MATHEMATICAL MODELING AND COMPUTING, Vol. 10, No. 3, pp. 651–659 (2023)

Statistical analysis of three new measures of relevance redundancy and complementarity

El Mourtji B., Chamlal H., Ouaderhman T.

Department of Mathematics and Computer Science,
Fundamental and Applied Mathematics Laboratory,
Faculty of Sciences Ain Chock, Hassan II University of Casablanca, Morocco

(Received 15 February 2023; Accepted 3 July 2023)

Discriminant analysis is part of statistical learning; its goal is to separate classes defined a priori on a population and involves predicting the class of given data points. Discriminant analysis is applied in various fields such as pattern recognition, DNA microarray etc. In recent years, the discrimination problem remains a challenging task that has received increasing attention, especially for high-dimensional data sets. Indeed, in such a case, the feature selection is necessary, which implies the use of criteria of relevance, redundancy and complementarity of explanatory variables. The aim of this paper is to present an analysis of three new criteria proposed in this sense, more precisely based on the Principal Component Analysis we have been able to achieve a double objective: that of studying the harmony of these three criteria and also visualizing the class of candidate variables for a more in-depth selection in addition to eliminating the noise variables in a discriminant model.

Keywords: *relevance; redundancy; complementarity; preordonnances theory; discriminant analysis; principal component analysis.*

2010 MSC: 62-07 **DOI:** 10.23939/mmc2023.03.651

1. Introduction

The objective of a discrimination problem based on a set of individuals is to infer a link between the characteristics of each individual and their class, usually defined by a label. Currently, due to the ease of data acquisition and storage, real problems of pattern recognition or discrimination are increasingly complex and involve a significant number of variables, often heterogeneous characterizing an example. Intuitively it seems natural that the increase in the number of variables should not affect the quality of the discrimination, in practice, it turns out to be a major problem. Variable selection (VS for short) is therefore useful in this context even if this reduction can lead to a slight loss of information. Since the relevant variables are not known a priori, the selection is justified, in the presence of a large number of attributes, by the possibility of the existence of interrelated variables and/or noise or redundant variables which generally give high error rates. The selection of variables essentially makes it possible to improve the performance of the classification models by using only the important variables for the problem studied, to reduce the time and the cost of calculation and to facilitate the understanding of the process generating information. VS can be categorized as filter methods (FM), wrapper methods (WM), and embedded methods (EM). The (WM) use the learning models to evaluate the feature subset by the classification accuracy rates. Regardless of being computationally intensive, there is a tendency to cause overfitting [6]. (EM) use learning models to guide variable selection and is often evaluated as classification rather than variable selection. The well-known Lasso and its variants are examples of (EM). (FM) separate the learning model and variable selection and weigh features based on their characteristics [8]. There exist some filters measuring the redundancy, defined as the overlapping information shared among variables toward predicting the target class, in addition to variable relevance, such as minimal-redundancy-maximal-relevance (mRMR) [11], conditional mutual information maximization (CMIM) [13], and minimum conditional relevance-minimum conditional redundancy

(MCRMCR) and minimum conditional relevance-minimum intra-class redundancy (MCRMICR) [14], aimed at finding the variable subset with the maximum relevance to the target class and minimum redundancy.

To define a powerful feature subset, some filters simultaneously take advantage of feature relevance, redundancy, and complementarity. This later is justified by the fact that the dependence among variables may not always affect negatively the discriminative power. Because a feature with a low relevance but highly dependent on other variables can be useful to enhance the discriminative power of the variable subset. Hence, neglecting the complementarity between variables can lose some valuable information for discrimination problems. In [7], the redundancy-complementariness dispersion is taken into account to adjust the measurement of pairwise inter-correlation of features. In [12], authors proposed an approach to VS that explicitly characterizes and uses feature complementarity in the search process where an adaptive cost function that uses redundancy-complementarity ratio to automatically update the trade-off rule between relevance, redundancy, and complementarity.

