Improved LDA by using Distributing Distances and Boundary Patterns

Ali Yaghoubi, Hamid Reza Ghaffari

Abstract- One of the statistical methods of class discriminant is linear discriminant analysis. This method, by using statistical parameters, obtain a space which by using available discriminating information among class means does classification act . By using distributing Distances, extended analysis linear discriminant to its heteroscedastic state. At this state ,to make classes more separating of available separating information among covariance matrix classes including classes mean is using. In this article ,because of using new scattering matrices which are defined based on boundary and non- boundary patterns, classes overlapping in Spaces which obtains has been reduced . On the other hand ,using new scattering matrices brings about increasing classification rate so, the done experiments confirm improvement of classification rate.

Keywords: boundary linear discriminant analysis, Boundary and non-boundary patterns, CHernoff criteria, linear discriminant analysis.

I. INTRODUCTION

In distinguishing pattern ,a capital step is classification of them and a current technique as well as capital stage is extracting feature .One of the techniques of feature extracting is reduction of linear dimension which is often used to decreasing dimension size ,data and statistical models as well as overcoming the problems which comes out in this field. The reduction of data dimension should not cause discriminating information which exists in main feature space to be destroyed. LDA is from common methods which is known as classifying method [1]. This classic method by Fisher for two-class problem and by RAO to solve multi-class problem has been developed [16]-[17]. In LDA a transformation matrix converts n-dimension of data to d (d<n) dimension space also it tries in given lower-dimensional space as Fisher criteria which does maximize the proportion of scattering matrices between-classes and within-classes and does classification act [12]-[13]-[15]. LDA is a fast and easy method to determine a good feature and need simple calculating matrixes so in different articles the most relating problems to LDA has been reported. One of LDA disadvantages is that its most care is about maximum separation between means in projection space and knows it as the best method. This is in a sense that the available separating information which in opposition among covariance matrix has ignored classes and to some extent LDA face problem and most methods to solve this problem has used distributing distances [3]-[4].

Manuscript Received on January 2015.

Ali Yaghoubi, M.Sc., Department of Engineering, Islamic Azad University, Ferdows Branch, Ferdows, Iran.

Dr. Hamid Reza Ghaffari, Department of Engineering, Islamic Azad University, Ferdows Branch, Ferdows, Iran.

In one of these methods Chernoff's distributive distance was used which is known as Chernoff's criteria [2]. In Several methods by using this criteria, classification rate was extended in different applications [6]-[5]. Also in [6], of this criteria to extend heteroscedasticity state of non-interdependent and extracted features has been used . The shortcoming of this criteria is that in original space the classes which is close to each other ,in a projection done by the criteria as overlapping ,too much of the classes are projected to this space as a cluster so, this process causes severe reduction of classification rate . In [6]-[7] by using measurement criteria reduces the effects of classes which are far from each other but, measurement criteria was repeatedly done and extracted features had been limited to the class numbers. In this article, we want to use new matrixes which is defined in [8] and treat Chernoff's shortcoming criteria so that we increase classification rate criteria as well as the numbers of extracted features .In the second section of the article we will have article definition, in the third section we will have the definition method and using of scattering patterns based on boundary and non- boundary patterns will be expressed and finally in the forth section of this article ,we will discuss the existing results on downloaded datasets of UCI.

II. DEFINING THE PROBLEM

If in a statistical classifying problem c classes is being considered as $\omega_1 \cdots \omega_2$ which have m label and n dimension, and then c class is as in

$$D_1 = \{x_{1,1}, \ldots, x_{1,m_1}\}, \ldots, D_c = \{x_{c,1}, \ldots, x_{c,m_c}\}.$$

Based on parametrical form which is considering for classifying, *c* class has primary probability $P_1, \ldots, P_2$ so $x_1 \sim N(m_1, S_1), \ldots, x_c \sim N(m_c, S_c)$ is *n*-dimension of accidental distributive vector that out of them *S* and *m* is considered as class covariance and mean.

