**A comparison of quadratic discriminant function with discriminant function based on absolute deviation from the mean**

S. Ganeslingam [1,*]

A. Nanthakumar [2]

Siva Ganesh [1,†]

[1] *Institute of Information Sciences and Technology*
*College of Sciences*
*Massey University*
*Private Bag 11 222*
*Palmerston North*
*New Zealand*

[2] *Department of Mathematics*
*State University of New York*
*Oswego, New York, 13126*
*U.S.A.*

**Abstract**

Consider the problem of statistical discrimination involving two multivariate normal distributions with equal means but different covariance matrices. Traditionally we use a *quadratic discriminant function* (QDF) to separate two such populations. A simple model for such situations is to perform a linear discriminant analysis on the absolute value of deviations from the mean. In this paper some theoretical results on this alternative approach are reported.

We introduce a linear discriminant function called *Absolute Euclidean Distance Classifier* (AEDC) and compare its performance with that of QDF. A real life case study was carried out to illustrate the superior performance of AEDC.

***Keywords :*** *Multivariate normal distributions, linear discriminant function, quadratic discriminant function, Euclidean distance classifier, contaminated data.*

* *E-mail*: s.ganeslingam@massey.ac.nz
† *E-mail*: s.ganesh@massey.ac.nz

***Journal of Statistics & Management Systems***
Vol. 9 (2006), No. 2, pp. 441–457

## 1. Introduction

Consider the problem of statistical discrimination involving two multivariate normal populations $\Pi_1$ and $\Pi_2$ with mean vectors $\mu_1$ and $\mu_2$ and covariance matrices $\Sigma_1$ and $\Sigma_2$, respectively. These parameters are not generally known. The discriminant function which would normally be used in such a situation is the *quadratic discriminant function* (QDF), which allocates an object with observation vector $x$ to $\Pi_1$ if

$$\ln\left\{\frac{|\Sigma_1|}{|\Sigma_2|}\right\} - \left\{(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)\right\} + \left\{(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right\} > 0\,, \tag{1}$$

otherwise it is allocated to $\Pi_2$ (see for example Morrison (1990)). In the above allocation rule and throughout this paper, we assume that the prior probabilities of both populations are equal.

If $\Sigma_1 = \Sigma_2 = \Sigma$ the allocation is based on the linear discriminant function, so that the object with observation vector $x$ is allocated to population $\Pi_1$ if

$$\left\{x - \frac{1}{2}(\mu_1+\mu_2)\right\}^T \Sigma^{-1}(\mu_1-\mu_2) > 0\,, \tag{2}$$

otherwise it is allocated to $\Pi_2$.

Friedman (1989) proposed a *Regularised Discriminant Function* (RDF) as a compromise between the normal-based linear and quadratic discriminant functions, by considering alternatives to the usual maximum likelihood estimates for the covariance matrices. These alternatives are characterised by two regularisation parameters, the values of which are customised to individual situations by jointly minimising a sample-based future misclassification risk; see Section 5.5 of McLachlan (1992), and Koolaard et al (1998, 2002). At this point it is worth mentioning the logistic regression classifier, which is more robust than the QDF for departure from normality. Asymptotic properties of the logistic discriminant function have been studied by Efron (1975) for cases of equal variance normal populations and generally by O'Neill (1980, 1992, 1994), and Press and Wilson (1978).

It is clear that when the data is from a normal distribution and $\mu_1 = \mu_2$ in expression (2), the linear discriminant function cannot be applied and one would resort to using the QDF if the covariance matrices are unequal; see Section 3.2.4 of McLachlan (1992). In this situation, Lachenbruch

(1975) suggested that a linear discriminant analysis be performed on the absolute values of the deviations of the observations from the (sample) mean. He obtained the *Absolute Linear Discriminant Function* (ALDF) for univariate normal data and showed (via simulation) that the ALDF was more robust (than the QDF) when long-tail contamination present in the data. It was also observed that in other situations the QDF seemed to perform better.

The motivation for Lachenbruch's work was the fact that at the time, because of limited computing facilities, the QDF was considered to demand too much computation, so the ALDF (being linear in $x$) could be used as an alternative. This reason is not really valid any more, since computing resources are now widely available. It is still of interest however, to investigate the properties and behaviour of a discriminant function which is linear in $x$ and to compare its properties with QDF. One such discriminant function (i.e. linear in $x$) can be obtained through the Euclidean distance classifier, a rarely used competitor to linear discriminant function; see for example Marco, Young and Turner (1987).

The *Euclidean Distance Classifier* (EDC) ignores the covariance matrix $\Sigma$ and allocates an individual with observation vector $x$ according to the following rule:

Allocate the observation vector $x$ to $\Pi_1$ if

$$\left\{x - \frac{1}{2}(\mu_1 + \mu_2)\right\}^T (\mu_1 - \mu_2) > 0 \tag{3}$$

otherwise allocate it to $\Pi_2$.

