Asymptotic properties of a distance-based classifier for high-dimensional data

筑波大学・数学域 矢田 和善 (Kazuyoshi Yata)
Institute of Mathematics
University of Tsukuba

筑波大学・数学域 青嶋 誠 (Makoto Aoshima)
Institute of Mathematics
University of Tsukuba

Abstract: We consider multiclass classification for high-dimensional and non-Gaussian data. We consider a distance-based classifier given by Yata and Aoshima (2012b). We show that the classifier is verified by the asymptotic normality as $p \to \infty$ either when n_is are fixed or $n_i \to \infty$ for some i. We give a simulation result of the classifier under high-dimensional settings.

Key words and phrases: Asymptotic normality, Distance-based classifier, HDLSS, Large p small n.

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 small. This is the so-called "HDLSS" or "large p, small n" situation where $p/n \to \infty$; here p is the data dimension and n is the sample size. The asymptotic behaviors of high-dimensional, low-sample-size (HDLSS) data were studied by Hall et al. (2005), Ahn et al. (2007) and Yata and Aoshima (2012a) when $p \to \infty$ while n is fixed. They explored conditions to give a geometric representation of HDLSS data. The HDLSS asymptotic study usually assumes either the normality for the population distribution or a ρ-mixing condition for the dependency of random variables in a sphered data matrix. See also Jung and Marron (2009). However, Yata and Aoshima (2009) succeeded in investigating consistency properties of both eigenvalues and eigenvectors of the sample covariance matrix in general settings including the case when all eigenvalues are in the range of sphericity. In addition, Yata and Aoshima (2010a,b) created the *cross-data-matrix (CDM) methodology* that provides effective inference on the eigenspace of HDLSS data. Recently, Aoshima and Yata (2011a,b) have developed a variety of inference for high-dimensional data along with sample size determination to assure prespecified accuracy. Aoshima and Yata (2011c) applied those inference procedures to microarray studies.

Suppose we have independent and p-variate populations, $\pi_i,\ i = 1, ..., k$, having unknown mean vector $\boldsymbol{\mu}_i = (\mu_{i1}, ..., \mu_{ip})^T$ and unknown covariance matrix $\boldsymbol{\Sigma}_i (> \boldsymbol{O})$ for each i. *We do not assume that* $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$. Let $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k)$. The eigen-decomposition of $\boldsymbol{\Sigma}_i$ is given by $\boldsymbol{\Sigma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i \boldsymbol{H}_i^T$, where $\boldsymbol{\Lambda}_i$ is a diagonal matrix of eigenvalues, $\lambda_{i1} \ge \cdots \ge \lambda_{ip} > 0$, and $\boldsymbol{H}_i = (\boldsymbol{h}_{i1}, ..., \boldsymbol{h}_{ip})$ is an orthogonal matrix of the corresponding eigenvectors. We have independent and identically distributed observations, $\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i}$, from each π_i, where $\boldsymbol{x}_{ij} = (x_{i1j}, ..., x_{ipj})^T$, $j = 1, ..., n_i$. We assume $n_i \ge 2$, $i = 1, ..., k$. Then, $\boldsymbol{z}_{ij} = \boldsymbol{\Lambda}_i^{-1/2} \boldsymbol{H}_i^T (\boldsymbol{x}_{ij} - \boldsymbol{\mu}_i)$ is a sphered data vector from a distribution with the identity covariance matrix. Here, we write $\boldsymbol{z}_{ij} = (z_{i1j}, ..., z_{ipj})^T$, $j = 1, ..., n_i$; $i = 1, ..., k$. Note that $E(z_{ijl}^2) = 1$ and $E(z_{ijl} z_{ij'l}) = 0$ for $i = 1, ..., k$; $j (\neq j') = 1, ..., p$; $l = 1, ..., n_i$. We assume for $i = 1, ..., k$, that the fourth moments of each variable in $\boldsymbol{z}_{ij}$ are uniformly bounded. We assume the following assumptions for $\boldsymbol{\Sigma}_i$s as necessary:

$$\textbf{(A-i)} \quad \frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^2 \boldsymbol{\Sigma}_j^2)}{\mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)^2} \to 0 \text{ and } \frac{\mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_l)}{\mathrm{tr}(\boldsymbol{\Sigma}_j^2)} \in (0, \infty) \text{ as } p \to \infty \text{ for } i, j, l = 1, ..., k.$$

Here, $f(p) \in (0, \infty)$ as $p \to \infty$ denotes that $\liminf_{p \to \infty} f(p) > 0$ and $\limsup_{p \to \infty} f(p) < \infty$ for a function $f(\cdot)$.

Remark 1.1. If all λ_{ij}s are bounded such as $\lambda_{ij} \in (0, \infty)$ as $p \to \infty$, (A-i) trivially holds. For a spiked model such as $\lambda_{ij} = a_{ij} p^{\alpha_{ij}}$ $(j = 1, ..., r_i)$ and $\lambda_{ij} = c_{ij}$ $(j = r_i + 1, ..., p)$ with positive constants, a_{ij}s, c_{ij}s and α_{ij}s, and positive integers r_is, (A-i) holds under the condition that $\alpha_{ij} < 1/2$ for $j = 1, ..., r_i (< \infty)$; $i = 1, ..., k$. See Yata and Aoshima (2010b) for the details of a spiked model. As an interesting example, (A-i) holds for $\boldsymbol{\Sigma}_{i'} = c_{i'} (\rho_{i'}^{|i-j|^{q_{i'}}})$, $i' = 1, ..., k$, where $c_{i'}$ and $q_{i'}$ are positive constants and $0 < \rho_{i'} < 1$.

