

Discriminant Analysis of Interval Data: An Assessment of Parametric and Distance-Based Approaches

A. Pedro Duarte Silva · Paula Brito

Received: 17th October, 2011 / Accepted: 24th October, 2014

Abstract Building on probabilistic models for interval-valued variables, parametric classification rules, based on Normal or Skew-Normal distributions, are derived for interval data. The performance of such rules is then compared with distance-based methods previously investigated. The results show that Gaussian parametric approaches outperform Skew-Normal parametric and distance-based ones in most conditions analyzed. In particular, with heterocedastic data a quadratic Gaussian rule always performs best. Moreover, restricted cases of the variance-covariance matrix lead to parsimonious rules which for small training samples in heterocedastic problems can outperform unrestricted quadratic rules, even in some cases where the model assumed by these rules is not true. These restrictions take into account the particular nature of interval data, where observations are defined by both MidPoints and Ranges, which may or may not be correlated. Under homocedastic conditions linear Gaussian rules are often the best rules, but distance-based methods may perform better in very specific conditions.

Keywords Discriminant analysis · Interval data · Parametric modelling of interval data · Symbolic Data Analysis

1 Introduction

In multivariate data analysis often each single observation contains some intrinsic variability. This is the case when analyzing a group rather than a single individual, where within group variability should not be overlooked, but taken into account, to avoid an important loss of information. Consider, for instance, that we are interested in analyzing basketball teams, in terms of age, height, points scored and nationality of players. If we just take averages or mode values, much information is lost; in such case, the range or even the distribution of

A. Pedro Duarte Silva
Faculdade de Economia e Gestão & CEGE, Catholic University of Portugal, Porto, Portugal
E-mail: psilva@porto.ucp.pt

Paula Brito
Faculdade de Economia & LIAAD-INESC Porto LA, Universidade do Porto, Porto, Portugal
E-mail: mpbrito@fep.up.pt

the players' values in each team is of utmost importance. Also, when we observe a variable along time and wish to record the set of observed values rather than just a specific one (e.g., mean, maximum), then again a set rather than a single value must be recorded. The same issue arises when analysing concepts instead of single specimen - a tree species and not the specific tree in our garden; an accident scenario, and not the particular accident we have witnessed. Therefore, restrictive summarizations, e.g., by means, medians or modes, to impose a fit to the classical representation structure should be avoided. Symbolic Data Analysis (see, e.g., [8], [9], [13], [27]) provides a framework where variability may explicitly be taken into account in the data representation and analysis. To this aim, new variable types have been introduced, whose realizations are no longer single real or categorical values, as in the classical case, but finite sets, intervals or distributions on an underlying set. To this day, many methods for *symbolic data*, have been developed (see [27]); however, most of those methods rely on exploratory non-parametric approaches.

In this paper, we are interested in the discriminant analysis of interval data, i.e., where elements are characterized by variables whose values are intervals on \mathcal{R} . Interval data may occur in many different situations. When describing ranges of variable values - for example, daily stock prices or temperature ranges - we obtain *native* interval data; in the aggregation of huge data bases, when real values describe the individual observations we obtain intervals for the description of the aggregated data.

Discriminant analysis of interval data has been investigated in different contexts. Ishibuchi, Tanaka and Noriko Fukuoka, in [17], determine interval representations in a discriminant space for such data, using a mathematical programming formulation. The proposed method is then applied to a chemical sensing problem. Jahanshahloo *et al* [18] rely on mathematical programming and goal programming to develop a data envelopment analysis-discriminant analysis methodology designed for interval data. In [26] and [32] discriminant analysis of interval data is developed based on imprecise probability theory. In [19], a generalization of classical Factorial Discriminant Analysis to symbolic data (see, for instance, [27]) is proposed. This method is based on a numerical analysis of the transformed symbolic data, followed by a symbolic interpretation of the results; it allows considering numerical, qualitative nominal or distribution-valued variables; classification rules are then based on proximities in the factorial plane (see also [21]). This method is available in the Symbolic Data Analysis software package *SODAS*. Bayesian decision trees for the case when predictors are interval variables are presented in [28]. Discriminant analysis of interval data has also been addressed using Support Vector Machines ([14], [11], [1]), as well as Artificial Neural Networks, see [30], [31], [6] and [29]. Appice *et al* [2] study the performance of the k-nearest neighbor method for different types of data and using different distance measures.

Distance-based approaches to linear discriminant analysis of interval data are discussed in [15]. These approaches rely on representations of interval data which are used in Symbolic Data Analysis for different methodologies; they lead to representations in the discriminant space in the form of intervals or single points, from which distance-based allocation rules are derived. In [10] a parametric modelling for interval data, assuming multivariate Normal or Skew-Normal distributions for the MidPoints and Log-Ranges of the interval variables, is proposed. The intrinsic nature of the interval variables may lead to special structures of the variance-covariance matrix, represented by five different possible cases. From these models parametric classification rules may be derived. This approach is implemented in an *R*-package, *MAINT.DATA*, which also includes maximum likelihood estimation and statis-

tical tests for the different considered cases ¹.

This paper evaluates the relative performance of different classification rules in the discriminant analysis of interval-valued data, focusing on the classification performance. We note that many proposals for the discriminant analysis in the context of interval data have a dual objective of both obtaining a symbolic representation in a discriminant space and classifying new elements with unknown origin. While both those goals are certainly worthwhile, here we are only concerned with comparison of classification accuracy for different methods (under varying data conditions), which may be the most important goal in many applications.

The remainder of the paper is organized as follows. Section 2 briefly introduces interval data, and its representations, and proceeds presenting different methods for discriminant analysis of interval data. In Section 3, an application of different methods to real data, on interval quarterly temperatures registered in chinese meteorological stations, is presented and discussed. Section 4 reports a simulation study designed to compare the performance of distance and parametric based approaches under different setups. Finally, Section 5 discusses the results and draws some conclusions.

2 Discriminant methods for interval data

2.1 Interval data

Let $S = \{s_1, \dots, s_n\}$ be the set of n entities under analysis. An interval variable is defined by an application

$$Y : S \rightarrow T \text{ such that } s_i \rightarrow Y(s_i) = [l_i, u_i],$$

where T is the set of intervals of an underlying set $O \subseteq \mathcal{R}$. Let I be an $n \times p$ matrix containing the values of p interval variables on S . Each $s_i \in S$ is hence represented by a p -dimensional vector of intervals, $I_i = (I_{i1}, \dots, I_{ip}), i = 1, \dots, n$, with $I_{ij} = [l_{ij}, u_{ij}], j = 1, \dots, p$ (see Table 1).

Table 1 Matrix I of interval data

	Y_1	...	Y_j	...	Y_p
s_1	$[l_{11}, u_{11}]$...	$[l_{1j}, u_{1j}]$...	$[l_{1p}, u_{1p}]$
...
s_i	$[l_{i1}, u_{i1}]$...	$[l_{ij}, u_{ij}]$...	$[l_{ip}, u_{ip}]$
...
s_n	$[l_{n1}, u_{n1}]$...	$[l_{nj}, u_{nj}]$...	$[l_{np}, u_{np}]$

The value of an interval variable Y_j for each $s_i \in S$ is defined by the lower and upper bounds l_{ij} and u_{ij} of $I_{ij} = Y_j(s_i)$. For modelling purposes, however, an alternative parameterization consisting in representing $Y_j(s_i)$ by the MidPoint $c_{ij} = \frac{l_{ij} + u_{ij}}{2}$ and Range $r_{ij} = u_{ij} - l_{ij}$ of I_{ij} may be useful.

¹ MAINT.DATA is available at the CRAN repository.

To extend linear classical discriminant analysis to the case of interval data, appropriate definitions of linear combinations, dispersion and association measures must be established. However, as discussed in [15], there is no unequivocal manner of defining these concepts and not all choices satisfy usual properties.

Let $Z = I \otimes \beta$ be r appropriately defined linear combinations of I based on $p \times r$ real coefficients $\beta_{j\ell}$.

Definitions of linear combinations, dispersion and association measures ideally should satisfy the following properties, for any $p \times r$ real matrix β :

$$\text{P1: } I_i \otimes \beta_\ell = \sum_{j=1}^p \beta_{j\ell} \times I_{ij} \text{ where } \beta_\ell \text{ denotes the } \ell\text{-th column of matrix } \beta$$

and

$\text{P2: } S_{I \otimes \beta} = \beta^t S_I \beta$, i.e., the covariance between interval variables should be a symmetric bilinear operator.

In [15] two distinct definitions of linear combinations are considered:

$LC1 : I_i \otimes_A \beta_\ell = z_{i\ell A} = [z_{i\ell A}, \bar{z}_{i\ell A}], i = 1, \dots, n$, with

$$\begin{cases} z_{i\ell A} = \sum_{j=1}^p \beta_{j\ell} l_{ij} \\ \bar{z}_{i\ell A} = \sum_{j=1}^p \beta_{j\ell} u_{ij} \end{cases}$$

It is clear that while $LC1$ satisfies property $P2$, it does not satisfy property $P1$ if at least one element of β_ℓ is negative.

On the other hand, the following alternative definition of linear combination of interval variables always respects $P1$, and respects $P2$ for suitable definitions of dispersion and association measures:

$$LC2 : I_i \otimes_B \beta_\ell = z_{i\ell B} = [z_{i\ell B}, \bar{z}_{i\ell B}], i = 1, \dots, n, \text{ with}$$

$$\begin{cases} z_{i\ell B} = \sum_{\beta_{j\ell} > 0} \beta_{j\ell} l_{ij} + \sum_{\beta_{j\ell} < 0} \beta_{j\ell} u_{ij} \\ \bar{z}_{i\ell B} = \sum_{\beta_{j\ell} > 0} \beta_{j\ell} u_{ij} + \sum_{\beta_{j\ell} < 0} \beta_{j\ell} l_{ij} \end{cases}$$

Note that $LC2$ is the definition that results from applying the rules of Interval Calculus [23], since the resulting intervals include all possible values that are scalar linear combinations of the values within the intervals I_{ij} .