It has been shown that the EDC may perform better than the linear discriminant function under certain circumstances. Note that in its original form, the Euclidean distance classifier cannot be used when $\mu_1 = \mu_2$. We thus consider the *Absolute Euclidean Distance Classifier* (AEDC), whereby the absolute values of the components of the observation vector are used in the EDC. The expectation is that it may do well, particularly in high dimensional settings, since it is also a form of regularisation.

In the literature, the problem of discrimination between two populations with zero mean vectors has been considered by Bartlett and Please (1963), Penrose (1946), Okamoto (1961), Geisser (1964) and Marco, Young and Turner (1987), among others. A predictive Bayesian approach to this problem was presented by Geisser and Desu (1968). It is stated in some of these papers that the assumption of zero means is particularly appropriate for the random vector representing each paired difference of a set of characteristics from either monozygotic or dizygotic twin populations. Raudys and Pikelis (1980) compared (under the assumption of equal variance covariance) the sample LDF with three other discriminant functions, including the sample EDC, when classifying individuals from two spherical normal populations. They concluded that the sample EDC outperforms the sample LDF when $p$ is large relative to the training sample size. Further, they concluded that in many practical situations, the sample EDC performs as well as or better than the sample LDF, even for non-spherical covariance configurations. Unfortunately these EDC, and LDF become inapplicable when the means of the two populations are equal. In real practice $\Sigma_1 \neq \Sigma_2$ and in such a situation the only alternative is to use the QDF, or AEDC which is based on absolute values of the deviations of the observations from the mean value. Recently Ganesalingam and Ganesh (2004) studied the performance of QDF and AEDC in discriminating two bivariate normal populations. They noted via a large scale simulations, in 89 of the 99 different covariance structures considered AEDC out performs QDF.

The main objective of this paper is therefore to extend the theoretical results to any dimension $p$, and to compare the performance of the AEDC and the usual QDF in the situation where $\mu_1 = \mu_2$ and $\Sigma_1 \neq \Sigma_2$. It will be seen that the AEDC ignores completely the covariance information and forms a discriminant function purely on the basis of the absolute deviations. In this paper we consider trivariate normal distributions and compare the performances of the two discriminant functions using error rates (misclassification probabilities) obtained via a cross-validation technique as well as based on the apparent error rates. In Section 2, we give the theoretical derivation of the distribution function of the vector of absolute values in the case of trivariate normal populations. It is interesting to note that there exists a general pattern and a symmetrical nature in the expression for the mean vector and the variance-covariance matrix. Therefore, the results are extended for any $p$ dimensional normal population. In Section 3 we derive the allocation rule based on the absolute deviations and a real life case study is reported in Section 4, followed by concluding remarks in Section 5.

## 2. Distribution of absolute vaues of a $p$-dimensional observation

Ganesalingam and Ganesh (2004) studied the distribution of absolute values for the case $p = 2$ and noted the superior performance of AEDC compared to the traditional QDF classifier. Further, in this paper the authors declared that the probability density function $f_Y(y)$ of $Y = |X|$, where $X = (x_1, \dots, x_p)$ is a $p$-dimensional vector of observation with spherically symmetrical distribution, can be written as

$$f_Y(y) = \Sigma g_X(a_1x_1, a_2x_2, \dots, a_px_p)$$

where the summation is taken over all $(a_1, a_2, \dots, a_p)$ such that $a_i = 1$ or $-1$ for $i = 1, 2, \dots p$; and $g_X(x)$ is the probability density function of the vector of observation $X$.

### 2.1 *Probability density function* (*pdf*) *of* $Y = |X|$ *for* $p = 3$

In this section our attention is focused on deriving the pdf of $Y = |X|$ for $p = 3$, with the hope of seeing a pattern in the expression for the mean vector and the variance-covariance matrix, so that we can write these down for a general $p$.

Let us consider the trivariate normal observation $X = (x_1, x_2, x_3)^T$ and say that this vector has a probability density function $g(x)$ with mean zero and variance-covariance matrix $\Sigma_x$.

If $Y = (y_1, y_2, y_3) = (|x_1|, |x_2|, |x_3|)$ then

$$\begin{aligned} f_Y(y) &= g(x_1, x_2, x_3) + g(-x_1, -x_2, -x_3) + g(x_1, -x_2, -x_3) \\ &\quad + g(-x_1, x_2, -x_3) + g(-x_1, -x_2, x_3) + g(-x_1, x_2, x_3) \\ &\quad + g(x_1, -x_2, x_3) + g(x_1, x_2, -x_3)\,. \end{aligned}$$

Note that

$$\begin{aligned} g(x_1, x_2, x_3) &= g(-x_1, -x_2, -x_3)\,, \\ g(x_1, x_2, -x_3) &= g(-x_1, -x_2, x_3)\,, \\ g(x_1, -x_2, x_3) &= g(-x_1, x_2, -x_3)\,, \\ g(-x_1, x_2, x_3) &= g(x_1, -x_2, -x_3)\,. \end{aligned}$$

So this implies

$$\begin{aligned} f_Y(y) &= 2g(x_1, x_2, x_3) + 2g(-x_1, x_2, x_3) + 2g(x_1, -x_2, x_3) \\ &\quad + 2g(x_1, x_2, -x_3)\,. \end{aligned}$$

Here, the observation vector $X$ has a trivariate normal distribution with mean $= 0$ and covariance matrix $\Sigma_x$ with entries of $\sigma_{ij}$, $i, j = 1, 2, 3$.