Let $\boldsymbol{x}_0$ be an observation vector of an individual belonging to one of the k populations. We estimate $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ by $\overline{\boldsymbol{x}}_{in_i} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}/n_i$ and $\boldsymbol{S}_{in_i} = \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{in_i})(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{in_i})^T/(n_i - 1)$. When $k = 2$, a typical classification rule is that one classifies an individual into π_1 if

$$\begin{aligned} &(\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_{1n_1})^T \boldsymbol{S}_{1n_1}^{-1} (\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_{1n_1}) - \log \left\{ \frac{\det(\boldsymbol{S}_{2n_2})}{\det(\boldsymbol{S}_{1n_1})} \right\} \\ &\quad < (\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_{2n_2})^T \boldsymbol{S}_{2n_2}^{-1} (\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_{2n_2}), \end{aligned} \tag{1.1}$$

and into π_2 otherwise. However, the inverse matrix of $\boldsymbol{S}_{in_i}$ does not exist in the HDLSS context ($p > n_i$). When $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, Saranadasa (1993) considered substituting the identity matrix $\boldsymbol{I}_p$ for $\boldsymbol{S}_{in_i}$. Bickel and Levina (2004) considered the inverse matrix defined by only diagonal elements of the pooled sample covariance matrix. Yata and Aoshima (2012a) considered using a ridge-type inverse covariance matrix derived by the *noise reduction methodology*. When $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, Dudoit et al. (2002) considered using the inverse matrix defined by only diagonal elements of $\boldsymbol{S}_{in_i}$. Aoshima and Yata (2011a) proposed a quadratic classification rule substituting $(\mathrm{tr}(\boldsymbol{S}_{in_i})/p)\boldsymbol{I}_p$ for $\boldsymbol{S}_{in_i}$ followed by a bias correction and showed the asymptotic normality of the classifier so that the sample size can be determined to assure prespecified accuracy. On the other hand, Hall et al. (2005, 2008), Ahn et al. (2007), and Chan and Hall (2009) considered distance-based classifiers. The above literatures mainly discussed two-class classification in high-dimensional, low sample size settings.

Recently, Yata and Aoshima (2012b) considered a classification rule given by using the identity matrix $\boldsymbol{I}_p$ instead of $\boldsymbol{S}_{in_i}$ in (1.1) as follows: One classifies an individual into π_1 if

$$\Big(\boldsymbol{x}_0 - \frac{\overline{\boldsymbol{x}}_{1n_1} + \overline{\boldsymbol{x}}_{2n_2}}{2}\Big)^T (\overline{\boldsymbol{x}}_{2n_2} - \overline{\boldsymbol{x}}_{1n_1}) - \frac{\mathrm{tr}(\boldsymbol{S}_{1n_1})}{2n_1} + \frac{\mathrm{tr}(\boldsymbol{S}_{2n_2})}{2n_2} < 0 \tag{1.2}$$

and into π_2 otherwise. Here, $-\mathrm{tr}(\boldsymbol{S}_{1n_1})/(2n_1) + \mathrm{tr}(\boldsymbol{S}_{2n_2})/(2n_2)$ is a bias-correction term. They showed the asymptotic normality of the classifier and gave a sample size determination so as to control misclassification rates being no more than a prespecified value. They further developed the classifier to multiclass classification when $k \ge 3$.

Remark 1.2. Chan and Hall (2009) considered a scale adjusted distance-based classifier as follows: One classifies an individual into π_1 if

$$\begin{aligned}&\sum_{j=1}^{n_1} \frac{||\boldsymbol{x}_0 - \boldsymbol{x}_{1j}||^2}{n_1} - \sum_{j=1}^{n_2} \frac{||\boldsymbol{x}_0 - \boldsymbol{x}_{2j}||^2}{n_2} - \sum_{i=1}^{n_1}\sum_{j=1}^{n_1} \frac{||\boldsymbol{x}_{1i} - \boldsymbol{x}_{1j}||^2}{2n_1(n_1-1)} \\ &+ \sum_{i=1}^{n_2}\sum_{j=1}^{n_2} \frac{||\boldsymbol{x}_{2i} - \boldsymbol{x}_{2j}||^2}{2n_2(n_2-1)} < 0\end{aligned} \tag{1.3}$$

and into π_2 otherwise. We note that the classifier given by (1.2) is equivalent to the one given by (1.3), though the description of (1.2) is much simpler than (1.3).

In this paper, we assume the following assumption for π_is as necessary:

(A-ii) z_{ijl}, $j = 1, ..., p$, are independent for $i = 1, ..., k$.

Yata and Aoshima (2012b) gave the asymptotic normality of the classifier given by (1.2) as $p \to \infty$ and $n_i \to \infty$, $i = 1, 2$, under a condition milder than (A-ii). In the present paper, under (A-ii), we show the asymptotic normality of the classifier holds as $p \to \infty$ even either when n_is are fixed or $n_i \to \infty$ for some i. We evaluate asymptotic error rates for the classifier by using the asymptotic normality. Further, we verify that similar arguments can be applied in multiclass classification when $k \ge 3$.

2. Asymptotic properties for two-class classification

We denote the error of misclassifying an individual from π_1 (into π_2) or π_2 (into π_1) by $e(2|1)$ or $e(1|2)$, respectively. Let $\Delta = ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2$ and

$$w(\boldsymbol{x}_0|n_1, n_2) = \Big(\boldsymbol{x}_0 - \frac{\overline{\boldsymbol{x}}_{1n_1} + \overline{\boldsymbol{x}}_{2n_2}}{2}\Big)^T (\overline{\boldsymbol{x}}_{2n_2} - \overline{\boldsymbol{x}}_{1n_1}) - \frac{\mathrm{tr}(\boldsymbol{S}_{1n_1})}{2n_1} + \frac{\mathrm{tr}(\boldsymbol{S}_{2n_2})}{2n_2}.$$

Yata and Aoshima (2012b) considered asymptotic properties of $w(\boldsymbol{x}_0|n_1, n_2)$ under the following assumptions:

(A-iii) $\dfrac{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_i (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\Delta^2} \to 0$ as $p \to \infty$ for $i = 1, 2$;

(A-iv) $\frac{\max_{j=1,2}\mathrm{tr}(\boldsymbol{\Sigma}_j^2)}{n_i\Delta^2} \to 0$ as $p \to \infty$ either when n_i is fixed or $n_i \to \infty$ for $i = 1, 2$.

