On the Error Rate Comparison of the Quadratic Discriminant Function, Euclidean Distance Classifier, Fisher's Linear Discriminant Function and the Vine Copulas

A. Nanthakumar

Correspondence: A. Nanthakumar, State University of New York at Oswego, USA

Abstract

The estimation of the error rates is of vital importance in classification problems as this is used as a basis to choose the best discriminant function; that is, the one with a minimum misclassification error. The quadratic discriminant function (QDF), Euclidean Distance Classifier (EDC), and Fisher's Linear Discriminant Function (FLDC) have been in use for a long time for the purpose of classification. In this paper, we compare the misclassification error rate of the QDF, EDC, and FLDC with the Vine Copulas based on Gaussian and Clayton models. The results were obtained for the general case where the means are unequal and the covariance matrices are unequal.

Keywords: multivariate normal distributions, quadratic discriminant function, misclassification error, Vine Copula

1. Introduction

In discriminant analysis, the selection of the discriminant function is solely decided by the associated error rates and hence the estimation of error rates is of importance in classification problems. During the past fifty years, several discriminant functions have been proposed for the purpose of classification since the advent of the "Separation Theorem" and the "Supporting hyper-plane Theorem" in Mathematics. Foremost among these discriminants is the quadratic discriminant function (QDF). In addition to this QDF, one can use the other discriminants such as the linear discriminant (LDF), Fisher's Linear Discriminant Function (FLDC), Euclidean Distance Classifier (EDC) and the Absolute Euclidean Distance Classifier (AEDC) (see Ganesalingam et al 2006), and many others. QDF is seen to outperform the other existing methods when the covariance matrices are non-singular. However, QDF is not applicable for discrimination when the covariance matrices are singular. Note that AEDC is not applicable when the population means are unequal. Similarly, LDC is not applicable when the population variance-covariance matrices are unequal.

On the other hand, the pairwise Vine Copulas enable higher dimensional multivariate distributions to be expressed in a simple two dimensional format. The Vine Copulas are used increasingly in financial and engineering modelling as these are useful in reducing the dimension. In this paper, we compare the performance of the Vine Copulas with the QDF, EDC and FLDC. The Vine based Copulas are seen to perform reasonably well when compared to QDF, EDC and FLDC. The Vine based Copulas are very useful for discrimination when the covariance matrix in the full dimension is singular while the covariance matrix in the two dimensional (pairwise) analysis is non-singular. In this paper, we consider a three dimensional (full dimension) problem for discrimination.

1.1 Classical Discriminant Methods

Consider the problem of statistical discrimination involving two multivariate normal populations Π_1 and Π_2 with mean vectors μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2 respectively. Here, we assume that $\mu_1 \neq \mu_2$ and $\Sigma_1 \neq \Sigma_2$. 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 Π_1, if

$$\ln\left\{\frac{|\Sigma_2|}{|\Sigma_1|}\right\} - \left\{(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)\right\} + \left\{(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2)\right\} > 0 \qquad (1.1)$$

Otherwise it is allocated to Π_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.

However, if $\Sigma_1 = \Sigma_2 = \Sigma$ then the object with observation vector x is allocated to population Π_1 if

$$\left\{ x - \frac{1}{2}(\mu_1 + \mu_2) \right\}^T \sum^{-1} (\mu_1 - \mu_2) > 0 \tag{1.2}$$

Otherwise, it is allocated to Π_2. Note that (1.2) is the linear discriminant (LDF) rule.

It is clear that when the data is normal and $\mu_1 = \mu_2$ in expression (1.2), the linear discriminant function cannot be applied. In this case, one could resort to using either the Fisher Linear Discriminant Function (FLDF) or the quadratic discriminant function (QDF) if the covariance matrices are unequal; see Section 3.2.4 of McLachlan (1992). In the case of Fisher Linear Discriminant Function (FLDF), the observation vector x is allocated to Population Π_1 if

$$\left\{ (\mu_1 - \mu_2)^T (n_1 \Sigma_1 + n_2 \Sigma_2)^{-1} x \right\} > 0 \tag{1.3}$$

Otherwise, it is allocated to Population Π_2. Here n_1 and n_2 are the sample sizes for the training data from Populations Π_1 and Π_2 respectively.

The Euclidean Distance Classifier (EDC) will allocate an individual observation vector to population Π_1 if

$$\left\{ x - \frac{1}{2}(\mu_1 + \mu_2) \right\}^T (\mu_1 - \mu_2) > 0 \tag{1.4}$$

Otherwise, it is allocated to Population Π_2.