Assume that the inverse of this covariance matrix $\Sigma^{-1}$ has as its entries $a_{ij}$, $i, j = 1, 2, 3$.

Then it follows

$$f(y)=\frac{2}{(2\pi)^{\frac{3}{2}}\sqrt{D}}$$

$$\times\Big\{\exp^{-\frac{1}{2}}\Big[-\frac{1}{2}(a_{11}x_1^2+a_{22}x_2^2+a_{33}x_3^2+2a_{12}x_1x_2+2a_{13}x_1x_3+2a_{23}x_2x_3)\Big]$$
$$+\exp^{-\frac{1}{2}}\Big[-\frac{1}{2}(a_{11}x_1^2+a_{22}x_2^2+a_{33}x_3^2-2a_{12}x_1x_2-2a_{13}x_1x_3+2a_{23}x_2x_3)\Big]$$
$$+\exp^{-\frac{1}{2}}\Big[-\frac{1}{2}(a_{11}x_1^2+a_{22}x_2^2+a_{33}x_3^2-2a_{12}x_1x_2+2a_{13}x_1x_3-2a_{23}x_2x_3)\Big]$$
$$+\exp^{-\frac{1}{2}}\Big[-\frac{1}{2}(a_{11}x_1^2+a_{22}x_2^2+a_{33}x_3^2+2a_{12}x_1x_2-2a_{13}x_1x_3-2a_{23}x_2x_3)\Big]\Big\}$$

where $D=|\Sigma|=\det(\Sigma)$, and $\Sigma^{-1}=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$.

Due to Cholesky decomposition $\exists$ a positive definite matrix $\Gamma$ such that $\Gamma'\Gamma=\Sigma^{-1}$ where

$$\Gamma = \begin{pmatrix} \sqrt{a_{11}-\dfrac{\left(a_{12}-\dfrac{a_{23}a_{13}}{a_{33}}\right)^2}{\left(a_{22}-\dfrac{a_{23}^2}{a_{33}}\right)}-\dfrac{a_{13}^2}{a_{33}}} & 0 & 0 \\ \dfrac{a_{12}-\dfrac{a_{23}a_{13}}{a_{33}}}{\sqrt{a_{22}-\dfrac{a_{23}^2}{a_{33}}}} & \sqrt{a_{22}-\dfrac{a_{23}^2}{a_{33}}} & 0 \\ \dfrac{a_{13}}{\sqrt{a_{33}}} & \dfrac{a_{23}}{\sqrt{a_{33}}} & \sqrt{a_{33}} \end{pmatrix}$$
$$= (\Gamma_{ij}), \quad i,j=1,2,3\,.$$

Using this, it is easy to see

$$D=\det(\Sigma)=\left\{a_{11}-\frac{\left(a_{12}-\dfrac{a_{23}a_{13}}{a_{33}}\right)^2}{\left(a_{22}-\dfrac{a_{23}^2}{a_{33}}\right)}-\frac{a_{13}^2}{a_{33}}\right\}^{-1}\left(a_{22}-\frac{a_{23}^2}{a_{33}}\right)^{-1}a_{33}^{-1}$$

and

$$\begin{aligned} E(y_1) &= E(|x_1|) \\ &= \sqrt{\frac{2}{\pi}}\left\{a_{11} - \frac{\left(a_{12} - \frac{a_{23}a_{13}}{a_{33}}\right)^2}{\left(a_{22} - \frac{a_{23}^2}{a_{33}}\right)} - \frac{a_{13}^2}{a_{33}}\right\}^{-\frac{1}{2}} \\ &= \sqrt{\frac{2}{\pi}\sigma_{11}}\,. \end{aligned}$$

***Proof.*** Note that

$$\Sigma^{-1} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}^{-1} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

and $\det(\Sigma) = D$ .

Then

$$\begin{aligned} a_{11} &= \frac{(\sigma_{22}\sigma_{33} - \sigma_{23}^2)}{D} \\ a_{12} &= \frac{-(\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23})}{D} \\ a_{13} &= \frac{(\sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{23})}{D} \\ a_{22} &= \frac{(\sigma_{11}\sigma_{33} - \sigma_{13}^2)}{D} \\ a_{23} &= \frac{-(\sigma_{11}\sigma_{23} - \sigma_{13}\sigma_{22})}{D} \\ a_{33} &= \frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}{D}\,. \end{aligned}$$

Also, one can write

$$a_{22} - \frac{a_{23}^2}{a_{33}} = \frac{\begin{pmatrix} \sigma_{11}^2\sigma_{22}\sigma_{33} + 2\sigma_{11}\sigma_{12}\sigma_{13}\sigma_{23} \\ -\sigma_{11}\sigma_{22}\sigma_{13}^2 - \sigma_{11}\sigma_{12}^2\sigma_{33} - \sigma_{11}^2\sigma_{23}^2 \end{pmatrix}}{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)D}$$

