}{n_i} + \frac{\text{tr}(\boldsymbol{\Sigma}_i^2)}{2n_i(n_i - 1)} \right) + \frac{\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)}{n_j} + \frac{\text{tr}(\boldsymbol{\Sigma}_j^2)}{2n_j(n_j - 1)} \right\}^{1/2}$$

for all  $i \neq j$ . We assume extra assumptions as  $p \rightarrow \infty$  and  $n_i \rightarrow \infty$ ,  $i = 1, \dots, k$ :

$$\text{(A-v)} \quad \frac{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_i (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{\delta_{ij}^2} = o(1) \quad \text{and} \quad \frac{\text{tr}(\boldsymbol{\Sigma}_i^2) \text{tr}(\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_j)^2}{\text{tr}(\boldsymbol{\Sigma}_i)^2 \delta_{ij}^2} = o(1) \quad \text{for all } i \neq j.$$

Note that under (A-ii) it holds  $\delta_{ij} = O\{\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}\}$  for all  $i \neq j$ , so that  $\text{tr}(\boldsymbol{\Sigma}_i)/\text{tr}(\boldsymbol{\Sigma}_j) \rightarrow 1$  as  $p \rightarrow \infty$  for all  $i \neq j$  under (A-ii) and (A-v). Then, we have the following results.

**Theorem 2.2.** *Assume that  $\Delta_{ij(1)}/\text{tr}(\boldsymbol{\Sigma}_j) \rightarrow 0$  as  $p \rightarrow \infty$  for all  $i \neq j$ . Under (A-i), (A-ii) and (A-v), it holds that as  $p \rightarrow \infty$  and  $n_i \rightarrow \infty$ ,  $i = 1, \dots, k$*

$$\frac{W_i(\mathbf{x}_0|n_i) - W_j(\mathbf{x}_0|n_j) + \Delta_{ij}}{\delta_{ij}} \Rightarrow Y_{ij} \quad \text{when } \mathbf{x}_0 \in \pi_i \text{ for all } i \neq j,$$

where “ $\Rightarrow$ ” denotes the convergence in distribution and  $Y_{ij}$  denotes a random variable distributed as the standard normal distribution.

**Remark 2.3.** When  $k = 2$ , Aoshima and Yata (2011a) gave the asymptotic normality under some stronger conditions.

**Corollary 2.1.** Assume that  $\Delta_{ij(1)}/\text{tr}(\Sigma_j) \rightarrow 0$  as  $p \rightarrow \infty$  for all  $i \neq j$ . Under (A-i), (A-ii) and (A-v), the classification rule by (2.2) has that as  $p \rightarrow \infty$  and  $n_i \rightarrow \infty$ ,  $i = 1, \dots, k$

$$e(i) \leq \sum_{j(\neq i)=1}^k \Phi(-\Delta_{ij}/\delta_{ij}) + o(1) \quad \text{for } i = 1, \dots, k,$$

where  $\Phi(\cdot)$  denotes the cumulative distribution function of the standard normal distribution.

**Remark 2.4.** When  $k = 2$ , the above result is given as

$$e(1) = \Phi(-\Delta_{12}/\delta_{12}) + o(1) \quad \text{and} \quad e(2) = \Phi(-\Delta_{21}/\delta_{21}) + o(1).$$

### 3. SAMPLE SIZE DETERMINATION TO CONTROL MISCLASSIFICATION RATES

Let  $\Delta_{ij*} = \{\text{tr}(\Sigma_j)/p\}\Delta_{ij} = \Delta_{ij(1)} + \Delta_{ij(2)}$  for all  $i \neq j$ . Let  $\Delta_{i*} = \min_{j(\neq i)=1, \dots, k} \min\{\Delta_{ij*}, \Delta_{ji*}\}$  for  $i = 1, \dots, k$ . We are interested in determining the sample size for (2.2) to ensure the requirement:

$$e(i) \leq \alpha_i \quad \text{whenever } \Delta_{i*} \geq \Delta_{i*L} \quad \text{for } i = 1, \dots, k,$$

where  $\alpha_i \in (0, 1/2)$  and  $\Delta_{i*L} (> 0)$   $i = 1, \dots, k$ , are prespecified constants. We assume  $\Delta_{i*L} = o\{\text{tr}(\Sigma_i^2)^{1/2}\}$ ,  $i = 1, \dots, k$ .

#### 3.1. Sample Size Determination

Let  $z_\alpha$  be the upper  $\alpha$  point of the standard normal distribution. We consider  $n_i$ s satisfying

$$\delta_{ij} \leq \frac{2\Delta_{(ij)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}} \quad (3.1)$$

for all  $i \neq j$ , where  $\Delta_{(ij)} = p \max\{\Delta_{i*L}, \Delta_{j*L}\} / \max\{\text{tr}(\Sigma_i), \text{tr}(\Sigma_j)\}$  ( $i \neq j$ ). Note that  $\Delta_{(ij)} = \Delta_{(ji)}$  and  $\Delta_{(ij)} \leq \min\{\Delta_{ij}, \Delta_{ji}\}$  for all  $i \neq j$ . Under (3.1), we have that

$$\begin{aligned} \Delta_{ij} + \Delta_{(ij)} \frac{z_{\alpha_i/(k-1)} - z_{\alpha_j/(k-1)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}} &\geq \Delta_{(ij)} \left( 1 + \frac{z_{\alpha_i/(k-1)} - z_{\alpha_j/(k-1)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}} \right) = \frac{2z_{\alpha_i/(k-1)}\Delta_{(ij)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}} \\ &\geq z_{\alpha_i/(k-1)}\delta_{ij}; \\ \Delta_{ji} - \Delta_{(ij)} \frac{z_{\alpha_i/(k-1)} - z_{\alpha_j/(k-1)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}} &\geq \frac{2z_{\alpha_j/(k-1)}\Delta_{(ij)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}} \geq z_{\alpha_j/(k-1)}\delta_{ji}, \end{aligned}$$

so that from Theorem 2.2 it follows that for  $i = 1, \dots, k$

$$\sum_{i(\neq j)=1}^k P\left(W_i(\mathbf{x}_0|n_i) - W_j(\mathbf{x}_0|n_j) \geq \Delta_{(ij)} \frac{z_{\alpha_i/(k-1)} - z_{\alpha_j/(k-1)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}}\right) \leq \alpha_i + o(1) \quad \text{when } \mathbf{x}_0 \in \pi_i$$

under (3.1) and the assumptions of Theorem 2.2. First, we consider the case when  $\liminf_{p \rightarrow \infty} |\text{tr}(\Sigma_i)/\text{tr}(\Sigma_j) - 1| > 0$  for  $i \neq j$ . In the case, it holds  $\liminf_{p \rightarrow \infty} \Delta_{ij}/p > 0$ . Under (A-i) and (A-ii), from Theorem 2.1

we have that

$$\begin{aligned} & P\left(W_i(\mathbf{x}_0|n_i) - W_j(\mathbf{x}_0|n_j) \geq \Delta_{(ij)} \frac{z_{\alpha_i/(k-1)} - z_{\alpha_j/(k-1)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}}\right) \\ &= P\left(-1 + o_P(1) \geq \frac{\Delta_{(ij)} z_{\alpha_i/(k-1)} - z_{\alpha_j/(k-1)}}{\Delta_{ij} z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}}\right) = o(1) \quad \text{when } \mathbf{x}_0 \in \pi_i \end{aligned}$$