In this paper, three main concepts are considered: relevance, redundancy, and complementarity. The relevance is defined as the univariate association strength of a variable with the target class. The more relevant variable is, the larger relevance measure value is. The redundancy refers to the overlapping information shared among features toward predicting the target. Finally, complementarity quantifies the extent to which several variables are strongly associated with the target class jointly. There are several definitions of the relevance of a characteristic. According to [9], a characteristic (variable) is classified as strongly relevant, weakly relevant, or irrelevant. A strongly relevant characteristic implies that the variable is indispensable in the sense that its removal leads to a loss of prediction accuracy. A characteristic is weakly relevant if it is not 'strongly relevant' and there is a V subset such that the performance of $V \cup \{Xi\}$ is significantly better than the performance of V. Features that are neither 'strongly relevant' nor 'weakly relevant' represent irrelevant features. These features will generally be removed from the starting set of characteristics. For the redundancy notion, it is widely accepted that two features are redundant to each other if their values are completely correlated. In what follows, we describe the main preliminary concepts that are used in this framework using the preordonnances theory which has attracted the attention of many researchers lately [2–5].

2. Evaluation criteria based on preodonnances theory

Since feature selection, in the context of discriminant analysis, involves the elimination of irrelevant and redundant features and highlights the recognition of complementary features using preordonnance theoretics, we give the definitions of relevance, redundancy, and complementarity in this section.

2.1. Relevance measure

The relevance of a variable signifies its explanatory power to predict a target class, and is a measure of variable worthiness, separately from other variables. In the context of variable reduction, we can conclude that the larger relevance, the stronger discriminative power of the feature. Hence, the relevance between the feature and the target class can be defined as in Eq. (1). In particular, let $E = \{1, 2, \dots, n\}$: a sample of size n and E_p a set defined by $E_p = \{(i,j) \in E^2 / i < j\}$. A preordonnance is a transitive and reflexive binary relation defined on E_p.

Definition 1 (Refs. [2,4,5]). *The relevance of a variable f with respect to the target Y is given by:* $\forall((i,j),(k,l)) \in (E_p)^2$;

$$\begin{aligned}\Psi(f) &= \operatorname{cor}(T_{P_f}, T_{P_Y}) = \psi_{cor}(P_f, P_Y) \\ &= \frac{\sum T_{P_f}(i,j,k,l) T_{P_Y}(i,j,k,l)}{\sqrt{\sum T^2_{P_f}(i,j,k,l)}\sqrt{\sum T^2_{P_f}(i,j,k,l)}},\end{aligned} \tag{1}$$

where P_f and P_Y are the preordonnances induced respectively by the features f and Y, while T_{P_f} and T_{P_Y} their associated coding. The symbol $\sum$ covers all $((i,j)(k,l)) \in E_p \times (E_p \setminus (i,j))$.

A higher relevance value means that the feature will have a larger effect to predict the target, that is, the feature is relevant. A negative relevance value ($\Psi(f_d) \leqslant 0$) means that the feature is irrelevant. Furthermore, let P_{f_1} and P_{f_2} be two preordonances induced by two variables, f_1 and f_2, respectively, and the rank variables induced on E_p by these preordonances are denoted by $r_{P_{f_1}}$ and $r_{P_{f_2}}$, respectively. In [1] authors have demonstrated equality between the association measure Ψ applied to the preordonances, P_{f_1} and P_{f_2}, and the Kendall's τ coefficient applied to the rank variables, $r_{P_{f_1}}$ and $r_{f_{X_2}}$ and in [2], the same association measure Ψ is demonstrated equal to the Lerman association coefficient between two variables. Hence, the relevance of the variable f can be defined differently as in Eq. (2).

$$\Psi(f) = \tau(r_{P_f}, r_{P_Y}) = L(f, Y) \tag{2}$$

2.2. Redundancy

As mentioned previously, an optimal feature subset should not only contains the variables having greater relevance to the target class, but also should have lower redundancy within variables. For that, a partial correlation is used to define the redundancy measure.

Definition 2 (Ref. [2]). *The redundancy of a variable f_m to another variable f_a ($f_a \in X \setminus \{f_m\}$) is defined by the agreement intensity between the variable f_m and the target variable Y, after ignoring the impact of the variable f_a, and is evaluated as:*

$$\Psi_{.}(f_m, f_a) = \psi_{cor}(P_{f_m}, P_Y)_{.P_{f_a}} = \tau_1(r_{P_{f_m}}, r_{P_Y})_{.r_{P_{f_a}}}, \tag{3}$$

where $(\tau_1)_{.}$ is the partial rank correlation coefficient.

The greater the value of $\Psi_{.}(f_m, f_a)$ is, the less redundant the variable f_m to the variable f_a is.