Also

$$\left(a_{22} - \frac{a_{23}^2}{a_{33}}\right)a_{33} = \begin{pmatrix} \sigma_{11}^2\sigma_{22}\sigma_{33} + 2\sigma_{11}\sigma_{12}\sigma_{13}\sigma_{23} \\ -\sigma_{11}\sigma_{22}\sigma_{13}^2 - \sigma_{11}\sigma_{12}^2\sigma_{33} - \sigma_{11}^2\sigma_{23}^2) \end{pmatrix} D^{-1}$$

Note that

$$\left\{ a_{11} - \frac{\left(a_{12} - \frac{a_{23}a_{13}}{a_{33}}\right)^2}{\left(a_{22} - \frac{a_{23}^2}{a_{33}}\right)} \right\} - \frac{a_{13}^2}{a_{33}} \left(a_{22} - \frac{a_{23}^2}{a_{33}}\right) a_{33} = D^{-1} .$$

This implies

$$a_{11} - \frac{\left(a_{12} - \frac{a_{23}a_{13}}{a_{33}}\right)^2}{\left(a_{22} - \frac{a_{23}^2}{a_{33}}\right)} - \frac{a_{13}^2}{a_{33}} = \frac{D^{-1}D^2}{\begin{pmatrix} \sigma_{11}^2\sigma_{22}\sigma_{33} + 2\sigma_{11}\sigma_{12}\sigma_{13}\sigma_{23} \\ -\sigma_{11}\sigma_{22}\sigma_{13}^2 - \sigma_{11}\sigma_{12}^2\sigma_{33} - \sigma_{11}^2\sigma_{23}^2 \end{pmatrix}} .$$

Also

$$D = \det(\Sigma) = \sigma_{11}\sigma_{22}\sigma_{33} + 2\sigma_{12}\sigma_{13}\sigma_{23} - \sigma_{22}\sigma_{13}^2 - \sigma_{12}^2\sigma_{33} - \sigma_{11}\sigma_{23}^2 ,$$

$$a_{11} - \frac{\left(a_{12} - \frac{a_{23}a_{13}}{a_{33}}\right)^2}{\left(a_{22} - \frac{a_{23}^2}{a_{33}}\right)} - \frac{a_{13}^2}{a_{33}} = \frac{1}{\sigma_{11}} (= \Gamma_{11}^2) .$$

This means

$$\begin{aligned} E(y_1) &= E(|x_1|) \\ &= \frac{2}{(2\pi)^{\frac{3}{2}}\sqrt{D}} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\Gamma_{33}x_3 + \Gamma_{32}x_2 + \Gamma_{31}x_1)^2 dx_3} \\ &\quad \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\Gamma_{22}x_2 + \Gamma_{21}x_1)^2 dx_2} \cdot \int_0^{\infty} x_1 \cdot e^{-\frac{1}{2}(\Gamma_{11}x_1)^2 dx_1} \\ &= \frac{2}{(2\pi)^{\frac{3}{2}}\sqrt{D}} \cdot \frac{\sqrt{2\pi}}{\Gamma_{33}} \cdot \frac{\sqrt{2\pi}}{\Gamma_{22}} \cdot \frac{1}{\Gamma_{11}^2} \end{aligned}$$

where

$$\begin{aligned} \sqrt{D} &= [\Gamma_{33}^{-1}\Gamma_{22}^{-1}\Gamma_{11}^{-1}] = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\Gamma_{11}} \\ &= \sqrt{\frac{2}{\pi}} \cdot \left\{ a_{11} - \frac{\left(a_{12} - \frac{a_{23}a_{13}}{a_{33}}\right)^2}{\left(a_{22} - \frac{a_{23}^2}{a_{33}}\right)} - \frac{a_{13}^2}{a_{33}} \right\}^{-1} \\ &= \sqrt{\frac{2}{\pi}\sigma_{11}} . \end{aligned}$$

Similarly,

$$E(y_2) = E(|x_2|) = \sqrt{\frac{2}{\pi}\sigma_{22}} \quad \text{and} \quad E(y_3) = E(|x_3|) = \sqrt{\frac{2}{\pi}\sigma_{33}}\,.$$

Note that

$$\begin{aligned}
E(y_1 y_2) &= E(|x_1| \cdot |x_2|) \\
&= \frac{4}{(2\pi)^{\frac{3}{2}}\sqrt{D}} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\Gamma_{33}x_3+\Gamma_{32}x_2+\Gamma_{31}x_1)^2 dx_3} \\
&\quad \cdot \int_0^{\infty} x_2 \cdot e^{-\frac{1}{2}(\Gamma_{22}x_2+\Gamma_{21}x_1)^2 dx_2} \cdot \int_0^{\infty} x_1 \cdot e^{-\frac{1}{2}(\Gamma_{11}x_1)^2 dx_1} \\
&= \frac{4}{(2\pi)^{\frac{3}{2}}\sqrt{D}} \cdot \frac{\sqrt{2\pi}}{\Gamma_{33}} \cdot \frac{1}{\Gamma_{22}^2} \cdot \frac{1}{\Gamma_{11}^2} \\
&= \frac{2}{\pi\Gamma_{11}\cdot\Gamma_{22}}\,.
\end{aligned}$$