Then, they gave the asymptotic consistency:

Theorem 2.1 (Yata and Aoshima, 2012b). *Assume (A-iii) and (A-iv). It holds as $p \to \infty$ that*

$$\frac{w(\boldsymbol{x}_0|n_1,n_2)}{\Delta} = \frac{(-1)^i}{2} + o_p(1) \quad \text{when } \boldsymbol{x}_0 \in \pi_i$$

for $i = 1, 2$. Then, the classification rule given by (1.2) has as $p \to \infty$ that

$$e(2|1) \to 0 \quad \text{and} \quad e(1|2) \to 0. \tag{2.1}$$

Remark 2.1. Under the condition that $\max_{j=1,2}\{\mathrm{tr}(\boldsymbol{\Sigma}_j^2)\}/\Delta^2 \to 0$ as $p \to \infty$, one can claim Theorem 2.1 when either n_i is fixed or $n_i \to \infty$ for $i = 1, 2$.

Remark 2.2. Chan and Hall (2009) gave (2.1) for a different distance-based classifier under different assumptions.

We have for $\boldsymbol{x}_0 \in \pi_i$, $i = 1, 2$, that

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\theta}}\{w(\boldsymbol{x}_0|n_1,n_2)\} &= \frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}{n_i} + \frac{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}{n_j} + \sum_{i=1}^{2}\frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}{2n_i(n_i-1)} \\ &\quad + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T(\boldsymbol{\Sigma}_i+\boldsymbol{\Sigma}_j/n_j)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \quad (= \kappa_i, \text{ say}),\end{aligned}$$

where $j \neq i$. We assume the following assumption:

(A-v) $\frac{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}_i(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}{\kappa_i} = o(1)$ for $i = 1, 2$.

Then, we have the following result.

Theorem 2.2. *Assume (A-i), (A-ii) and (A-v). We have as $p \to \infty$ and at least one: $n_1 \to \infty$ or $n_2 \to \infty$, that*

$$\frac{w(\boldsymbol{x}_0|n_1,n_2) - (-1)^i\Delta/2}{\sqrt{\kappa_i}} \Rightarrow N(0,1) \quad \text{when } \boldsymbol{x}_0 \in \pi_i \text{ for } i = 1, 2, \tag{2.2}$$

where "$\Rightarrow$" denotes the convergence in distribution and $N(0,1)$ denotes a random variable distributed as the standard normal distribution.

We assume extra assumptions for $\boldsymbol{H}_i = (\boldsymbol{h}_{i1}, ..., \boldsymbol{h}_{ip})$, $i = 1, 2$:

(A-vi) $\frac{\sum_{j=1}^{p}\lambda_{ij}^2\{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{h}_{ij}\}^4}{\kappa_i^2} = o(1)$ for $i = 1, 2$;

(A-vii) There exists a permutation $\psi : \{1, ..., p\} \longmapsto \{1, ..., p\}$ such that $|\boldsymbol{h}_{1j}^T\boldsymbol{h}_{2\psi(j)}| = 1$ for $j = 1, ..., p$.

Note that (A-v) implies (A-vi) from the fact that $\sum_{j=1}^p \lambda_{ij}^2 \{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{h}_{ij}\}^4 \le \{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_i (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\}^2$. If $\boldsymbol{H}_1 = \boldsymbol{H}_2$, (A-vii) holds. Thus, (A-vii) naturally follows when $\boldsymbol{\Sigma}_1 = c\boldsymbol{\Sigma}_2$ with a positive constant $c$. Then, we have the following result.

Corollary 2.1. *Assume (A-i), (A-ii), (A-vi) and (A-vii). Then, we have (2.2) as $p \to \infty$ either when n_i is fixed or $n_i \to \infty$ for $i = 1, 2$.*

Remark 2.3. From Theorem 2.2, for the classifier given by (1.2), we have as $p \to \infty$ and at least one: $n_1 \to \infty$ or $n_2 \to \infty$, that

$$e(2|1) = \Phi\Big(\frac{-\Delta}{2\sqrt{\kappa_1}}\Big) + o(1) \quad \text{and} \quad e(1|2) = \Phi\Big(\frac{-\Delta}{2\sqrt{\kappa_2}}\Big) + o(1) \tag{2.3}$$

under (A-i), (A-ii) and (A-v), where $\Phi(\cdot)$ denotes the cumulative distribution function of a $N(0,1)$ random variable. Further, if one can assume (A-i), (A-ii), (A-vi) and (A-vii), it holds (2.3) as $p \to \infty$ either when n_i is fixed or $n_i \to \infty$ for $i = 1, 2$.

Remark 2.4. Chan and Hall (2009) gave the asymptotic normality for the distance-based classifier given by (1.3) (or (1.2)) under different assumptions.