2.3. Complementarity

The relevance and the non-redundancy are necessary conditions of optimality but not sufficient. Indeed, two interdependent features may be complementary to each other and possibly have a high discriminative power when they serve as a group. Then, to quantify the extent to which two or more features are strongly associated with the response variable jointly, we use Ψ_W as the association between more than two variables. It will be used to evaluate the agreement between preordonnances induced by the variables under study. In particular: Let $f_1, f_2, \ldots, f_m$ be m heterogeneous variables, $P_1, P_2, \ldots, P_m$ denote the induced preordonnances and let $r_1, r_2, \ldots, r_m$ be the associated rank variables defined on the set of pairs E_p. We define the multiple concordance coefficient between $P_1, P_2, \ldots, P_m$ as follows:

$$\Psi_W(P_1, P_2, \ldots, P_m) = W(r_1, r_2, \ldots, r_m), \tag{4}$$

where W is Kendall's multiple concordance coefficient.

The coefficients cited above can deal with continuous and categorical features either separately or in a mixed fashion. Their expression is not complicated and thus their implementation is not difficult. Kendall's empirical coefficients τ and W can be obtained from almost any statistical software.

3. Analysis of three evaluation criteria

To analyze the behavior of three evaluation criteria, the Principal Component Analysis (PCA) method [10], is used to summarize and visualize information for a given data set, in which quantitative variables may be inter-correlated. The principal component analysis is a dimensionality-reduction method that extracts the important information from the data and expresses this information as a set of summary indices called principal components. PCA is based on some mathematical concepts such as variance, covariance, and eigenvalues. The steps for PCA algorithm can be summarized as follows:

1. Standardizing the data.
2. Calculating the covariance.
3. Calculating the eigen values and eigen vectors.
4. Sorting the Eigen Vectors.

5. Calculating the new features Or Principal Components.
6. Remove fewer or unimportant features from the new dataset.
7. Recast the data along the axes of the principal component.

Based on the proposed evaluation criteria and PCA method with 2 principal components, the data points can be projected into a set of linearly uncorrelated criteria with the help of orthogonal transformation, where a criterion represents one coordinate axis and the remaining criteria represent the second coordinate axis. For example, if the x-axis represents the relevance and the redundancy, to some extent, and the y-axis represents the complementarity, we can categorize the variables into 4 subsets:

— **Strongly favorable**: subset containing variables strongly relevant, with lower redundancy within each other, and with high complementarity which means that the combination of considered variables can yield more information with respect to the target class.
— **Favorable but inhomogeneous**: subset containing variables strongly relevant to the target class and with lower redundancy within each other, but fail to explain the target class when serving as a group.
— **Homogeneous but infavorable**: subset containing variables with lower relevance, with a certain redundancy within each other, but their combination can provide additional classification information.
— **Noise**: subset containing irrelevant variables, with higher redundancy, and they can not add any additional classification information as a combination.

This categorization allows us to easily detect the subset containing candidate variables that predict as much as possible the target class (the subset in the top right in Figure 1*b*). On one hand, these variables are strongly relevant (with a positive relevance value) and not redundant, which promote the strength of their link with the target class. On the other hand, the combination of these variables is homogenous; this means that this group of variables has a high discriminative power to predict the target class jointly. On the other hand, the cluster (placed at the bottom left in Figure 1*b*) can be considered as the set of non-candidate variables to form the optimal or suboptimal subset of explanatory variables predicting the target class. Indeed, this cluster contains irrelevant variables, with considerable redundancy among them, and which do not accurately explain the target when represented as a group. This subset of variables must be firmly rejected. It should be noted that the redundancy in this study computes the agreement intensity between a variable f_m and the target variable Y, after ignoring the impact of the variable $f_{(1)}$ where $f_{(1)}$ is the more relevant variable (maximizing the Ψ measure) and the complementarity criteria computes the agreement intensity between a variable f_m, the variable $f_{(1)}$ and the target Y.