2.2 Distance-based approaches

In [15] three different approaches for discriminant analysis of interval data are compared. The first approach assumes an Uniform distribution in each observed interval (as in [7]), derives the corresponding measures of dispersion and association, and appropriately defines linear combinations of interval variables that maximize the usual discriminant criterion; a second approach expands the original data set into the set of all interval description vertices (following [12]), and proceeds with a classical analysis of the expanded set; finally, a third approach is based on the parametrization of each interval by its MidPoint and Range (as do [20] and [24]). Representations in the discriminant space may then take the form of intervals or single points from which distance-based allocation rules are derived.

The first considered approach assumes that each interval variable represents the possible values of an underlying real-valued variable; following Bertrand and Goupil [7] an equidistribution hypothesis is assumed, which consists in considering that the values of the underlying variable are uniformly distributed; the empirical distribution function of an interval variable is then defined as a uniform mixture of n uniform distributions.

If the n observations are partitioned into k groups, C_1, \dots, C_k , then the global empirical density functions are mixtures of the corresponding group specific functions. The global variance and covariance can be decomposed in a within group component and a between group component, W and B respectively. Variances and covariances obtained from these matrices satisfy property P2.

As in the classical case, the discriminant functions coefficients are given by the eigenvectors of $W^{-1}B$. Single point representations on a discriminant space are obtained directly, interval representations may be determined by an appropriate linear combination of the lower and upper bounds (see Section 2.1 above).

An alternative approach is investigated, which consists in considering all the vertices of the hypercube representing each of the n individuals in the p -dimensional space, and then perform a classical discriminant analysis of the resulting $n \times 2^p$ by p matrix, following [12] for Principal Component Analysis.

A new matrix of single real values M is created from the interval data matrix I (see Table 1), where to each row i of I correspond 2^p rows of M , obtained by all possible combinations of the limits of intervals $[l_{ij}, u_{ij}]$, $j = 1, \dots, p$.

A classical discriminant analysis on matrix M then leads to a factorial representation of points, one for each of the 2^p vertices, from which an interval representation may be obtained: let Q_i be the set of row indices q in matrix M which refer to the vertices of the hypercube corresponding to s_i ; for $q \in Q_i$ let $\zeta_{q\ell}$ be the value of the ℓ -th real-valued discriminant function for the vertex with row index q ; the value of the ℓ -th interval discriminant variate z for s_i , is then defined by $\underline{z}_{i\ell} = \text{Min} \{\zeta_{q\ell}, q \in Q_i\}$ and $\bar{z}_{i\ell} = \text{Max} \{\zeta_{q\ell}, q \in Q_i\}$.

The third explored approach uses the representation of each observed interval by its MidPoint and Range, as in [24] and [25] for Regression Analysis, and [20] for Principal Component Analysis. Two separate classical discriminant analysis on these values are then performed and the results combined in some appropriate way; alternatively MidPoints and Ranges may also be considered conjointly.

In all three approaches allocation rules are based on point distances or distances between intervals, according to whether the representations on the discriminant space assume the form of single points or intervals.

For the first considered approach, a natural rule based on point distances consists in allocating each observation to the group with nearest centroid in the discriminant space, using the Euclidean distance and correcting for distinct prior probabilities and/or misclassification costs.

Alternatively, linear combinations of the interval variables may be determined, leading to interval-valued discriminant variates, in which case allocation rules may be derived by using distances between interval vectors.

In the vertices approach, discriminant variates are interval-valued, so this same type of allocation rule is applied.

For the MidPoints and Ranges approach, only point distances are used to define allocation rules. When two separate analysis are performed for MidPoints and Ranges, the

discriminant variates are generally correlated, and the Mahalanobis distance should be used; when a single discriminant analysis is performed for both MidPoints and Ranges, the Euclidean distance is adequate.

2.3 Parametric modelling of interval data

Consider each interval I_{ij} represented by its MidPoint c_{ij} and Range r_{ij} . The Gaussian model (see [10]) consists in assuming a multivariate Normal distribution for MidPoints C and the logs of the Ranges R , $R^* = \ln(R)$, $(C, R^*) \sim N_{2p}(\mu, \Sigma)$, with $\mu = [\mu_C^t, \mu_{R^*}^t]^t$ and $\Sigma = \begin{pmatrix} \Sigma_{CC} & \Sigma_{CR^*} \\ \Sigma_{R^*C} & \Sigma_{R^*R^*} \end{pmatrix}$ where μ_C and μ_{R^*} are p -dimensional column vectors of the mean values of, respectively, the MidPoints and Log-Ranges, and $\Sigma_{CC}, \Sigma_{CR^*}, \Sigma_{R^*C}$ and $\Sigma_{R^*R^*}$ are $p \times p$ matrices with their variances and covariances.

We denote $X_i = [C_i^t, R_i^{*t}]^t$ the $2p$ dimensional column vector comprising all the MidPoints and Log-Ranges for $s_i, i = 1, \dots, n$.

This model has the advantage that it allows for a straightforward application of classical inference methods. It is important to keep in mind, however, that the MidPoint c_{ij} and the Range r_{ij} of the value of an interval variable $I_{ij} = Y_j(s_i)$ are two quantities related to one same variable, and must therefore be considered together. It follows that the global covariance matrix should take into account the link that may exist between MidPoints and Log-Ranges of the same or different variables. Intermediate parameterizations between the non-restricted and the non-correlation setup considered for real-valued data are relevant for the specific case of interval data.

The most general formulation allows for non-zero correlations among all MidPoints and Log-Ranges (Case 1); in another setup, interval variables Y_j are uncorrelated, but for each variable, the MidPoint may be correlated with its Range (Case 2); a third situation allows for MidPoints (respectively, Ranges) of different variables to be correlated, but no correlation between MidPoints and Ranges is allowed (Case 3); finally, all MidPoints and Ranges are uncorrelated, both among themselves and between each other (Case 4). In a full complete setup another case could still be considered, namely, allowing for non-null correlation between the MidPoint of each variable and its Log-Range, but not between MidPoints and Log-Ranges of different variables. This case appears to be less natural, and leads to considerably computational complexity, and will therefore not be considered in the present investigation. Table 2 summarizes the different cases considered in this paper.

Table 2 Different cases for the variance-covariance matrix

Case	Characterization	Σ
1	Non-restricted	Non-restricted
2	Y_j 's non correlated	$\Sigma_{CC}, \Sigma_{CR^*} = \Sigma_{R^*C},$ $\Sigma_{R^*R^*}$ all diagonal
3	C 's non-correlated with R^* 's	$\Sigma_{CR^*} = \Sigma_{R^*C} = 0$
4	All C 's and R^* 's are non-correlated	Σ diagonal

It should be remarked that in cases 2, 3 and 4, Σ can be written as a diagonal by blocks matrix, after a possible rearrangement of rows and columns. It follows that maximum likelihood estimates under these cases can be obtained directly from the classical non-restricted

estimates. Testing for the different models /configurations can be done in a straightforward manner, using the likelihood-ratio approach.

The Gaussian model has many advantages, which explains its generalized applicability in multivariate data analysis; in particular, it allows for a direct modelling of the covariance structure between the variables. Nevertheless it does present some limitations, namely the fact that it imposes a symmetrical distribution on the MidPoints and a specific relation between mean, variance and skewness for the Ranges. A more general model that overcomes these limitations may be obtained by considering the family of Skew-Normal distributions (see, for instance, [4]). The Skew-Normal generalizes the Gaussian distribution by introducing an additional shape parameter, while trying to preserve some of its mathematical properties.

A random vector X is said to follow a $2p$ -multivariate Skew-Normal distribution (see [5]) if its density is given by

$$f(x; \alpha, \xi, \Omega) = 2\phi_{2p}(x - \xi; \Omega)\Phi_{2p}(\alpha^t \omega^{-1}(x - \xi)), x \in \mathbb{R}^{2p} \quad (1)$$

where ξ and α are $2p$ -dimensional vectors, Ω is a symmetric $2p \times 2p$ positive-definite matrix, ω is a diagonal matrix formed by the square-roots of the diagonal elements of Ω and ϕ_{2p}, Φ_{2p} are, respectively, the density and the distribution function of a $2p$ -dimensional standard Gaussian vector.

As an alternative to the Normal model, it may be considered that (C, R^*) follow jointly a $2p$ -multivariate Skew-Normal distribution. Notice that the Skew-Normal model encompasses mixed models with marginal Normal random variables, for which the corresponding shape parameter is null.

In the Gaussian model, for each case, an estimate of the optimum classification rule can be obtained by directly generalisation of the classical linear (2) and quadratic (3) discriminant classification rules,

$$Y = \operatorname{argmax}_g (\hat{\mu}_g^t \hat{\Sigma}^{-1} X - \frac{1}{2} \hat{\mu}_g^t \hat{\Sigma}^{-1} \hat{\mu}_g + \log \hat{\pi}_g) \quad (2)$$

$$Y = \operatorname{argmax}_g (-\frac{1}{2} X^t \hat{\Sigma}_g^{-1} X + \hat{\mu}_g^t \hat{\Sigma}_g^{-1} X + \log \hat{\pi}_g - \frac{1}{2} (\log \det \hat{\Sigma}_g + \hat{\mu}_g^t \hat{\Sigma}_g^{-1} \hat{\mu}_g)) \quad (3)$$

where $Y \in 1, \dots, k$ denotes the group assignments, g is a group index, $\hat{\mu}_g, \hat{\Sigma}, \hat{\Sigma}_g$ and $\hat{\pi}_g$ are the maximum likelihood estimates of μ_g, Σ, Σ_g for the corresponding cases, and π_g are the prior probabilities of group membership.

