International Academic Institute for Science and Technology

International Academic Journal of Humanities
Vol. 9, No. 1, 2022, pp. 12-23.

ISSN 2454-2245

International Academic Journal of Humanities

www.iaiest.com

Comparing Some Methods of Kernel Discriminant Analysis Under Fuzzy Environment

Sackineh Shamil Jasim [a]

[a] *Department of Statistics, College of Administration and Economics, University of Karbala, Iraq.*
E-mail: sackineh.sh@uokerbala.edu.iq

Abstract

One of the most widely used statistical methods in the field of data analysis is the method of discriminant analysis, and the use of discriminant analysis requires the availability of a number of assumptions, the most important of which is that the data of explanatory variables have a multivariate normal distribution. The discriminant analysis often suffers from the problem of small sample size, especially when it is applied to classify several dimensions; we deal with this problem through the use of the Kernel Discriminant Analysis method (KDAM). But in many situations in real world suffuref from inaccuracy in measurements that is leads to inaccuracy in conducting the required statistical analyzes, so it is necessary to get out of the traditional pattern of data to that fuzzy pattern by finding a fuzzy information system based on the principle of cutting in the fuzzy group to find measurements of data that represent the phenomenon with certainty and which represent the group of important elements that represent those phenomenon. In this paper, we form the estimation function of the kernel density with an application in discriminant analysis to be a comparison in estimation density function depends on the choice of bandwidth, which controls the smoothing of the estimate and on the choice of the kernel function., some methods were used to estimate the bandwidth parameter, which is least square Cross validation, Biased cross validation and Bayes discriminant rule Method under fuzzy measurements. The estimation of Kernel density is important in multivariate data analysis, and it depends heavily on the appropriate selection of the bandwidth matrix (H) that reduces the mean integration square error (MISE) and the error rate ($\widehat{MR}$) through the common rule of sets which is KDR kernel discrimination. Finally, it was based on real data from the disease (general leukemia) and the 6 factors (variables) that affect the disease. We note that (KDA-Bay) method is better than (KDA-SCV) and (KDA-BCV) because it gives the lowest classification error rate from the results, and we note that all methods under fuzzy data better than the methods under traditional data.

Keywords: Kernel Function, Discriminant Analysis, Least Square Cross Validation, Biased Cross Validation, Bayes Discriminant Rule, Fuzzy Logic, Fuzzy Set, Alpha-cut.

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

Introduction

In many situations of daily life, and when we start statistical analysis, we encounter data that we suffer from inaccurate measurements, either because of observational or measurement errors or the lack of appropriate conditions for collecting such data, such as failure times of machines, equipment, electrical devices, medical, engineering and economic data. These data are often accompanied by values that may be far from the original data format and constitute solid tail data. When estimating these phenomena, we suffer from inaccuracy in their measurements, as well as the presence of anomalies in them, using the usual estimation method, which leads to misleading, inaccurate and unrepresentative estimates of the reality of the phenomenon being studied. Therefore, it is necessary to find appropriate and robust methods of estimation under these circumstances to ensure the accuracy of the obtained estimations.

The method of multivariate statistical analysis depends in its various ways on the description and analysis of phenomena with multiple dimensions and variables. Discriminant analysis is one of these statistical methods in data analysis, the discriminant analysis will be between two or more groups that are similar in many characteristics on the basis of several variables from Through the use of the characteristic function, which is a linear structure of the independent variables. It aims to form a linear or non-linear formula between the study variables to classify a single item into a specific group and based on some characteristics and measurements taken for the studied phenomenon variables.

The two parametric methods of linear and quadratic discriminant analysis are the most common and are characterized by the ease of their calculation. There is linear discrimination in the case of two groups and non-linear discrimination in the case of more than two groups. Non-linear discrimination is used in the case of unequal variances. Kernel Discriminant Estimation (KDE) estimations belong to a class of estimations called nonparametric density function estimations, which are dealing with data at least for one variable without any assumption of a distribution.

Kernel function methods are one of the important techniques in statistics and it is widely applied to various fields of data analysis, thanks to its flexibility and good performance. The estimation of core density is an important method in the analysis of multi-dimensional and variable data, as its use in discriminant analysis largely depends on the appropriate selection of the beam width matrix (H) that reduces the mean integration square error (MISE) and classification error rate ($\widehat{MR}$). The selection of smoothing parameters is a problem of fundamental importance in estimating core density. Choosing the Univariate beam width parameter for estimating core density (KDE) is the simplest form of this problem. If the selection of the Univariate beam width parameter was in progress and the multivariate case was much slower, however, the selection of beam width matrices is an important problem because of the usefulness of multivariate core density estimates in several areas, including (data visualization, nonparametric discriminant analysis, and good fit test).