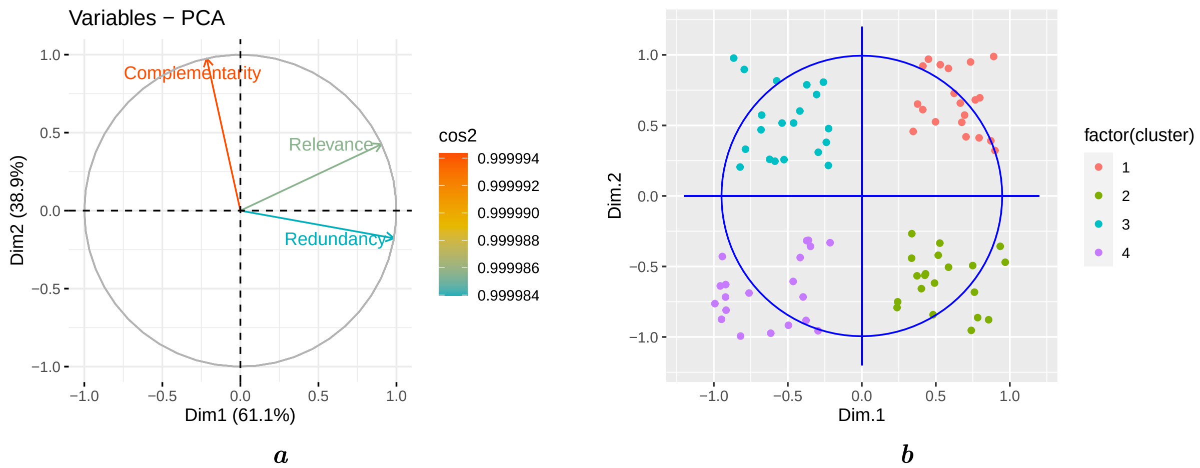


Fig. 1. (***a***) Representation of variables on factor map, (***b***) Variables categorization.

To better understand this categorization, two scenarios are analyzed in the following.

3.1. Scenario 1

A dataset is generated with 200 observations and 150 variables from this model:

$$Z = X\gamma + \alpha,$$

where the matrix's rows are generated from $N(0_p, \Pi)$ as a multivariate normal distribution, where $\Pi \in \mathbb{M}_{p\times p}$ is a block diagonal matrix set by:

$$\Pi_{jl} = \begin{cases} 1 & \text{if} \quad j = l, \\ \rho & \text{if} \quad j \leqslant 50, \quad l \leqslant 50 \ \text{ and } \ j \neq l, \\ \rho & \text{if} \quad 51 \leqslant i \leqslant 100, \quad 51 \leqslant j \leqslant 100 \ \text{ and } \ j \neq l, \\ 0 & \text{otherwise,} \end{cases}$$

where ρ is the correlation parameter between features. The elements of α, $\alpha_i' s$, for $i = 1, \ldots, n$ are simulated independently following a normal distribution with mean zero and variance $\sigma^2 = 2.5^2$. The vector of coefficients, γ, takes the form:

$$\begin{aligned} &\gamma_j \approx \text{Unif}[0.9, 1.1], && \text{if} \ \ 1 \leqslant j \leqslant 25; \\ &\gamma_j \approx \text{Unif}[-1.1, -0.9], && \text{if} \ \ 51 \leqslant j \leqslant 75; \\ &\gamma_j = 0 && \text{otherwise.} \end{aligned}$$

Finally, the binary response Y is simulated as follows:

$$\text{logit}(\mathbb{P}(Y = -1)) = Z.$$

In particular, we have 100 features distributed over two subsets of 50 features, 25 of which are correlated to the target variable, and the other 25 form noise. The remaining 50 features are generated independently of the first 100 features and the target variable.

From PCA outputs, we can extract a matrix containing the coordinates of the active variables. Table 1 summarizes the PCA coordinates for different values of ρ.

Table 1. Dimensions using PCA method on datasets with different ρ values.

Criteria	$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.8$		$\rho = 0.95$	
Dimensions	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2
Relevance	0.9852998	-0.1708330	0.9901019	-0.1403307	0.9978113	-0.06591778	0.9999410	0.001712412
Complementarity	0.8642104	0.5031306	0.7820116	0.6232636	0.7032425	0.71094605	0.7116317	-0.702508407
Redundancy	0.9607623	-0.2773724	0.9265917	-0.3760638	0.8648223	-0.50206146	0.6768299	0.736100714

According to Table 1, we can notice that the reading of the results depends heavily on the considered dataset. For example, for the dataset with $\rho = 0.4$, three variables have high values for the first dimension. Therefore, their reading is significantly reduced to Dim 1-axis. However, for the dataset with $\rho = 0.8$, the variable 'Complementarity' may be interpreted on Dim 2-axis and the remaining variables are interpreted on Dim 1-axis. The same case is for the variable 'Redundancy' in the dataset with $\rho = 0.95$.

