Improve Heteroscedastic Discriminant Analysis by Using CBP Algorithm

Jafar A. Alzubi[1] and Ali yaghoubi[2] and Mehdi Gheisari[3] and Yongrui Qin[4]

[1] Al-Balqa Applied University, Jordan
j.zubi@bau.edu.jo

[2] School of Computer Science and Technology, Guangzhou University, Guangzhou, China, 510006.
Yaghoubi_ali67@yahoo.com

[3] School of Computer Science and Technology, Guangzhou University, Guangzhou, China, 510006.
mehdi.gheisari61@gmail.com

[4] Department of Computer Science, University of Huddersfield, UK
y.qin2@hud.ac.uk

Abstract. Linear discriminant analysis is considered as current techniques in feature extraction so, LDA, by discriminant information which obtains in mapping space, does the classification act. When the classes' distribution is not normal, LDA, to perform classification, will face problem and will resulted the poor performance of criteria in performing the classification act. One of the proposed ways is the use of other measures, such as Chernoff's distance so, by using Chernoff's measure LDA has been spreading to its heterogeneous states and LDA in this state, in addition to use information among the medians, uses the information of the classes' Covariance matrices. By defining scattering matrix , based on Boundary and non- Boundary samples and using these matrices in Chernoff's criteria, the decrease of the classes' overlapping in the mapping space in as result, the rate of classification correctness increases. Using Boundary and non- Boundary samples in scattering matrices causes improvement over the result. In this article, we use a new discovering multi-stage Algorithm to choose Boundary and non- Boundary samples so, the results of the conducted experiments shows promising performance of the proposing method.

Keywords: Linear Discriminant Analysis, CBP Algorithm, Chernoff Criterion, Boundary Pattern.

1   Introduction

Classification data into groups is considered of important stages of pattern recognition that one of its major stages is feature extraction. One of the features of extraction is reducing the linear dimensions which often, to reduce data dimensions and statistical models and also, overcoming the problems which arise in this field can be used.

Reducing data dimensions shouldn't cause discriminant information that is in the original space of the main feature to be eliminated. From the usual methods in the field of discriminant extracting information (also known as a classification technique) is LDA [1]. This classical approach have been developed, by Fisher [13], for two-class classification and by Rao for multi- class classification [14]. In LDA a transformation matrix, changes a main n dimension space of the data to d dimension. The above conversion maximizes the proportion of scattering matrix between the classes and does the classification [10, 1, 11, and 20]. LDA is a fast and easy way to set a good character and requires simple matrices calculation. In several articles, many of the problems related to the LDA has been reported and solutions provided so, LDA concentration is to obtain a space where follow with the maximum average between the classes If so, imagining normal distribution to be existed for the classes this strategy can be used But, in the real world we cannot have such hypothesis, and this strategy is not always useful. In [21] it has been tried to measure the shortcoming of this Criteria so, the classes which are in the original space close to each other and this action along with a sharp decline in the rate of classification .in [3]-[4], This method, with the help of weight criteria, decreases the impact of the classes that, in the original space, are far apart each other. Weight criteria carry out repeatedly and that the number of extracted feature is limited to the number of classes. In [21], the use of new matrices to overcome this problem has been suggested. The problem is that the selection algorithm of Boundary and non-Boundary patterns and RPS (relevant pattern selection) in a series of the training data is not working properly. Because at the same time, either to the number of its neighbors and or a threshold level that RPS are used, is depended [22]. Out of the methods of Boundary and non- Boundary patterns, we can point to the Algorithm based on the Graph. A graph-based algorithm is Hit Miss Networks (HMN) which are directed graphs of instances in the training set [23]. This Algorithm of the Graph obtains orientation of the samples in training set so, for this for each sample, the nearest neighbor of it, is determined from all the classes and an edge between the targeted sample and each of its neighbor is defined so, a hit edge between the sample and its neighbor by the similar class label is defined. For this for each sample, the nearest neighbor of it, is determined from all the classes and an edge between the targeted sample and each of its neighbor is defined so, a hit edge between the sample and its neighbor by the similar class label is defined. A Miss edge, between sample and nearby its nearby sample is defined by different class label. The result is that each sample is defined as a node and the edges of Hit and Miss is calculated as the degree of each node and based on these classifications a detecting pattern is imposed on the sets of training data. Based on the concept of Algorithm HMN, two other kinds of the Algorithm have been suggested based on the Graph [24] [25]. In this paper we use a method of sample reduce (which has been as a powerful discovering technique) naming CBP (class boundary persevering) to separate Boundary and non- Boundary patterns [26]. CBP is an Algorithm which gives us the best description underlying distribution about class samples spread. This Algorithm has used some steps heuristically that by using four steps has pruned the primary training sets and divides in to subsets of Boundary and non- Boundary samples. In this paper, we investigate, theoretically, the effect of proposing Algorithm on the process and the number of extracted features. The

organization, in this paper continues to be as follow; hence, in the second part we will have a definition of the problem. In Section 3 , we will discuss the definition using the scattering matrix based on bordering as well as non-bordering patterns and in Section 4 of this article , ultimately , we will discussed on the results of downloaded dataset from UCI and will investigate them.