But $\Gamma_{11} = \dfrac{1}{\sqrt{\sigma_{11}}}$ and $\Gamma_{22} = \sqrt{\dfrac{\sigma_{11}}{\sigma_{11}\sigma_{12} - \sigma_{12}^2}}$

$$\therefore \qquad E(y_1 y_2) = \frac{2}{\pi\Gamma_{11}\cdot\Gamma_{22}} = \frac{2}{\pi}\sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\,.$$

This gives

$$\begin{aligned}
\operatorname{cov}(y_1, y_2) &= E(y_1 y_2) - E(y_1)\cdot E(y_2) \\
&= \frac{2}{\pi}\sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2} - \left(\sqrt{\frac{2}{\pi}\sigma_{11}}\right)\left(\sqrt{\frac{2}{\pi}\sigma_{22}}\right) \\
&= \frac{2}{\pi}\left\{\sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2} - \sqrt{\sigma_{11}\sigma_{22}}\right\}.
\end{aligned}$$

Similarly

$$\operatorname{cov}(y_1, y_3) = \frac{2}{\pi}\left\{\sqrt{\sigma_{11}\sigma_{33} - \sigma_{13}^2} - \sqrt{\sigma_{11}\sigma_{33}}\right\}$$

and

$$\operatorname{cov}(y_2, y_3) = \frac{2}{\pi}\left\{\sqrt{\sigma_{22}\sigma_{33} - \sigma_{23}^2} - \sqrt{\sigma_{22}\sigma_{33}}\right\}.$$

Moreover,

$$E(y_1^2) = \sigma_{11}\,.$$

So we have,

$$\operatorname{var}(y_1) = E(y_1^2) - (E(y_1))^2 = \left(1 - \frac{2}{\pi}\right)\sigma_{11}\,.$$

Similarly,

$$\mathrm{var}(y_2) = \left(1 - \frac{2}{\pi}\right)\sigma_{22} \quad \text{and} \quad \mathrm{var}(y_3) = \left(1 - \frac{2}{\pi}\right)\sigma_{33}\,.$$

2.2 *Extension: probability density function of $Y = |X|$ for any dimension $p$*

We can see from Section 2.1 a clear pattern in the expression for the mean vector and the variance covariance matrix of the folded normal distribution obtained for $p = 3$. Hence we can now generalise the result for any $p$.

*Generalisation*

If $Y = |X|$ where $X$ has a $p$-dimensional normal distribution with mean $o$ and variance-covariance matrix $\Sigma_x$, the distribution of $Y$ is a folded multivariate normal with mean $\mu_Y$, a $p \times 1$ vector given by:

$$\mu_Y = \left[\sqrt{\frac{2}{\pi}\sigma_{11}}\,, \sqrt{\frac{2}{\pi}\sigma_{22}}\,, \ldots\,, \sqrt{\frac{2}{\pi}\sigma_{pp}}\right]^T$$

and $\Sigma_Y$, a $p \times p$ symmetric matrix given by

$$\mathrm{diagonal}(\Sigma_Y) = \left(1 - \frac{2}{\pi}\right)\sigma_{ii}\,, \quad i = 1, 2, \ldots, p$$

with upper triangular elements of $\Sigma_Y$, for $i < j$ given by

$$\frac{2}{\pi}\left\{\sqrt{\sigma_{ii}\sigma_{jj} - \sigma_{ij}^2} - \sqrt{\sigma_{ii}\sigma_{jj}}\right\}; \quad i, j = 1, 2, \ldots, p\,.$$

For example, for $p = 2$

$$\begin{aligned}
\mathrm{cov}(Y_1, Y_2) &= \frac{2}{\pi}\left\{\sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2} - \sqrt{\sigma_{11}\sigma_{22}}\right\},\\
\mathrm{cov}(Y_1, Y_3) &= \frac{2}{\pi}\left\{\sqrt{\sigma_{11}\sigma_{33} - \sigma_{13}^2} - \sqrt{\sigma_{11}\sigma_{33}}\right\},\\
\mathrm{cov}(Y_2, Y_3) &= \frac{2}{\pi}\left\{\sqrt{\sigma_{22}\sigma_{33} - \sigma_{23}^2} - \sqrt{\sigma_{22}\sigma_{33}}\right\}.
\end{aligned}$$

The proof of this generalisation directly follows by mathematical induction via Cholesky decomposition of the inverse of the variance-covariance matrix $\Sigma_x$. The results for $p = 1$, was first proved by Lachenbruch (1975), for $p = 2$ by Ganesalingam and Ganesh (2004) and in this paper for $p = 3$. Therefore the result is true for any positive integer value of $p$. Note that there is no covariance term for the case $p = 1$.