In PCA method, the eigenvalues measure the amount of variation retained by each principal component. In our analysis, and for all ρ-values, two principal components explain 100% of the variation. Indeed, these two principal components correspond to the directions with the maximum amount of variation in the data set (see Figure 2). In the same context, the squared cosine determines the quality of representation for variables on the factor map, where a high squared cosine indicates a good representation of the variable on the principal component. In this case, the variable is positioned close to the circumference of the correlation circle, this is the case in our example (see Figure 2), which reconfirms the quality of the presentation built on three criteria: relevance, redundancy and complementarity based on preordonnnances theory.

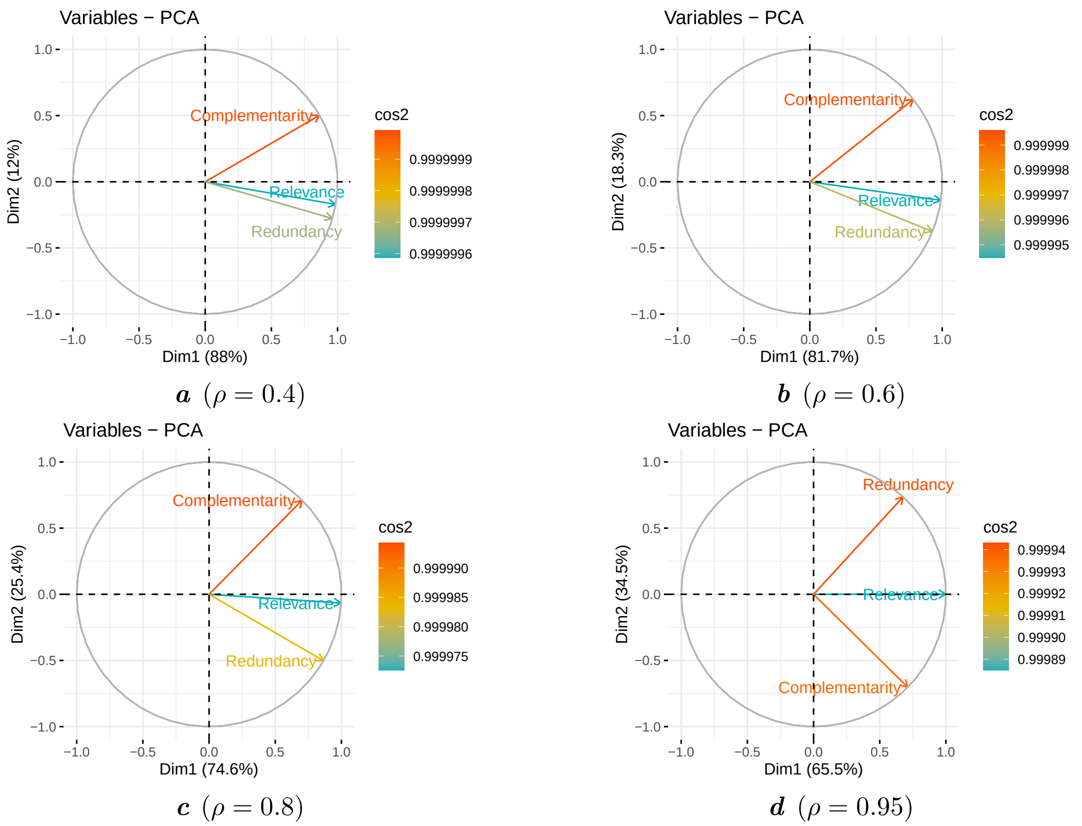
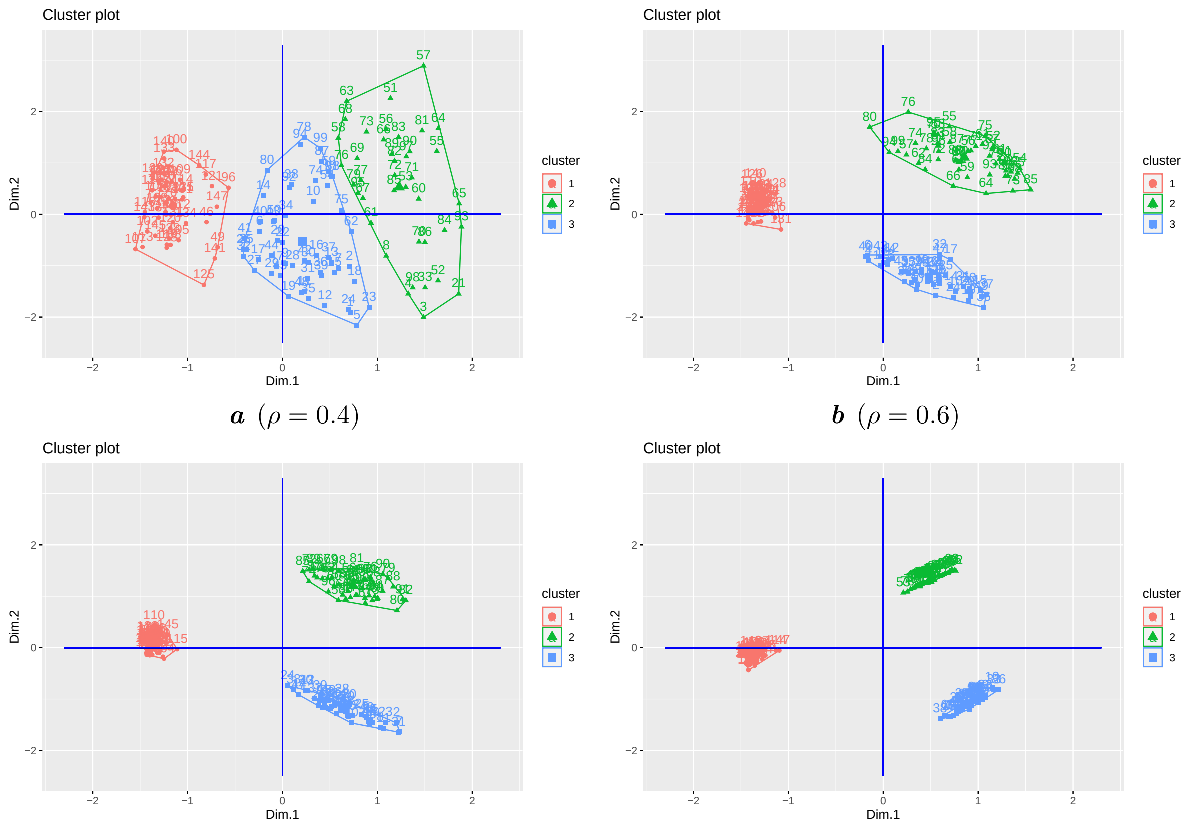