Let us consider an easy example such as $\pi_i :\ N_p(\boldsymbol{\mu}_i,\ \boldsymbol{\Sigma}_i)$, $i = 1, 2$, with $\boldsymbol{\mu}_1 = \boldsymbol{0}$, $\boldsymbol{\mu}_2 = (p^{-1/6}, ..., p^{-1/6})$, $\boldsymbol{\Sigma}_1 = (0.3^{|i-j|^{1/3}})$ and $\boldsymbol{\Sigma}_2 = 1.2\,(0.3^{|i-j|^{1/3}})$. Note that $\Delta = ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2 = p^{2/3}$ and $\mathrm{tr}(\boldsymbol{\Sigma}_i^2) = O(p)$, $i = 1, 2$. One can check that (A-i), (A-ii), (A-v) and (A-vii) hold for fixed n_is. We considered the cases of $p = 2^s$, $s = 1, ..., 10$. We set $n_1 = 5$ and $n_2 = 10$. Independent pseudorandom observations of $w(\boldsymbol{x}_0|n_1, n_2)$ were generated 2000 ($= R$, say) times when $\boldsymbol{x}_0 \in \pi_1$ or π_2, respectively. In the end of the rth replication, we checked whether the rule (1.2) does (or does not) classify $\boldsymbol{x}_0$ correctly (or not) and defined $P_{ir} = 0$ (or 1) accordingly for each π_i. We calculated $\overline{e}(2|1) = R^{-1}\sum_{r=1}^R P_{1r}$ and $\overline{e}(1|2) = R^{-1}\sum_{r=1}^R P_{2r}$ for the estimates of $e(2|1)$ and $e(1|2)$. Note that the standard deviation of the estimates are less than 0.011. In Figure 2.1, we plotted $\overline{e}(2|1)$ and $\overline{e}(1|2)$ together with $\Phi\{-\Delta/(2\sqrt{\kappa_i})\}$, $i = 1, 2$, for each $p$. Here, we calculated $\Phi\{-\Delta/(2\sqrt{\kappa_1})\}$ and $\Phi\{-\Delta/(2\sqrt{\kappa_2})\}$ from Remark 2.3. As expected theoretically, we observed that the plots became close to $\Phi\{-\Delta/(2\sqrt{\kappa_i})\}$ as $p$ increases.

3. Asymptotic properties for multiclass classification

In this section, we consider $k$ (≥ 3)-class classification for high-dimensional data. Let

$$Y_i(\boldsymbol{x}_0|n_i) = ||\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_{in_i}||^2 - \frac{\mathrm{tr}(\boldsymbol{S}_{in_i})}{n_i}$$

for $i = 1, ..., k$. We consider a classification rule given by Yata and Aoshima (2012b): One classifies an individual into π_i if

$$\max\Big\{\operatorname*{argmin}_{j=1,...,k} Y_j(\boldsymbol{x}_0|n_j)\Big\} = i. \tag{3.1}$$

When it holds that $\mathrm{argmin}_{j=1,...,k} Y_j(\boldsymbol{x}_0|n_j) = \{i_1, ..., i_l\}$ with integers $l \in [2, k]$ and $i_1 < \cdots < i_l$, we have $\max\{\mathrm{argmin}_{j=1,...,k} Y_j(\boldsymbol{x}_0|n_j)\} = i_l$. Note that the difference, $Y_1(\boldsymbol{x}_0|n_1)/2 - Y_2(\boldsymbol{x}_0|n_2)/2$, coincides with the classifier, $w(\boldsymbol{x}_0|n_1, n_2)$, given in Section 2. Let $\Delta_{ij} = ||\boldsymbol{\mu}_i - \boldsymbol{\mu}_j||^2$ for $i, j = 1, ..., k$; $i \ne j$. Yata and Aoshima (2012b) considered the following assumptions:

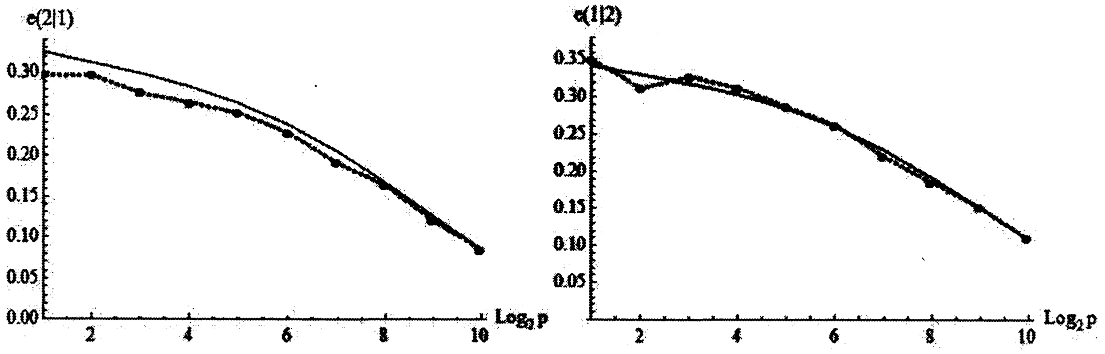

Figure 2.1: When $n_1 = 5$ and $n_2 = 10$, the left panel displays $\overline{e}(2|1)$ (dashed line) and $\Phi\{-\Delta/(2\sqrt{\kappa_1})\}$ (solid line) for $p = 2^s$ $(s = 1, ..., 10)$ and the right panel displays $\overline{e}(1|2)$ (dashed line) and $\Phi\{-\Delta/(2\sqrt{\kappa_2})\}$ (solid line) for $p = 2^s$ $(s = 1, ..., 10)$.

(A-viii) $\dfrac{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_i (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{\Delta_{ij}^2} \to 0$ as $p \to \infty$ for $i, j = 1, ..., k;\ i \neq j$;

(A-ix) $\dfrac{\max_{i'=1,...,k} \mathrm{tr}(\boldsymbol{\Sigma}_{i'}^2)}{n_i \Delta_{ij}^2} \to 0$ as $p \to \infty$ either when n_i is fixed or $n_i \to \infty$ for $i, j = 1, ..., k;\ i \neq j$.