2   Defining the problem

In a problem of classification, consider c number of classes as $w_1, \dots, w_c$ which has C label and N dimension and then, this C class is shown as $D_1 = \{x_{1,1}, \dots, x_{1,m_1}\}, \dots, D_c = \{x_{c,1}, \dots, x_{c,m_c}\}$ and on the basis of parametric form which is considered for the classification, then c class will have initial probability $p_1, \dots, p_c$ And random distribution vectors of n dimension $x_1 \sim N(m_1, S_1), \dots, x_c \sim N(m_c, S_c)$ and so that m, S are considered as the average and class covariance.

2.1   Chernoff's criteria

In general, in the methods of reducing the linear dimensions, the goal is to find the conversion matrix of W so that it can change the input normal distribution vectors $x_1, \dots, x_n$ from the n dimension to the d dimension so, the new normal distribution vectors can be obtained as. $y_1 = W x_1, \dots, y_{d \times 1} = W_{n \times 1} x_{n \times 1}$ In LDA, W is chosen in a way that the pattern of $tr\{(ww^T)^{-1} w s_B w^T\}$ maximizes, which S_B is the definer of scattering matrix of between classes and is as $S_B = \sum_{i=1}^{C} p_i (m_i - m)(m_i - m)$ and the process of classification is performed based on the process of Fischer's classification [6] [15]. Scattering matrix of S_B is the product of available separating information in the difference between the average of the classes and it does not benefit from separating information available in separating information which is in the difference of Covariance Matrices. in order to benefit available separating information in the Covariance difference, in the method of LDA, and in [6] instead of using Eglidoosi distance between the averages, the Chernoff's of distributing distance between the distribution of each pair of class is used so, the result is that, in addition to benefiting of discriminant information between the averages and discriminant information the difference between the Covariance is considered.
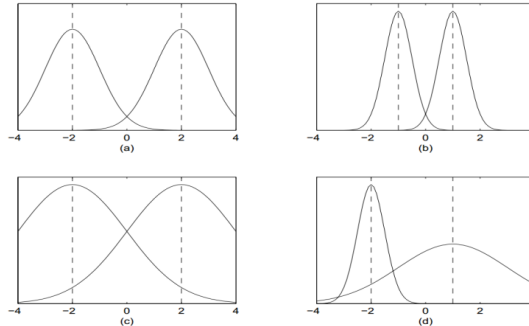
Fig. 1. Separability based on distribution of classes [7]

As it is observed in (Fig. 1) (parts of a, b). The averages are fixed and for the separation of the classes, two classes distribution has been used so, it is observed that the amount of the overlap has considerably reduced.

Chernoff's two-class criteria

Based on chernoff distance between two distributive class in original space so the solution for this optimizing problem is to search projection vector of W so that maximize the criteria of (1), as in

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

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

$$S_c(W) = S_W^{-1}[S_E - S_W^{1/2} \frac{p_1 \log(S_W^{-1/2} S_1 S_W^{-1/2}) + p_2 \log(S_W^{-1/2} S_2 S_W^{-1/2})}{p_1 p_2} S_W^{1/2}] \quad (2)$$

W, as especially corresponding Eigenvector is with the maximum amount of the matrix (2).

Chernoff's multi- class criteria

In order to vast Chernoff's criterion of two-class to state of multi-class, an certain decomposition of scattering matrices between the classes are used so, in this decomposition, between classes matrix , by using two-class blocks is built. Now, Chernoff's multi class criterion is shown as (3) Formula and the goal is finding W which maximize criterion of (3)

$$J_C(A) = \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} P_i P_j tr \left(\begin{array}{l} (WS_w W^t)^{-1} \times \\ WS_w^{1/2} \left[\begin{array}{l} (S_w^{-1/2} S_{ij} S_w^{-1/2})^{-1/2} \times S_w^{-1/2} S_{Eij} S_w^{-1/2} (S_w^{-1/2} S_{ij} S_w^{-1/2})^{-1/2} + \\ \frac{1}{\pi_i \pi_j} (\log(S_w^{-1/2} S_{ij} S_w^{-1/2}) - \pi_i \log(S_w^{-1/2} S_i S_w^{-1/2}) - \pi_j \log(S_w^{-1/2} S_j S_w^{-1/2})) \end{array} \right] WS_w^{1/2} \end{array} \right)$$

$$S_{Eij} = (m_i - m_j)(m_i - m_j) \quad (3)$$

In order to determine W, the decomposition of Eigenvalue is formed, also, W is the equivalent of Eigenvector with the largest value from Matrix (4).