Fig. 2. Correlation circles for datasets with (***a***) $\rho = 0.4$, (***b***) $\rho = 0.6$, (***c***) $\rho = 0.8$, and (***d***) $\rho = 0.95$.

To highlight the harmony and the discriminating power of the proposed criteria, k-means method with 3 clusters ($k = 3$) is used to determine and visualize the classes in the considered datasets, using coordinates of the variables from PCA method. It can be seen from Figure 3 that, for all ρ values, the point cloud forms three disjoint classes. As mentioned previously, the reading of results depends on the considered dataset. From Table 1, for the datasets with $\rho = 0.4$ and $\rho = 0.6$, three measures namely relevance, redundancy, and complementarity can be read in Dim 1-axis. Thus, the first cluster (green-colored) is the best subset of variables that are relevant, not-redundant, and with high complementarity, followed by the blue-colored subset containing weakly relevant variables, with a low redundancy within variables and which has an acceptable complementarity between variables. While the last subset (red-colored) forms the noise. For the dataset with $\rho = 0.8$, the Dim 1-axis represents the relevance and the redundancy simultaneously and the Dim 2-axis represents the complementarity criteria. Thus, the subset in the top right (green-colored) contains the variables with high relevance, low redundancy, and high complementarity, this kind of subset may strongly contain the optimal subset of variables predicting as much as possible the class target. Therefore, it can be used as input data for heuristic methods or local searches as examples. The second subset in the bottom right (blue-colored) contains variables that have a powerful potential individually but are not homogeneous when presented as a group. The third subset (red-colored) still forms the noise since it contains irrelevant variables with high redundancy. This later has positively contributed to its complementarity (a zero complementarity). For the dataset with $\rho = 0.95$, the Dim 1-axis represents the relevance and the complementarity simultaneously and the Dim 2-axis represents the redundancy criteria. Thus, the green-colored cluster is still the best subset containing variables with high relevance, low redundancy (a high value of the $\Psi_{.}$ measure), and high complementarity. The blue-colored subset contains variables with high relevance, and high complementarity but with a considerable redundancy within variables which weakens the discriminating power of the whole. The variables in the red-colored cluster are irrelevant and cannot predict the target class when featured as a group even if they have a lower redundancy between each other. It should be mentioned that the greater the value of ρ is, the more compactness the cluster is.