In this paper, we compare the performance of QDF, FLDF and EDC empirically with the Vine based Gaussian and Clayton Copula models, and focus on the error rate. Here, we use computer simulations to compare the performance.

In the next section, we will use Vine Copulas to compute the misclassification error rate.

2. Vine Copulas

2.1 Vine Copulas

Joe (1996, 2014), Bedford and Cooke (2001, 2002), Kurowicka and Cooke (2006) pioneered in introducing the concept of Vine Copulas in modelling high dimensional joint distributions by using sufficient number of pairwise copulas. This aspect of Copula modelling is known as Vine Copulas. There are two types of Vine Copulas; C-Vine copula and D-Vine copula. In D-Vine copula, no node in any tree is connected to more than two edges whereas in C-Vine copula every node in tree T_j has exactly $n - j$ edges. The D-Vine and the C-Vine copulas coincide when the dimension $p = 3$.

More recently, Acar et al (2012) and Stoeber and Czado (2013) looked at the limitations, extensions and applications of the Vine Copulas. This motivated us to consider the applicability of the Vine Copulas in the context of discriminant analysis. Here, we compare the performance of the Vine based Clayton and Gaussian Copulas against the existing discriminant methods such as Quadratic Discriminant Function (QDF), Fisher's Linear Discriminant Function (FLDC), and Euclidean Distance Classifier (EDC).

Let $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ be a three variable vector and $f(x, y, z)$ be the density function. Then, by using the properties of the Vine

copulas, one can write

$$f(x, y, z) = f_1(x).f_2(y).f_3(z).c_{12}.c_{23}.c_{13\backslash 2} \tag{2.1}$$

where $f_1(x)$ is the marginal density of X

　　　　$f_2(y)$ is the marginal density of Y

　　　　$f_3(z)$ is the marginal density of Z

　　　　c_{12} is the pairwise copula density of X and Y

　　　　c_{23} is the pairwise copula density of Y and Z

　　　　$c_{13\backslash 2}$ is the pairwise copula density of X and Z **given** Y

Vine based on the Gaussian Copula:

Here we investigate the construction of the vine based on the Gaussian Copula. The Gaussian Copula densities are

given as follows.

$$c_{12} = \frac{1}{\sqrt{1-\rho_{12}^2}} \cdot e^{-\frac{1}{2.(1-\rho_{12}^2)}\left\{\rho_{12}^2 \cdot \left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \rho_{12}^2 \cdot \left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2.\rho_{12} \cdot \left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right\}} \tag{2.2}$$

$$c_{23} = \frac{1}{\sqrt{1-\rho_{23}^2}} \cdot e^{-\frac{1}{2.(1-\rho_{23}^2)}\left\{\rho_{23}^2 \cdot \left(\frac{y-\mu_2}{\sigma_2}\right)^2 + \rho_{23}^2 \cdot \left(\frac{z-\mu_3}{\sigma_3}\right)^2 - 2.\rho_{23} \cdot \left(\frac{y-\mu_2}{\sigma_2}\right)\left(\frac{z-\mu_3}{\sigma_3}\right)\right\}} \tag{2.3}$$

$$c_{13\backslash 2} = \frac{1}{\sqrt{1-\rho_{13\backslash 2}^2}} \cdot e^{-\frac{1}{2.(1-\rho_{13\backslash 2}^2)}\left\{\rho_{13\backslash 2}^2 \cdot \left(\frac{x-\mu_{1\backslash 2}}{\sigma_{1\backslash 2}}\right)^2 + \rho_{13\backslash 2}^2 \cdot \left(\frac{z-\mu_{3\backslash 2}}{\sigma_{3\backslash 2}}\right)^2 - 2.\rho_{13\backslash 2} \cdot \left(\frac{x-\mu_{1\backslash 2}}{\sigma_{1\backslash 2}}\right)\left(\frac{z-\mu_{3\backslash 2}}{\sigma_{3\backslash 2}}\right)\right\}} \tag{2.4}$$

For an arbitrary trivariate normal population, the conditional density

$$f_{13\backslash 2}(x,z \backslash y) = \frac{f(x,y,z)}{f_2(y)}$$

$$= \frac{\sqrt{\sigma_{22}}}{2\pi.|\Sigma|^{1/2}} \cdot e^{-\frac{1}{2}\left\{a_{11}.(x-\mu_1)^2 + \left(a_{22} - \frac{1}{\sigma_{22}}\right).(y-\mu_2)^2 + a_{33}.(z-\mu_3)^2 + 2.a_{12}.(x-\mu_1).(y-\mu_2) + 2.a_{13}.(x-\mu_1).(z-\mu_3) + 2.a_{23}.(y-\mu_2).(z-\mu_3)\right\}} \tag{2.5}$$

where a_{ij} is the $(i,j)^{th}$ entry of the inverse covariance matrix of Σ, and σ_{ij} is the $(i,j)^{th}$ entry of the covariance matrix Σ.