Parametric and Nonparametric Discriminant Analysis

There are two parametric methods commonly used in discriminant analysis. The linear discrimination function is based on a linear structure of the variables, and in the event that the data do not meet the conditions of linear discrimination, i.e. the inequality of the covariance matrix and the covariance of the sets, the quadratic discrimination function will be used and according to the discrimination rule.

$$x \text{ is allocated to group } j_0 \text{ if } j_0 = \underset{j \in \{1,\ldots,v\}}{argmax}\ \pi_j f_j(x) \quad (1)$$

The linear discriminant function will be:

$$x \text{ is allocated to group } j_0 \text{ if } j_0 = \underset{j \in \{1,\ldots,v\}}{argmax}\ log(\pi_j) - \frac{1}{2}(x-\mu_j)^T \sum^{-1} (x-\mu_j) \quad (2)$$

As for the nonlinear discriminant analysis, it will be as follows:

$$x \text{ is allocated to group } j_0 \text{ if } j_0 = \underset{j \in \{1,\ldots,v\}}{argmax}\ log(\pi_j) - \frac{1}{2}log|\sum_j| - \frac{1}{2}(x-\mu_j)^T \sum_j^{-1}(x-\mu_j \quad (4)$$

And in the case of using non-normal or non-linear data, the density function can be estimated directly from the data, and this direct method is known as the kernel estimator. As the core estimation depends on the density of the population to obtain a more common rule for allocation which is the Bayesian Discrimination Rule (BDR).

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

The core discriminant rule (KDR) is obtained from the Bayesian discriminant rule by replacing the probability density function with the estimation of the core density function $\hat{f}_J$ (x ; H_j) and as follows:

$G_i = \pi_i f_i(\underline{x})$ is Bayesian discriminant rule

KDR: $d_j(x) = \text{argmax}\ \hat{\pi}_j \hat{f}_j(x\ ;\ H_J)$ is kernel discriminant rule

Where $j \in \{1,\ldots,v\}$

j = 1,2 is the number of groups used

$\hat{\pi}$ is the previous possibilities

$\hat{f}_j(x\ ;\ H_j)$ is probability density function

For the purpose of clarifying the concept of Kernel Discriminant Analysis, it is necessary to clarify some of the following basic concepts.

Kernel Estimator

It defines the kernel estimator, which depends on the weight function used in estimating the nonparametric function, and it gives weights to the adjacent data points in the estimation procedure. On the practical side, the kernel functions are continuous, finite, and symmetric around the zero point, and their value is real and its integral is equal to one.

1. K(x) = K(-x) (The function is continuous and symmetric around zero)
2. $\int_R K(x)\,dx = 1$
3. $\int_{-\infty}^{\infty} x\,K(x)\,dx = 0$
4. $\int_{-\infty}^{\infty} x^2 K(x)\,dx = K > 0$ for some constant k.
5. $K(x) \geq 0$ for all $x \geq 0$

This means that the function is of the second degree and used in applications. One of the characteristics of the core estimator is the simplicity and ease of programming it on the computer and drawing its capabilities graphically.

This estimator is characterized as a smooth estimator and works to modify the observations and reduce the spread of them.

It also does not depend on difficult assumptions, which leads as a result to give a clear and pure picture of the function.

The most commonly used cortical functions are the caustic function with mean (0) and variance (1). And the researcher (Tarn Duong, in 2004) explained, "The reason for choosing the (Gaussian) function is due to the presence of many desirable properties, which makes choosing the bandwidth easier, and these characteristics are not found for other (kernel) functions in the R package, For example, packages (sm, np, evmix kern smooth,) that cannot perform core discriminant analysis but only estimate core density.

Classification of Data

The number of correct classifications and the misclassifications rate is calculated. It is called the classification error rate. It is also called the apparent error rate. The results can be displayed in the classification table or the confusion matrix.

Table No. (1) shows the classification process into two groups:- [9]

Table 1: The Classification Results for Two Groups

Actual Group	Number of observation	Predicted Group	
		1	2
1	n_1	n_{11}	n_{12}
2	n_2	n_{21}	n_{22}

As (n1) represents the number of observations in the first group (G1) and (n11) represents the correct classification in the first group (G1) and (n12) represents the classification error in the first group (G1), and that (n1 = n11 + n12 and The same for (n2) represents the number of observations in the second group (G2) and (n21) represents the

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

classification error in the second group (G2) and (n22) represents the correct classification in the second group (G2) and that,

$$(n_2 = n_{21} + n_{22})$$

$$\text{Misclassification Rate} = \frac{n_{12}+n_{21}}{n_{11}+n_{12}+n_{21}+n_{22}} \quad (5)$$

And,

$$\text{Apparent Correct Classification} = \frac{n_{11}+n_{22}}{n_1+n_2} \quad (6)$$

The apparent error rate (rated error rate (MR)ˆ) can be calculated as follows:

MR = 1- (obvious correct rating rate).

Relying on Table No. (1), the classification error rate (MR) is obtained, and using the previous probabilities ((π_i)ˆ) and the estimation functions of the core density (KDE) for two populations ((.; H_2 fˆ and ((. ; H_1 fˆ The Optimal Classification Rate is to reduce the classification error rate, in the following manner:

Set the view (i) to the first group if $\widehat{\pi_1}\,\hat{f}(.\,;H_1) > \widehat{\pi_2}\,\hat{f}(.\,;H_2)$.

And the other cases assign observation (i) to the second group (G2).

There are two important points to consider in the case of discriminant analysis using KDE estimates:

1. Choosing the bandwidth (H) should depend on certain (specific) views to classify it in addition to its dependence on the density of the community.
2. In the problem of distinguishing multiple layers (ie, several variables), instead of using a single (constant) bandwidth for all density estimates for the community, it is useful to use a variable bandwidth to estimate the classified density.