For the Skew-Normal model different alternatives may be considered. In particular, we consider a Location model in which the groups differ only in terms of the location parameter ξ , and a General Model, where the groups differ in terms of all parameters. The corresponding classification rules are, respectively,

$$Y = \operatorname{argmax}_g (\hat{\xi}_g^t \hat{\Omega}^{-1} X - \frac{1}{2} \hat{\xi}_g^t \hat{\Omega}^{-1} \hat{\xi}_g + \log \hat{\pi}_g + \zeta_0(\hat{\alpha}^t \hat{\omega}^{-1}(X - \hat{\xi}_g))) \quad (4)$$

$$Y = \operatorname{argmax}_g (-\frac{1}{2} X^t \hat{\Omega}_g^{-1} X + \hat{\xi}_g^t \hat{\Omega}_g^{-1} X + \log \hat{\pi}_g - \frac{1}{2} (\log \det \hat{\Omega}_g + \hat{\xi}_g^t \hat{\Omega}_g^{-1} \hat{\xi}_g) + \zeta_0(\hat{\alpha}_g^t \hat{\omega}_g^{-1}(X - \hat{\xi}_g))) \quad (5)$$

where $\hat{\xi}_g, \hat{\Omega}, \hat{\Omega}_g, \hat{\alpha}, \hat{\alpha}_g$ are maximum likelihood estimates and $\zeta_0(w) = \ln(2\Phi(w))$.

3 Application: China temperatures data

This data set gathers temperatures measured in meteorological stations in China. The analysis is based on data consisting of the intervals of observed temperatures (Celsius scale) in each of the four quarters, Q_1 to Q_4 , of the years 1974 to 1988 in 60 stations. Table 3 reproduces the original data for some stations and years. The full table comprises $n = 60 \times 15 - 1$ outlier² = 899 rows and 4 columns.

Table 3 China temperatures interval data

Station	Region	Q_1	Q_2	Q_3	Q_4
Beijing-1974	North	[-9.5, 10.6]	[6.5, 29.8]	[12.6, 29.6]	[-10.44, 9.06]
Beijing-1975	North	[-8.6, 12.9]	[7.9, 30.2]	[15.0, 31.6]	[-7.0, 19.2]
...
ZhangYe-1988	Northwest	[-15.4, 7.2]	[2.3, 26.4]	[8.6, 30.2]	[-12.0, 15.1]

The 60 meteorological stations belong to 6 different regions in China (North, Northwest, Northeast, East, South Central, Southwest), which define a partition of the 899 stations-year combinations. Figure 1 depicts the 60 stations according to the defined partition. To control for possible temporal auto-correlation, the global yearly average temperature was subtracted to the corresponding original values.

On a preliminary analysis to assess deviations from Normality, Q-Q plots did not reveal any strong deviations, although for a few variables and classes Normality was rejected by the Kolmogorov-Smirnov test (which is to be expected given the large sample sizes).

Twenty five different discriminant methods are applied and their results compared, namely: nine distance-based approaches - classical linear discriminant analysis based on the intervals' MidPoints only, with allocation defined by Euclidean point-distances; linear discriminant analysis using MidPoints and Ranges separately, with allocation defined by Mahalanobis point-distances, and using MidPoints and Ranges simultaneously, with allocation defined by Euclidean point-distances; uniformity-based $LC1$, uniformity-based $LC2$, Vertices method and Factorial Discriminant Analysis (FCA) (our implementation, following [21]), all with allocation rules using Hausdorff interval distances, for FCA, versions with single, average and complete linkage are considered; and sixteen parametric-based approaches, eight using the Gaussian model - Linear and Quadratic Discriminant Analysis, and eight using Skew-Normal discriminant Analysis - Location and General model - always considering cases 1 to 4 for the variance-covariance matrix (see Table 2). Table 4 presents the cross-validation estimates (ten-fold cross-validation replicated 20 times) of the error-rates for the first six distance-based methods; Table 5 for the FCA method, Tables 6 to 9 present the cross validation estimates of the error-rates for the parametric methods and the four different covariance structure cases, Linear and Quadratic Discriminant Analysis for the Gaussian model (Tables 6 and 7, respectively), and Skew-Normal Discriminant Analysis - Location and General models (Tables 8 and 9, respectively).

As it is usually the case, estimates of the error rates differ from group to group; in particular, the East and Northeast regions are in general better identified than the remaining ones. In terms of global error rates, in general, parametric methods perform better than distance-based ones, Gaussian Quadratic Discriminant Analysis better than Gaussian Linear

² The outlier is YinChuan in 1982, where a value of -999.99°C is reported in July

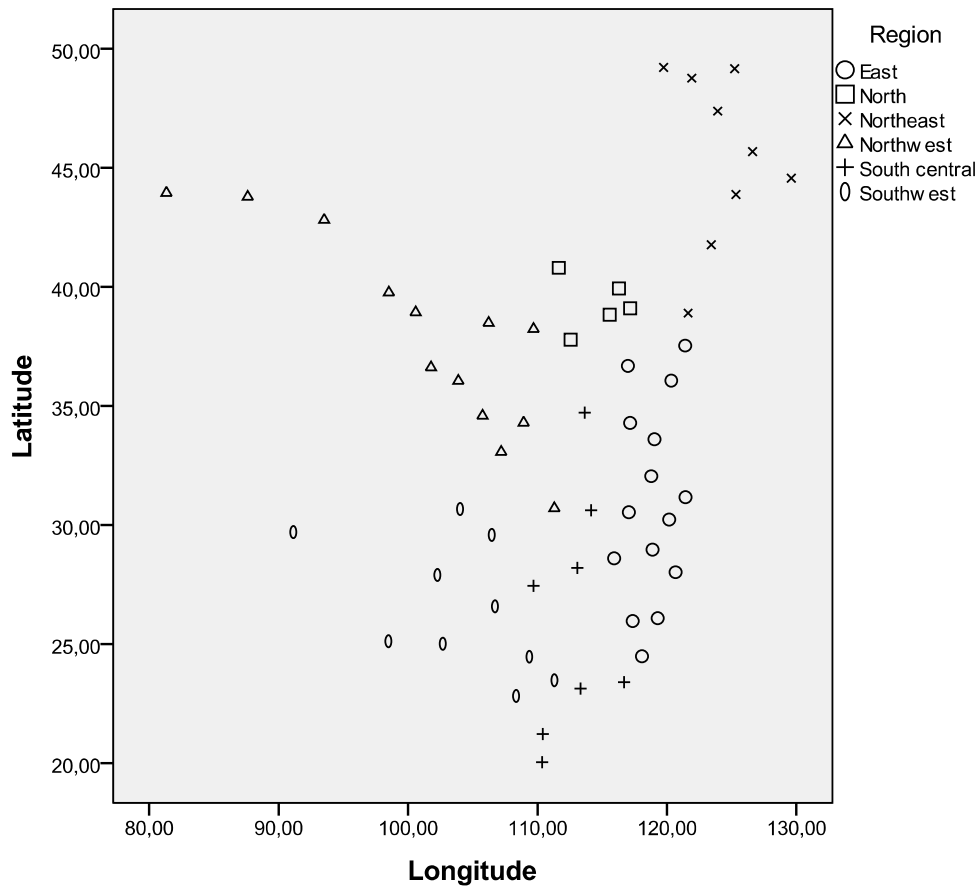


Fig. 1 The 60 meteorological stations, according to region.

Table 4 China data: Error-rates (cross-validation) for distance-based methods

Region	MidPoints Point dist.	Unif. <i>LC1</i>	Unif. <i>LC2</i>	Vertices	Mid. Rng. Sep.	Mid. Rng. Simul.
North	0.9882	0.0370	0.0583	0.0883	0.5724	0.5684
Northeast	0.1927	0.2220	0.2226	0.2222	0.1108	0.1135
Northwest	0.2558	0.9176	0.8741	0.7523	0.3191	0.3161
East	0.1750	0.2142	0.1740	0.2686	0.1453	0.1439
South Central	0.5667	0.5000	0.5000	0.5000	0.5471	0.5521
Southwest	0.4143	0.6670	0.6740	0.5893	0.4673	0.4650
GLOBAL	0.3551	0.4661	0.4496	0.4354	0.3206	0.3200

Discriminant Analysis and General Skew-Normal Model better than Location Skew-Normal Model. More general models give best results, which was to be expected given the large sample size. Nevertheless, the shape parameters introduced by the Skew-Normal Model made almost no difference in classification results. Also, methods using interval distances in the discriminant space, i.e. the *LC1*, *LC2*, the Vertices method and FCA perform poorly, tending

Table 5 China data: Error-rates (cross-validation) for the FDA method

Region	Single linkage	Average linkage	Complete linkage
North	0.8661	0.4227	0.0148
Northeast	0.4717	0.5338	0.9702
Northwest	0.5433	0.6168	0.9935
East	0.6572	0.8165	0.9700
South Central	0.8233	0.8875	0.8367
Southwest	0.5890	0.8020	0.9233
GLOBAL	0.6330	0.7050	0.8699

Table 6 China data: Error-rates (cross-validation) for the Gaussian parametric methods: Linear Discriminant Analysis