Note that,

$$E(X.Z \backslash Y) = \mu_1.\mu_3 + \frac{(a_{12}.a_{13} - a_{11}.a_{23})}{(a_{11}.a_{33} - a_{13}^2)}.\mu_1.(y-\mu_2) - \frac{(a_{12}.a_{33} - a_{13}.a_{23})}{(a_{11}.a_{33} - a_{13}^2)}.(y-\mu_2)$$
$$+ \frac{(a_{12}.a_{33} - a_{13}.a_{23})(a_{11}.a_{23} - a_{12}.a_{13})(y-\mu_2)^2}{(a_{11}.a_{33} - a_{13}^2)^2} - \frac{a_{13}}{(a_{11}.a_{33} - a_{13}^2)} \tag{2.6}$$

Also, note that

$$E(X \backslash Y) = \mu_1 + \rho_{12}.\frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}.(y-\mu_2)$$

$$= \mu_1 + \frac{(a_{13}.a_{23} - a_{22}.a_{12})}{(a_{11}.a_{33} - a_{13}^2)}.(y-\mu_2) \tag{2.7}$$

and

$$E(Z \backslash Y) = \mu_3 + \rho_{23}.\frac{\sqrt{\sigma_{33}}}{\sqrt{\sigma_{22}}}.(y-\mu_2)$$

$$= \mu_3 + \frac{(a_{13}.a_{12} - a_{11}.a_{23})}{(a_{11}.a_{33} - a_{13}^2)}.(y-\mu_2) \tag{2.8}$$

Also,

$$Cov(X,Z \backslash Y) = \frac{-a_{13}}{(a_{11}.a_{33} - a_{13}^2)}$$

$$= \sigma_{13} - \frac{\sigma_{12}.\sigma_{23}}{\sigma_{22}} \tag{2.9}$$

$$Var(X \backslash Y) = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} \tag{2.10}$$

$$Var(Z \setminus Y) = \sigma_{33} - \frac{\sigma_{23}^{2}}{\sigma_{22}} \tag{2.11}$$

$$\text{Note that } \rho_{13\setminus 2} = \frac{\text{cov}(X, Z \setminus Y)}{\sqrt{Var(X \setminus Y)} . \sqrt{Var(Z \setminus Y)}} \tag{2.12}$$

For the Vine Copula based approach, the likelihood ratio which depends on the Gaussian model is

$$L_G = \frac{f_1^{(2)}(x).f_2^{(2)}(y).f_3^{(2)}(z).c_{12}^{(2)}.c_{23}^{(2)}.c_{13\setminus 2}^{(2)}}{f_1^{(1)}(x).f_2^{(1)}(y).f_3^{(1)}(z).c_{12}^{(1)}.c_{23}^{(1)}.c_{13\setminus 2}^{(1)}} \tag{2.13}$$

The log-likelihood ratio is given by

$$\ln(L_G) = \ln\left(\frac{f_1^{(2)}(x)}{f_1^{(1)}(x)}\right) + \ln\left(\frac{f_2^{(2)}(y)}{f_2^{(1)}(y)}\right) + \ln\left(\frac{f_3^{(2)}(z)}{f_3^{(1)}(z)}\right) + \ln\left(\frac{c_{12}^{(2)}}{c_{12}^{(1)}}\right) + \ln\left(\frac{c_{23}^{(2)}}{c_{23}^{(1)}}\right) + \ln\left(\frac{c_{13\setminus 2}^{(2)}}{c_{13\setminus 2}^{(1)}}\right) \tag{2.14}$$

where the superscripts (1) and (2) represent populations (1) and (2) respectively.