## 3. Discrimination using absolute values

The usual sample LDF is known to perform poorly when the number of features $p$ is large relative to the size of the training sample. As noted earlier, the simple and rarely applied alternative to the LDF is the sample *Euclidean Distance Classifier* (EDC). Raudys and Pikelis (1980) concluded that the sample EDC when classifying individuals from two spherical normal populations outperforms the sample LDF when $p$ is large relative to the training sample size. Further, they showed that the sample EDC performs reasonably well when compared to the sample LDF, even for non-spherical covariance configurations.

The LDF is the most popular discriminant function in use today, due in part to its optimal properties when the parameters are known, Anderson (1984). However, when the parameters are unknown and must be estimated, it is well known that the sample LDF is no longer uniformly optimal and behaves poorly when the dimension is large relative to the training sample size. This is mainly due to the poor quality of the parameter estimates in high dimension; see for example Van Ness and Simpson (1976), and Van Ness (1979). Note also that the sample EDC requires no matrix inversions and no estimation of the covariance matrix and therefore avoids the difficulties of estimating the covariance matrix when the data dimension is large relative to the training sample size.

Recently, Marco et al (1987) studied the conditions under which the EDC becomes a Bayes classifier, in terms of error rates. They suggested that in many practical situations the sample EDC may perform better than the sample LDF, since considerably fewer parameters are estimated for the sample EDC. In this section we develop the EDC based on the absolute values of the original observation vector for a trivariate normal data and we call this the *Absolute Euclidean Distance Classifier* (AEDC).

The *Euclidean Distance Classifier* (EDC) will allocate an individual observation vector $X$ according to the following rule (3) (re-written below) to population 1: if

$$\left\{x - \frac{1}{2}(\mu_1 + \mu_2)\right\}^T \cdot (\mu_1 - \mu_2) > 0$$

otherwise, to population 2.

In the event of equal means, using the absolute values, $Y = |X|$, this rule takes the form: allocate a three dimensional observation vector $X$ to population 1 if:

$$y_1(\mu_1^{(1)} - \mu_1^{(2)}) - \frac{1}{2}\Big((\mu_1^{(1)})^2 - (\mu_1^{(2)})^2\Big) + y_2(\mu_2^{(1)} - \mu_2^{(2)}) \\ -\frac{1}{2}\Big((\mu_2^{(1)})^2 - (\mu_2^{(2)})^2\Big) + y_3(\mu_3^{(1)} - \mu_3^{(2)}) - \frac{1}{2}\Big((\mu_3^{(1)})^2 - (\mu_3^{(2)})^2\Big) > 0 \quad (4)$$

where $\mu_i^{(k)}$ is the mean of the $i$th component of $Y$ in the $k$th population $i = 1, 2, 3$ and $k = 1, 2$. So, the *Absolute Euclidean Distance Classifier* (AEDC) reads as:

Allocate the observation vector $X$ to population 1 if,

$$y_1\left(\sqrt{\frac{2}{\pi}\sigma_{11}^{(1)}} - \sqrt{\frac{2}{\pi}\sigma_{11}^{(2)}}\right) - \frac{1}{2}\left(\frac{2}{\pi}(\sigma_{11}^{(1)} - \sigma_{11}^{(2)})\right) \\ + y_2\left(\sqrt{\frac{2}{\pi}\sigma_{22}^{(1)}} - \sqrt{\frac{2}{\pi}\sigma_{22}^{(2)}}\right) - \frac{1}{2}\left(\frac{2}{\pi}(\sigma_{22}^{(1)} - \sigma_{22}^{(2)})\right) \\ + y_3\left(\sqrt{\frac{2}{\pi}\sigma_{33}^{(1)}} - \sqrt{\frac{2}{\pi}\sigma_{33}^{(2)}}\right) - \frac{1}{2}\left(\frac{2}{\pi}(\sigma_{33}^{(1)} - \sigma_{33}^{(2)})\right) > 0$$

else to population 2.

This means, allocate an observation $X$ to population 1 if

$$\Big[y_1\Big(\sqrt{\sigma_{11}^{(1)}} - \sqrt{\sigma_{11}^{(2)}}\Big) + y_2\Big(\sqrt{\sigma_{22}^{(1)}} - \sqrt{\sigma_{22}^{(2)}}\Big) + y_3\Big(\sqrt{\sigma_{33}^{(1)}} - \sqrt{\sigma_{33}^{(2)}}\Big)\Big] \\ \geq \frac{1}{2}\sqrt{\frac{2}{\pi}}\Big[\Big(\sigma_{11}^{(1)} - \sigma_{11}^{(2)}\Big) + \Big(\sigma_{22}^{(1)} - \sigma_{22}^{(2)}\Big) + \Big(\sigma_{33}^{(1)} - \sigma_{33}^{(2)}\Big)\Big] \quad (5)$$

otherwise to population 2.