Misclassification Rate

The two researchers (Hall and Wand) (1988) mentioned in an experiment with two-layer problems that finding the best bandwidth directly is by using the classification error rate or the criterion $\widehat{MR}$ and its formula is as follows:
$\widehat{MR}$ = 1- $n^{-1}\sum_{j=1}^{m} 1\,\{Y_j$; KDR Is the correct classification using}

n is sample size,

d the number of dimensions of the variables

$$Xi = \begin{pmatrix} xi1 \\ . \\ . \\ . \\ xid \end{pmatrix} \quad i= 1, \ldots, n$$

The estimation of the density f(x) from a random sample $X= (x_1, \ldots, x_d)^T$ as follows:

$$f(x) = f\,(x_1, \ldots, x_d)$$

The general form of the multivariate density estimator with the (H) diagonal bandwidth matrix is as follows:

$$\hat{f}\,(x\,;H) = n^{-1}\sum_{i=1}^{n} K_H\,(x - X_i) \quad (7)$$

Where $X= (x_1, \ldots, x_d)^T$ Is random size with P dimensions
d Dimensions are the number of variables used in the research.

$$Xi = (X_{i1}, X_{i2}, \ldots, X_{id})^T ; i = 1,2, \ldots,n \quad (8)$$

H is represents a symmetric, positively defined (d*d) beam width matrix parameter of the dimensions according to the selection of the influencing variables.

The Kernel function is as follows:

$$K_H\,(x) = |H|^{\frac{-1}{2}} K\,(H^{-1/2}\,x) \quad (9)$$

The function used to find the estimate of kernel density is Gaussian function

$$K\,(x) = (2\pi)^{-d/2} \exp\,(-\frac{1}{2}x^T x) \quad (10)$$

The kernel function often takes from the variables the probability density function, and there are two common methods

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

$$1 - K^P(X) = \prod_{i=1}^{d} k(xi) \text{ and } 2)\ K^S(X) = c_{x,d}\, k\{ (X^T X)^{1/2}$$

Where:

$$2 - C_{k,d}^{-1} = \int k\{ (X^T X)^{1/2}\}\, dx$$

The first is called (knockout the kernel) and the second is radially analogue. The most common choice for K is (Standard d_variate) for normal density.

But when we have a density estimate for a multivariate with a fixed bandwidth (one value), its formula is as follows:

$$\hat{f}_{\mathrm{h}}(\mathrm{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d}\, \mathrm{k}\left(\frac{x - X_i}{h}\right) (11)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d}\, \mathrm{k}\left(\frac{x_1 - X_{i1}}{h}, \ldots, \frac{x_d - X_{id}}{h}\right) (12)$$

The value of bandwidth (h) in all vehicles is a constant value.

The quality of the core estimation does not depend largely on the shape of the function (K), but mostly depends on the value of the bandwidth for it.

Kernel Discriminant Analysis

The use of core density estimates in discriminant analysis is well known by researchers to identify statistical patterns [12]. Kernel discriminant analysis (KDA) is an effective approach for non-linear dimensions under special conditions, as probabilistic methods can be used with the (KDA) method. In order to improve it.

Smoothed Cross – Validation (SCV)

The Univariate smooth crossover (SCV) method was introduced by the researchers (Hall, Marron and Park (1992)) with the best experimental smoother g independent of h (obtained by the program). Then, the SCV smoother for multivariate was presented by researchers (Sain, Baggerly and Scott (1994). He worked with it initially from a slightly modified formula from the LSCV method in equation (40-2) known as the Leave-in-diagonal formula to get on samples of data that do not have repeating values.

$$\mathrm{Lscv}(\mathrm{H}) = \mathrm{n}^{-1}(4\pi)^{-\mathrm{d}/2}|\mathrm{H}|^{-1/2} + \mathrm{n}^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} (K_{2\mathrm{H}} - 2K_{\mathrm{H}} + \mathrm{K}_0)(\mathrm{X_i} - \mathrm{X_j}) (13)$$

Where $K_0 \rightarrow$ Dirac delta function

To form (SCV) before smoothing the data differences (Xi - Xj) by K2G, ie replacing (Xi - Xj) with convolution with K2G (Xi - Xj):

$$\mathrm{SCV}(\mathrm{H}) = \mathrm{n}^{-1}(4\pi)^{-\mathrm{d}/2}|\mathrm{H}|^{-1/2} + \mathrm{n}^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} (K_{2\mathrm{H}+2\mathrm{G}} - 2K_{\mathrm{H}+2\mathrm{G}} + \mathrm{K}_{2\mathrm{G}})(\mathrm{X_i} - \mathrm{X_j}) (14)$$

G: represents the trial bandwidth matrix

The algorithm of this method is also the same as in the previous method, except for step (1), which uses the bandwidth matrix (H), which was obtained by the method (KDA _ SCV).

Biased Cross Validation (BCV)

The LSCV method is based on the MISE estimate, while the bias method is based on the AMISE method.