Region	LDA C1	LDA C2	LDA C3	LDA C4
North	0.5989	0.7204	0.6456	0.8637
Northeast	0.1248	0.2219	0.1577	0.2219
Northwest	0.2998	0.2799	0.2854	0.2284
East	0.1532	0.2446	0.1875	0.2340
South Central	0.5483	0.5000	0.5429	0.5000
Southwest	0.4570	0.6147	0.4873	0.6330
GLOBAL	0.3213	0.3844	0.3398	0.3857

Table 7 China data: Error-rates (cross-validation) for the Gaussian parametric methods: Quadratic Discriminant Analysis

Region	QDA C1	QDA C2	QDA C3	QDA C4
North	0.3371	0.3540	0.2638	0.3189
Northeast	0.1025	0.1926	0.1565	0.2221
Northwest	0.3203	0.3486	0.3688	0.3451
East	0.1079	0.2116	0.1585	0.2126
South Central	0.4463	0.5079	0.5000	0.5000
Southwest	0.2543	0.5310	0.3333	0.6263
GLOBAL	0.2414	0.3430	0.2814	0.3589

Table 8 China data: Error-rates (cross-validation) for the Skew-Normal Discriminant Analysis: Location Model

Region	SKLM C1	SKLM C2	SKLM C3	SKLM C4
North	0.8114	0.7135	0.6504	0.8952
Northeast	0.1058	0.2221	0.1521	0.2221
Northwest	0.2629	0.2789	0.2886	0.2214
East	0.1410	0.2470	0.1918	0.2484
South Central	0.5367	0.5008	0.5504	0.5000
Southwest	0.4507	0.6150	0.4880	0.6087
GLOBAL	0.3222	0.3844	0.3424	0.3862

to classify too many stations in one given region (varying from method to method). Point distance approaches using MidPoints and Ranges provide results comparable to those of Linear Discriminant Analysis.

Table 9 China data: Error-rates (cross-validation) for the Skew-Normal Discriminant Analysis: General Model

Region	SKGM C1	SKGM C2	SKGM C3	SKGM C4
North	0.3121	0.3387	0.2574	0.3543
Northeast	0.1202	0.1684	0.1508	0.2291
Northwest	0.3059	0.3576	0.3771	0.3028
East	0.1000	0.2071	0.1581	0.2164
South Central	0.4621	0.5050	0.4738	0.5021
Southwest	0.2577	0.5107	0.3187	0.5720
GLOBAL	0.2397	0.3352	0.2816	0.3459

4 Simulation

To better understand the factors affecting the relative performance of the methods under comparison, we performed a controlled simulation experiment. We considered a full factorial design for problems with two groups, three interval variables, and the following seven factors:

Classification method (**CM** – 22 levels): the 6 distance-based considered in [15], and the 16 parametric methods under comparison.

Data Generating Process (**DGP** – 2 levels): MidPoints generated by transformations using Gaussian and Skew-Normal variables.

Separation (**Sep** - 2 levels): the Mahalanobis distance between group centroids set at 1.0 (poorly separated groups) and 3.0 (well separated groups).

Range heterogeneity (**RHet** - 2 levels): In one level (homogeneous Ranges) the Log-Ranges had the same distribution across groups, and in the other one (heterogeneous Ranges) the groups differed in terms of both the MidPoint and Log-Range distributions.

Training sample size (**TSS** - 4 levels): Total number of training sample observations, set at 30, 60, 100 and 150.

Variance ratios (**VR** - 2 levels): Ratio between the variances associated with different distributions across groups. Set at 1 (homocedastic problems) and 9 (heterocedastic problems).

True case (**TConf** - 4 levels): case of true covariance of the within-group distributions of MidPoints and Log-Ranges. Set at the levels 1 (unrestricted), 2 (Uncorrelated Interval Variables), 3 (MidPoints uncorrelated with Log-Ranges) and 4 (all MidPoints and Log-Ranges uncorrelated with each other).

For each data condition, defined by a combination of factors **DGP**, **Sep**, **TSS**, **RHet**, **VR** and **TConf**, we generated 100 independent balanced training samples, used them to establish the empirical rules, and evaluated these rules on one balanced validation sample with 1000 observations, independently generated.

Each observation was defined from p independent Gaussian or Skew-Normal (X) and p independent Uniform variates (U) where the means in the first group were set to zero and the means in the second group ensured the desired level of Mahalanobis distance. In the Skew-Normal case, one of the MidPoint variables has Fisher's skewness coefficient set to 0.75, for the remaining variables this coefficient is null. We note that for multivariate Skew-Normal distributions, the admissible domain of the parameters is somehow restricted, in particular

not allowing for high skewness in many variables [3]. We chose to concentrate skewness on one variable, to maximize the deviation from the Gaussian model.

In the case of heterocedastic problems we used the average covariance matrix to find the required Mahalanobis distances.

For Case 4, the X variables define the intervals' MidPoints, and the U variables the Log-Ranges. For Case 1, MidPoints (MP) and Log-Ranges (LR) were defined through the transformations $MP = [L1|L2] [X^t U^t]^t$, $LR = L3 U^t$, where $[L1|L2]$ and $L3$ are orthonormal matrices of independently generated Uniform loadings. For Case 2, MP and LR were generated in the same way, but placing all required zeros in matrix $L2$ to ensure that only MidPoints and Log-Ranges of the same interval variables had non-null correlations. Finally, for Case 3 they were defined through $MP = L1 X^t$, $LR = L3 U^t$. Notice that none of these models coincides with the models adjusted to the data by the parametric classification rules.

Table 10 presents, for each data condition, the method with lowest estimated expected error rate; those indicated in bold perform better than all other methods by more than one percentage point. Tables 11 and 12 gather the corresponding estimated error rates and standard errors. We note that the methods in bold have always error estimates lower than the second best method by more than two standard errors.

Complete tables with average validation sample misclassification rates are available from the authors upon request.

As expected, the average error rates for heterocedastic cases are always lower than the corresponding values for the homocedastic values, the differences being particularly striking when the group centroids are badly separated - which may be explained by the fact that in the former case differences in variances-covariances also contribute to the separation between the groups.

The relative standings of the different methods may be summarized as follows :

1. *Heterocedastic conditions*: Some of the most interesting results occur for heterocedastic conditions. Under these conditions, one parametric Gaussian quadratic rule always performs the best while linear rules and distance-based methods perform poorly. Remarkably, for small training samples, the best results are often achieved by a model with a more restricted case than the true generating process. In particular, under Case 1, for samples with 30 or 60 observations, the quadratic Gaussian rule based on Case 3 (Mid-Points uncorrelated with Log-Ranges) performs the best in all but one data condition. More general quadratic rules work well for large samples but can be disappointing in small samples. Figure 2 illustrates this behaviour. Furthermore, the parametric Gaussian quadratic rules usually perform better than the corresponding Skew-Normal rules even when data is generated from a Skew-Normal distribution - as illustrated in Figure 3.
2. *Homocedastic conditions*: When MidPoints are correlated with Log-Ranges (Cases C1 and C2), the parametric Gaussian linear rule with the correct covariance case generally performs the best. However, for large training samples the corresponding quadratic rule comes close. Overparametrized but correct linear rules are usually also competitive.

This situation is illustrated by Figure 4. However, the Skew-Normal parametric models with non-satisfied restrictions may present a rather erratic behaviour; otherwise, their performance exhibits a similar pattern - see Figure 5.

Among the distance-based methods, the Vertices and the Central-Points methods appear to be the most competitive ones, the former particularly when interval-valued variables are non-correlated (Cases 2 and 4), and the latter when Midpoints and LogRanges are uncorrelated (Cases 3 and 4). However, in this latter case the differences are particularly large for small sample sizes and tend to become negligible for large samples. Figures 6 and 7 illustrate these points. In some specific situations, the Centers and the Vertices methods are even the best ones. However, the Centers method deals badly with correlation between MidPoints and Log-Ranges, and does not take much advantage of larger samples. This is illustrated in Figure 8 where we can see that the Centers method is the best one under Case C3 with 30 observations in the training sample and the second best (after the Vertices method) under Case C4 with 30 observations in the training sample; on the other hand its relative performance deteriorates under Cases C3 and C4 for large samples, and is rather poor under Cases C1 and C2 for any sample size. The Vertices method deals badly with correlation among different interval-valued variables. In Figure 9 we can see that while the performance of this method is the best one under Cases C2 and C4 with large training samples, it is not even competitive in cases C1 and C3.

5 Conclusions

In this paper, the relative performance of distance-based and parametric classification rules, for discriminant analysis of interval-valued data, is evaluated. A thorough simulation study as well as an application to real data showed that parametric based rules outperform distance-based ones for most data conditions, with a few exceptions in very specific cases. Vertices and Centers methods can be competitive, or even the best, in some specific situations, but they tend to require strong restrictions in the covariance configurations, and do not respond well to heterocedastic data conditions. Furthermore, the parametric Gaussian rules usually perform better than the corresponding Skew-Normal rules even when data is generated from a Skew-Normal distribution. As expected, in homocedastic problems, linear discriminant rules perform the best, while for large training samples and heterocedastic conditions quadratic methods are usually superior. However, for small training samples in heterocedastic problems, unrestricted quadratic rules can be outperformed by restricted rules, even in some cases where the model assumed by these rules is not true. It is known (see e.g. [22]) that traditional quadratic discrimination perform poorly with small training samples, even when covariance matrices clear differ, due to the large number of parameters that need to be estimated. For that reason Flury *et al* [16] proposed parsimonius quadratic rules that constraint some parameters of the covariance matrices. In the case of interval data, Cases 2 through 4 provide a natural way of imposing constraints which, according to our simulation results, appear to be quite effective in reducing expected error rates for heterocedastic problems with small or moderate training samples.