even if  $n_i$ s are fixed for  $i \neq j$ . Next, we consider the case when  $\text{tr}(\mathbf{\Sigma}_1) = \dots = \text{tr}(\mathbf{\Sigma}_k)$ . Let  $\sigma_i = \text{tr}(\mathbf{\Sigma}_i^2)^{1/2}$  for  $i = 1, \dots, k$ . From the fact that  $\text{tr}(\mathbf{\Sigma}_i \mathbf{\Sigma}_j) \leq \{\text{tr}(\mathbf{\Sigma}_i^2) \text{tr}(\mathbf{\Sigma}_j^2)\}^{1/2}$  ( $i \neq j$ ), it holds that for  $i \neq j$

$$\delta_{ij} \leq \frac{2p}{\text{tr}(\mathbf{\Sigma}_j)} \text{tr}\left(\frac{\text{tr}(\mathbf{\Sigma}_i^2)}{n_i - 1} + \frac{\text{tr}(\mathbf{\Sigma}_j^2)^{1/2} \max_{l=i,j} \text{tr}(\mathbf{\Sigma}_l^2)^{1/2}}{n_j - 1}\right)^{1/2} \leq \frac{2p}{\text{tr}(\mathbf{\Sigma}_j)} \max_{l=i,j} \sigma_l^{1/2} \left(\frac{\sigma_i}{n_i - 1} + \frac{\sigma_j}{n_j - 1}\right)^{1/2}.$$

Let us write  $\sigma_{(i)} = \max_{j(\neq i)=1, \dots, k} \sigma_j$  and  $\alpha_{(i)} = \min_{j(\neq i)=1, \dots, k} \alpha_j$  for  $i = 1, \dots, k$ . From the above arguments, we can find  $n_i$ ,  $i = 1, \dots, k$ , to satisfy (3.1) by

$$n_i \geq \frac{(z_{\alpha_i/(k-1)} + z_{\alpha_{(i)}/(k-1)})^2 \max_{l=1, \dots, k} \sigma_l}{\Delta_{i^*L}^2} \sigma_i^{1/2} (\sigma_i^{1/2} + \sigma_{(i)}^{1/2}) + 1 \quad (\text{hereafter called } C_i). \quad (3.2)$$

Note that  $n_i \rightarrow \infty$ ,  $i = 1, \dots, k$ , as  $p \rightarrow \infty$  from the fact that  $\Delta_{i^*L} = o\{\text{tr}(\mathbf{\Sigma}_i^2)^{1/2}\}$ ,  $i = 1, \dots, k$ . For example, when  $k = 2$ ,  $\text{tr}(\mathbf{\Sigma}_1) = \text{tr}(\mathbf{\Sigma}_2)$  and  $\Delta_{1^*L} = \Delta_{2^*L}$ , the smallest integer  $(n_1, n_2)$  satisfying (3.2) holds the following optimality:

$$\min \sum_{i=1}^2 n_i \quad \text{subject to} \quad \frac{2p}{\text{tr}(\mathbf{\Sigma}_j)} \max_{l=1,2} \sigma_l^{1/2} \left(\frac{\sigma_1}{n_1 - 1} + \frac{\sigma_2}{n_2 - 1}\right)^{1/2} \leq \frac{2\Delta_{(ij)}}{z_{\alpha_i} + z_{\alpha_j}} \quad \text{for } i \neq j.$$

According to (3.2), we take samples from each  $\pi_i$  and calculate  $W_i(\mathbf{x}_0|n_i)$ ,  $i = 1, \dots, k$ , in (2.1). We consider the following classification procedure based on the *misclassification rate adjusted classifier* by Aoshima and Yata (2014a):

### Misclassification rate adjusted classifier (MRAC)

**Step 1:** Set  $i = 0$ .

**Step 2:** Put  $i = i + 1$ . If  $i = k$ , go to Step 4; otherwise go to Step 3.

**Step 3:** If it holds that

$$W_i(\mathbf{x}_0|n_i) - W_j(\mathbf{x}_0|n_j) < p \frac{\max\{\Delta_{i^*L}, \Delta_{j^*L}\}}{\max\{\text{tr}(\mathbf{S}_{in_i}), \text{tr}(\mathbf{S}_{jn_j})\}} \frac{z_{\alpha_i/(k-1)} - z_{\alpha_j/(k-1)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}}$$

for all  $j = i + 1, \dots, k$ , go to Step 4; otherwise go to Step 2.

**Step 4:** Classify  $\mathbf{x}_0$  into  $\pi_i$ .

We have the following result.

**Theorem 3.1.** Under (A-i) to (A-iii), for the MRAC with (3.2), it holds that as  $p \rightarrow \infty$

$$\limsup e(i) \leq \alpha_i \quad \text{whenever } \Delta_{i^*} \geq \Delta_{i^*L} \text{ for } i = 1, \dots, k. \quad (3.3)$$

### 3.2. Designing a Lower Bound, $\Delta_{i^*L}$

First, we consider a lower bound of  $\Delta_{ij(1)}$ . Let  $\widehat{\Delta}_{ij(1)} = \|\bar{\mathbf{x}}_{in_i} - \bar{\mathbf{x}}_{jn_j}\|^2 - \text{tr}(\mathbf{S}_{in_i})/n_i - \text{tr}(\mathbf{S}_{jn_j})/n_j$ . By using the two sample test by Aoshima and Yata (2015), under certain regularity conditions, it holds that as  $p \rightarrow \infty$  and  $n_i \rightarrow \infty$ ,  $i = 1, \dots, k$

$$\frac{\widehat{\Delta}_{ij(1)} - \Delta_{ij(1)}}{\kappa_{ij}} \Rightarrow Y_{ij} \quad \text{for all } i \neq j,$$

where  $Y_{ij}$  denotes a random variable distributed as the standard normal distribution and

$$\kappa_{ij} = \left\{ 2 \frac{W_{in_i}}{n_i(n_i - 1)} + 2 \frac{W_{jn_j}}{n_j(n_j - 1)} + 4 \frac{\text{tr}(\mathbf{S}_{in_i} \mathbf{S}_{jn_j})}{n_i n_j} \right\}^{1/2}$$

having  $W_{in_i}$ s defined by (9) in Yata and Aoshima (2013). Here,  $W_{in_i}$  is an unbiased estimator of  $\text{tr}(\boldsymbol{\Sigma}_i^2)$  and  $\text{Var}\{W_{in_i}/\text{tr}(\boldsymbol{\Sigma}_i^2)\} \rightarrow 0$  as  $p \rightarrow \infty$  and  $n_i \rightarrow \infty$  under (A-i). See Aoshima and Yata (2014a) for the details. It follows that  $P(\widehat{\Delta}_{ij(1)} - \kappa_{ij} z_{\alpha'} \leq \Delta_{ij(1)}) \rightarrow 1 - \alpha'$  for given  $\alpha' \in (0, 1/2)$ . Thus, one may design a lower bound of  $\Delta_{ij(1)}$  by

$$\Delta_{ij(1)L} = \widehat{\Delta}_{ij(1)} - \kappa_{ij} z_{\alpha'} \quad (3.4)$$

for sufficiently small  $\alpha'$ . Next, we consider a lower bound of  $\Delta_{ij(2)}$ . For  $i \neq j$  it holds that

$$\Delta_{ij(2)} \geq \frac{\text{tr}(\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_j)^2}{2 \max\{\text{tr}(\boldsymbol{\Sigma}_i), \text{tr}(\boldsymbol{\Sigma}_j)\}}$$

with equality if and only if  $\text{tr}(\boldsymbol{\Sigma}_i) = \text{tr}(\boldsymbol{\Sigma}_j)$ . We note that as  $p \rightarrow \infty$  and  $n_i \rightarrow \infty$ ,  $i = 1, \dots, k$