Vine based on the Clayton Copula:

Next, we investigate the Vine based on the Clayton Copula. The Clayton Copula based densities are given as follows.

$$\tilde{c}_{12} = (1 + \alpha_{12}) \left\{ \Phi\left(\frac{x - \mu_1}{\sigma_1}\right).\Phi\left(\frac{y - \mu_2}{\sigma_2}\right) \right\}^{-(1+\alpha_{12})} . \left\{ \left(\Phi\left(\frac{x - \mu_1}{\sigma_1}\right)^{-\alpha_{12}} + \Phi\left(\frac{y - \mu_2}{\sigma_2}\right)^{-\alpha_{12}} - 1 \right)^{-\left(2 + \frac{1}{\alpha_{12}}\right)} \right\} \tag{2.15}$$

$$\tilde{c}_{23} = (1 + \alpha_{23}) \left\{ \Phi\left(\frac{y - \mu_2}{\sigma_2}\right).\Phi\left(\frac{z - \mu_3}{\sigma_3}\right) \right\}^{-(1+\alpha_{23})} . \left\{ \left(\Phi\left(\frac{y - \mu_2}{\sigma_2}\right)^{-\alpha_{23}} + \Phi\left(\frac{z - \mu_3}{\sigma_3}\right)^{-\alpha_{23}} - 1 \right)^{-\left(2 + \frac{1}{\alpha_{23}}\right)} \right\} \tag{2.16}$$

$$\tilde{c}_{13\setminus 2} = (1 + \alpha_{13\setminus 2}) \left\{ \Phi\left(\frac{x - \mu_{1\setminus 2}}{\sigma_{1\setminus 2}}\right).\Phi\left(\frac{z - \mu_{3\setminus 2}}{\sigma_{3\setminus 2}}\right) \right\}^{-(1+\alpha_{13\setminus 2})} . \left\{ \left(\Phi\left(\frac{x - \mu_{1\setminus 2}}{\sigma_{1\setminus 2}}\right)^{-\alpha_{13\setminus 2}} + \Phi\left(\frac{z - \mu_{3\setminus 2}}{\sigma_{3\setminus 2}}\right)^{-\alpha_{13\setminus 2}} - 1 \right)^{-\left(2 + \frac{1}{\alpha_{13\setminus 2}}\right)} \right\} \tag{2.17}$$

So, the likelihood ratio based on the Clayton Copula is

$$L_C = \frac{f_1^{(2)}(x).f_2^{(2)}(y).f_3^{(2)}(z).\tilde{c}_{12}^{(2)}.\tilde{c}_{23}^{(2)}.\tilde{c}_{13\setminus 2}^{(2)}}{f_1^{(1)}(x).f_2^{(1)}(y).f_3^{(1)}(z).\tilde{c}_{12}^{(1)}.\tilde{c}_{23}^{(1)}.\tilde{c}_{13\setminus 2}^{(1)}} \tag{2.18}$$

where the superscripts (1) and (2) represent populations (1) and (2) respectively.

3. Numerical Results

In this section, we report the results for some covariance matrices based on 1000 simulation runs each containing simulated samples of size = 1000. Note that the mean vectors and the covariance matrices for the two populations are unequal. We generate these random samples from a multivariate normal population by using the mean vector listed below and the covariance matrices that are listed in the table. In other words, the mean vector and the covariance matrix are assumed to be known in our simulation. We decompose the covariance structure by using the well-known Cholesky decomposition. Next, we generate the vector components according to a normal distribution by using the statistical software SAS. The error rate P12 (classifying Population Π_1 as Population Π_2) was calculated empirically by averaging the error rate when we apply the rule given by (1.1). Similarly, the error rate P12 (classifying Population Π_1 as Population Π_2) was calculated empirically by averaging the error rate when we apply the rule given by (2.14) for the vine Copula based on the Clayton model and the Gaussian model.

In a similar fashion, we calculate the error rate P21 (classifying Population Π_2 as Population Π_1) empirically by interchanging the corresponding terms. We present the numerical results for p = 3. The mean vectors and the covariance matrices are assumed to be known. The mean vectors are as follows for all the covariance matrices studied for the case

p=3.

$\mu_1^{(1)} = 2.2, \quad \mu_2^{(1)} = 2.4, \quad \mu_3^{(1)} = 1.9, \quad \mu_1^{(2)} = 1.7, \quad \mu_2^{(2)} = 1.1, \quad \mu_3^{(2)} = 1.4$