According to Figure 3, we can infer that the considered criteria, by their projection on a two-dimension axis, allow finding the true classification of the explanatory variables: the green-colored cluster, using any rho value, contains the variables from 51 to 100, the blue-colored cluster contains the variables from 1 to 50, and the remaining variables are in the noisy subset as the scenario indicates. This categorization shows the potential power of the considered criteria.



a $(\rho = 0.4)$ ***b*** $(\rho = 0.6)$

c $(\rho = 0.8)$ ***d*** $(\rho = 0.95)$

Fig. 3. Visualization of clusters for datasets with (***a***) $\rho = 0.4$, (***b***) $\rho = 0.6$, (***c***) $\rho = 0.8$, and (***d***) $\rho = 0.95$.

3.2. Scenario 2

This scenario evaluates the proposed criteria to separate classes defined a priori, to determine the complementary variables and the redundant ones, and also to identify the irrelevant ones from multi-class dataset. The dataset contains 100 observations and fourteen variables $\{f_1, f_2, \ldots, f_{14}\}$ constructed as follows:

$$\begin{cases} \{f_1, f_2, \ldots, f_{10}\} \sim U[0,1], \\ f_{11} = f_1 + 0.1, \\ f_{12} = f_2 - 0.2, \\ f_{13} = 2 \times f_1, \\ f_{14} = 2 \times f_2. \end{cases} \tag{5}$$

The multi-label output Y is constructed by concatenating four binary outputs $[Y^1, Y^2, Y^3, Y^4]$ evaluated as follows:

$$Y = [Y^1, Y^2, Y^3, Y^4] \quad \text{such that:} \quad \begin{cases} Y^1 = 1 & \text{if} \quad f_1 > 0.5 \quad \text{and} \quad f_2 > 0.5, \\ Y^2 = 1 & \text{if} \quad f_1 < 0.5 \quad \text{and} \quad f_2 < 0.5, \\ Y^3 = 1 & \text{if} \quad f_1 < 0.5 \quad \text{and} \quad f_2 > 0.5, \\ Y^4 = 1 & \text{if} \quad f_1 > 0.5 \quad \text{and} \quad f_2 < 0.5, \\ Y^i = 0 & \text{otherwise} \quad (i = 1, 2, 3, 4). \end{cases} \tag{6}$$

The relevant features are $\{f_1, f_2, f_{11}, f_{12}, f_{13}, f_{14}\}$. Remaining features are irrelevant for the class variable. It should be noted that the variable f_1 is redundant to f_{11} and to f_{13}, the variable f_2 is redundant to f_{12} and to f_{14}, and variables f_1 and f_2 are complementary. The objective of studying this scenario is to illustrate the potential power of the considered criteria to, on the one hand, separate classes into relevant or noisy and, on the other hand, to identify the complementary and redundant variables. The PCA correlation circle and the cluster plot are summarized in Figure 4. In Figure 4*b*, two poles are shown: the pole of the relevant variables, namely f_1, f_2, f_{11}, f_{12}, f_{13}, f_{14} (red-colored), where the variable f_1, f_{11} and f_{13} are redundant (similarly for f_2, f_{12} and f_{14}), and the variables f_1 and f_2 are complementary. The second pole is that of variables that do not meet any criteria (blue-colored). Thus, these variables form noise. We can infer, from the correlation circle in Figure 4, that two principal components explain 100% of the variation and that all criteria can be read in Dim 1-axis in this case. Then, according to Figure 4*b*, we can notice that three criteria, namely relevance, redundancy and complementarity based on preordonnaces theory and on PCA method, are able to separate classes properly, relevant variables (red-colored) from the noise (blue-colored), to show the complementarity of f_1 to f_2, and to merge confused variables (redundant ones) as the scenario indicates, and this shows the importance of combining these three criteria to predict a target class.

a $(\rho = 0.4)$ ***b***

Fig. 4. (***a***) PCA correlation circle, (***b***) Cluster plot according to $Dim.1$ and $Dim.2$.