$$\frac{\text{tr}(\mathbf{S}_{in_i})}{\text{tr}(\boldsymbol{\Sigma}_i)} = 1 + O_P\left\{ \frac{\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}}{n_i^{1/2} \text{tr}(\boldsymbol{\Sigma}_i)} \right\} = 1 + o_P(1)$$

under (A-i). Thus, one may design a lower bound of  $\Delta_{ij(2)}$  by

$$\Delta_{ij(2)L} = \frac{\text{tr}(\mathbf{S}_{in_i} - \mathbf{S}_{jn_j})^2}{2 \max\{\text{tr}(\mathbf{S}_{in_i}), \text{tr}(\mathbf{S}_{jn_j})\}}$$

for  $i \neq j$ . Let  $\Delta_{ij^*L} = \Delta_{ij(1)L} + \Delta_{ij(2)L}$  for all  $i \neq j$ . Note that  $\Delta_{ij^*L} = \Delta_{ji^*L}$  for  $i \neq j$ . Finally, we choose a lower bound,  $\Delta_{i^*L}$ , by  $\Delta_{i^*L} = \min_{j(\neq i)=1, \dots, k} \Delta_{ij^*L}$  for sufficiently small  $\alpha'$ .

### 3.3. Two-Stage Procedure

In order to estimate  $C_i$ s in (3.2), we proceed with the following two steps:

1. Choose  $m_i (\geq 4)$  satisfying

$$\frac{m_i}{C_i} \leq 1, \quad \frac{C_i}{m_i^2} \rightarrow 0 \quad \text{and} \quad \frac{C_i}{m_i} \frac{\text{tr}(\boldsymbol{\Sigma}_i^4)}{\text{tr}(\boldsymbol{\Sigma}_i^2)^2} \rightarrow 0 \quad \text{as } p \rightarrow \infty \quad \text{under (A-ii)} \quad (3.5)$$

for  $i = 1, \dots, k$ . Note that (3.5) holds when  $m_i/C_i \in (0, 1)$  as  $p \rightarrow \infty$ . Take pilot samples,  $\mathbf{x}_{ij}$ ,  $j = 1, \dots, m_i$ , of size  $m_i$  from each  $\pi_i$ . Then, calculate  $W_{im_i}$  for each  $\pi_i$  according to (9) in Yata and Aoshima



(2013). Let  $\hat{\sigma}_i = W_{im_i}^{1/2}$  and  $\hat{\sigma}_{(i)} = \max_{j(\neq i)=1, \dots, k} \hat{\sigma}_j$  for  $i = 1, \dots, k$ . Define the total sample size for each  $\pi_i$  by

$$N_i = \max \left\{ m_i, \left\lceil \frac{(z_{\alpha_i/(k-1)} + z_{\alpha_{(i)}/(k-1)})^2 \max_{l=1, \dots, k} \hat{\sigma}_l}{\Delta_{i^*L}^2} \hat{\sigma}_i^{1/2} (\hat{\sigma}_i^{1/2} + \hat{\sigma}_{(i)}^{1/2}) + 1 \right\rceil \right\}, \quad (3.6)$$

where  $\lceil x \rceil$  denotes the smallest integer  $\geq x$ .

2. For each  $i$ , if  $N_i = m_i$ , do not take any additional samples from  $\pi_i$  and otherwise, that is if  $N_i > m_i$ , take additional samples,  $\mathbf{x}_{ij}$ ,  $j = m_i + 1, \dots, N_i$ , of size  $N_i - m_i$  from  $\pi_i$ . By combining the initial samples and the additional samples, calculate  $\bar{\mathbf{x}}_{iN_i}$  and  $\mathbf{S}_{iN_i}$ ,  $i = 1, \dots, k$ . Then, follow MRAC by using  $W_i(\mathbf{x}_0|N_i)$  and  $\text{tr}(\mathbf{S}_{iN_i})$  instead of  $W_i(\mathbf{x}_0|n_i)$  and  $\text{tr}(\mathbf{S}_{in_i})$ .

**Theorem 3.2.** Under (A-i) to (A-iii), (3.3) holds for the MRAC with (3.5) and (3.6).

**Remark 3.1.** When  $k = 2$ , Aoshima and Yata (2011a) gave a two-stage classification rule based on the geometric classifier. See Theorem 4.3 in Aoshima and Yata (2011a) for the details. We emphasize that the MRAC can claim (3.3) for  $k \geq 2$  even under milder conditions than the original one by Aoshima and Yata (2011a).

**Remark 3.2.** Under (A-i), (A-ii) and (3.5), it holds  $N_i/C_i = 1 + o_P(1)$  as  $p \rightarrow \infty$ , which is in the HDLSS situation in the sense that  $N_i/p = o_P(1)$  under the condition that  $\max_{j=1, \dots, k} \{\text{tr}(\boldsymbol{\Sigma}_j^2)\}/\Delta_{i^*L}^2 = o(p)$ .

**Remark 3.3.** Even when  $m_i/C_i > 1$  for some  $i$ , the assertion in Theorem 3.2 is still claimed. However, it may cause oversampling in the sense that  $N_i/C_i > 1$  w.p.1.

## 4. SIMULATION

In order to examine the performance of the MRAC with (3.5) and (3.6), we used computer simulations. First, we considered 2 classes having Gaussian distributions. Independent pseudo random observations were generated from  $\pi_i : N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ ,  $i = 1, 2$ . We considered  $\boldsymbol{\Sigma}_1 = \mathbf{B}\{(-1)^{|i-j|} 0.3^{|i-j|^{1/3}}\} \mathbf{B}$  and  $\boldsymbol{\Sigma}_2 = c\{(-1)^{|i-j|} 0.4^{|i-j|^{1/3}}\}$ , where  $\mathbf{B} = \text{diag}[\{0.5 + 1/(p+1)\}^{1/2}, \dots, \{0.5 + p/(p+1)\}^{1/2}]$ . Note that  $\text{tr}(\boldsymbol{\Sigma}_1) = p$  and  $\text{tr}(\boldsymbol{\Sigma}_2) = cp$ . We set  $\boldsymbol{\mu}_1 = (1, \dots, 1, 0, \dots, 0)^T$  whose the first 30 elements are 1 and  $\boldsymbol{\mu}_2 = (0, \dots, 0)^T$ , so that  $\Delta_{ij(1)} = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 = 30$ . We prespecified  $\Delta_{1^*L} = \Delta_{2^*L} = \Delta_{12(1)} = 30$ . We set  $(\alpha_1, \alpha_2) = (0.05, 0.15)$  and  $m_i = \lceil 0.5 \times (C_i - 1) \rceil + 1$ ,  $i = 1, 2$ , where  $C_i$  is defined by (3.2). We considered four cases: (a)  $p = 500$  when  $c = 1$ , (b)  $p = 1000$  when  $c = 1$ , (c)  $p = 500$  when  $c = 1.2$ , and (d)  $p = 1000$  when  $c = 1.2$ . By averaging the outcomes from 2000 ( $= R$ , say) replications, the findings were summarized in Table 1. Under a fixed scenario, suppose that the  $r$ th replication ends with  $N_i = n_{ir}$  ( $i = 1, 2$ ) observations for  $r = 1, \dots, R$ . Let  $\bar{n}_i = R^{-1} \sum_{r=1}^R n_{ir}$  and  $V(n_i) = (R-1)^{-1} \sum_{r=1}^R (n_{ir} - \bar{n}_i)^2$ . In the end of the  $r$ th replication, we checked whether the classifier does (or does not) classify  $\mathbf{x}_0$  from  $\pi_i$  correctly and defined  $P_{ir} = 0$  (or 1) accordingly for each  $i$ . We calculated  $\bar{e}(i) = R^{-1} \sum_{r=1}^R P_{ir}$  for each  $i$  as an estimate of  $e(i)$ . Their estimated standard errors were given by  $s\{\bar{e}(i)\}$  for each  $i$ , where  $s^2\{\bar{e}(i)\} = R^{-1} \bar{e}(i) \{1 - \bar{e}(i)\}$ . As observed in Table 1, the two-class MRAC with (3.5) and (3.6) gave adequate performances for all the cases when considered those standard errors. Especially, when  $\text{tr}(\boldsymbol{\Sigma}_1) \neq \text{tr}(\boldsymbol{\Sigma}_2)$  such as in (c) and (d), the MRAC gave good performances because  $\Delta_{i^*} > \Delta_{i^*L}$ ,  $i = 1, 2$ .