Table 1. Misclassification Error Rate

Σ_1	Σ_2	QDF P12 (P21)	Vine Copula Clayton P12 (P21)	Vine Copula Gaussian P12 (P21)	EDC P12(P21)	FLD P12(P21)
$\begin{pmatrix} 19 & 18 & 13 \\ 18 & 26 & 10 \\ 13 & 10 & 35 \end{pmatrix}$	$\begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{pmatrix}$	0.435 (0.476)	0.373(0.415)	0.311(0.350)	0.432(0.441)	0.494(0.470)
$\begin{pmatrix} 30 & 16 & 6 \\ 16 & 10 & 4 \\ 6 & 4 & 2 \end{pmatrix}$	$\begin{pmatrix} 6 & 6 & 1 \\ 6 & 8 & 2 \\ 1 & 2 & 1 \end{pmatrix}$	0.412 (0.390)	0.417(0.432)	0.319(0.315)	0.449(0.505)	0.366(0.383)
$\begin{pmatrix} 19 & 18 & 13 \\ 18 & 26 & 10 \\ 13 & 10 & 35 \end{pmatrix}$	$\begin{pmatrix} 6 & 6 & 1 \\ 6 & 8 & 2 \\ 1 & 2 & 1 \end{pmatrix}$	0.480 (0.482)	0.371(0.305)	0.156 (0.112)	0.437(0.400)	0.406(0.375)
$\begin{pmatrix} 6 & 1 & -2 \\ 1 & 13 & 4 \\ -2 & 4 & 4 \end{pmatrix}$	$\begin{pmatrix} 11 & -8 & 3 \\ -8 & 9 & -3 \\ 3 & -3 & 2 \end{pmatrix}$	0.272 (0.285)	0.494 (0.485)	0.470 (0.535)	0.355(0.358)	0.518(0.548)
$\begin{pmatrix} 30 & 16 & 6 \\ 16 & 10 & 4 \\ 6 & 4 & 2 \end{pmatrix}$	$\begin{pmatrix} 11 & -8 & 3 \\ -8 & 9 & -3 \\ 3 & -3 & 2 \end{pmatrix}$	0.250 (0.240)	0.344 (0.320)	0.212 (0.140)	0.401(0.356)	0.470(0.473)
$\begin{pmatrix} 6 & 6 & 1 \\ 6 & 8 & 2 \\ 1 & 2 & 1 \end{pmatrix}$	$\begin{pmatrix} 11 & -8 & 3 \\ -8 & 9 & -3 \\ 3 & -3 & 2 \end{pmatrix}$	0.277 (0.272)	0.487 (0.396)	0.275 (0.339)	0.405(0.416)	0.529(0.521)
$\begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{pmatrix}$	$\begin{pmatrix} 11 & -8 & 3 \\ -8 & 9 & -3 \\ 3 & -3 & 2 \end{pmatrix}$	0.393 (0.392)	0.471 (0.464)	0.376 (0.381)	0.430 (0.352)	0.531(0.525)

4. Discussion and Conclusion

As can be seen from Table1, the Vine based Gaussian Copula seem to outperform the existing discriminant methods such as the Quadratic Discriminant Function (QDF), Euclidean Distance Classifier (EDC), Fisher's Linear Discriminant Function (FLDC) on the error rate P12 (classifying Population Π_1 as Population Π_2) and also on the error rate P21 (classifying Population Π_2 as Population Π_1). We did not consider the Absolute Euclidean Distance Classifier (AEDC) for the reason that it requires the population means to be equal. Similarly, we did not consider the Linear Discriminant Function (LDF) as it requires the covariance matrices to be equal. The reason is that in this study, the populations have unequal means and unequal covariance matrices. So, in place of LDF, we considered Fisher's Linear Discriminant Function (FLDF) as it does not require the covariance matrices to be equal. Although, the use of QDF is acceptable as long as the covariance matrices are non-singular, in real life problems with high dimensions, the variables are often correlated and hence the covariance matrix exhibits singularity. This may render the QDF method unsuitable for discriminant analysis. In such cases, the alternative is to use the two dimensional pairwise Vine Copulas for the discriminant analysis. In fact, the two dimensional Vine based Gaussian Copula is a better choice as according to our study, the Vine based Gaussian Copula seemed to perform better than QDF, EDC, and FLDC as a discriminant.

References

Acar, E. F., Genest, C., & NešLehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, *110*, 74-90. https://doi.org/10.1016/j.jmva.2012.02.001

Anderson, T. W. (1984). An Introduction to Multivariate Statistical Analysis, Wiley, New York. *TW Anderson*, *25*, 643-665.