We denote the error of misclassifying an individual from π_i (into another class) by $e(i)$. Then, they gave the consistency property.

Theorem 3.1. (Yata and Aoshima, 2012b). *Assume (A-viii) and (A-ix). Then, the classification rule given by (3.1) has as* $p \to \infty$ *that*

$$e(i) \to 0 \quad \textit{for}\ i = 1, ..., k.$$

Remark 3.1. Under the condition that $\max_{i'=1,...,k}\{\mathrm{tr}(\boldsymbol{\Sigma}_{i'}^2)\}/\Delta_{ij}^2 \to 0$ as $p \to \infty$ for $i, j = 1, ..., k;\ i \neq j$, one can claim Theorem 3.1 when either n_i is fixed or $n_i \to \infty$ for $i = 1, ..., k$.

We have for $\boldsymbol{x}_0 \in \pi_i$, $i = 1, ..., k$, and for $j (\neq i) = 1, ..., k$, that

$$\begin{aligned}
&\mathrm{Var}_{\boldsymbol{\theta}}\{Y_i(\boldsymbol{x}_0|n_i)/2 - Y_j(\boldsymbol{x}_0|n_j)/2\} \\
&= \frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}{n_i} + \frac{\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)}{n_j} + \frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}{2n_i(n_i-1)} + \frac{\mathrm{tr}(\boldsymbol{\Sigma}_j^2)}{2n_j(n_j-1)} \\
&\quad + (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T(\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j/n_j)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \quad (= \kappa_{ij},\ \text{say}).
\end{aligned}$$

We assume the following assumption:

(A-x) $\dfrac{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_i (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{\kappa_{ij}} = o(1)$ for $i, j = 1, ..., k; i \neq j$.

Then, we have the following results under at least $k-1$ out of the k conditions that $n_i \to \infty$, $i=1,...,k$, that is, one of the n_is might be fixed.

Theorem 3.2. *Assume (A-i), (A-ii) and (A-x). We have that*

$$\frac{Y_i(\boldsymbol{x}_0|n_i) - Y_j(\boldsymbol{x}_0|n_j) + \Delta_{ij}}{2\sqrt{\kappa_{ij}}} \Rightarrow N(0,1) \quad \textit{when } \boldsymbol{x}_0 \in \pi_i$$

for $i,j=1,...,k$; $i \neq j$, as $p \to \infty$ under at least $k-1$ out of the k conditions that $n_i \to \infty$, $i=1,...,k$.

Corollary 3.1. *Assume (A-i), (A-ii) and (A-x). For the classification rule given by (3.1), we have that*

$$e(i) \le \sum_{j(\neq i)=1}^{k} \Phi\Big(\frac{-\Delta_{ij}}{2\sqrt{\kappa_{ij}}}\Big) + o(1) \quad \textit{for } i=1,...,k \tag{3.2}$$

as $p \to \infty$ under at least $k-1$ out of the k conditions that $n_i \to \infty$, $i=1,...,k$.

We assume extra assumptions for $\boldsymbol{H}_i = (\boldsymbol{h}_{i1},...,\boldsymbol{h}_{ip})$, $i=1,...,k$:

(A-xi) $\quad \dfrac{\sum_{l=1}^{p} \lambda_{il}^2\{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{h}_{il}\}^4}{\kappa_{ij}^2} = o(1)$ for $i,j=1,...,k$; $i \neq j$;

(A-xii) There exists a permutation $\psi_{ij} : \{1,...,p\} \longmapsto \{1,...,p\}$ such that $|\boldsymbol{h}_{il}^T \boldsymbol{h}_{j\psi_{ij}(l)}|$ $=1$, $l=1,...,p$, for $i,j=1,...,k$; $i \neq j$.

Note that (A-x) implies (A-xi). If $\boldsymbol{H}_1 = \cdots = \boldsymbol{H}_k$, (A-xii) holds. Then, we have the following result.

Corollary 3.2. *Assume (A-i), (A-ii), (A-xi) and (A-xii). Then, no matter whether n_i is fixed or $n_i \to \infty$ for $i=1,...,k$, one can claim the results given by Theorem 3.2 and Corollary 3.1.*

Remark 3.3. Yata and Aoshima (2012b) gave the asymptotic normality and (3.2) for (3.1) under different assumptions.

Appendix

Proof of Theorem 2.2. We assume $\boldsymbol{x}_0 \in \pi_1$ without loss of generality. We first consider the case when $n_1, n_2 \to \infty$. We have from (A-i) and (A-v) that

$$w(\boldsymbol{x}_0|n_1,n_2) + \Delta/2 = (\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T\{(\overline{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2) - (\overline{\boldsymbol{x}}_{1n_1} - \boldsymbol{\mu}_1)\} + o_p(\kappa_1^{1/2}). \tag{A.1}$$

Let us write that $\boldsymbol{H}_1^T(\boldsymbol{x}_0 - \boldsymbol{\mu}_1) = (\lambda_{11}^{1/2} z_{01},...,\lambda_{1p}^{1/2} z_{0p})^T$. Then, we have that $(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T\{(\overline{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2) - (\overline{\boldsymbol{x}}_{1n_1} - \boldsymbol{\mu}_1)\} = \sum_{j=1}^{p} \lambda_{1j}^{1/2} z_{0j}\{\boldsymbol{h}_{1j}^T(\overline{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2) - \lambda_{1j}^{1/2}\overline{z}_{1jn_1}\}$, where $\overline{z}_{ijn_i} = \sum_{l=1}^{n_i} z_{ijl}/n_i$. Let

$$v_j = \frac{\lambda_{1j}^{1/2} z_{0j}\{\boldsymbol{h}_{1j}^T(\overline{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2) - \lambda_{1j}^{1/2}\overline{z}_{1jn_1}\}}{\{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)/n_1 + \mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)/n_2\}^{1/2}}, \quad j=1,...,p.$$