Next, we considered 3 classes having non-Gaussian distributions generated by  $y_{ijl} = (8/10)^{1/2} w_{ijl}$ , where  $w_{ijl}$ ,  $j = 1, \dots, p$  ( $l = 1, 2, \dots$ ) are independently distributed as  $t$ -distribution with 10 degrees of freedom for each  $\pi_i$  ( $i = 1, 2, 3$ ). Note that  $E(y_{ijl}) = 0$ ,  $E(y_{ijl}^2) = 1$ , and  $y_{ijl}$ ,  $j = 1, \dots, p$  ( $i = 1, 2, 3$ ;  $l = 1, 2, \dots$ ) are independent. Let  $\mathbf{x}_{il} = \mathbf{H}_i \boldsymbol{\Lambda}_i^{1/2} (y_{i1l}, \dots, y_{ipl})^T + \boldsymbol{\mu}_i$  ( $i = 1, 2, 3$ ;  $l = 1, 2, \dots$ ), where  $\boldsymbol{\Lambda}_i = \mathbf{H}_i^T \boldsymbol{\Sigma}_i \mathbf{H}_i$ . Then, the distribution of  $\mathbf{x}_{il}$  satisfies (A-i) for each  $\pi_i$ . We considered  $\boldsymbol{\Sigma}_1 = \mathbf{B}\{(-1)^{|i-j|} 0.3^{|i-j|^{1/3}}\} \mathbf{B}$ ,  $\boldsymbol{\Sigma}_2 = \mathbf{B}\{(-1)^{|i-j|} 0.4^{|i-j|^{1/3}}\} \mathbf{B}$  and  $\boldsymbol{\Sigma}_3 = 1.2\{(-1)^{|i-j|} 0.4^{|i-j|^{1/3}}\}$ . We set

**Table 1.** Accuracy of the two-class MRAC with (3.5) and (3.6)

	$C_i$	$\bar{n}_i$	$\bar{n}_i - C_i$	$V(n_i)$	$\bar{e}(i)$	$s\{\bar{e}(i)\}$
When $\text{tr}(\Sigma_1) = \text{tr}(\Sigma_2)$ ( $c = 1$ )						
$p = 500: (m_1, m_2) = (10, 11)$						
$\pi_1$	18.55	19.15	0.6	16.16	0.047	0.00471
$\pi_2$	20.37	21.19	0.81	27.85	0.151	0.00801
$p = 1000: (m_1, m_2) = (19, 21)$						
$\pi_1$	36.29	36.87	0.58	14.1	0.038	0.00428
$\pi_2$	40.01	40.74	0.73	26.86	0.17	0.00839
When $\text{tr}(\Sigma_1) \neq \text{tr}(\Sigma_2)$ ( $c = 1.2$ )						
$p = 500: (m_1, m_2) = (13, 15)$						
$\pi_1$	23.11	23.28	0.17	13.86	0.027	0.00362
$\pi_2$	27.74	27.95	0.21	30.8	0.072	0.00576
$p = 1000: (m_1, m_2) = (24, 28)$						
$\pi_1$	45.47	45.91	0.44	16.38	0.013	0.00253
$\pi_2$	54.85	55.46	0.61	36.86	0.048	0.00476

$\mu_1 = (1, \dots, 1, 0, \dots, 0)^T$  whose the first 40 elements are 1,  $\mu_2 = (0, \dots, 0, 1, \dots, 1, 0, \dots, 0)^T$  whose the 21st to the 60th elements are 1, and  $\mu_3 = (0, \dots, 0)^T$ . Then, we had  $\Delta_{i*} \geq 40$  for  $i = 1, 2, 3$ . We prespecified  $\Delta_{i*L} = 40$ ,  $i = 1, 2, 3$ . We set  $m_i = \lceil 0.5 \times (C_i - 1) \rceil + 1$  for each  $\pi_i$ . We considered four cases: (a)  $p = 500$  when  $(\alpha_1, \alpha_2, \alpha_3) = (0.1, 0.1, 0.1)$ , (b)  $p = 1000$  when  $(\alpha_1, \alpha_2, \alpha_3) = (0.1, 0.1, 0.1)$ , (c)  $p = 500$  when  $(\alpha_1, \alpha_2, \alpha_3) = (0.05, 0.1, 0.15)$ , and (d)  $p = 1000$  when  $(\alpha_1, \alpha_2, \alpha_3) = (0.05, 0.1, 0.15)$ . By averaging the outcomes from 2000 ( $= R$ , say) replications, the findings were summarized in Table 2. Throughout, the three-class MRAC with (3.5) and (3.6) gave adequate performances for all the cases when considered those standard errors.

## 5. EXAMPLE

We analyzed gene expression data by Armstrong et al. (2002) in which the data set consisted of 12582 ( $= p$ ) genes. We had 3 classes of leukemia subtypes, that is,  $\pi_1$ : acute lymphoblastic leukemia (24 samples),  $\pi_2$ : mixed-lineage leukemia (20 samples), and  $\pi_3$ : acute myeloid leukemia (28 samples). We used the MRAC and compared the geometric classifier by (3.5) and (3.6) with the distance-based classifier by Aoshima and Yata (2014a). The total sample size of the distance-based classifier is defined by

$$N_{i*} = \max \left\{ m_i, \left\lceil \frac{(z_{\alpha_i/(k-1)} + z_{\alpha_{(i)/(k-1)}})^2 \max_{l=1, \dots, k} \hat{\sigma}_l}{\Delta_{i(1)L}^2} \hat{\sigma}_i^{1/2} (\hat{\sigma}_i^{1/2} + \hat{\sigma}_{(i)}^{1/2}) + 1 \right\rceil \right\}$$

for each  $\pi_i$ , where  $\Delta_{i(1)} = \min_{j(\neq i)=1, \dots, k} \Delta_{ij(1)}$  for  $i = 1, \dots, k$ , and  $\Delta_{i(1)L}$  is a lower bound of  $\Delta_{i(1)}$  such as  $\Delta_{i(1)} \geq \Delta_{i(1)L}$ . Since  $\Delta_{i*} \geq \Delta_{i(1)}$ ,  $N_{i*}$ s are larger than  $N_i$ s in (3.6) w.p.1 when  $\Delta_{i*} > \Delta_{i(1)L}$ .