Bartlett, M. S., & Please, N. W. (1963). Discrimination in the case of zero mean differences. *Biometrika*, *50*(1/2), 17-21. https://doi.org/10.1093/biomet/50.1-2.17

Bedford, T., & Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, *32*(1-4), 245-268. https://doi.org/10.1023/A:1016725902970

Bedford, T., & Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 1031-1068. https://doi.org/10.1214/aos/1031689016

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association*, *70*(352), 892-898. https://doi.org/10.1080/01621459.1975.10480319

Flury, B. (1997). *A First Course in Multivariate Statistic.* Springer. New York. https://doi.org/10.1007/978-1-4757-2765-4

Ganesalingam, G. S., & Ganesh, S. (2004). Statistical discrimination based on absolute deviations from the mean. *Journal of Statistics and Management Systems*, *7*(1), 25-40. https://doi.org/10.1080/09720510.2004.10701101

Ganesalingam, S., Ganesh, S., & Nanthakumar, A. (2008). Approximation for error rates associated with the discriminant function based on absolute deviation from the mean. *Journal of Statistics and Management Systems*, *11*(5), 861-881. https://doi.org/10.1080/09720510.2008.10701348

Ganeslingam, S., Nanthakumar, A., & Ganesh, S. (2006). A comparison of quadratic discriminant function with discriminant function based on absolute deviation from the mean. *Journal of Statistics and Management Systems*, *9*(2), 441-457. https://doi.org/10.1080/09720510.2006.10701216

Geisser, S. (1964). Posterior odds for multivariate normal classifications. *Journal of the Royal Statistical Society. Series B (Methodological)*, 69-76. https://doi.org/10.1111/j.2517-6161.1964.tb00540.x

Geisser, S., & Desu, M. M. (1968). Predictive zero-mean uniform discrimination. *Biometrika*, *55*(3), 519-524. https://doi.org/10.1093/biomet/55.3.519

Glick, N. (1978). Additive estimators for probabilities of correct classification. *Pattern recognition*, *10*(3), 211-222. https://doi.org/10.1016/0031-3203(78)90029-8

Hand, D. J. (1986). Recent advances in error rate estimation. *Pattern Recognition Letters*, *4*(5), 335-346. https://doi.org/10.1016/0167-8655(86)90054-1

Joe, H. (1996). Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series*, 120-141. https://doi.org/10.1214/lnms/1215452614

Joe, H. (2014). *Dependence modeling with copulas*. Chapman and Hall/CRC. https://doi.org/10.1201/b17116

Kurowicka, D., & Cooke, R. M. (2006). *Uncertainty analysis with high dimensional dependence modelling*. John Wiley & Sons. https://doi.org/10.1002/0470863072

Lachenbruch, P. A. (1967). An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, 639-645. https://doi.org/10.2307/2528418

Lachenbruch, P. A. (1975). Zero-mean difference discrimination and the absolute linear discriminant function. *Biometrika*, *62*(2), 397-401. https://doi.org/10.1093/biomet/62.2.397

Marco, V. R., Young, D. M., & Turner, D. W. (1987). The Euclidean distance classifier: an alternative to the linear discriminant function. *Communications in Statistics-Simulation and Computation*, *16*(2), 485-505. https://doi.org/10.1080/03610918708812601

McLachlan, G. (2004). *Discriminant analysis and statistical pattern recognition* (Vol. 544). John Wiley & Sons.

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

Nanthakumar, A., Ganesh, S., & Ganesalingam, S. (2009). On triangular approximation to some symmetric probability distributions. *Journal of Statistics and Management Systems*, *12*(6), 1183-1193. https://doi.org/10.1080/09720510.2009.10701450

Press, S. J., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, *73*(364), 699-705. https://doi.org/10.1080/01621459.1978.10480080

Raudys, S., & Pikelis, V. (1980). On dimensionality, sample size, classification error, and complexity of classification algorithm in pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (3), 242-252. https://doi.org/10.1109/TPAMI.1980.4767011

Scherer, W. T., Pomroy, T. A., & Fuller, D. N. (2003). The triangular density to approximate the normal density: decision rules-of-thumb. *Reliability Engineering & System Safety*, *82*(3), 331-341. https://doi.org/10.1016/j.ress.2003.08.003

Stoeber, J., Joe, H., & Czado, C. (2013). Simplified pair copula constructions—Limitations and extensions. *Journal of Multivariate Analysis*, *119*, 101-118. https://doi.org/10.1016/j.jmva.2013.04.014

Van Ness, J. (1979). On the effects of dimension in discriminant analysis for unequal covariance populations. *Technometrics*, 119-127. https://doi.org/10.2307/1268588

Van Ness, J. W., & Simpson, C. (1976). On the effects of dimension in discriminant analysis. *Technometrics*, *18*(2), 175-187. https://doi.org/10.1080/00401706.1976.10489422

Copyrights