This is the rule we employed with our case study reported in Section 4 where the data we considered is 3 dimensional. Note that this discriminant function is a linear function of $Y$ and the elements of the variance-covariance matrix $\Sigma_x$ of the original observation vector $X$. By the symmetrical nature, this result now generalizes to any dimension $p$ as follows. The observation vector $X$ is allocated to population 1 if

$$\sum_{i=1}^{p}\left[y_i\Big(\sqrt{\sigma_{ii}^{(1)}} - \sqrt{\sigma_{ii}^{(2)}}\Big) - \frac{1}{2}\sqrt{\frac{2}{\pi}}(\sigma_{ii}^{(1)} - \sigma_{ii}^{(2)})\right] > 0 \quad (6)$$

otherwise to population 2. It is interesting to note that for $p = 1$, allocation rule (6) above reduces to

$$y - \frac{1}{2}\sqrt{\frac{2}{\pi}}\Big(\sqrt{\sigma_{11}^{(1)}} + \sqrt{\sigma_{11}^{(2)}}\Big) > 0\,.$$

This is exactly same as Lachenbruch (1975), where he used $\sqrt{\sigma_{11}^{(1)}} = \sigma_1$, $\sqrt{\sigma_{11}^{(2)}} = \sigma_2$.

For $p = 2$, the allocation rule (6) takes the form

$$y_1\left(\sqrt{\sigma_{11}^{(1)}} - \sqrt{\sigma_{11}^{(2)}}\right) + y_2\left(\sqrt{\sigma_{22}^{(1)}} - \sqrt{\sigma_{22}^{(2)}}\right) \geq \frac{1}{2}\sqrt{\frac{2}{\pi}}\left[\left(\sigma_{11}^{(1)} - \sigma_{11}^{(2)}\right) + \left(\sigma_{22}^{(1)} - \sigma_{22}^{(2)}\right)\right]$$

which, when expressed in terms of $\mu_Y$ takes the form

$$y_1(\mu_{Y_{11}} - \mu_{Y_{21}}) + y_2(\mu_{Y_{12}} - \mu_{Y_{22}}) > \frac{1}{2}\left\{(\mu^2_{Y_{11}} - \mu^2_{Y_{21}}) + (\mu^2_{Y_{12}} - \mu^2_{Y_{22}})\right\}$$

where $\mu_{Y_{kj}}$ is the mean of the $j$th component of $Y$ in the $k$th population $(k,\ j = 1, 2)$. The sample analogue of the above expression can be obtained by replacing $\mu_{Y_{kj}}$ with $\bar{Y}_{kj}$, $(k,\ j = 1, 2)$. This agrees perfectly with Ganesalingam and Ganesh (2004). Note that with the AEDC, the inversion of $\Sigma_x$ is not required at any point of the allocation process.

## 4. Case study

A large-scale Monte Carlo simulation study was carried out by Ganesalingam and Ganesh (2004) for the bivariate case. They considered 99 different covariate structures and established that the AEDC outperforms QDF in 89 of the 99 cases. The advantage of using AEDC is that it is simpler and computationally easy to implement and further, it does not involve the inversion of the variance-covariance matrix. It was noted that AEDC also outperforms QDF in the case of contaminated data, where the variance is larger than that of uncontaminated data. Overall it was noted that the error rates associated with AEDC are much less than that of those with QDF.

### 4.1 *Data set considered for the case study*

Now we wish to support our earlier findings, Ganesalingam and Ganesh (2004), by a real life case study. The data set is from an anthropological study undertaken in the University of Hamburg, Germany and is reported in Flury (1997).

The data set consists of 89 pairs of male twins. Of the 89 pairs, $N_1 = 49$ are monozygoitc and $N_2 = 40$ are dizygotic. There are six variables for each pair of twins. These are stature, hip width and chest circumference for each of the two brothers. Taking the difference between the first and the second twins we used three variables, namely difference in stature, difference in hip width and difference in chest circumference, and considered as a three dimensional classification problem. Thus we have two populations, namely monozygotic and dizygotic twins, on each of whom we have measured three variables, namely difference in stature, difference in hip width and difference in chest circumference.

One can intuitively see that these three variables will have zero means or means close to zero. This is understandable because by nature, the twins are bound to have similar (closer) values for each of the six variables, hence the absolute difference will be zero or near zero and thus the means of the difference will be zero or close to zero. This is the situation where our linear discriminant function fails and we appeal to QDF or AEDC.

### 4.2 *Application of QDF and AEDC to the male twin data*

The AEDC as given in (5) and the QDF as given in (1) are formed using these two samples (49 monozygotic and 40 dizygotic twins) and the confusion matrices obtained are given in Tables 1 and 2, with the results of apparent error rates and cross-validated error rates given below within brackets. It can be noted from the Tables 1 and 2 that the apparent error rates of the AEDC are same as the cross validated error rates while QDF exhibits a small difference.