We prespecified  $(\alpha_1, \alpha_2, \alpha_3) = (0.05, 0.15, 0.1)$ , so that  $\alpha_{(1)} = 0.1$ ,  $\alpha_{(2)} = 0.05$  and  $\alpha_{(3)} = 0.05$ . We set  $m_1 = m_2 = m_3 = 10$ . According to Section 3.2, by setting  $\alpha' = 0.05$  and  $n_i = m_i (= 10)$ ,  $i = 1, 2, 3$ , we had  $\Delta_{12*L} = 6.11 \times 10^9$ ,  $\Delta_{13*L} = 2.45 \times 10^{10}$  and  $\Delta_{23*L} = 8.09 \times 10^9$ . Thus, we prespecified  $\Delta_{1L*} = \min(\Delta_{12L*}, \Delta_{13L*}) = 6.11 \times 10^9$ ,  $\Delta_{2L*} = \min(\Delta_{12L*}, \Delta_{23L*}) = 6.11 \times 10^9$  and  $\Delta_{3L*} = \min(\Delta_{13L*}, \Delta_{23L*}) = 8.09 \times 10^9$ . Also, we had  $\Delta_{12(1)L} = 5.96 \times 10^9$ ,  $\Delta_{13(1)L} = 2.37 \times 10^{10}$  and

**Table 2.** Accuracy of the three-class MRAC with (3.5) and (3.6)

	$C_i$	$\bar{n}_i$	$\bar{n}_i - C_i$	$V(n_i)$	$\bar{e}(i)$	$s\{\bar{e}(i)\}$
When $(\alpha_1, \alpha_2, \alpha_3) = (0.1, 0.1, 0.1)$						
$p = 500: (m_1, m_2, m_3) = (11, 13, 13)$						
$\pi_1$	19.72	20.63	0.91	14.96	0.05	0.00485
$\pi_2$	23.28	24.15	0.88	28.75	0.075	0.00587
$\pi_3$	24.93	26.03	1.09	38.36	0.106	0.00688
$p = 1000: (m_1, m_2, m_3) = (20, 24, 26)$						
$\pi_1$	38.66	39.24	0.59	15.42	0.038	0.00428
$\pi_2$	45.9	46.56	0.66	26.17	0.067	0.00559
$\pi_3$	49.23	50.0	0.77	39.49	0.107	0.0069
When $(\alpha_1, \alpha_2, \alpha_3) = (0.05, 0.1, 0.15)$						
$p = 500: (m_1, m_2, m_3) = (13, 15, 14)$						
$\pi_1$	23.48	24.47	0.99	20.58	0.02	0.00313
$\pi_2$	27.75	28.65	0.9	34.94	0.074	0.00585
$\pi_3$	26.56	27.65	1.1	43.55	0.105	0.00685
$p = 1000: (m_1, m_2, m_3) = (24, 28, 27)$						
$\pi_1$	46.21	46.86	0.65	19.26	0.016	0.00281
$\pi_2$	54.92	55.54	0.62	31.61	0.057	0.00516
$\pi_3$	52.5	53.29	0.79	39.49	0.126	0.00742

$\Delta_{23(1)L} = 7.81 \times 10^9$  according to (3.4). Thus, we prespecified  $\Delta_{1(1)L} = 5.96 \times 10^9$ ,  $\Delta_{2(1)L} = 5.96 \times 10^9$  and  $\Delta_{3(1)L} = 7.81 \times 10^9$ .

By using pilot samples of size  $m_1 = m_2 = m_3 = 10$ , we calculated  $W_{1m_1} = 2.59 \times 10^{19}$ ,  $W_{2m_2} = 2.16 \times 10^{19}$  and  $W_{3m_3} = 2.51 \times 10^{19}$ . From (3.6), the total sample size for  $\pi_1$  was calculated by

$$N_1 = \max \left\{ 10, \left\lceil \frac{(z_{\alpha_1/2} + z_{\alpha_{(1)}/2})^2 \max_{l=1,2,3} \hat{\sigma}_l}{\Delta_{1^*L}^2} \hat{\sigma}_1^{1/2} (\hat{\sigma}_1^{1/2} + \hat{\sigma}_{(1)}^{1/2}) + 1 \right\rceil \right\} = 19.$$

Similarly, we had  $N_2 = 16$  and  $N_3 = 12$ . We considered constructing the geometric classifier,  $W_i(\mathbf{x}_0|N_i)$ ,  $i = 1, 2, 3$ , by  $(N_1, N_2, N_3) = (19, 16, 12)$  samples and checking the accuracy of the MRAC by using remaining  $(24 - N_1, 20 - N_2, 28 - N_3) = (5, 4, 16)$  samples. We randomly split the data set from each  $\pi_i$  into training sets of sizes  $(N_1, N_2, N_3) = (19, 16, 12)$  and test sets of sizes  $(5, 4, 16)$ . We constructed  $W_i(\mathbf{x}_0|N_i)$ ,  $i = 1, 2, 3$ , by the training sets and checked the accuracy of the MRAC by using the test sets. We repeated this procedure 100 times. Then, we had the average of misclassification rates as  $\bar{e}(1) = 0.044$ ,  $\bar{e}(2) = 0.09$  and  $\bar{e}(3) = 0.064$ . Also, for the distance-based classifier by Aoshima and Yata (2014a), we calculated the total sample sizes as  $(N_{1^*}, N_{2^*}, N_{3^*}) = (20, 17, 12)$  and had the average of misclassification rates as  $\bar{e}(1) = 0.023$ ,  $\bar{e}(2) = 0.08$  and  $\bar{e}(3) = 0.066$ . Similarly, for various settings of  $\alpha_i$ s, we investigated the performances of the geometric classifier and the distance-based classifier in the MRAC. Throughout, we used the same settings as  $m_1 = m_2 = m_3 = 10$  and  $(\Delta_{1^*L}, \Delta_{2^*L}, \Delta_{3^*L}) = (6.11 \times 10^9, 6.11 \times 10^9, 8.09 \times 10^9)$  or  $(\Delta_{1(1)L}, \Delta_{2(1)L}, \Delta_{3(1)L}) = (5.96 \times 10^9, 5.96 \times 10^9, 7.81 \times 10^9)$ . We summarized the results in Table 3. Both the classifiers seem to give adequate performances in such a HDLSS situation. The geometric classifier would save more observations compared to the distance-based classifier specially in small sample size settings. On the other hand, the distance-based classifier is very versatile and it holds (3.3) under milder conditions than the geometric classifier. See Sections 3 and 4 in Aoshima and Yata (2014a) for the details.

**Table 3.** Average misclassification rates of the MRAC by the geometric classifier with (3.5) and (3.6) and by the distance-based classifier by Aoshima and Yata (2014a). We set  $m_1 = m_2 = m_3 = 10$  and  $(\Delta_{1*L}, \Delta_{2*L}, \Delta_{3*L}) = (6.11 \times 10^9, 6.11 \times 10^9, 8.09 \times 10^9)$  or  $(\Delta_{1(1)L}, \Delta_{2(1)L}, \Delta_{3(1)L}) = (5.96 \times 10^9, 5.96 \times 10^9, 7.81 \times 10^9)$ . When  $\alpha_i \leq 0.05$  at least for two  $\pi_i$ s, the result was not available within the data sets