Then, it holds for $j = 2, ..., p$, that $E_{\boldsymbol{\theta}}(v_j | v_{j-1}, ..., v_1) = 0$ under (A-ii). Note that $\sum_{j=1}^{p} E_{\boldsymbol{\theta}}(v_j^2) = 1$. We consider applying the martingale central limit theorem. Refer to Section 2.6 in Ghosh et al. (1997) for the details of the martingale central limit theorem. Let $I(\cdot)$ be the indicator function. Note that $E_{\boldsymbol{\theta}}[\{\boldsymbol{h}_{1j}^T(\overline{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2)\}^4] = O\{(\boldsymbol{h}_{1j}^T\boldsymbol{\Sigma}_2\boldsymbol{h}_{1j})^2/n_2^2\}$. Note that $\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) \le \mathrm{tr}(\boldsymbol{\Sigma}_1^2\boldsymbol{\Sigma}_2^2)$, $\sum_{j=1}^{p} \lambda_{1j}^2(\boldsymbol{h}_{1j}^T\boldsymbol{\Sigma}_2\boldsymbol{h}_{1j})^2 \le \mathrm{tr}(\boldsymbol{\Sigma}_1^2\boldsymbol{\Sigma}_2^2)$, and $\mathrm{tr}(\boldsymbol{\Sigma}_1^3\boldsymbol{\Sigma}_2) \le \{\mathrm{tr}(\boldsymbol{\Sigma}_1^4)\mathrm{tr}(\boldsymbol{\Sigma}_1^2\boldsymbol{\Sigma}_2^2)\}^{1/2}$. Then, by using Chebyshev's inequality and Schwarz's inequality, from (A-i), we have for Lindeberg's condition that

$$
\begin{aligned}
&\sum_{j=1}^{p} E_{\boldsymbol{\theta}}\{v_j^2 I(v_j^2 \ge \tau)\} \\
&\le \sum_{j=1}^{p} \frac{E_{\boldsymbol{\theta}}(v_j^4)}{\tau} = \sum_{j=1}^{p} O\Big[\frac{\lambda_{1j}^2(\boldsymbol{h}_{1j}^T\boldsymbol{\Sigma}_2\boldsymbol{h}_{1j}/n_2 + \lambda_{1j}/n_1)^2}{\{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)/n_1 + \mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)/n_2\}^2}\Big] \\
&\qquad = O\Big[\frac{\mathrm{tr}(\boldsymbol{\Sigma}_1^4)/n_1^2 + \mathrm{tr}(\boldsymbol{\Sigma}_1^3\boldsymbol{\Sigma}_2)/(n_1n_2) + \mathrm{tr}(\boldsymbol{\Sigma}_1^2\boldsymbol{\Sigma}_2^2)/n_2^2}{\{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)/n_1 + \mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)/n_2\}^2}\Big] \to 0 \qquad \text{(A.2)}
\end{aligned}
$$

for any $\tau > 0$. Here, in a way similar to (A.2), we claim that

$$
P_{\boldsymbol{\theta}}\Big(\Big|\sum_{j=1}^{p} v_j^2 - 1\Big| \ge \tau\Big) \le \tau^{-2} E_{\boldsymbol{\theta}}\Big\{\Big(\sum_{j=1}^{p} v_j^2 - 1\Big)^2\Big\} \to 0
$$

for any $\tau > 0$. Thus it holds that $\sum_{j=1}^{p} v_j^2 = 1 + o_p(1)$. Hence, by using the martingale central limit theorem, we obtain that

$$
\sum_{j=1}^{p} v_j \Rightarrow N(0, 1). \qquad \text{(A.3)}
$$

Note that $\kappa_1/\{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)/n_1 + \mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)/n_2\} \to 1$ under (A-i) and (A-v). Then, by combining (A.1) with (A.3), we conclude the result when $\boldsymbol{x}_0 \in \pi_1$ and $n_1, n_2 \to \infty$.

Next, we consider the case when $n_1 \to \infty$ but n_2 is fixed. We have that

$$
\begin{aligned}
&w(\boldsymbol{x}_0 | n_1, n_2) + \frac{\Delta}{2} \\
&= (\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T(\overline{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2) - \frac{\sum_{i \ne i'}(\boldsymbol{x}_{2i} - \boldsymbol{\mu}_2)^T(\boldsymbol{x}_{2i'} - \boldsymbol{\mu}_2)}{2n_2(n_2 - 1)} + o_p(\kappa_1^{1/2}) \\
&= \sum_{j=1}^{p} \boldsymbol{h}_{2j}^T(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)\lambda_{2j}^{1/2}\overline{z}_{2jn_2} - \sum_{j=1}^{p} \frac{\sum_{i \ne i'} \lambda_{2j} z_{2ji} z_{2ji'}}{2n_2(n_2 - 1)} + o_p(\kappa_1^{1/2}).
\end{aligned}
$$

Let us rewrite that

$$
v_j = \frac{\boldsymbol{h}_{2j}^T(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)\lambda_{2j}^{1/2}\overline{z}_{2jn_2} - \sum_{i \ne i'} \lambda_{2j} z_{2ji} z_{2ji'}\{2n_2(n_2 - 1)\}^{-1}}{[\mathrm{tr}(\boldsymbol{\Sigma}_2^2)/\{2n_2(n_2 - 1)\} + \mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)/n_2]^{1/2}}, \quad j = 1, ..., p.
$$