Acknowledgements This research is supported by the Project NORTE-07-0124-FEDER-000059 within the North Portugal Regional Operational Programme (ON.2 - O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by national funds, through the Portuguese funding agency, Fundação para a Ciência e a Tecnologia (FCT). One author

Table 10 Best methods by data condition
(in bold if all other methods present average error rates larger by more than 0.01).

	Gaussian				Skew-Normal			
	Homocedastic ; Equal Ranges ; Good Separation							
	30	60	100	150	30	60	100	150
C1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1
C2	ldaC2	ldaC2	ldaC2	ldaC2	ldaC2	ldaC2	qdaC2	ldaC2
C3	ldaC4	ldaC4	Vrt	LocC3	CP	CP	CP	ldaC3
C4	Vrt	ldaC4	Vrt	Vrt	Vrt	ldaC4	Vrt	LocC4
	Heterocedastic ; Equal Ranges ; Good Separation							
	30	60	100	150	30	60	100	150
C1	qdaC3	qdaC1	qdaC1	qdaC1	qdaC3	qdaC3	qdaC1	qdaC1
C2	qdaC2	qdaC2	qdaC2	qdaC2	qdaC2	qdaC2	qdaC2	qdaC2
C3	qdaC4	qdaC3	qdaC3	qdaC3	qdaC3	qdaC3	qdaC3	qdaC3
C4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4
	Homocedastic ; Different Ranges ; Good Separation							
	30	60	100	150	30	60	100	150
C1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1
C2	ldaC2	ldaC2	ldaC2	ldaC2	ldaC2	ldaC2	ldaC2	ldaC2
C3	ldaC3	ldaC3	ldaC3	ldaC3	ldaC3	qdaC3	qdaC3	ldaC3
C4	ldaC4	ldaC4	ldaC4	ldaC4	ldaC4	ldaC4	ldaC4	ldaC4
	Heterocedastic ; Different Ranges ; Good Separation							
	30	60	100	150	30	60	100	150
C1	qdaC3	qdaC3	qdaC1	qdaC1	qdaC3	qdaC3	qdaC1	qdaC1
C2	qdaC2	qdaC2	qdaC2	qdaC2	qdaC2	qdaC2	qdaC2	qdaC2
C3	qdaC3	qdaC3	qdaC3	qdaC3	qdaC3	qdaC3	qdaC3	qdaC3
C4	qdaC4	qdaC4	qdaC2	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4
	Homocedastic ; Equal Ranges ; Bad Separation							
	30	60	100	150	30	60	100	150
C1	ldaC1	ldaC1	ldaC1	Vrt	ldaC1	ldaC1	ldaC1	ldaC1
C2	ldaC2	ldaC2	ldaC2	ldaC2	ldaC2	ldaC2	ldaC2	ldaC2
C3	CP	ldaC3	CP	CP	CP	CP	LocC4	ldaC3
C4	ldaC4	ldaC4	Vrt	ldaC4	Vrt	CP	Vrt	ldaC4
	Heterocedastic ; Equal Ranges ; Bad Separation							
	30	60	100	150	30	60	100	150
C1	qdaC3	qdaC3	qdaC3	qdaC1	qdaC3	qdaC3	qdaC1	qdaC1
C2	qdaC2	qdaC2	qdaC2	qdaC2	qdaC4	qdaC2	qdaC2	qdaC2
C3	qdaC4	qdaC3	qdaC3	qdaC3	qdaC4	qdaC3	qdaC3	qdaC3
C4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4
	Homocedastic ; Different Ranges ; Bad Separation							
	30	60	100	150	30	60	100	150
C1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1	ldaC1
C2	Vrt	Vrt	Vrt	Vrt	ldaC2	ldaC2	ldaC2	CpRgCmb
C3	ldaC3	ldaC3	ldaC3	ldaC3	ldaC3	LocC3	ldaC3	qdaC3
C4	ldaC4	ldaC2	Vrt	Vrt	CpRgSep	ldaC4	qdaC4	CpRgSep
	Heterocedastic ; Different Ranges ; Bad Separation							
	30	60	100	150	30	60	100	150
C1	qdaC3	qdaC3	qdaC1	qdaC1	qdaC3	qdaC3	qdaC1	qdaC1
C2	qdaC4	qdaC2	qdaC2	qdaC2	qdaC4	qdaC2	qdaC2	qdaC2
C3	qdaC4	qdaC3	qdaC3	qdaC3	qdaC3	qdaC3	qdaC3	qdaC3
C4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4	qdaC4

would like to thank the Fundação para a Ciência e a Tecnologia for its financial support through project PEst-OE/EGE/UI0731/2014.

Table 11 Average and standard error (in brackets) error rates for the best methods by data condition: a) Good separation of group centroids (in bold if all other methods present average error rates larger by more than 0.01).

	Gaussian				Skew-Normal			
	30	60	100	150	30	60	100	150
	Homocedastic ; Equal Ranges ; Good Separation							
C1	0.1015 (.00237)	0.0800 (.00137)	0.0626 (.00069)	0.0741 (.00063)	0.0924 (.00247)	0.0800 (.00116)	0.0585 (.00086)	0.0621 (.00063)
C2	0.0858 (.00145)	0.0705 (.00084)	0.0727 (.00048)	0.0505 (.00055)	0.0732 (.00147)	0.0666 (.00084)	0.0643 (.00059)	0.0635 (.00039)
C3	0.0685 (.00353)	0.0743 (.00414)	0.0721 (.00047)	0.0726 (.00095)	0.0792 (.00122)	0.0658 (.00071)	0.0574 (.00072)	0.0634 (.00036)
C4	0.1670 (.00142)	0.1505 (9e-04)	0.1491 (6e-04)	0.1295 (.00064)	0.1566 (.00152)	0.1538 (.00158)	0.1311 (.00062)	0.1414 (.00100)
	Heterocedastic ; Equal Ranges ; Good Separation							
C1	0.0784 (.00236)	0.0369 (.00196)	0.0211 (.00088)	0.0168 (.00077)	0.0660 (.00439)	0.0326 (.00237)	0.0350 (.00108)	0.0212 (7e-04)
C2	0.0485 (.00263)	0.0241 (.00103)	0.0178 (.00048)	0.0118 (.00062)	0.0381 (.00194)	0.0180 (.00085)	0.0228 (.00059)	0.0096 (.00046)
C3	0.0537 (.00244)	0.0226 (.00096)	0.0155 (.00068)	0.0148 (.00043)	0.0733 (.00290)	0.0327 (.00101)	0.0215 (.00071)	0.0218 (5e-04)
C4	0.0207 (.00243)	0.0145 (.00105)	0.0111 (.00098)	0.0093 (.00062)	0.0290 (.00274)	0.0170 (.00178)	0.0237 (.00125)	0.0207 (.00116)
	Homocedastic ; Different Ranges ; Good Separation							
C1	0.0981 (.00217)	0.0760 (.00108)	0.0774 (.00071)	0.0739 (.00048)	0.0998 (.00220)	0.0844 (.00098)	0.0870 (.00049)	0.0726 (.00069)
C2	0.0680 (.00150)	0.0785 (.00083)	0.0669 (.00037)	0.0703 (.00028)	0.0922 (.00135)	0.0718 (.00055)	0.0653 (.00045)	0.0646 (5e-04)
C3	0.0868 (.00156)	0.0874 (.00084)	0.0650 (.00058)	0.0640 (.00047)	0.0852 (.00186)	0.0837 (.00113)	0.0751 (.00079)	0.0680 (.00053)
C4	0.0865 (.00132)	0.0763 (.00064)	0.0790 (.00052)	0.0777 (.00037)	0.0772 (.00138)	0.0682 (.00077)	0.0638 (.00051)	0.0580 (.00032)
	Heterocedastic ; Different Ranges ; Good Separation							
C1	0.0757 (.00337)	0.0343 (.00155)	0.0246 (.00104)	0.0259 (.00071)	0.0770 (.00227)	0.0343 (.00222)	0.0298 (8e-04)	0.0171 (.00084)
C2	0.0364 (.00225)	0.0233 (.00092)	0.0193 (.00048)	0.0148 (.00037)	0.0391 (.00228)	0.0224 (.00088)	0.0143 (.00059)	0.0120 (.00054)
C3	0.0518 (.00254)	0.0222 (.00115)	0.0158 (.00073)	0.0196 (.00059)	0.0738 (.00334)	0.0391 (.00113)	0.0224 (.00079)	0.0232 (.00067)
C4	0.0168 (.00184)	0.0119 (.00068)	0.0078 (.00055)	0.0049 (4e-04)	0.0274 (.00206)	0.0135 (.00095)	0.0185 (.00071)	0.0093 (.00046)

References

1. Angulo, C., Anguita, D. and González, L., Interval discriminant analysis using Support Vector Machines in ESANN'2007 Proceedings - European Symposium on Artificial Neural Networks, Bruges (2007).
2. Appice, A., D'Amato, C., Esposito, F., Malerba, D., Classification of symbolic objects: A lazy learning approach, *Intelligent Data Analysis*, 10 (4), 301–324 (2006).
3. Arellano-Valle, R. and Azzalini, A., The centred parametrization for the multivariate skew-normal distribution, *Journal of Multivariate Analysis*, 99, 1362-1382 (2008).
4. Azzalini, A. and Dalla Valle, A., The multivariate Skew-Normal distribution, *Biometrika*, 83 (4), 715-726 (1996).