4. Conclusion

In this paper, a statistical analysis of three new criteria of relevance, redundancy, and complementarity was established. To summarize and visualize information for a given data for a more in-depth selection and to eliminate the noise variables in a discriminant model, the PCA method with 2 principal components was used. On this basis, the considered categorization was able to divide the correlation circle into classes. This categorization resulted in the selection of the optimal or the suboptimal subset of explanatory variables predicting the target class and also in the determination of the noisy variables. Both scenarios have shown the ability of the considered criteria to select the relevant variables, the redundant ones, and to determine the most homogeneous subset of variables predicting the target jointly.

[1] Chah Slaoui S., Chamlal H. Nouvelles approches pour la sélection de variables discriminantes. Revue de statistique appliquée. **48** (4), 59–82 (2000).

[2] Chamlal H., Ouaderhman T., Aaboub F. A graph based preordonnances theoretic supervised feature selection in high dimensional data. Knowledge-Based Systems. **257**, 109899 (2022).

[3] Chamlal H., Ouaderhman T., El Mourtji B. Feature selection in high dimensional data: A specific preordonnances-based memetic algorithm. Knowledge-Based Systems. **266**, 110420 (2023).

[4] Chamlal H., Ouaderhman T., El Mourtji B. Multicriteria approaches based on a new discrimination criterions for feature selection. In: 2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS). 1–7 (2021).

[5] Chamlal H., Ouaderhman T., Rebbah F. E. A hybrid feature selection approach for Microarray datasets using graph theoretic-based method. Information Sciences. **615**, 449–474 (2022).

[6] Chen Z., Chen Q., Zhang Y., Zhou L., Jiang J., Wu C., Huang Z. Clustering-based feature subset selection with analysis on the redundancy–complementarity dimension. Computer Communications. **168**, 65–74 (2021).

[7] Chen Z., Wu C., Zhang Y., Huang Z., Bin R., Ming Z., Nengchao L. Feature selection with redundancy-complementariness dispersion. Knowledge-Based Systems. **89**, 203–217 (2015).

[8] Ferreira A. J., Figueiredo M. A. T. Efficient feature selection filters for high-dimensional data. Pattern Recognition Letters. **33** (13), 1794–1804 (2012).

[9] John G. H., Kohavi R., Pfleger K. Irrelevant Features and the Subset Selection Problem. In: Machine Learning Proceedings 1994. 121–129 (1994).

[10] Kurita T. Principal Component Analysis (PCA). In: Computer Vision: A Reference Guide. 1–4 (2019).

[11] Radovic M., Ghalwash M., Filipovic N., Obradovic Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC Bioinformatics. **18** (1), 9 (2017).

[12] Singha S., Shenoy P. P. An adaptive heuristic for feature selection based on complementarity. Machine Learning. **107** (12), 2027–2071 (2018).

[13] Souza F., Premebida C., Araújo R. High-order conditional mutual information maximization for dealing with high-order dependencies in feature selection. Pattern Recognition. **131**, 108895 (2022).

[14] Zhou H., Zhang Y., Zhang Y., Liu H. Feature selection based on conditional mutual information: minimum conditional relevance and minimum conditional redundancy. Applied Intelligence. **49** (3), 883–896 (2019).

Статистичний аналіз трьох нових мір релевантності, надмірності та комплементарності

Ель Мурджі Б., Чамлал Х., Уадерман Т.

Кафедра математики та інформатики,
лабораторія фундаментальної та прикладної математики,
факультет наук Айн Чок, Університет Хасана II Касабланки, Марокко

Дискримінантний аналіз є частиною статистичного навчання; його мета полягає в тому, щоб розділити класи, визначені апріорі в популяції, і передбачає прогнозування класу заданих точок даних. Дискримінантний аналіз застосовується в різних областях, таких як розпізнавання образів, мікрочипи ДНК тощо. В останні роки проблема дискримінації залишається складною задачею, якій приділяється все більше уваги, особливо для масивів даних великої вимірності. Дійсно, у такому разі необхідний вибір ознак, що передбачає використання критеріїв релевантності, надмірності та комплементарності пояснювальних змінних. Метою цієї статті є представити аналіз трьох нових критеріїв, запропонованих у цьому сенсі, точніше, на основі аналізу основних компонентів вдалося досягнути подвійної мети: вивчити гармонію цих трьох критеріїв, а також візуалізувати клас змінних–кандидатів для більш поглибленого вибору на додаток до усунення шумових змінних у дискримінантній моделі.

Ключові слова: *актуальність; надмірність; компліментарність; теорія передпорядків; дискримінантний аналіз; аналіз головних компонент.*