AMISE $\hat{f}(\cdot; \mathbf{H}) = n\text{-}1R(K)|\mathbf{H}|\text{-}1/2 + 1\ 4\mu 2(K)2(\mathrm{vech}T\ \mathbf{H})\Psi 4(\mathrm{vech}\ \mathbf{H})$

- **BCV1(H)** = n−1R(K)|H|−1/2 + 14μ2(K)2(vechT H)Ψˇ 4(vech H)

The function BCV2 with Ψˇ4 formula

- **BCV2(H)** = $n\text{-}1R(K)|\mathbf{H}|\text{-}1/2 + 1\ 4\mu 2(K)2(\mathrm{vech}T\ \mathbf{H})\Psi\check{}\ 4(\mathrm{vech}\ \mathbf{H})$

The Bayes Discriminant Rule (BDR)

We have v populations, each associated with a density fj and a prior probability π_j, j = 1; 2; … ;. We wish to allocate a point x in the sample space to one (and only one) of these populations. A common allocation rule is the Bayes discriminant rule Allocate x to group.

$$\mathrm{j}_0 \text{ where } \mathrm{j}_0 = \mathrm{argmax}\ \pi\mathrm{j}_{\ j\in\{1,2,\ldots\ldots v\}}\ \pi_{\mathrm{j}}\ \mathrm{fj(x)}: (15)$$

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

From each fj, we have a random sample Xj1;Xj2; : : : ;Xj;nj. These Xji are collectively known as the training data. In the cases where we wish only to classify a data sample, rather than the entire sample space, the random sample Y 1;Y 2; : : : ;Y m drawn from $\sum_{j=1}^{v} \pi j\ fj(x)$ is known as the test data.

The kernel discriminant rule (KDR) is obtained from the Bayes discriminant rule by replacing fj by its kernel density estimate

$$\hat{f}\ (\ x\ ;\ H) = n^{-1} \sum_{i=1}^{n} \ K_H\ (x - X_i\ )\ (16)$$

and πj is usually replaced by the sample proportion nj/n where n =$\sum_{j=1}^{v} nij$ that is:

KDR: Allocate x to group j_0 where

$$j_0 = \text{argmax}\ \pi j\ _{j \in \{1,2,\ldots\ldots v\}}\ \hat{\pi} j\ \hat{f}j(x): (17)$$

Crisp and Fuzzy Set

Let Ψ is Universal set, A subset from it, then each element in A may be belonging or not belonging to A. (H. Garg et al, 2013) (A. Ibrahim, A. Mohammed, 2017).

Let $\mu_A(x)$ is a characteristic function for A give the membership in Ψ to A, it is a binary function, {0, 1}, where,

$$\mu_A(x) = \begin{cases} 1, if\ x\ \in\ A \\ 0, if\ x\ \notin\ A \end{cases} (18)$$

If $\mu_A(x)$ = 1, then the element x has full belonging to the set A. If $\mu_A(x)$= 0, then the element x does not belong to the set A.

As for the fuzzy set, it is a set of ambiguous boundaries, each element in the fuzzy set has a certain degree of membership, and the fuzzy set is characterized by a membership function that assigns each element in the set a degree of membership in the interval [0, 1]. In which the element or object is allowed to belong partly. (Pak, 2017).

Let Ψ is Universal set, a fuzzy subset $\widetilde{A}$ from it that distinguished with the membership function $\mu_{\widetilde{A}}(x)$ which produce values in the interval [0, 1] for each values in the fuzzy sample space, then the fuzzy set is, (Danyaro & et al., 2010).

$$\widetilde{A} = \{(x_i, \mu_{\widetilde{A}}(x_i)), x \in\ \Psi, i = 1,2,3, \ldots\ldots n, 0 < \mu_{\widetilde{A}}(x) < 1\}\ (19)$$

Membership Functions

The membership function is one of the basic and important functions in fuzzy set theory, which is used to generate the affiliation of elements within the fuzzy set. (Abboudi & et al, 2020). In other words, it is the function that draws the degree of importance of the element (the degree of belonging) in the comprehensive group to the fuzzy group, and it is a function with a positive value (Rutkowski, 2004).

The affiliation functions are represented by a y-axis diagram representing the degree of affiliation to the group and the x-axis represents the normal values of the fuzzy variable, and the basic condition for these functions is that their range is between zero and one, when the degree of affiliation 1 means that the value is definitely It belongs to the group and the degree of affiliation is zero, which means that the value does not belong to the group, and the value between the two values (1,0) defines the degrees of affiliation of the elements in the group. The affiliation function in the fuzzy group is a generalization of the characteristic function of the traditional group, and depending on the type of affiliation function, different types of fuzzy groups are obtained (H. Garg et al, 2013).

Alpha-Cut

The principle of cutting in the fuzzy group was first introduced by Zadeha (Zadeha, 1971), and α is defined as the lowest degree of belonging to any element in the fuzzy group A˜ and its value falls within the period [0 1]. (H. Garg et al, 2013, 398), which represents the degree of affiliation of the important elements, since the important affiliation is limited to two values ((a_(1,) a_m) on the fulcrum of the fuzzy group (Support A˜), and except for those values, it is of little importance and outside the scope of work (cut-out).