A. *Chernoff's two-class patterns*

Based on chernoff distance between two distributive class in original space so, the solution to this problem is optimizing and searching projection vector *W* in order to maximize criterion (1), as in

$$J_c(W) = \mathrm{tr}\{(WS_W W^t)^{-1}[WS_E W^t - WS_W^{1/2} \frac{p_1 \log(S_W^{-1/2} S_1 S_W^{-1/2}) + p_2 \log(S_W^{-1/2} S_2 S_W^{-1/2})}{p_1 p_2} S_W^{1/2} W^t]\} \quad (1)$$

W is obtained based on the Eigenvalue decomposition of the matrix:

$$S_c(W) = S_W^{-1}[S_E - S_W^{1/2}\frac{p_1\log(S_W^{-1/2}S_1S_W^{-1/2}) + p_2\log(S_W^{-1/2}S_2S_W^{-1/2})}{p_1p_2}S_W^{1/2}] \quad (2)$$

W is Eigenvector corresponding to the largest Eigenvalues of the matrix (2).

B. Chernoff's multi- class patterns

To extend chernoff's two-class criterion into multi-class ones, a certain decomposition of between-class scattering matrix is used. At this decomposition, between-classes scattering matrix is built by using double class blocks. Now chernoff's multi-class is expressed as (3) formula and the goal is finding *W* which maximize criterion:

$$J_C(A) = \sum_{i=1}^{K-1}\sum_{j=i+1}^{K} P_iP_j \mathrm{tr}\begin{pmatrix} (WS_wW^t)^{-1}\times \\ WS_w^{1/2}\begin{bmatrix} (S_w^{-1/2}S_{ij}S_w^{-1/2})^{-1/2}\times S_w^{-1/2}S_{Eij}S_w^{-1/2}(S_w^{-1/2}S_{ij}S_w^{-1/2})^{-1/2} + \\ \frac{1}{\pi_i\pi_j}(\log(S_w^{-1/2}S_{ij}S_w^{-1/2}) - \\ \pi_i\log(S_w^{-1/2}S_iS_w^{-1/2}) - \\ \pi_j\log(S_w^{-1/2}S_jS_w^{-1/2})) \end{bmatrix}WS_w^{1/2} \end{pmatrix} \quad (3)$$

$$S_{Eij} = (m_i - m_j)(m_i - m_j)$$

To determine *W*, Eigenvalue decomposition of the matrix is formed, *W* is equivalent to Eigenvector and largest Eigenvalue of matrix :

$$S_C = \sum_{i=1}^{C-1}\sum_{j=i+1}^{C} P_iP_j \mathrm{tr}\begin{pmatrix} (S_w)^{-1}\times \\ S_w^{1/2}\begin{bmatrix} (S_w^{-1/2}S_{ij}S_w^{-1/2})^{-1/2}\times S_w^{-1/2}S_{Eij}S_w^{-1/2}(S_w^{-1/2}S_{ij}S_w^{-1/2})^{-1/2} + \\ \frac{1}{\pi_i\pi_j}(\log(S_w^{-1/2}S_{ij}S_w^{-1/2}) - \pi_i\log(S_w^{-1/2}S_iS_w^{-1/2}) - \pi_j\log(S_w^{-1/2}S_jS_w^{-1/2})) \end{bmatrix}S_w^{1/2} \end{pmatrix} \quad (4)$$

δ_{ij}^2 The point is that $\mathrm{tr}(S_{Cij}) = \delta_{ij}^2$.

δ_{ij}^2 is expressed as Eigenvalue so, Eigenvector which is relating to the most Eigenvalue is considered as projection vector of W . The Eigenvector equivalent vector by Eigenvalue δ_{ij}^2 is considered as Eigenvector between two class i and j, so to this reason it is distinguished as the biggest Eigenvalue as well as projection vector of W thus, to perceive more , consider Fig .2

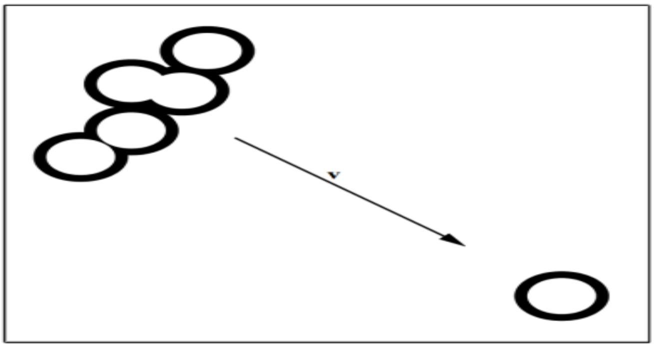

Fig. 2 Map of classes using a vector V [7]

Fig.2, is a kind of six-class model in which each circle is considered as one class so, these circles have similar radius that shows within-class scattering matrix has been equally assumed . If the right class on the corner down in Fig.2 being considered as j_0 on the condition to be far from the rest on original space the contributions of Eigenvalue $\delta_{ij}^2, 1 \le i \le C, i \ne j$ will dominate on between-class scatter. the result of direction distinguished with ***v*** symbol has been shown in Fig.2 which will be known as principal discriminant figure so, the result is that for *V* projection *i* and *j* classes which will bring the projection $i \ne j$ to a cluster with high among classes overlapping that necessitate rising classification rate faults in projection space . By this example, we can conclude that in estimating chernoff's criterion between-class scattering matrix, class pairs without their separation in real space ,are considered which causes bad performance in separating classes .