$(\alpha_1, \alpha_2, \alpha_3)$	Geometric classifier				Distance-based classifier			
	$\bar{e}(1)$	$\bar{e}(2)$	$\bar{e}(3)$	$(N_1, N_2, N_3)$	$\bar{e}(1)$	$\bar{e}(2)$	$\bar{e}(3)$	$(N_{1*}, N_{2*}, N_{3*})$
(0.15, 0.15, 0.15)	0.097	0.135	0.071	(13, 12, 10)	0.058	0.113	0.069	(14, 13, 10)
(0.1, 0.15, 0.15)	0.081	0.1	0.071	(15, 14, 10)	0.047	0.095	0.071	(15, 14, 10)
(0.1, 0.1, 0.15)	0.08	0.08	0.088	(16, 16, 10)	0.054	0.055	0.084	(17, 16, 10)
(0.1, 0.1, 0.1)	0.074	0.085	0.084	(16, 16, 10)	0.04	0.088	0.071	(17, 16, 11)
(0.05, 0.15, 0.1)	0.044	0.09	0.064	(19, 16, 12)	0.023	0.08	0.066	(20, 17, 12)
(0.05, 0.1, 0.1)	0.064	0.065	0.071	(19, 18, 12)	0.035	0.08	0.067	(20, 19, 12)
(0.1, 0.05, 0.1)	0.104	0.06	0.086	(19, 18, 12)	0.07	0.05	0.086	(20, 19, 12)
(0.1, 0.1, 0.05)	0.066	0.075	0.06	(19, 18, 12)	0.038	0.07	0.062	(20, 19, 12)

## A. APPENDIX

*Proof of Theorem 2.1.* Under (A-iv), it holds that  $\text{Var}\{\|\bar{\mathbf{x}}_{in_i} - \boldsymbol{\mu}_i\|^2 - \text{tr}(\mathbf{S}_{in_i})/n_i\} = O\{\text{tr}(\boldsymbol{\Sigma}_i^2)/n_i^2\} = o(\Delta_{ij}^2)$  and  $\text{Var}\{(\bar{\mathbf{x}}_{jn_j} - \boldsymbol{\mu}_j)^T(\mathbf{x}_0 - \boldsymbol{\mu}_i) | \mathbf{x}_0 \in \pi_i\} = O\{\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)/n_j\} = O\{\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2} \text{tr}(\boldsymbol{\Sigma}_j^2)^{1/2}/n_j\} = o(\Delta_{ij}^2)$  for all  $i, j$ . Note that  $(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)/n_j \leq \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2 \lambda_{j1}/n_j \leq \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2 \text{tr}(\boldsymbol{\Sigma}_j^2)^{1/2}/n_j = o(\Delta_{ij}^2)$  for all  $i \neq j$ , under (A-iv). Then, it holds that  $\text{Var}\{(\mathbf{x}_0 - \boldsymbol{\mu}_i) - (\bar{\mathbf{x}}_{jn_j} - \boldsymbol{\mu}_j)\}^T (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) | \mathbf{x}_0 \in \pi_i\} = (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T (\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j/n_j) (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) = o(\Delta_{ij}^2)$  for all  $i \neq j$ , under (A-iii) and (A-iv). Thus by using Chebyshev's inequality, under (A-iii) and (A-iv) we obtain that

$$\begin{aligned} \|\mathbf{x}_0 - \boldsymbol{\mu}_i - (\bar{\mathbf{x}}_{in_i} - \boldsymbol{\mu}_i)\|^2 - \text{tr}(\mathbf{S}_{in_i})/n_i &= \|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 + o_P(\Delta_{ij}); \\ \|\mathbf{x}_0 - \boldsymbol{\mu}_i - (\bar{\mathbf{x}}_{jn_j} - \boldsymbol{\mu}_j) + \boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2 - \text{tr}(\mathbf{S}_{jn_j})/n_j &= \|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 + \Delta_{ij(1)} + o_P(\Delta_{ij}) \end{aligned}$$

when  $\mathbf{x}_0 \in \pi_i$  for all  $i \neq j$ . Under (A-i) and (A-iv) we have that  $\text{Var}\{\text{tr}(\mathbf{S}_{in_i})\} = O\{\text{tr}(\boldsymbol{\Sigma}_i^2)/n_i\} = o(\Delta_{ij}^2)$  and  $\text{Var}(\|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 | \mathbf{x}_0 \in \pi_i) = O\{\text{tr}(\boldsymbol{\Sigma}_i^2)\}$  for all  $i \neq j$ , so that  $\text{tr}(\mathbf{S}_{in_i}) = \text{tr}(\boldsymbol{\Sigma}_i) + o_P(\Delta_{ij})$  and  $\|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 = \text{tr}(\boldsymbol{\Sigma}_i) + O_P\{\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}\}$  when  $\mathbf{x}_0 \in \pi_i$  for all  $i \neq j$ . Note that  $\text{tr}(\boldsymbol{\Sigma}_i)/p \in (0, \infty)$  as  $p \rightarrow \infty$  for  $i = 1, \dots, k$ . Then, under (A-i), (A-iii) and (A-iv), we have that

$$\begin{aligned} \frac{W_j(\mathbf{x}_0 | n_j) - W_i(\mathbf{x}_0 | n_i)}{\Delta_{ij}} &= p \frac{\|\mathbf{x}_0 - \boldsymbol{\mu}_i - (\bar{\mathbf{x}}_{jn_j} - \boldsymbol{\mu}_j) + \boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2 - \text{tr}(\mathbf{S}_{jn_j})/n_j}{\text{tr}(\mathbf{S}_{jn_j}) \Delta_{ij}} \\ &\quad - p \frac{\|\mathbf{x}_0 - \boldsymbol{\mu}_i - (\bar{\mathbf{x}}_{in_i} - \boldsymbol{\mu}_i)\|^2 - \text{tr}(\mathbf{S}_{in_i})/n_i}{\text{tr}(\mathbf{S}_{in_i}) \Delta_{ij}} + p \log \left\{ \frac{\text{tr}(\mathbf{S}_{jn_j})}{\text{tr}(\mathbf{S}_{in_i})} \right\} / \Delta_{ij} \\ &= p \frac{\|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 + \Delta_{ij(1)}}{\text{tr}(\boldsymbol{\Sigma}_j) \Delta_{ij}} - p \frac{\|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2}{\text{tr}(\boldsymbol{\Sigma}_i) \Delta_{ij}} + p \log \left\{ \frac{\text{tr}(\boldsymbol{\Sigma}_j)}{\text{tr}(\boldsymbol{\Sigma}_i)} \right\} / \Delta_{ij} + o_P(1) \\ &= p \frac{\{\|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 - \text{tr}(\text{tr}(\boldsymbol{\Sigma}_i))\} \{\text{tr}(\boldsymbol{\Sigma}_i) - \text{tr}(\boldsymbol{\Sigma}_j)\}}{\text{tr}(\boldsymbol{\Sigma}_i) \text{tr}(\boldsymbol{\Sigma}_j) \Delta_{ij}} + 1 + o_P(1) \\ &= 1 + o_P(1) \end{aligned} \tag{A.1}$$

when  $\mathbf{x}_0 \in \pi_i$  for all  $i \neq j$ . Hence, we conclude the results.  $\square$

*Proof of Theorem 2.2.* We note that  $\max\{\text{tr}(\boldsymbol{\Sigma}_i^2)/n_i^2, \text{tr}(\boldsymbol{\Sigma}_j^2)/n_j^2\} = o(\delta_{ij}^2)$  for all  $i \neq j$  under (A-ii). Also, note that  $(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T (\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j/n_j) (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) = o(\delta_{ij}^2)$  for all  $i \neq j$  under (A-ii) and (A-v) since  $\delta_{ji}/(n_j \delta_{ij}) =$