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

α-cut Set

It defines the alpha segments of the fuzzy set $\tilde{A}$, which is the traditional set whose elements belong to the fuzzy set $\tilde{A}$ and which has a degree of affiliation greater than or equal to α, and mathematically it can be written as follows: (H. Garg et al, 2013).

$$\tilde{A}\alpha=\{ \tilde{x}\in X;\ \mu_{\tilde{A}}(x)\geq\alpha\} \quad (20)$$

So $0 < \alpha < 1$, X the Universal set)

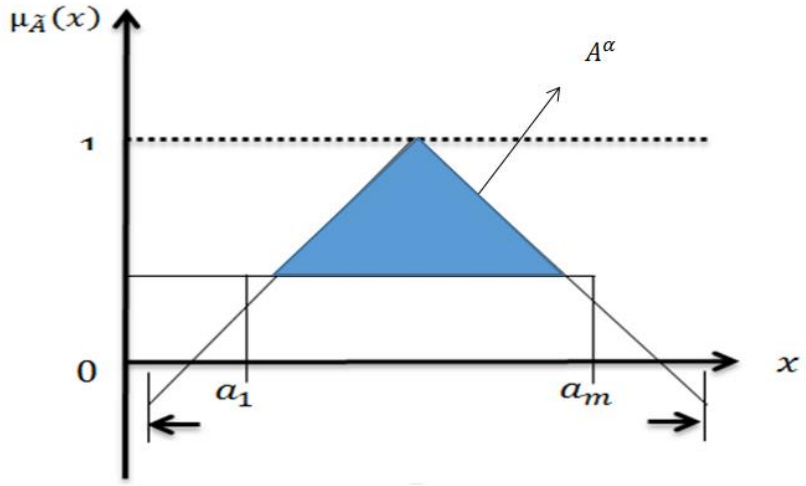


Figure 1: Fuzzy set and alpha cut

Applied Side

In this application, real data was used to compare the methods of estimating the preliminary parameters:

The method of data collection was by relying on laboratory records of patients (infected and uninfected), which included a study on the disease (general leukemia) at Al-Huusien Hospital in Karbala Governorate. Two groups as follows:-

1. **The First Group:** It included people who were not infected with the disease, with a size of (50) views, and we symbolized them with the symbol (1).
2. **The Second Group:** It included people with the disease with a size of (50) views, and we symbolized them with the symbol (2).

By conducting a simple laboratory analysis called after complete blood cells (CBC), which is a common medical examination that is used periodically in medical examinations, and it is important in that the health care provider can identify the number of the main blood components, red and white cells, hemoglobin and others, as a deficiency or increase Any of these components may be Accompanied by a disease, and thus, through this analysis, the most important components of blood will be obtained. Most of the medical studies that dealt with general leukemia mentioned several factors that affect the incidence of the disease, as shown below and the measurements that are included from them, which are as follows:-

Y:- It is an indicator variable that indicates the presence of a disease or not, as we symbolize (1) if the person is not infected, and we symbolize (2) if the person has the disease.

As for the explanatory variables, they are: - (Natural ratios)

X1:- represents the sex (Sex) where (1) for males and (2) for females.

X2 :- represents WBC (White Blood Cells). 4.00 – 11.0))

X3 :- represents the red blood cells RBC (Red Blood Cells). (3.90 - 6.50)

X4 :- represents the percentage of hemoglobin in the blood (HGB). (11.5 - 17.5)

X5 :- represents the percentage of hematocrit (HCT) (36.0 - 50.0)

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

X6 :- represents the percentage of PLT (Blood Platelets). (150.0 - 450)

Also, $\widehat{MR}$ represents the classification error ratio for the group, where $\widehat{MR}$ 1 represents the classification error rate for the first group, and $\widehat{MR}$ 2 represents the classification error rate for the second group.

Table 2: Represents the Classification Error Ratio ($\widehat{MR}$) for the Real Data and for the Three Methods Under Traditional Data

KDA — BCV			$\widehat{MR}$	
Estimate / True	Estimate 1	Estimate 2	$\widehat{MR}$ 1	$\widehat{MR}$ 2
True 1	43	7	0.16	0.26
True 2	10	40	0.20	
KDA — scv			$\widehat{MR}$	
Estimate / True	Estimate 1	Estimate 2	$\widehat{MR}$ 1	$\widehat{MR}$ 2
True 1	35	15	0.3	0.58
True 2	29	21	0.44	
KDA — Bay			$\widehat{MR}$	
Estimate / True	Estimate 1	Estimate 2	$\widehat{MR}$ 1	$\widehat{MR}$ 2
True 1	40	6	0.14	0.21
True 2	9	38	0.18	

Table (2) shows the classification error rate ($\widehat{MR}$) under traditional data that the value between true 1 and estimate 1 is the correct value and the value between True 1 and Estimate 2 is the false value. Also, the value between true 2 and estimate 1 is the false value. A value between true 2 and estimate 2 is the correct value.

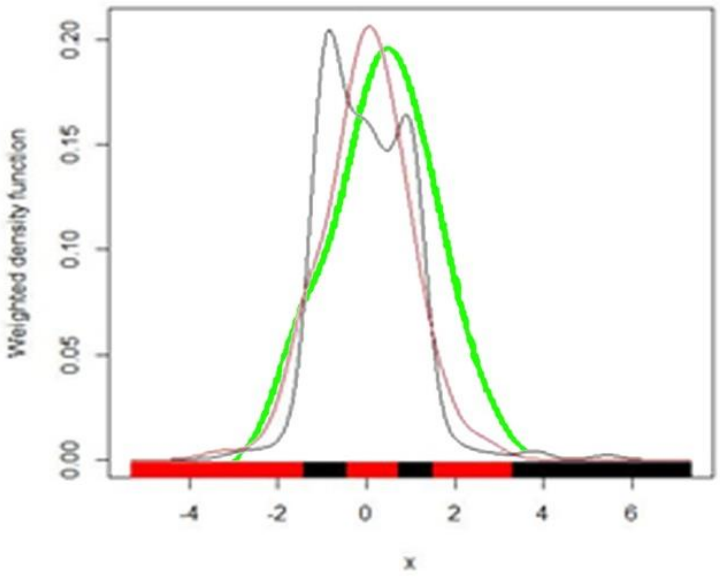


Figure 2: Represents the Interaction between the Two Groups of Patients and Non-patients for the Traditional Real Data

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

The data was plotted by multiplying the kernel function with each variable and then taking an average in order to get its value and plot it. The figure shows the overlap between the two groups and the rate of classification error.