Table 12 Average and standard error (in brackets) error rates for the best methods by data condition: b) Bad separation of group centroids (in bold if all other methods present average error rates larger by more than 0.01).

	Gaussian				Skew-Normal			
	Homocedastic ; Equal Ranges ; Bad Separation							
	30	60	100	150	30	60	100	150
C1	0.3696 (.00418)	0.3465 (.00201)	0.3225 (.00182)	0.3393 (.00277)	0.3691 (.00393)	0.3256 (.00315)	0.3255 (.00177)	0.3176 (.00128)
C2	0.3613 (.00325)	0.3422 (.00256)	0.3371 (.00166)	0.3114 (.00116)	0.3656 (.00396)	0.3448 (.00246)	0.3038 (.00176)	0.3065 (.00178)
C3	0.3314 (.00437)	0.3270 (.00267)	0.3153 (.00125)	0.3025 (.00111)	0.3216 (.00465)	0.3199 (.00178)	0.3166 (.00077)	0.3064 (.00132)
C4	0.4168 (.00301)	0.3939 (.00236)	0.3779 (.00174)	0.3638 (.00187)	0.4034 (.00346)	0.3574 (.00287)	0.3707 (.00174)	0.3800 (.00133)
	Heterocedastic ; Equal Ranges ; Bad Separation							
	30	60	100	150	30	60	100	150
C1	0.0950 (.00192)	0.0462 (.00232)	0.0321 (.00158)	0.0273 (.00217)	0.0781 (.00336)	0.0463 (.00321)	0.0400 (.00246)	0.0215 (.00186)
C2	0.0615 (.00374)	0.0293 (.00293)	0.0231 (.00215)	0.0209 (.00149)	0.0500 (.00274)	0.0325 (.00276)	0.0197 (.00205)	0.0183 (.00189)
C3	0.0630 (.00205)	0.0364 (.00322)	0.0257 (.00238)	0.0299 (.00176)	0.0779 (.00182)	0.0330 (.00277)	0.0260 (.00188)	0.0307 (.00195)
C4	0.0401 (.00291)	0.0221 (.00269)	0.0226 (.00259)	0.0258 (.00208)	0.0591 (.00332)	0.0380 (.00273)	0.0327 (.00227)	0.0220 (.00164)
	Homocedastic ; Different Ranges ; Bad Separation							
	30	60	100	150	30	60	100	150
C1	0.3676 (.00357)	0.3600 (.00275)	0.3295 (.00181)	0.3367 (.00165)	0.3796 (.00417)	0.3305 (.00215)	0.3246 (.00194)	0.3194 (.00157)
C2	0.3424 (.00259)	0.3489 (.00185)	0.3101 (.00135)	0.3027 (8e-04)	0.3430 (.00380)	0.3420 (.00197)	0.3180 (.00152)	0.3019 (.00156)
C3	0.3607 (.00470)	0.3329 (.00267)	0.3451 (.00134)	0.3319 (.00151)	0.3653 (.00343)	0.3576 (.00445)	0.3609 (.00116)	0.3174 (.00155)
C4	0.3516 (.00406)	0.3250 (.00197)	0.3243 (.00211)	0.3275 (.00169)	0.3579 (.00470)	0.3417 (.00170)	0.3266 (.00165)	0.3150 (.00108)
	Heterocedastic ; Different Ranges ; Bad Separation							
	30	60	100	150	30	60	100	150
C1	0.0898 (.00127)	0.0525 (.00214)	0.0336 (.00243)	0.0286 (.00226)	0.0822 (.00137)	0.0471 (.00153)	0.0309 (.00276)	0.0245 (.00193)
C2	0.0533 (.00263)	0.0236 (.00244)	0.0242 (.00159)	0.0132 (.00131)	0.0502 (.00353)	0.0307 (.00267)	0.0161 (.00218)	0.0247 (.00157)
C3	0.0521 (.00168)	0.0263 (.00300)	0.0192 (.00178)	0.0256 (.00192)	0.0595 (.00380)	0.0377 (.00220)	0.0319 (.00171)	0.0252 (.00155)
C4	0.0336 (.00414)	0.0211 (.00241)	0.0238 (.00185)	0.0207 (.00144)	0.0498 (.00357)	0.0344 (.00204)	0.0377 (.00165)	0.0321 (.00150)

- Azzalini, A. and Capitanio, A., Statistical applications of the multivariate Skew-Normal distribution, *J. R. Statist. Soc. B*, 61 (3) (1999), 579–602 (1999).
- Beheshti, M., Berrached, A., de Korvin, A., Hu, C. and Sirisaengtaksin, O., On interval weighted freelay neural networks, *in Proc. of the 31st Annual Simulation Symposium*, IEEE Computer Society Press, pp. 188-194 (1998).
- Bertrand, P. and Goupil, F., Descriptive Statistics for Symbolic Data, *in Bock, H.-H. and Diday, E. (Eds.), Analysis of Symbolic Data, Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer, Heidelberg, pp. 106-124 (2000).
- Bock, H.-H. and Diday, E. (Editors), *Analysis of Symbolic Data, Exploratory methods for extracting statistical information from complex data*, Springer-Verlag, Berlin-Heidelberg (2000).

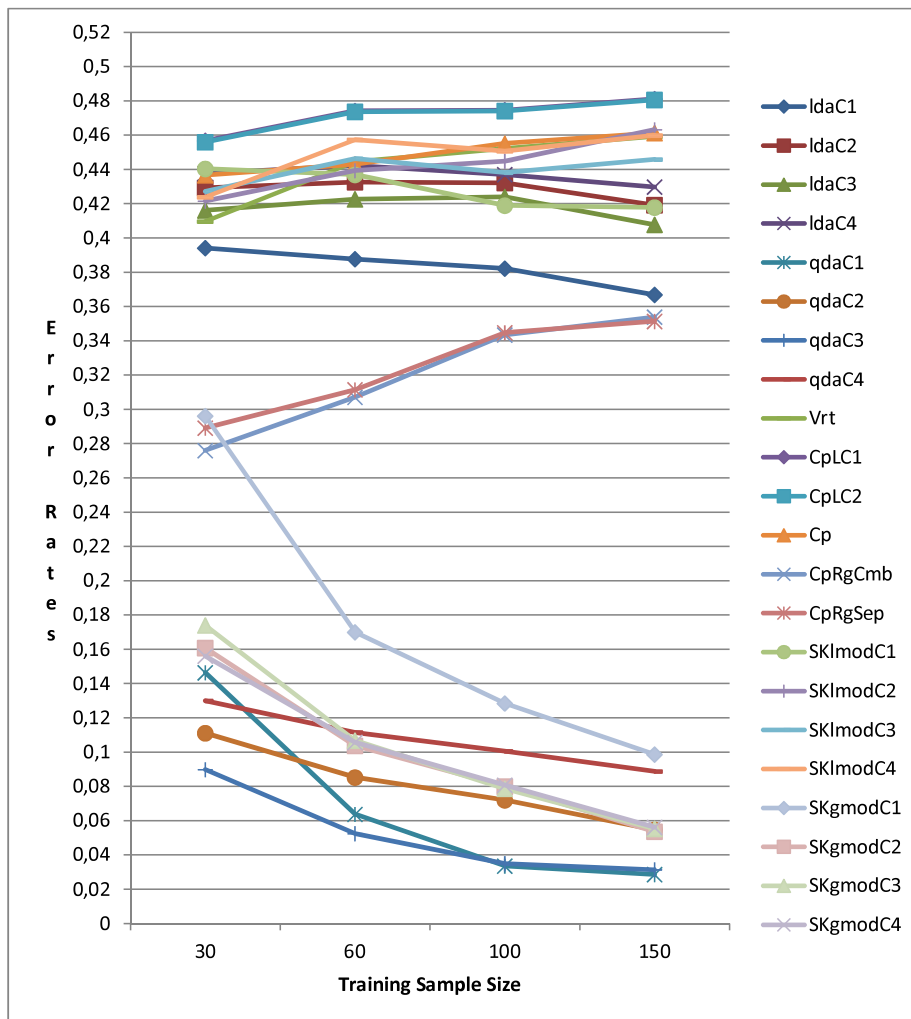


Fig. 2 Error rates of all methods on Gaussian data, in the validation sample for the heterocedastic condition under Case 1, with heterogeneous Ranges and bad-separated classes.

9. Billard, L. and Diday, E., From the statistics of data to the statistics of knowledge: Symbolic Data Analysis, *Journal of the American Statistical Association*, 98 (462), 470-487, (2003).
10. Brito, P. and Duarte Silva, A.P., Modelling interval data with Normal and Skew-Normal distributions, *Journal of Applied Statistics*, 39 (1), 3-20. (2012).
11. Carrizosa, E., Gordillo, J. and Plastria, F., Classification problems with imprecise data through separating hyperplanes, Available at http://www.optimization-online.org/DB_HTML/2007/09/1781.pdf (2007).
12. Chouakria, A., Cazes, P. and Diday, E., Symbolic Principal Component Analysis, in Bock, H.-H. and Diday, E. (Eds.), *Analysis of Symbolic Data, Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer, Heidelberg, pp. 200-212 (2000).
13. Diday, E. and Noirhomme-Fraiture, M., *Symbolic Data Analysis and the SODAS Software*, Wiley, Chichester (2008).
14. Do, T.-N and Poulet, F., Kernel methods and visualization for interval data mining, in Janssen, J. and Lenca, P. (Eds.), *Proc. of the Conf. on Applied Stochastic Models and Data Analysis, ASMDA 2005, ENST Bretagne* (2005).