Then, it holds for $j = 2, \dots, p$, that $E_{\boldsymbol{\theta}}(v_j | v_{j-1}, \dots, v_1) = 0$ under (A-ii). Note that $\sum_{j=1}^{p} E_{\boldsymbol{\theta}}(v_j^2) = 1$. Also, note that

$$\begin{aligned}
&\Big| \sum_{j=1}^{p} E_{\boldsymbol{\theta}}[\{\boldsymbol{h}_{2j}^T(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)\lambda_{2j}^{1/2}\overline{z}_{2jn_2}\}^3 \sum_{i \neq i'} \lambda_{2j} z_{2ji} z_{2ji'}] \Big| = O\{\mathrm{tr}(\boldsymbol{\Sigma}_1^{3/2}\boldsymbol{\Sigma}_2^{5/2})\} \\
&= O\{\mathrm{tr}(\boldsymbol{\Sigma}_1^3\boldsymbol{\Sigma}_2)^{1/2}\mathrm{tr}(\boldsymbol{\Sigma}_2^4)^{1/2}\} = O[\{\mathrm{tr}(\boldsymbol{\Sigma}_1^6)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)\}^{1/4}\mathrm{tr}(\boldsymbol{\Sigma}_2^4)^{1/2}] \\
&= O\{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)^{3/4}\mathrm{tr}(\boldsymbol{\Sigma}_2^2)^{1/4}\mathrm{tr}(\boldsymbol{\Sigma}_2^4)^{1/2}\} = o\{\mathrm{tr}(\boldsymbol{\Sigma}_2^2)^2\}
\end{aligned}$$

under (A-i) from the fact that $|E_{\boldsymbol{\theta}}(z_{ijl}^3)| < \infty$, $i = 1, 2$; $j = 1, \dots, p$. Then, similar to the case when $n_1, n_2 \to \infty$, we can claim the result. For the case when n_1 is fixed but $n_2 \to \infty$, similar arguments follow. The proof is completed.

Proof of Corollary 2.1. We assume $\boldsymbol{x}_0 \in \pi_1$ without loss of generality. Under (A-vii), we assume that $\boldsymbol{h}_{1j}^T\boldsymbol{h}_{2\psi(j)} = 1$ for $j = 1, \dots, p$, without loss of generality. Under (A-vii), we have that

$$\begin{aligned}
&w(\boldsymbol{x}_0|n_1, n_2) + \Delta/2 \\
&= (\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T\{(\overline{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2) - (\overline{\boldsymbol{x}}_{1n_1} - \boldsymbol{\mu}_1)\} + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T(\overline{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2 - \boldsymbol{x}_0 + \boldsymbol{\mu}_1) \\
&\quad + \frac{\sum_{i \neq i'}(\boldsymbol{x}_{1i} - \boldsymbol{\mu}_1)^T(\boldsymbol{x}_{1i'} - \boldsymbol{\mu}_1)}{2n_1(n_1 - 1)} - \frac{\sum_{i \neq i'}(\boldsymbol{x}_{2i} - \boldsymbol{\mu}_2)^T(\boldsymbol{x}_{2i'} - \boldsymbol{\mu}_2)}{2n_2(n_2 - 1)} \\
&= \sum_{j=1}^{p} \lambda_{1j}^{1/2} z_{0j}(\lambda_{2\psi(j)}^{1/2}\overline{z}_{2\psi(j)n_2} - \lambda_{1j}^{1/2}\overline{z}_{1jn_1}) + \sum_{j=1}^{p} \frac{\sum_{i \neq i'} \lambda_{1j} z_{1ji} z_{1ji'}}{2n_1(n_1 - 1)} \\
&\quad - \sum_{j=1}^{p} \frac{\sum_{i \neq i'} \lambda_{2\psi(j)} z_{2\psi(j)i} z_{2\psi(j)i'}}{2n_2(n_2 - 1)} + \sum_{j=1}^{p}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\boldsymbol{h}_{1j}(\lambda_{2\psi(j)}^{1/2}\overline{z}_{2\psi(j)n_2} - \lambda_{1j}^{1/2} z_{0j}), \quad \text{(A.4)}
\end{aligned}$$

where $\overline{z}_{1jn_1}$ and z_{0j}s are the ones given in the proof of Theorem 2.2 and $\overline{z}_{2\psi(j)n_2} = \sum_{l=1}^{n_2} z_{2\psi(j)l}/n_2$. Let

$$\begin{aligned}
u_j = \Big\{&\lambda_{1j}^{1/2} z_{0j}(\lambda_{2\psi(j)}^{1/2}\overline{z}_{2\psi(j)n_2} - \lambda_{1j}^{1/2}\overline{z}_{1jn_1}) - \frac{\sum_{i \neq i'} \lambda_{2\psi(j)} z_{2\psi(j)i} z_{2\psi(j)i'}}{2n_2(n_2 - 1)} \\
&+ \frac{\sum_{i \neq i'} \lambda_{1j} z_{1ji} z_{1ji'}}{2n_1(n_1 - 1)} + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\boldsymbol{h}_{1j}(\lambda_{2\psi(j)}^{1/2}\overline{z}_{2\psi(j)n_2} - \lambda_{1j}^{1/2} z_{0j})\Big\}/\kappa_1^{1/2},
\end{aligned}$$

$j = 1, \dots, p$. Note that $E_{\boldsymbol{\theta}}(u_j) = 0$, $j = 1, \dots, p$, and $\mathrm{Var}_{\boldsymbol{\theta}}(\sum_{j=1}^{p} u_j) = 1$. Note that u_j, $j = 1, \dots, p$, are independent under (A-ii). In a way similar to (A.2), from (A-i) and