III. SEPARATION BASED ON BOUNDARY AND NON-BOUNDARY PATTERNS

In this article we are going to use new scattering matrix to solve this problem so, that is the case that these matrices have been built instead of considering all data based on pattern placement situation. some data has been well separated by different class label and some other near to decision making class has been mixed so ,to this reason ,their discrepancy must be considered in making scattering matrix thus, based on this ,boundary and non- boundary patterns is being defined.

C. Boundary patterns

It is a data that its located is near to decision making boundary and causes neighboring *k* with dissimilar class labels.

D. Non- Boundary patterns

It is a data that its located is far from decision making boundary and causes neighboring *k* with similar class labels.

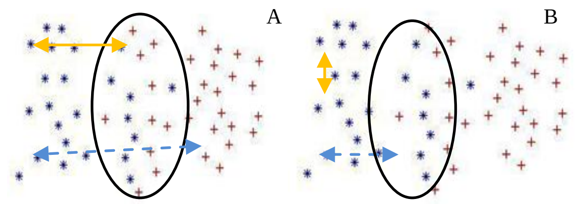

Fig. 3 Effect of boundary and non-boundary patterns on scattering matrices

In Fig.3, in terms of between-class matrix scattering matrices (Fig.3 (A)) and within-classes (Fig.3, (B)) the effect of boundary and non boundary patterns have been surveyed (the area which has been shown by an oval shape and is as boundary area between two class.

In Fig.3 (A) , *Influence on between-class matrix:* the dotted line indicates show the pair of non boundary patterns with different class labels however, at the moment these pattern pairs have well separated each other and maximizing among them have no effect on maximizing between-class scattering matrix. On the other hand, The solid line show pattern pair boundary and non-boundary patterns so, minimizing the difference among them has directly related on class separation because reduces pattern difference with different label hence maximizing their difference is significant.

In Fig.3 (B) , *Influence on within-class matrix:* the dotted line indicates show the pair of boundary and non-boundary patterns with similar labels class however, minimizing their difference is significant. as it can reduce the difference between different class-labeled patterns. minimizing their difference among them for within-class

Retrieval Number: F2486014615/2015©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

reduction scatterness. On the other hand, , the solid line shows both of non-boundary patterns with similar label. minimizing their difference for within-class scattering matrix considerable because they are clear and representative patterns. based on the above research , we divide the set of inputting patterns in two subsets of boundary patterns and non-boundary ones.

IV. THE SELECTION OF BOUNDARY AS WELL AS NON- BOUNDARY PATTERNS

To design new scattering matrix , it is needed that the patterns which exist in boundary decision making class being selected so, the boundary area is where distributing different classes overlap . The key idea of boundary pattern selection is Proximity pattern [9]. proximity criterion show pattern's close it indicates how data are placed close to the boundary which is :

$$\mathrm{Proximity}(x,k)=\sum_{i=1}^{c}p_i(x)\log_l\frac{1}{p_i(x)} \qquad (5)$$

$$p_i(x)=k_i/(k+1)$$

c symbolize the class number and *k* symbolize the number of pattern neighboring also k_i is the number of neighboring patterns belonging to *i* class so, According to the value of proximity, the non-boundary patterns determines which is:

$$X^{(NB)}=\left\{x_i \,\middle|\, \begin{array}{l}\mathrm{Proximity}(x_i,k)\le\theta(c)\\ ,x_i\in X,0<\theta(c)<1\end{array}\right\} \qquad (6)$$