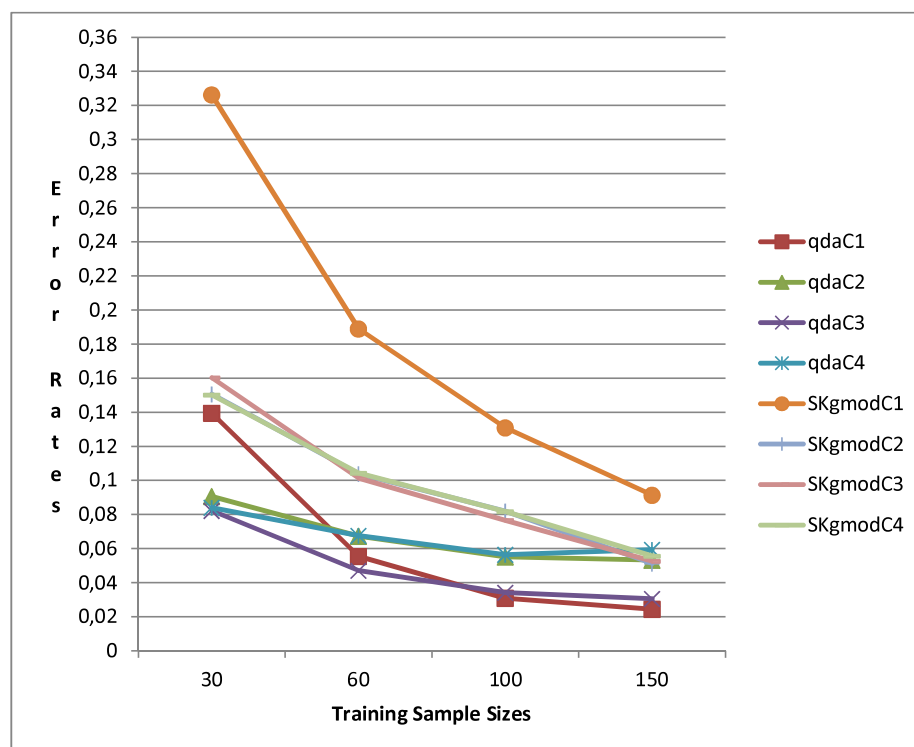


Fig. 3 Error rates of Gaussian quadratic and Skew-Normal general methods on Skew-Normal data, in the validation sample for the heterocedastic condition under Case 1, with heterogenous Ranges and bad-separated classes.

15. Duarte Silva, A.P. and Brito, P., Linear discriminant analysis for interval data, *Computational Statistics*, 21 (2), 289-308 (2006).
16. Flury, B.W., Schmid, M.J. and Narayanan, A., Error rates in quadratic discrimination with constraints on the covariance matrices, *Journal of Classification*, 11, 101-120 (1994).
17. Ishibuchi, H., Tanaka, H. and Noriko Fukuoka, N., Discriminant analysis of multi-dimensional interval data and its application to chemical sensing, *International Journal of General Systems*, 16 (4), 311-329 (1990).
18. Jahanshahlooa, G.R., Lotfib, F.H., Balfc, F.R. and Rezaid, H.Z., Discriminant analysis of interval data using Monte Carlo method in assessment of overlap, *Applied Mathematics and Computation*, 191 (2), 521-532 (2007).
19. Lauro, N.C., Verde, R. and Palumbo, F., Factorial discriminant analysis on symbolic objects, in Bock, H.-H. and Diday, E. (Eds.), *Analysis of Symbolic Data, Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer, Heidelberg, pp. 212-233 (2000).
20. Lauro, C. and Palumbo, F., Principal Component Analysis for non-precise data, in Vichi, M. *et al* (Eds.), *New Developments in Classification and Data Analysis*, Springer, pp. 173-184 (2005).
21. Lauro, N.C., Verde, R. and Irpino, A., Factorial Discriminant Analysis, in Diday, E. and Noirhomme-Fraiture, M. (Eds.), *Symbolic Data Analysis and the Sodas Software*, Wiley, Chichester, pp. 341-358 (2008).
22. Marks, S. and Dunn, O.J., Discriminant functions when the covariance matrices are unequal, *Journal of the American Statistical Association* 69, 555-559 (1974).
23. Moore, R.E., *Interval Analysis*. Prentice Hall, New Jersey (1966).
24. Neto, E.A.L. and De Carvalho, F.A.T., Centre and range method for fitting a linear regression model to symbolic intervalar data, *Computational Statistics & Data Analysis*, 52, 1500-1515 (2008).
25. Neto, E.A.L. and De Carvalho, F.A.T., Constrained linear regression models for symbolic interval-valued variables, *Computational Statistics & Data Analysis*, 54, 333-347 (2010).

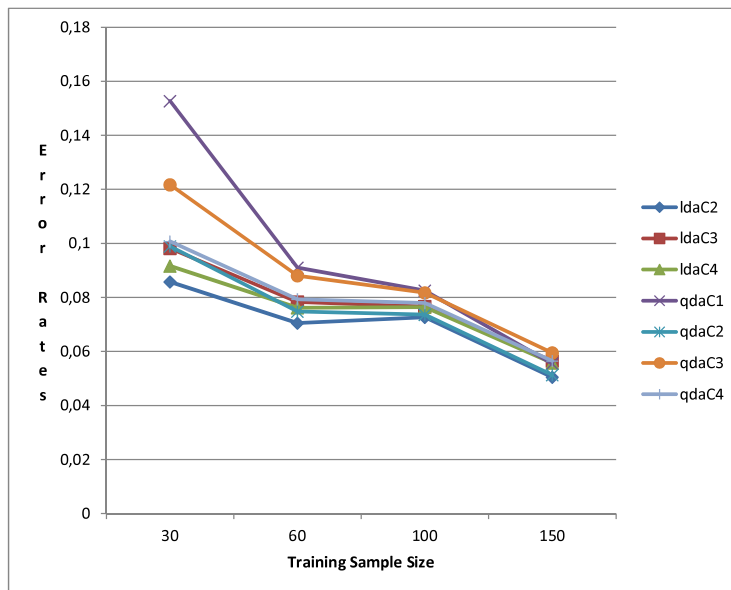


Fig. 4 Error rates of the parametric Gaussian based methods on Gaussian data, in the validation sample for the homocedastic condition under Case 2, with homogenous Ranges and well-separated classes.

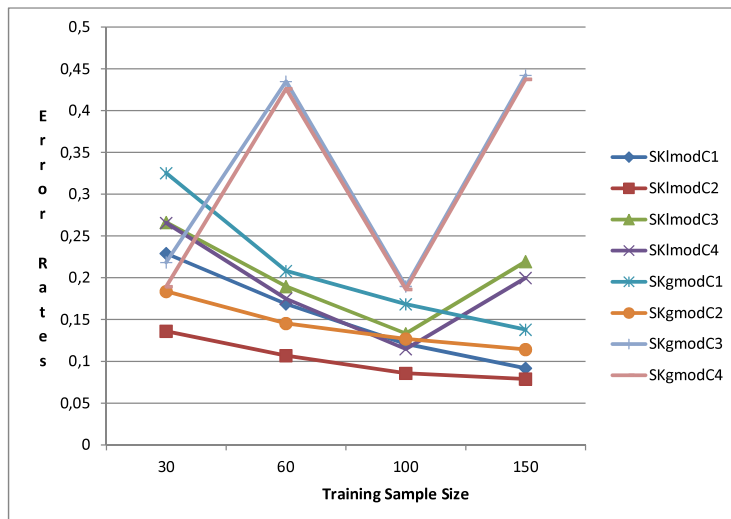


Fig. 5 Error rates for parametric Skew-Normal methods on Skew-Normal data, in the validation sample for the Location Model under Case 2, with homogenous Ranges and well-separated classes.

26. Nivlet, P., Fournier, F. and Royer, J.J., Interval discriminant analysis: An efficient method to integrate errors in supervised pattern recognition, *in* ISIPTA'01, pp. 284-292 (2001).
27. Noirhomme-Fraiture, M. and Brito, P., Far beyond the classical data models: Symbolic Data Analysis, *Statistical Analysis and Data Mining*, 4 (2), 157-170 (2011).
28. Rasson, J.P., Piron, J.-Y., Lallemand, P. and Adans, S., Unsupervised divisive classification, *in* Diday, E. and Noirhomme-Fraiture, M. (Eds.), *Symbolic Data Analysis and the Sodas Software*, Wiley, Chichester, pp. 149-156 (2008).

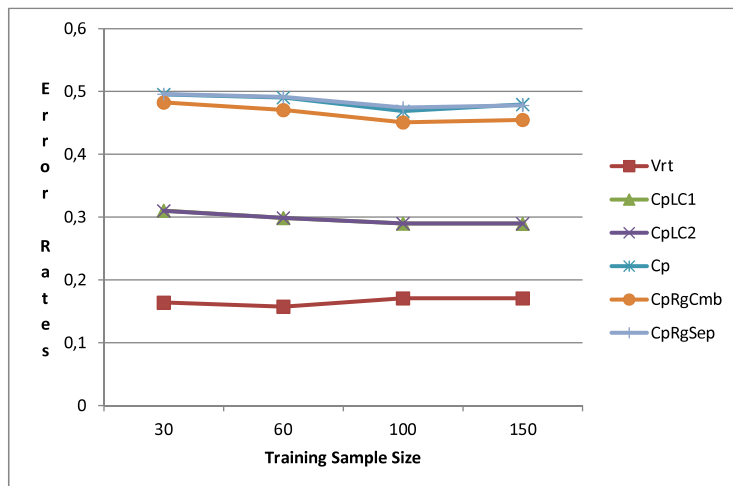


Fig. 6 Error rates of the distance based methods on Gaussian data, in the validation sample for the homocedastic condition under Case 2, with homogenous Ranges and well-separated classes.