Table (2) shows the classification error rate ($\widehat{MR}$) and the estimated values for the number of valid and incorrect observations for each group under traditional data. The classification error rate is calculated for each group and for all the methods used for it, and the classification points for each group are as follows:

classification.point. scv group1

0.1012743

classification.point. scv group2

0.1609851

classification.point.Bcv group1

0.1114534

classification.point.Bcv group2

0.1722548

classification.point.Bay group1

0.1003422

classification.point.Bay group2

0.1538977

The groups are classified with the values obtained from the equation (KDR) and the bandwidth matrix for the three methods under traditional data is as follows:

$$\text{H Scv} = \begin{bmatrix} 0.3199 & -0.0407 & -0.0014 & -0.0159 & -0.0060 & -0.014 \\ -0.0407 & 0.3837 & 0.0932 & 0.0698 & 0.0774 & 0.1357 \\ -0.0014 & 0.0932 & 0.4886 & 0.3848 & 0.4212 & 0.1110 \\ -0.0159 & 0.0698 & 0.3848 & 0.5140 & 0.4942 & -0.0199 \\ -0.0060 & 0.0774 & 0.4212 & 0.4942 & 0.5119 & 0.0156 \\ -0.0141 & 0.1357 & 0.1110 & -0.0199 & 0.0156 & 0.4606 \end{bmatrix}$$

$$\text{H Bcv} = \begin{bmatrix} 0.4287 & -0.0512 & -0.0122 & -0.0256 & -0.0078 & -0.025 \\ -0.0512 & 0.3975 & 0.0831 & 0.0677 & 0.0879 & 0.2254 \\ -0.0122 & 0.0932 & 0.5689 & 0.4433 & 0.5655 & 0.1455 \\ -0.0256 & 0.0677 & 0.4433 & 0.6241 & 0.5672 & -0.0277 \\ -0.0078 & 0.0879 & 0.5655 & 0.5672 & 0.6231 & 0.0254 \\ -0.0251 & 0.2254 & 0.1455 & -0.0277 & 0.0254 & 0.6681 \end{bmatrix}$$

$$\text{H Bay} = \begin{bmatrix} 0.1123 & -0.0201 & -0.0012 & -0.0133 & -0.0050 & -0.0111 \\ -0.0201 & 0.1433 & 0.0811 & 0.0596 & 0.0571 & 0.1254 \\ -0.0012 & 0.0811 & 0.2877 & 0.3349 & 0.3232 & 0.1004 \\ -0.0133 & 0.0596 & 0.3349 & 0.4111 & 0.3553 & -0.0186 \\ -0.0050 & 0.0571 & 0.3232 & 0.3553 & 0.1112 & 0.0146 \\ -0.0111 & 0.1254 & 0.1004 & -0.0186 & 0.0146 & 0.2211 \end{bmatrix}$$

The previous results represent the results of the kernel discriminant analysis under traditional real data. Assuming that the data suffers from inaccuracy in its measurements, i.e. suffers from uncertainty, we will use the principle of alpha-cut in the fuzzy set, considering the selection of the elements that belong to each set and using a trigonometric membership function that has a degree of significance (cut off) α=0.90 Therefore, the results of the Bayesian discriminant analysis will be as follows:

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

Table 3: Represents the Classification Error Ratio ($\widehat{MR}$) for the Real Data and for the Three Methods Under Fuzzy Data

KDA — BCV			$\widehat{MR}$	
Estimate / True	Estimate 1	Estimate 2	$\widehat{MR}$ 1	$\widehat{MR}$ 2
True 1	40	5	0.14	0.22
True 2	3	42	0.11	
KDA — scv			$\widehat{MR}$	
Estimate / True	Estimate 1	Estimate 2	$\widehat{MR}$ 1	$\widehat{MR}$ 2
True 1	31	14	0.21	0.41
True 2	25	20	0.31	
KDA — Bay			$\widehat{MR}$	
Estimate / True	Estimate 1	Estimate 2	$\widehat{MR}$ 1	$\widehat{MR}$ 2
True 1	40	5	0.11	0.19
True 2	5	40	0.10	

Table (2) shows the classification error rate ($\widehat{MR}$) under fuzzy data that the value between true 1 and estimate 1 is the correct value and the value between True 1 and estimate 2 is the false value. Also, the value between true 2 and estimate 1 is the false value. A value between true 2 and estimate 2 is the correct value.