$X^{(NB)}$ is as the sets of non-boundary patterns and $\theta(c)$ is a non-zero value depending on *c*. When the class numbers and pattern ones increase, the probability of patterns mixture with a different class is too much and these pattern numbers are known as boundary pattern. To reduce boundary patterns threshold level $\theta(c)$ is $\theta(c)=1-1/c$ defined as boundary pattern in order to satisfy the requirement [10]. Based on this definition, if the class number increases the $\theta(c)$ value goes near one and causes more data places in non- boundary pattern set so, by such a threshold level a suitable non-boundary pattern number is obtained. According to this process, the set of input samples sets is divided in two boundary as well as non- boundary set . In (7) formula, scattering matrices which has been built based on boundary and non-boundary patterns is shown.

$$S^{(b)}\equiv\sum_{i=1}^{c}\sum_{j=1}^{n^{(B)}}\left(x_j^{(B)}-m(i)\right)\left(x_j^{(B)}-m(i)\right)^T, \qquad (7)$$

$$S^{(w)}\equiv\sum_{i=1}^{c}\sum_{j:y_j=i}\left(x_j^{(NB)}-m(i)\right)\left(x_j^{(NB)}-m(i)\right)^T$$

In making matrices which is between-class scatter , of boundary and class mean difference for making scattering matrix has been used، and those between non- boundary patterns and class means are used in within-class scatter matrix. Fig 3. diagram shows a concept of scattering matrix.

A

Boundary Region

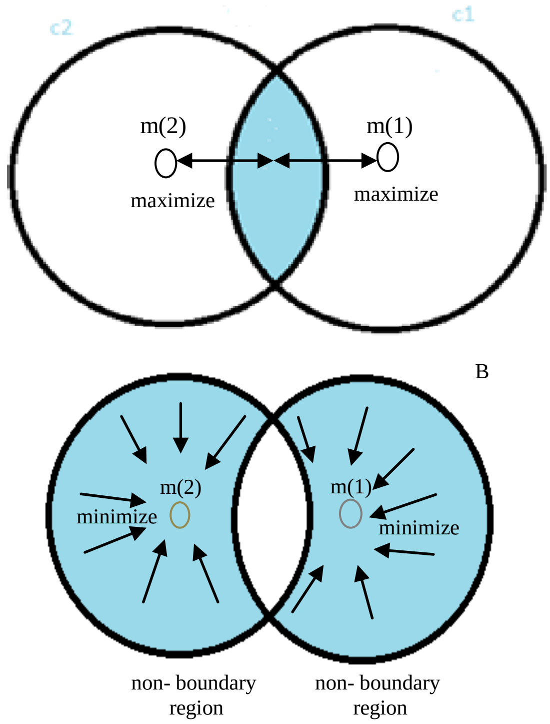

Fig. 4 The concept of geometric new scattering matrix

Fig.4 (A) shows scattering matrix diagram of between-class scatter matrix $S^{(b)}$. This matrix tries to separate boundary patterns which is far from all-class center.

Fig.4 (B) shows scattering matrix diagram of within-classes matrix. This matrix tries to concentrate non-boundary patterns which is around their classes center.

To express purpose, we can change among classes scattering matrix formula which its prove is mentioned in [21] so, we have change of between-class scatter matrix as:

$$S^{(b)}\equiv\sum_{i=1}^{l}\sum_{j=1}^{n^{(B)}}\left(x_j^{(B)}-m(i)\right)\left(x_j^{(B)}-m(i)\right)^T=S^{(b)}\equiv\sum_{i=1}^{l}\sum_{j=1}^{n^{(B)}}\left(x_j^{(B)}-\sum_{u:y_u=i}x_u\right)\left(x_j^{(B)}-\sum_{v:y_v=i}x_v\right)^T\equiv \qquad (8)$$

$$\sum_{i=1}^{l}\sum_{j=1}^{n^{(B)}}\frac{1}{n(i)^2}\left(\sum_{u:y_u=i}(x_j^{(B)}-x_u)\right)\left(\sum_{v:y_v=i}(x_j^{(B)}-x_v)\right)^T\equiv\sum_{i=1}^{l}\sum_{j=1}^{n^{(B)}}\frac{1}{n(i)^2}\sum_{u,v:y_u=y_v=i}(x_j^{(B)}-x_u)(x_j^{(B)}-x_v)^T$$

$$\equiv\sum_{j=1}^{n}\sum_{u,v=1}^{n}\tilde{a}_{juv}^{(b)}(x_j-x_u)(x_j-x_u)^T$$

$$\tilde{a}_{juv}^{(b)}=\begin{cases}\dfrac{1}{\{n(y_u)\}^2} & if \quad x_j\in X^{(B)} \quad and \quad y_u=y_v\\ 0 & otherwise\end{cases}$$

the number of samples belonging to the class y_u is $n(y_u)$.