(A-vi), we have for Lyapunov's condition that

$$\begin{aligned}
&\sum_{j=1}^{p} E_{\boldsymbol{\theta}}(u_j^4) \\
&= \kappa_1^{-2} \times O\Big[\mathrm{tr}(\boldsymbol{\Sigma}_1^4)/n_1^2 + \mathrm{tr}(\boldsymbol{\Sigma}_2^4)/n_2^4 + \mathrm{tr}(\boldsymbol{\Sigma}_1^3\boldsymbol{\Sigma}_2)/(n_1n_2) + \mathrm{tr}(\boldsymbol{\Sigma}_1^2\boldsymbol{\Sigma}_2^2)/n_2^2 \\
&\qquad + \sum_{j=1}^{p}\{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{h}_{1j}\}^4(\lambda_{2\psi(j)}^2/n_2^2 + \lambda_{1j}^2 + \lambda_{1j}\lambda_{2\psi(j)}/n_2)\Big] \\
&= \kappa_1^{-2} \times O\Big[\mathrm{tr}(\boldsymbol{\Sigma}_1^4)/n_1^2 + \mathrm{tr}(\boldsymbol{\Sigma}_2^4)/n_2^4 + \mathrm{tr}(\boldsymbol{\Sigma}_1^2\boldsymbol{\Sigma}_2^2)/n_2^2 \\
&\qquad + \sum_{j=1}^{p}\{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{h}_{1j}\}^4\lambda_{1j}^2 + \sum_{j=1}^{p}\{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{h}_{2j}\}^4\lambda_{2j}^2/n_2^2\Big] \to 0 \quad \text{as } p\to\infty
\end{aligned}$$

from the fact that

$$\kappa_1^{-2}\sum_{j=1}^{p}\{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{h}_{2j}\}^4\lambda_{2j}^2/n_2^2 = O\Big[\kappa_2^{-2}\sum_{j=1}^{p}\{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{h}_{2j}\}^4\lambda_{2j}^2\Big].$$

Hence, by using Lyapunov's central limit theorem, it holds that

$$\sum_{j=1}^{p} u_j \Rightarrow N(0,1) \tag{A.5}$$

as $p \to \infty$ either when n_i is fixed or $n_i \to \infty$ for $i = 1, 2$. Then, by combining (A.4) with (A.5), we conclude the results. The proof is completed.

Proofs of Theorem 3.2 and Corollary 3.1. From Theorem 2.2, we have under (A-i), (A-ii) and (A-x) that $\{Y_i(\boldsymbol{x}_0|n_i) - Y_j(\boldsymbol{x}_0|n_j) + \Delta_{ij}\}/(2\kappa_{ij}^{1/2}) \Rightarrow N(0,1)$ when $\boldsymbol{x}_0 \in \pi_i$ for $j = 1, ..., k$; $j \neq i$. Then, from Bonferroni's inequality, it holds that $1 - e(i) \geq 1 - \sum_{j(\neq i)=1}^{k} \Phi\{-\Delta_{ij}/(2\kappa_{ij}^{1/2})\} + o(1)$ when $\boldsymbol{x}_0 \in \pi_i$. This concludes the proofs.

Proof of Corollary 3.2. In a way similar to the proof of Corollary 2.1, we can conclude the results.

Acknowledgment

Research of the first author was partially supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 23740066. Research of the second author was partially supported by Grant-in-Aid for Scientific Research (B) and Challenging Exploratory Research, JSPS, under Contract Numbers 22300094 and 23650142.

References

Ahn, J., Marron, J. S., Muller, K. M. and Chi, Y.-Y. (2007). The high-dimension, low-sample-size geometric representation holds under mild conditions. *Biometrika* **94**, 760-766.

Aoshima, M. and Yata, K. (2011a). Two-stage procedures for high-dimensional data. *Sequential Anal. (Editor's special invited paper)* **30**, 356-399.

Aoshima, M. and Yata, K. (2011b). Author's response. *Sequential Anal.* **30**, 432-440.

Aoshima, M. and Yata, K. (2011c). Effective methodologies for statistical inference on microarray studies. *In: P.E. Spiess (Ed.), Prostate Cancer - From Bench to Bedside*, InTech, 13-32.

Bickel, P. 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.

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. *J. Am. Statist. Assoc.* **97**, 77-87.

Ghosh, M., Mukhopadhyay, N. and Sen, P. K. (1997). *Sequential Estimation.* Wiley, New York.

Hall, P., Marron, J. S. and Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *J. R. Statist. Soc.* B **67**, 427-444.

Hall, P., Pittelkow, Y. and Ghosh, M. (2008). Theoretical measures of relative performance of classifiers for high dimensional data with small sample sizes. *J. R. Statist. Soc.* B **70**, 159-173.

Jung, S. and Marron, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37**, 4104-4130.

Saranadasa, H. (1993). Asymptotic expansion of the misclassification probabilities of D-and A-criteria for discrimination from two high dimensional populations using the theory of large dimensional random matrices. *J. Multivariate Anal.* **46**, 154-174.

Yata, K. and Aoshima, M. (2009). PCA consistency for non-Gaussian data in high dimension, low sample size context. *Commun. Statist. Theory Methods, Special Issue Honoring Zacks, S. (ed. Mukhopadhyay, N)* **38**, 2634-2652.

Yata, K. and Aoshima, M. (2010a). Intrinsic dimensionality estimation of high-dimension, low sample size data with d-asymptotics. *Commun. Statist. Theory Methods, Special Issue Honoring Akahira, M. (ed. Aoshima, M.)* **39**, 1511-1521.

Yata, K. and Aoshima, M. (2010b). Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. *J. Multivariate Anal.* **101**, 2060-2077.

Yata, K. and Aoshima, M. (2012a). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivariate Anal.* **105**, 193-215.

Yata, K. and Aoshima, M. (2012b). Misclassification rate adjusted classifier for multiclass, high-dimensional data, submitted.