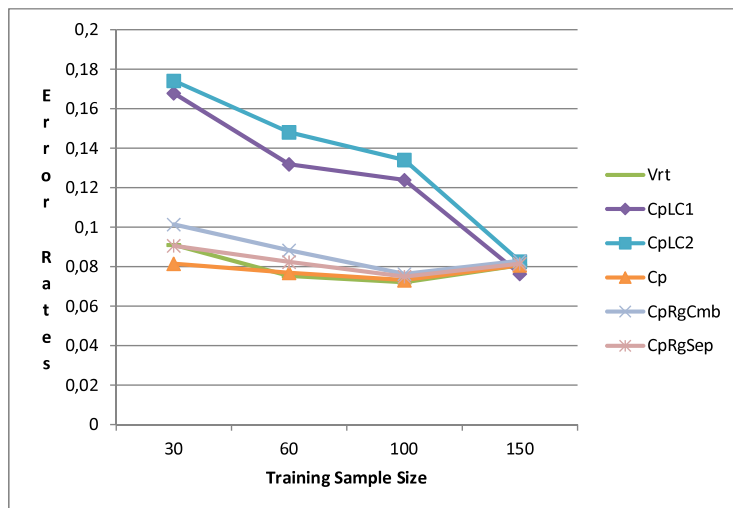


Fig. 7 Error rates of the distance based methods on Gaussian data, in the validation sample for the homocedastic condition under Case 3, with homogenous Ranges and well-separated classes.

29. Rossi, F. and Conan Guez, B., Multilayer perceptron on interval data, in Jajuga, K., Sokolowski, A. and Bock, H.-H. (Eds.), Classification, Clustering and Data Analysis, Springer, Berlin, Heidelberg, New York, pp. 427-434 (2002).
30. Síma, J., Neural expert systems, Neural Networks, 8 (2), 261-271 (1995).
31. Simoff, S.J., Handling uncertainty in neural networks: an interval approach, in Proc. of the IEEE International Conference on Neural Networks, IEEE, Washington D. C., pp. 606-610 (1996).
32. Utkin, L.V. and Coolen, F.P.A., Interval-valued regression and classification models in the framework of machine learning, in 7th International Symposium on Imprecise Probability: Theories and Applications, Innsbruck (2011).

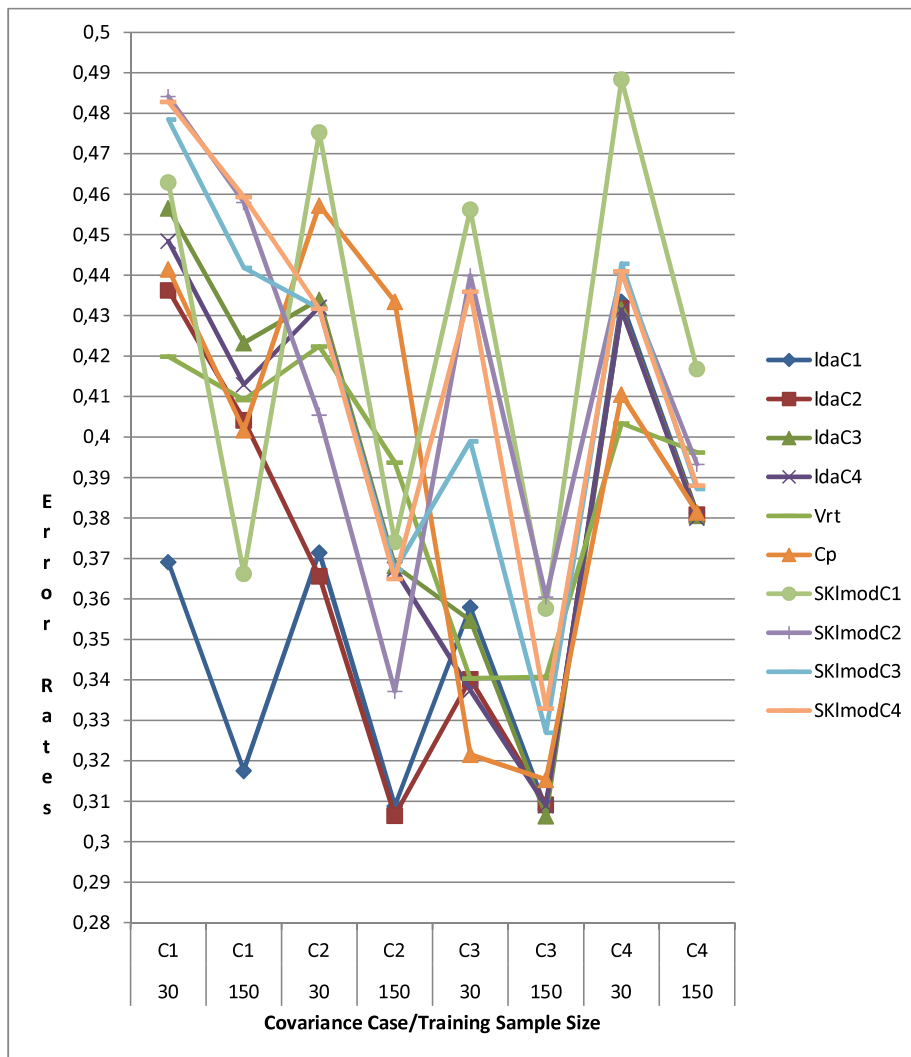


Fig. 8 Error rates for selected methods on Skew-Normal data, in the validation sample for the homocedastic condition for all cases, with homogenous Ranges and bad-separated classes.

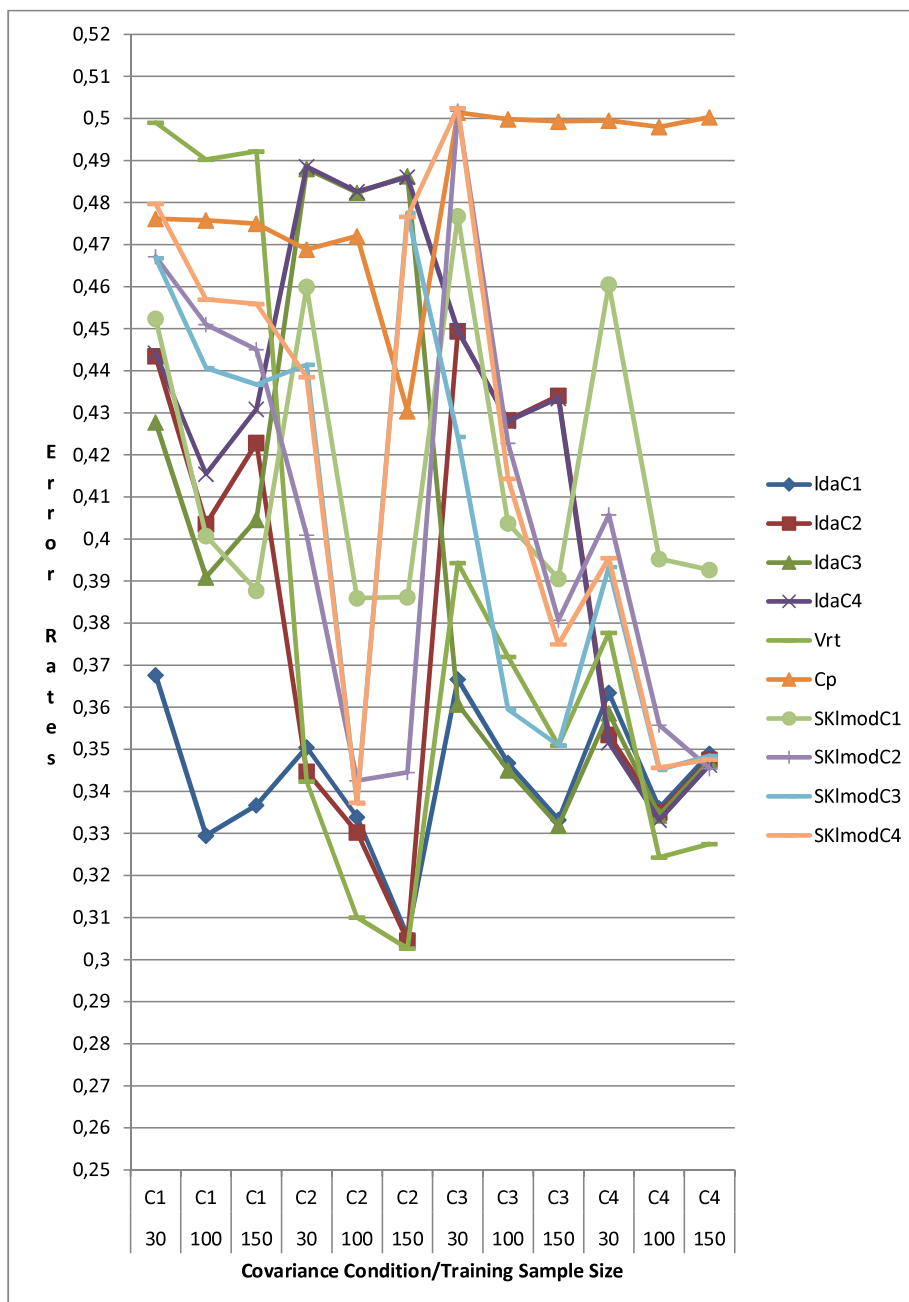


Fig. 9 Error rates for selected methods on Gaussian data, in the validation sample for the homocedastic condition for all cases, with heterogenous Ranges and bad-separated classes.