$o(1)$  for all  $i \neq j$  under (A-ii). Let  $\omega(\mathbf{x}_0|n_i, n_j) = 2\{p/\text{tr}(\boldsymbol{\Sigma}_j)\}(\mathbf{x}_0 - \boldsymbol{\mu}_i)^T\{(\bar{\mathbf{x}}_{jn_j} - \boldsymbol{\mu}_j) - \text{tr}(\boldsymbol{\Sigma}_j)(\bar{\mathbf{x}}_{in_i} - \boldsymbol{\mu}_i)/\text{tr}(\boldsymbol{\Sigma}_i)\}$  for  $i \neq j$ . Then, similar to (A.1), under (A-i), (A-ii), (A-v) and  $\Delta_{ij(1)}/\text{tr}(\boldsymbol{\Sigma}_j) = o(1)$  for all  $i \neq j$ , we have that

$$\begin{aligned}
& W_j(\mathbf{x}_0|n_j) - W_i(\mathbf{x}_0|n_i) \\
&= p \frac{\|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 + \Delta_{ij(1)} - 2(\mathbf{x}_0 - \boldsymbol{\mu}_i)^T(\bar{\mathbf{x}}_{jn_j} - \boldsymbol{\mu}_j)}{\text{tr}(\mathbf{S}_{jn_j})} - p \frac{\|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 - 2(\mathbf{x}_0 - \boldsymbol{\mu}_i)^T(\bar{\mathbf{x}}_{in_i} - \boldsymbol{\mu}_i)}{\text{tr}(\mathbf{S}_{in_i})} \\
&+ p \log \left\{ \frac{\text{tr}(\mathbf{S}_{jn_j})}{\text{tr}(\mathbf{S}_{in_i})} \right\} + o_P(\delta_{ij}) \\
&= p \frac{\{\|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 - \text{tr}(\boldsymbol{\Sigma}_i)\}\{\text{tr}(\mathbf{S}_{in_i} - \mathbf{S}_{jn_j})\}}{\text{tr}(\mathbf{S}_{in_i})\text{tr}(\mathbf{S}_{jn_j})} + p \left\{ \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{\text{tr}(\mathbf{S}_{jn_j})} - \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{\text{tr}(\mathbf{S}_{in_i})} \right\} - p \log \left\{ \frac{\text{tr}(\mathbf{S}_{in_i})}{\text{tr}(\mathbf{S}_{jn_j})} \right\} \\
&- \omega(\mathbf{x}_0|n_i, n_j) + p \frac{\Delta_{ij(1)}}{\text{tr}(\boldsymbol{\Sigma}_j)} + o_P(\delta_{ij}) \tag{A.2}
\end{aligned}$$

when  $\mathbf{x}_0 \in \pi_i$  for all  $i \neq j$  since  $\text{tr}(\mathbf{S}_{in_i})/\text{tr}(\boldsymbol{\Sigma}_i) - 1 = O_P(\delta_{ij}/p) = o_P(1)$ . Here, we note that  $\text{tr}(\boldsymbol{\Sigma}_i^2)/p = o\{\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}\}$ ,  $i = 1, \dots, k$ , under (A-ii) from the fact that  $\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}/\text{tr}(\boldsymbol{\Sigma}_i) = o(1)$  under (A-ii). It holds that  $\|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 = \text{tr}(\boldsymbol{\Sigma}_i) + O_P\{\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}\}$  when  $\mathbf{x}_0 \in \pi_i$  and  $\text{tr}(\mathbf{S}_{in_i}) = \text{tr}(\boldsymbol{\Sigma}_i) + O_P[\{\text{tr}(\boldsymbol{\Sigma}_i^2)/n_i\}^{1/2}]$ ,  $i = 1, \dots, k$ . Then, under (A-ii) and (A-v), we have that

$$\begin{aligned}
p \frac{\{\|\mathbf{x}_0 - \boldsymbol{\mu}_i\|^2 - \text{tr}(\boldsymbol{\Sigma}_i)\}\{\text{tr}(\mathbf{S}_{in_i} - \mathbf{S}_{jn_j})\}}{\text{tr}(\mathbf{S}_{in_i})\text{tr}(\mathbf{S}_{jn_j})} &= O_P \left\{ \frac{\max_{l=i,j} \text{tr}(\boldsymbol{\Sigma}_l^2)}{\min\{n_i, n_j\}^{1/2}p} + \frac{\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}|\text{tr}(\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_j)|}{p} \right\} \\
&= o_P(\delta_{ij}) \tag{A.3}
\end{aligned}$$

when  $\mathbf{x}_0 \in \pi_i$  for all  $i \neq j$ . On the other hand, under (A-i) and (A-ii), it holds that

$$\begin{aligned}
p \log \left\{ \frac{\text{tr}(\mathbf{S}_{in_i})}{\text{tr}(\mathbf{S}_{jn_j})} \right\} - p \log \left\{ \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{\text{tr}(\boldsymbol{\Sigma}_j)} \right\} &= p \log \left\{ \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{\text{tr}(\mathbf{S}_{jn_j})} \right\} - p \log \left\{ \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{\text{tr}(\mathbf{S}_{in_i})} \right\} - p \log \left\{ \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{\text{tr}(\boldsymbol{\Sigma}_j)} \right\} \\
&= p \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{\text{tr}(\mathbf{S}_{jn_j})} + p - p \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{\text{tr}(\mathbf{S}_{in_i})} - p \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{\text{tr}(\boldsymbol{\Sigma}_j)} + o_P(\delta_{ij}) \tag{A.4}
\end{aligned}$$

for all  $i \neq j$ . Then, by combining (A.2) with (A.3) and (A.4), under the assumptions of Theorem 2.2 we have that

$$W_j(\mathbf{x}_0|n_j) - W_i(\mathbf{x}_0|n_i) = \omega(\mathbf{x}_0|n_i, n_j) + \Delta_{ij} + o_P(\delta_{ij})$$

when  $\mathbf{x}_0 \in \pi_i$ . Note that  $\text{Var}\{\omega(\mathbf{x}_0|n_i, n_j)\}/\delta_{ij}^2 = 1 + o(1)$  for all  $i \neq j$  under (A-ii). Then, in a way similar to the proof of Theorem 3 in Aoshima and Yata (2014a), under (A-i) and (A-ii) we can claim that  $\omega(\mathbf{x}_0|n_i, n_j)/\delta_{ij} \Rightarrow Y_{ij}$  for all  $i \neq j$ . Thus it concludes the result.  $\square$

*Proof of Corollary 2.1.* By using Theorem 2.2 and Bonferroni's inequality, we have that  $1 - e(i) \geq 1 - \sum_{j(\neq i)=1}^k \Phi\{-\Delta_{ij}/\delta_{ij}\} + o(1)$  when  $\mathbf{x}_0 \in \pi_i$ . This concludes the proof.  $\square$