This kind of formulization change is based on measure distance among sampling pattern thus, the point is that the values of non-zero measures depend on scattering among boundary patterns with similar labels. in the other hand; we can say that the zero measure is the opposition between non-boundary patterns of one class and non- boundary ones of other class. For this sake, the patterns existing in non-boundary area does not have any effect on calculating class separation capacity because they have been well separated before . The other point is that based on boundary and non-boundary pattern between two classes matrices which has been separated each other in real space don't place in boundary area and don't categorize as non-boundary pattern. The result is that based on between-class scattering matrix, this kind of patterns don't have any influence on between-class scattering matrix estimation.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

V. CHERNOFF'S IMPROVEMENT PATTERN BY USING NEW SCATTERING MATRIX

As it was said in equation (8) , the zero measure pertaining to difference between non-boundary patterns from one class to non-boundary from another ones. Based on this, these patterns don't have any effect on classes separation because they have well been separated. the result is that the difference between these patterns do not have any on class scattering matric estimation. The goal of matrices design (7) is finding a W direction so that by using emphasis on inattentive differences as well as attentive one we can find a direction in chernoff's pattern (3) in which by data projection it can maximize chernoff's optimizing criterion (J_C) and as a result we reduce classification data rate . It is clear that the amount of class overlapping considerably are reduced in the place of dimension drop. based on this, by using new scattering matrices in Chernoff's matrices (3) we replace them with primary scattering matrices:

$$J_C(A)=\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}P_iP_j tr\left(\begin{array}{l}\left(AS_wA^t\right)^{-1}\times \\ AS_w^{1/2}\left[\begin{array}{l}\left(S_w^{-1/2}S_{ij}^{(w)}S_w^{-1/2}\right)^{-1/2}\times S_w^{-1/2}S_{ij}^{(b)}S_w^{-1/2}\left(S_w^{-1/2}S_{ij}^{(w)}S_w^{-1/2}\right)^{-1/2}+\\ \frac{1}{\pi_i\pi_j}(\log(S_w^{-1/2}S_{ij}^{(w)}S_w^{-1/2})-\\ \pi_i\log(S_w^{-1/2}S_i^{(w)}S_w^{-1/2})-\\ \pi_j\log(S_w^{-1/2}S_j^{(w)}S_w^{-1/2}))\end{array}\right]AS_w^{1/2}\end{array}\right) \quad (9)$$

$$S^{(b)}\equiv\sum_{i=1}^{l}\sum_{j=1}^{n^{(B)}}\left(x_j^{(B)}-m(i)\right)\left(x_j^{(B)}-m(i)\right)^T$$

$$S^{(w)}=\sum_{i=1}^{l}\sum_{j:y_j=i}(x_j^{(NB)}-m(i))(x_j^{(NB)}-m(i))^T$$

To determine optimizing W, specific value matrix analysis (10) is happening, however; W vector projection and specific vector equal to the most specific value matrices is determined.

$$S_C=\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}P_iP_j tr\left(\begin{array}{l}\left(S_w\right)^{-1}\times \\ S_w^{1/2}\left[\begin{array}{l}\left(S_w^{-1/2}S_{ij}^{(w)}S_w^{-1/2}\right)^{-1/2}\times S_w^{-1/2}S_{ij}^{(b)}S_w^{-1/2}\left(S_w^{-1/2}S_{ij}^{(w)}S_w^{-1/2}\right)^{-1/2}+\\ \frac{1}{\pi_i\pi_j}(\log(S_w^{-1/2}S_{ij}^{(w)}S_w^{-1/2})-\pi_i\log(S_w^{-1/2}S_i^{(w)}S_w^{-1/2})-\pi_j\log(S_w^{-1/2}S_j^{(w)}S_w^{-1/2}))\end{array}\right]S_w^{1/2}\end{array}\right)$$

$$S_cW=\lambda S_wW \quad (10)$$

$$A=[W_1,W_2,\ldots,W_d]$$

W is equal to W_d optimizing vector and also Eigenvector equal to the most Eigenvalue in Eigenvalues matrix .

VI. EXPERIMENTS

At this part we offer experiment results to the applicatory of suggestive method . The experiments are done on different data sets which have been downloaded UCI Machine Learning website. [18].These datasets have been listed in Table.1, The majority of these data sets have been used in [17,15,14,2] articles . the unclear available values in datasets have been replaced with relative medium feature value [8]. The output of proposed method has been compared with CDA,LDA that imply better output of proposed method in compare with the other two methods.