**Table 1**
**Confusion matrix for AEDC**

| | | Predicted membership | | |
|---|---|---|---|---|
| | | 1 | 2 | |
| Actual membership | 1 | 46 (46) | 3 (3) | 49 |
| | 2 | 16 (16) | 24 (24) | 40 |
| | | 62 | 27 | 89 |

**Table 2**
**Confusion matrix for QDF**

| | | Predicted membership 1 | Predicted membership 2 | |
|---|---|---|---|---|
| Actual membership | 1 | 45 (42) | 4 (7) | 49 |
| | 2 | 9 (9) | 31 (31) | 40 |
| | | 54 (51) | 35 (38) | 89 |

## 5. Conclusion

It can be seen from Tables 1 and 2 that in this case study with a small population size of 89; and dimension $p = 3$, AEDC and QDF behave almost equally in discriminating the two populations with an edge in performance exhibited by the QDF. This is understandable because the QDF uses a complete covariance information, whereas, the AEDC completely ignores it. This is the main advantage of the AEDC in which the inversion of the covariance matrices is not required compared to the QDF where we sometimes encounter singular covariance matrices. Hence AEDC is computationally quicker and user friendly.

Ganesalingam and Ganesh (2004) considered various covariance structures in their simulation study for the two dimensional case and noted that in 89 of the 99 different covariance structures AEDC out performs QDF. Based on this large scale simulation study and this case study considered in this paper we recommend the use of AEDC in all situations where the difference of two population means is zero. This is most common with twin data. Further, the performance of AEDC is not much affected by the dimensionality as the inversion of the covariance matrix is not required, which makes again the AEDC more user friendly.

## References

[1] T. W. Anderson (1984), *An Introduction to Multivariate Statistical Analysis*, 2nd edn., John Wiley, New York.

[2] M. S. Bartlett and N. W. Please (1963), Discrimination in the case of zero-mean differences, *Biometrica*, Vol. 50, pp. 17–21.

[3] S. Geisser (1964), Posterior odds for multivariate normal classifications, *J. R. Statist Soc.*, Vol. B26, pp. 69–76.

[4] B. Efron (1975), The efficiency of logistic regression compared to normal discriminant analysis, *Journal of the American Statistical Association*, Vol. 70, pp. 892–898.

[5] B. Flury (1997), *A First Course in Multivariate Statistic*, Springer, New York.

[6] S. Ganesalingam and S. Ganesh (2004), Statistical discrimination based on absolute deviation from the mean, *Journal of Statistics and Management Systems*, Vol. 7 (1), pp. 25–40.

[7] S. Geisser and M. M. Desu (1968), Predictive zero-mean uniform discrimination, *Biometrika*, Vol. 55, pp. 509–524.

[8] J. P. Koolaard, S. Ganeshlingam and C. R. O. Lawoko (1998), Comparison of regularized discriminant analysis with the standard discrimination methods, *Computational Statistics*, Vol. 13, pp. 495–509.

[9] J. P. Koolaard, S. Ganeshlingam and C. R. O. Lawoko (1998), The use of a distance measure in regularized discriminant analysis, *Computational Statistics*, Vol. 17 (2), pp. 185–202.

[10] P. A. Lachenbruch (1967), An almost unbiasesd method of obtaining confidence intervals for the probability of misclassification in discriminant analysis, *Biometics*, Vol. 23, pp. 639–645.

[11] P. A. Lachenbruch (1975), Zero-mean difference discrimination and the absolute linear discriminant function, *Biometrika*, Vol. 62 (2), pp. 397–401.

[12] V. R. Macro, D. M. Young and D. W. Turner (1987), The Euclidean classifier: an alternative to linear discriminant function, *Commun. Statist. Simulation*, Vol. 16, pp. 485–505.

[13] G. J. McLachlan (1992), *Discriminant Analysis and Statistical Pattern Recognition*, New York, John Wiley.

[14] D. F. Morrison (1990), *Multivariate Statistical Methods*, McGraw Hill.

[15] M. Okamoto (1961), Discrimination for variance matrices, *Osaka Math J.*, pp. 131–139.

[16] T. J. O'Neill (1980), The general distribution of the error rate of a classification procedure with application to logistic regression discrimination, *Journal of American Statistical Association*, Vol. 75, pp. 154–160.

[17] T. J. O'Neill (1992), Error rates of non-Bayes classification rules and the robustness of Fisher's linear discriminant function, *Biometrika*, Vol. 79, pp. 177–184.

[18] T. J. O'Neill (1994), The bias of estimating equations with applications to the error rate of logistic discrimination, *Journal of the American Statistical Association*, Vol. 89, pp. 1492–1498.

[19] L. S. Penrose (1946), Some notes on discrimination, *Ann. Eugin. Lond.*, Vol. 13, pp. 228–237.

[20] S. J. Press and S. Wilson (1978), Choosing between logistic regression and discriminant analysis, *Journal of the American Statistical Association*, Vol. 73, pp. 699–705.

[21] S. Raudys and V. Pikelis (1980), On dimensionality, sample size, classification error and complexity of classification algorithm in pattern recognition, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. PAMI2(3), pp. 242–252.

[22] J. W. Van Ness (1979), On the effects of dimension in discriminant analysis for unequal covariance populations, *Technometrics*, Vol. 21, pp. 119–127.

[23] J. W. Van Ness and C. Simpson (1976), On the effects of dimension in discriminant analysis, *Technometrics*, Vol. 18, pp. 175–187.

*Received September, 2005*