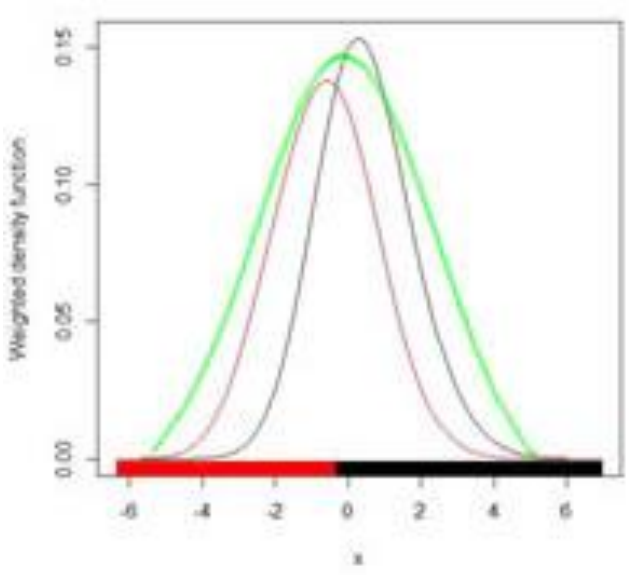


Figure 3: Represents the Interaction between the Two Groups of Patients and Non-patients for the Fuzzy Real Data

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

The data was plotted by multiplying the kernel function with each variable and then taking an average in order to get its value and plot it. The figure shows the overlap between the two groups and the rate of classification error.

Table (2) shows the classification error rate ($\widehat{MR}$) and the estimated values for the number of valid and incorrect observations for each group under traditional data. The classification error rate is calculated for each group and for all the methods used for it, and the classification points for each group are as follows:

classification.point. scv group1
0.1011113
classification.point. scv group2
0.1301333
classification.point.Bcv group1
0.1211321
classification.point.Bcv group2
0.1134511
classification.point.Bay group1
0.111001
classification.point.Bay group2
0.110075

The groups are classified with the values obtained from the equation (KDR) and the bandwidth matrix for the three methods under fuzzy data is as follows:

$$\text{H Scv} = \begin{bmatrix} 0.3199 & -0.0101 & -0.0012 & -0.0122 & -0.0030 & -0.0131 \\ -0.0101 & 0.3422 & 0.0531 & 0.0444 & 0.0711 & 0.1311 \\ -0.0012 & 0.0531 & 0.4886 & 0.3141 & 0.4111 & 0.1100 \\ -0.0122 & 0.0444 & 0.3141 & 0.5111 & 0.4913 & -0.0144 \\ -0.0030 & 0.0711 & 0.4111 & 0.4913 & 0.5111 & 0.0121 \\ -0.0131 & 0.1311 & 0.1100 & -0.0144 & 0.0121 & 0.4133 \end{bmatrix}$$

$$\text{H Bcv} = \begin{bmatrix} 0.3111 & -0.0100 & -0.0011 & -0.0121 & -0.0020 & -0.0121 \\ -0.0100 & 0.2421 & 0.0431 & 0.0244 & 0.0511 & 0.1211 \\ -0.0011 & 0.0431 & 0.3881 & 0.2141 & 0.3111 & 0.1101 \\ -0.0121 & 0.0244 & 0.2141 & 0.3111 & 0.4913 & -0.0141 \\ -0.0020 & 0.0511 & 0.3111 & 0.4913 & 0.4111 & 0.0111 \\ -0.0121 & 0.1211 & 0.1101 & -0.0141 & 0.0111 & 0.3132 \end{bmatrix}$$

$$\text{H Bay} = \begin{bmatrix} 0.1101 & -0.0100 & -0.0011 & -0.0111 & -0.0020 & -0.0011 \\ -0.0100 & 0.1211 & 0.0311 & 0.0191 & 0.0121 & 0.1112 \\ -0.0011 & 0.0311 & 0.1125 & 0.3011 & 0.3111 & 0.1004 \\ -0.0111 & 0.0191 & 0.3011 & 0.2161 & 0.2112 & -0.0113 \\ -0.0020 & 0.0121 & 0.3111 & 0.2112 & 0.1101 & 0.0111 \\ -0.0011 & 0.1112 & 0.1004 & -0.0113 & 0.0111 & 0.2100 \end{bmatrix}$$

Results and Discussions

The first step: (for the KDA _ scv method)

It was found that this method gives a classification error rate for the first group $\widehat{MR}1$ (0.3) and for the second group $\widehat{MR}2$ (0.58), and thus the overall classification error rate ($\widehat{MR}$) is: (0.44) under traditional data, the first group $\widehat{MR}1$ (0.21) and for the second group $\widehat{MR}2$ (0.41), and thus the overall classification error rate ($\widehat{MR}$) is: (0.31) under fuzzy data

The second step: (for the KDA _ bcv method)

When using the method, the classification error was found to be $\widehat{MR}1$ for the first group (0.16) and for the second group, $\widehat{MR}2$ (0.26) under traditional data, When using the method, the classification error was found to be $\widehat{MR}1$ for the first group (0.14) and for the second group, $\widehat{MR}2$ (0.22) under fuzzy data,

Thus, the total classification error ratio ($\widehat{MR}$) for this method is: (0.11).

The third step: (for the KDA _ Bay method):

When using the method, the classification error was found to be $\widehat{MR}1$ for the first group (0.14) and for the second group, $\widehat{MR}2$ (0.26) Thus, the total classification error ratio ($\widehat{MR}$) for this method is: (0.18) under traditional data,

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022

When using the method, the classification error was found to be $\widehat{MR}1$ for the first group (0.11) and for the second group, $\widehat{MR}2$ (0.19) Thus, the total classification error ratio ($\widehat{MR}$) for this method is: (0.10) under fuzzy data.