*Proof of Theorem 3.1.* From (3.2), it holds that  $\delta_{ij} \leq 2\Delta_{(ij)}\{1 + o(1)\}/(z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)})$  when  $\text{tr}(\boldsymbol{\Sigma}_i)/\text{tr}(\boldsymbol{\Sigma}_j) = 1 + o(1)$  for all  $i \neq j$ . We denote the error of misclassifying an individual from  $\pi_i$  into  $\pi_j$  by  $e(j|i)$  for  $i \neq j$ . Then, under (3.2) and the assumptions of Theorem 2.2, we have that

$$\begin{aligned}
e(j|i) &= P \left\{ \frac{W_i(\mathbf{x}_0|n_i) - W_j(\mathbf{x}_0|n_j)}{\delta_{ij}} \geq \frac{p \max\{\Delta_{i^*L}/\text{tr}(\mathbf{S}_{jn_j}), \Delta_{j^*L}/\text{tr}(\mathbf{S}_{in_i})\}}{\delta_{ij}} \frac{z_{\alpha_i/(k-1)} - z_{\alpha_j/(k-1)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}} \right\} \\
&\leq P \left\{ Y_{ij} \geq \frac{\Delta_{(ij)}}{\delta_{ij}} \frac{2z_{\alpha_i/(k-1)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}} \right\} + o(1) \leq P(Y_{ij} \geq z_{\alpha_i/(k-1)}) + o(1) = \frac{\alpha_i}{k-1} + o(1)
\end{aligned}$$

when  $\mathbf{x}_0 \in \pi_i$  for  $i \neq j$ , where  $Y_{ij}$  denotes a random variable distributed as the standard normal distribution. We note that (A-v) holds under (A-iii) when  $\liminf \delta_{ij}/\Delta_{ij} > 0$  for all  $i \neq j$ . On the other hand, when  $\delta_{ij}/\Delta_{ij} = o(1)$  for  $i \neq j$ , from Theorem 2.1 it holds that for  $\mathbf{x}_0 \in \pi_i$

$$\begin{aligned} e(j|i) &= P\left\{ \frac{W_i(\mathbf{x}_0|n_i) - W_j(\mathbf{x}_0|n_j)}{\Delta_{ij}} \geq \frac{p \max\{\Delta_{i^*L}/\text{tr}(\mathbf{S}_{jn_j}), \Delta_{j^*L}/\text{tr}(\mathbf{S}_{in_i})\}}{\Delta_{ij}} \frac{z_{\alpha_i/(k-1)} - z_{\alpha_j/(k-1)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}} \right\} \\ &= P\left\{ -1 + o_P(1) \geq \frac{\Delta_{(ij)}\{1 + o_P(1)\}}{\Delta_{ij}} \frac{z_{\alpha_i/(k-1)} - z_{\alpha_j/(k-1)}}{z_{\alpha_i/(k-1)} + z_{\alpha_j/(k-1)}} \right\} = o(1) \end{aligned}$$

under (A-i) to (A-iii) without (A-v). We note that  $\delta_{ij}/\Delta_{ij} = o(1)$  for  $i \neq j$  under (A-ii) when it holds that  $\liminf_{p \rightarrow \infty} \Delta_{ij(1)}/\text{tr}(\mathbf{\Sigma}_j) > 0$  or  $\liminf_{p \rightarrow \infty} |\text{tr}(\mathbf{\Sigma}_i)/\text{tr}(\mathbf{\Sigma}_j) - 1| > 0$ . Thus one can claim  $e(j|i) \leq \alpha_i/(k-1) + o(1)$  for all  $i \neq j$  under (3.2) and (A-i) to (A-iii). Then, from Bonferroni's inequality, we have that  $1 - e(i) \geq 1 - \sum_{j(\neq i)=1}^k e(j|i) \geq 1 - \alpha_i + o(1)$  when  $\mathbf{x}_0 \in \pi_i$ . This concludes the proof.  $\square$

*Proof of Theorem 3.2.* Let  $C_{iL} = \lfloor C_i - (\omega C_i)^{1/2} \rfloor$ ,  $i = 1, \dots, k$ , where  $\omega (> 0)$  is a variable such that  $\omega \rightarrow 0$  as  $p \rightarrow \infty$ . Then, from the proof of Theorem 5 in Aoshima and Yata (2014a), it holds that  $\max\{m_i, C_{iL}\} \leq N_i < C_i + (\omega C_i)^{1/2}$  as  $p \rightarrow \infty$  w.p.1. Then, in a way similar to the proofs of Theorems 2.4 and 2.5 in Aoshima and Yata (2011a), under (A-i) to (A-iii) we have that for all  $i \neq j$

$$\begin{aligned} \text{tr}(\mathbf{S}_{iN_i}) &= \text{tr}(\mathbf{\Sigma}_i) + O_P[\{\text{tr}(\mathbf{\Sigma}_i^2)/C_{iL}\}^{1/2}]; \\ \|\bar{\mathbf{x}}_{iN_i} - \boldsymbol{\mu}_i\|^2 - \text{tr}(\mathbf{S}_{iN_i})/N_i &= o_P(\Delta_{ij}); \\ (\bar{\mathbf{x}}_{jN_j} - \boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) &= o_P(\Delta_{ij}); \text{ and} \\ \omega(\mathbf{x}_0|N_i, N_j) &= \omega(\mathbf{x}_0|C_{iL}, C_{jL}) + o_P(\Delta_{ij}) \text{ when } \mathbf{x}_0 \in \pi_i, \end{aligned}$$

where  $\omega(\mathbf{x}_0|N_i, N_j)$  is given in the proof of Theorem 2.2. Similar to the proof of Theorem 2.2, under (A-i) to (A-iii) we have that

$$W_j(\mathbf{x}_0|N_i) - W_i(\mathbf{x}_0|N_j) = \omega(\mathbf{x}_0|C_{iL}, C_{jL}) + \Delta_{ij} + o_P(\Delta_{ij})$$

when  $\mathbf{x}_0 \in \pi_i$  for all  $i \neq j$ . Then, in a way similar to the proof of Theorem 3.1, we can conclude the results.  $\square$

## ACKNOWLEDGEMENTS

The authors would like to thank the Editor-in-Chief, Prof. Nitis Mukhopadhyay, for giving us the opportunity to contribute to Stein (1945)'s 70 Year Celebration Issue. Research of the first author was partially supported by Grants-in-Aid for Scientific Research (B) and Challenging Exploratory Research, Japan Society for the Promotion of Science (JSPS), under Contract Numbers 22300094 and 26540010. Research of the second author was partially supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 26800078.

## REFERENCES

- Aoshima, M. and Yata, K. (2011a). Two-Stage Procedures for High-Dimensional Data, *Editor's special invited paper, Sequential Analysis* 30: 356–399.
- Aoshima, M. and Yata, K. (2011b). Authors' Response, *Sequential Analysis* 30: 432–440.
- Aoshima, M. and Yata, K. (2014a). A Distance-Based, Misclassification Rate Adjusted Classifier for Multi-class, High-Dimensional Data, *Annals of Institute of Statistical Mathematics* 66: 983–1010.

- Aoshima, M. and Yata, K. (2014b). High-Dimensional Quadratic Classifiers in Non-Sparse Settings, submitted.
- Aoshima, M. and Yata, K. (2015). Asymptotic Normality for Inference on Multisample, High-Dimensional Mean Vectors under Mild Conditions, *Methodology and Computing in Applied Probability* 17: 419–439 .
- Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J. (2002). MLL Translocations Specify a Distinct Gene Expression Profile That Distinguishes a Unique Leukemia, *Nature Genetics* 30: 41–47.
- Chan, Y. -B. and Hall, P. (2009). Scale Adjustments for Classifiers in High-Dimensional, Low Sample Size Settings, *Biometrika* 96: 469–478.
- Dudoit, S., Fridlyand, J., and Speed, T. P. (2002). Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data, *Journal of American Statistical Association* 97: 77–87.
- Stein, C. (1945). A Two-Sample Test for a Linear Hypothesis Whose Power is Independent of the Variance, *Annals of Mathematical Statistics* 16: 243–258.
- Stein, C. (1949). Some Problems in Sequential Estimation (abstract), *Econometrica* 17: 77–78.
- Yata, K. and Aoshima, M. (2013). Correlation Tests for High-Dimensional Data Using Extended Cross-Data-Matrix Methodology, *Journal of Multivariate Analysis* 117: 313–331.