Table 1. The UCI dataset used for the experiments [19]

Dataset name	Number data	Number class	Number feature(number dimension)
Haberman	31	2	3
Australian credit	653	2	51
German credit	1000	2	38
Primary tumor	336	2	15
Banknote authentication	1370	2	3
Vote	435	2	16
Hepatitis	137	2	34
Liver	345	2	6
Zoo	101	2	16
Wine	178	3	13
new-thyroid	215	3	6
balance-scale	150	3	5
Iris	150	3	5
1189	1092	4	12
Breast cancer Wisconsin	699	2	11
Hayes Roth	132	3	5

E. *Experimental setup*

To assess the performance of proposed method ,fisher's two classifying strategy leave-one-out(LOO) cross-validation and 10-fold cross validation is used . In leave-one-out(LOO) cross-validation Strategy ,n-1 is input data for training and the rest of the data are used for identification and testing . Although leave-one-out(LOO) cross-validation is a good method to evaluate performance [20]-[21] , it has been criticized by researchers [21] for this sake, in this article from 10-fold cross-validation Has been used for classification . In order to avoid poverty of covariance matrix as well as noise exclusion ,in pre-processing phase , PCA has been used therefore, the dimensions whose specific values have been lesser than one- hundred-thousandth has been omitted from dataset . In order to avoid of the between classes specifying scattering matrix which is current problem in using linear classification analysis and its derivation ,a lawful method mentioned in [18] has been used . Firstly, the in-classes scattering matrix mentioned in (7) ranking was investigated and then if its ranking was incomplete a*I value was added to them in which a=0.001 and I is like unit matrix. to avoid problems with log and square root of matrix A inverse , of mentioned method in [13] has been used, therefore; to calculate F function, of specific A matrix has been used . A matrix is analyzed as specific analysis $\left(\mathrm{VDV}^{-1}\right)$ in which V are as specific vectors of D and A matrix and also as specific matrix of A respectively, and then we apply f function on the main elements of specific value which contains these values and placed them in specific value matrix that resulted $\mathrm{f(A)}=\mathrm{Vf(D)V}^{-1}$ change. if Eigenvalues in applying log function to be reverse ,negative or zero then ,the number result will be equal to zero so, to stop this , a small fixed amount must be added [22]. In order to do this, a positive small fixed amount has been added to specific matrix D either negative or zero.

F. *Discussion on experimental method's outputs*

This part has focused on discussing observations derived from experiments done on datasets. The experiments showed that ,in Iris dataset between class 1,2 and 1,3 there is no boundary sample but between class 2 and 3, 20 boundary sample were observed therefore, we can conclude that class 1 in proportion to class 2 and 3 is so far , to this reason scattering matrix between classes 1 and the other two classes doesn't have any effect on designing chernoff's classification because matrix has been defined based on boundary samples and instead of them we can use space matrix estimation between

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

class pairs which owned by fisher linear classification and based on mean intervals. This is because the best method for classifying the classes which are far apart is fisher's classification. also, in 1189 dataset it was observed that 2 and 3 class don't have any non- boundary sample and two class near decision-making border between two class have been classes patterns which are severely mixed with different label show that these two classes don't have overlapping therefore , for this reason out of classes, new scattering matrix which is based on boundary samples for chernoff's approximate criteria and classes classification is used. The reported results, in table 2, symbolize improvement of using researcher's proposed method in proportion to other two methods. Also in

Table. 2, comparison between 3 methods which have been done on Table.1, database , have been shown .

Table 2.The output results of experiments on the dataset used three methods

Dataset name	LDA	Proposed Method	HDA
Haberman	61.8065	74.9247	55.6999
Australian credit	67.6713	74.4988	66.9161
German credit	63.2000	65.3000	58.4000
Primary tumor	61.3387	65.9777	63.9947
Banknote authentication	59.2669	97.3348	95.2618
Vote	75.7558	954979	94.2653
hepatitis	64.9451	65.3956	64.1758
Liver	57.4118	67.7059	61.1513
Zoo	81.2727	86.2727	79.4545
Wine	85.8448	92.5050	87.5678
new-thyroid	42.4026	76.2338	45.9870
balance-scale	57.3725	68.2567	55.7557
Iris	87.8711	92.7222	72.6418
1189	65.1720	82.2272	70.7988
Breast cancer Wisconsin	88.5300	96.5714	87.2298
Hayes Roth	82.4176	86.8702	79.4505

VII. CONCLUSION

In this article, using new designed matrix, the possibility of overcoming current problem for chernoff's criterion was obtained and the effect of far apart class pair on chernoff's criterion was lost so, on the other side of the coin , the effect of both close classes increased . The conducted experiments show that using these matrixes boost the rate of data classification.