From the results, we note that (KDA-Bay) method is better than (KDA-SCV) and (KDA-BCV) because it gives the lowest classification error rate from the results, and we note that all methods under fuzzy data better than the methods under traditional data.

References

A. Ibrahim, Nathier & A. Mohammed, Hussein, (2017), "Parameters and Reliability Estimation for the Fuzzy Exponential Distribution", *American Journal of Mathematics and Statistics, 7*(4): 143-151.

Al-Mikhlafi, Fouad Abdo Ismail, *"Categorizing and Distinguishing Yemeni Governorates According to Individual Income Sources Using Cluster Analysis and Discriminant Analysis",* Master's Thesis, Faculty of Commerce and Economics, Sana'a University.

Anil, K. G. and Chaudhuri,P. and Sengupta, D.(2006), *"Classification using kernel density estimators : Multi – scale analysis and visualization",* Applied statistics unit, Indian statistical Institute, Calcutta India, PP. 1-22.

Buch, T. K. (2007), *Nomparametric density estimation multidimension.* University of Copenhagen.

Duong, T. (2004), *"Bandwidth selectors for multivariate kernel density Estimation",* This thesis is presented for the degree of Doctor of Philosophy at the University of Western.

Duong, T. (2007), "Ks: Kernel density estimation and kernsl discriminant analysis for multivariate data in R", *Published by the American statistical Association in Journal of Statistical Software,* Vol. 21, Issue. 7.

Duong, T., Hazelton, M. (2005), "Cross – validation Bandwith matrices for multivariate kernel density estimation", university of Western Australia. *Published in Scandinavian Journal of Statistics,* vol. 32, Issue, 3, Pp. 485 – 506.

Duong, T. Hazelton, M. (2004), "Convergence rate for unconstrained bandwidth matrix selectors in multivariate kernel density estimation", *University of Western Australia, Published in Journal of Multivariate Analysis,* vol. 93, Issue. 2, PP. 417-433.

Garg, Harish, Sharma, S.P. & Rani, Monica, (2013)," Weibull fuzzy probability distribution for analyzing the behavior of pulping unit in a paper industry". *Int. J. Industrial and Systems Engineering,* Vol. 14, No. 4, pp 395-413.

KAMALUDDEEN USMAN DANYARO, JAFREEZAL JAAFAR, SHAHIR LIEW, (2010), " COMPARABILITY BETWEEN FUZZY SETS AND CRISP SETS: A SEMANTIC WEB APPROACH", SELECTED TOPICS in EDUCATION and EDUCATIONAL TECHNOLOGY, ISSN: 1792-5061, ISBN: 978-960-474-232-5, 240-244.

Ledl, T. (2004), "Kernel density estimation: Theory and Application in discriminant analysis", university by of Vienna. *Published in Austrian Journal of statistics,* vol. 33, NO. 3, PP. 267 – 279.

Ledl, T. (2002), "Kernel Density Estimation Theory, Aspects of Dimension and Application in Discriminant Analysis", To obtain the academic degree Magister, Faculty of Economics and Computer Science, University of Vienna.

Liberati, C. and Andrew, J.H., and Bozdogan, H. (2009), "Data Adaptive Simultaneous Parameter and Kernel (KDA) Using Information Complexity", *Journal of Pattern Recognition Research,* 1, PP. 119-132.

Lu,J. Plataniotis, K.N., Venetsanopoulos, A.N. and Wang, J. (2005), *"An efficient kernel discriminant analysis method",* University of Toronto, Ontario, Canada, Published by Elsevier B.V, vol. 38, Issue. 10, PP. 1788 – 1790.

Mitras, Ban Ahmed, Muhammad, Alaa Mahmoud, (2013), "Using the core estimator and the k-type clustering method to recognize hand gesture", *Al-Rafidain Journal of Computer Science and Mathematics, College of Computer Science and Mathematics, University of Mosul,* Volume (10), issue (1).

Pak, Abbas; (2017), "Statistical inference for the parameter of Lindley distribution based on fuzzy data" *Brazilian Journal of Probability and Statistics,* Vol. 31, No. 3, 502–515.

Pekalska, E., Haasdonk, B. (2009), "Kernel Discriminant Analysis for Positive Definite and Indefinite Kernels", *Published by the IEEE Computer Society,* vol. 31, No. 6.PP.1017-1031.

You, D., Onur, C.H. and Aleix M.M. (2011), "Kernel optimization in discriminant analysis", *Published in Xplore Digital Library,* Volume: 33, Issue: 3.

Zhang, X., Maxwell L.K., Rob. J.H. (2004), "Bandwidth selection for multivariate kernel density estimation using CMC", Monash University, Australia.

Zhao, Z., Sun, L., Yu, SH., Liu, L. and Ye, J. (2009), "Multiclass Probabilistic Kernel Discriminant Analysis", *Published in : IJCAI 0g Proceedings of the 21st international jont conference on Artifical intelligence, San Francisco CA, USA,* pp. 1363-1368.

DOI:10.9756/IAJH/V9I1/IAJH0902
Received:18 July 2022/Accepted:27 August 2022/Published:08 September 2022