REFERENCES

[1] Fukunaga, Keinosuke. *Introduction to statistical pattern recognition*. Academic press, 1990.
[2] Duin, R. P. W., and M. Loog. "Linear dimensionality reduction via a heteroscedastic extension of LDA: the Chernoff criterion." Pattern Analysis and Machine Intelligence, IEEE Transactions on 26.6 (2004): 732-739.
[3] Loog, Marco, and Robert PW Duin. "Non-iterative Heteroscedastic Linear Dimension Reduction for Two-Class Data." Structural, Syntactic, and Statistical Pattern Recognition. Springer Berlin Heidelberg, 2002. 508-517.
[4] Zhu, Xinzhong. "Super-class Discriminant Analysis: A novel solution for heteroscedasticity." Pattern Recognition Letters 34.5 (2013): 545-551.
[5] Safayani, Mehran, and Mohammad Taghi Manzuri Shalmani. "Two-Dimensional Heteroscedastic Feature Extraction Technique for Face Recognition."Computing and Informatics 30.5 (2012): 965-986.
[6] Sugiyama, Masashi. "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis." The Journal of Machine Learning Research 8 (2007): 1027-1061.
[7] ReinholdHaeb-Umbach, MarcoLoog. "MULTI-CLASS LINEAR DIMENSION REDUCTION BY GENERALIZED FISHER CRITERIA." The Proceedings of the 6~(th) International Conference on Spoken Language Processing (Volume II). 2000.
[8] Na, Jin Hee, Myoung Soo Park, and Jin Young Choi. "Linear boundary discriminant analysis." Pattern Recognition 43.3 (2010): 929-936.
[9] Shin, Hyunjung, and Sungzoon Cho. "Neighborhood property–based pattern selection for support vector machines." Neural Computation 19.3 (2007): 816-855.
[10] Na, Jin Hee, et al. "Relevant pattern selection for subspace learning." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.
[11] Sugiyama, Masashi. "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis." The Journal of Machine Learning Research 8 (2007): 1027-1061.
[12] McLachlan, Geoffrey. Discriminant analysis and statistical pattern recognition. Vol. 544. John Wiley & Sons, 2004.
[13] Masip, David, Ludmila I. Kuncheva, and Jordi Vitrià. "An ensemble-based method for linear feature extraction for two-class problems." Pattern Analysis and Applications 8.3 (2005): 227-237.
[14] Fukunaga, Keinosuke. Introduction to statistical pattern recognition. Academic press, 1990.
[15] Jain, Anil K., Robert P. W. Duin, and Jianchang Mao. "Statistical pattern recognition: A review." Pattern Analysis and Machine Intelligence, IEEE Transactions on 22.1 (2000): 4-37.
[16] Yang, Jian-Yi, et al. "Prediction of protein structural classes by recurrence quantification analysis based on chaos game representation." Journal of theoretical biology 257.4 (2009): 618-626.
[17] Fisher, Ronald A. "The use of multiple measurements in taxonomic problems."Annals of eugenics 7.2 (1936): 179-188.
[18] Friedman, Jerome H. "Regularized discriminant analysis." Journal of the American statistical association 84.405 (1989): 165-175.
[19] D.J. Newman, S. Hettich, C.L. Blake, C.J. Merz, UCI repository of machine learning databases, 1998 <http://archive.ics.uci.edu/ml>.
[20] Vapnik, Vladimir, and Olivier Chapelle. "Bounds on error expectation for support vector machines." Neural computation 12.9 (2000): 2013-2036.
[21] Chapelle, Olivier, et al. "Choosing multiple parameters for support vector machines." Machine learning 46.1-3 (2002): 131-159.
[22] Loog, Marco, R. P. W. Duin, and Reinhold Haeb-Umbach. "Multiclass linear dimension reduction by weighted pairwise Fisher criteria." IEEE Transactions on Pattern Analysis and Machine Intelligence 23.7 (2001): 762-766.