$$S_C = \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} P_i P_j tr \left(\begin{array}{l} (S_w)^{-1} \times \\ S_w^{1/2} \left[\begin{array}{l} (S_w^{-1/2} S_{ij} S_w^{-1/2})^{1/2} \times S_w^{-1/2} S_{Eij} S_w^{-1/2} (S_w^{-1/2} S_{ij} S_w^{-1/2})^{-1/2} + \\ \frac{1}{\pi_i \pi_j} (\log(S_w^{-1/2} S_{ij} S_w^{-1/2}) - \pi_i \log(S_w^{-1/2} S_i S_w^{-1/2}) - \pi_j \log(S_w^{-1/2} S_j S_w^{-1/2})) \end{array} \right] S_w^{1/2} \end{array} \right) \quad (4)$$

The point that$(S_{Cij}) = \delta_{ij}^2 . \delta_{ij}^2$ is expressed as the Eigenvalue of $S_{Cij}$and Eigenvector with the largest 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 have further understand, (see Fig. 2).
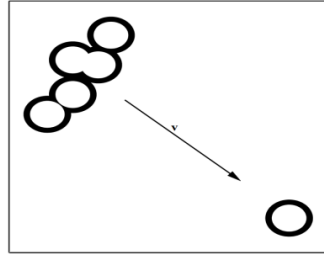
Fig. 2. Map of classes using a vector V [4]

Fig. 2. Is a six – class model that each circle is considered as a class and the circles have a similar radius which shows that of within-class scattering matrix has been equally assumed If the class of on the down right corner, in Fig. 2. Is considered as j_0, on the condition that this class to be well far away, from the rest in the original space, the share of Eigenvalue, $1 \leq i \leq C, i \neq j$ will be dominant on scattering matrix between the classes. Therefore, the result of the direction that has been shown by the V arrow would be known as principal discriminant of between the classes. The result is that to map V the classes of i and j that will projection $i \neq j$ to one cluster high overlapping of classification between the classes that is required to increase the rate of classification error in the projection space. Therefore, from this example, we can conclude that in the estimate of between the class matrixes of Chernoff's criterion all pairs of classes, without being separate in original space are considered that this process leads to a bad performance in separating the classes [35-43].

3   Making scattering matrices based on Boundary and non-Boundary patterns

Boundary pattern is a Datum, by having k neighbor on different class position near the Boundary of decision-making and non-Boundary pattern Datum with neighboring k and label of similar class, its position is away from the decision- making Boundary. Boundary patterns contains sufficient information to have an accurate description of the level classes' separation.

While non-Boundary patterns does not effect on the rate of classification [27], we can use the difference of these patterns in making new scattering matrices [28] so, most of algorithms often suffers from storage of a large number of training samples. The result is that high involvement of memory in response time and also high sensitivity to noise is raised thus, in order to overcome these problems, we use a new algorithm naming class boundary persevering CBP [26, 45].

In proposing framework, first by using the algorithm of (CBP) we divide training sets of X into two sub - bordering of X_B and non-bordering of X_{NB}.

3.1 Smoothing the class boundaries

Placing noises on the borders of the class causes reducing the rate of classification so, Placing noises on the borders of the class causes reducing the rate of classification so, in order to deal the noises [29], [30] filter (Wilson ENN) is normally used which is often known as noisy [31] also, in tests, in order to discard harmful instances misclassified by ENN, we use a KNN classification with (k=3).

3.2 distinguishing between Boundary and non- Boundary instances

After the implementation of the first step, a new scheme using geometrical characteristics of class underlying distribution to partition the initial set to two sub sets Boundary and non- Boundary is used. First each series of pattern x a reachable R(x) is formed which containing samples that belong to $\omega(x)$ that lay to the nearest enemy (a sample by different class label) [32]. R(x) is an available set of X which is defined in respect to the ith the nearest the enemy $\xi_i(x)$ of x defined as:

$$R_i(x) = \left\{ y \in X : \psi(x) = \psi(y) \wedge \|x - y\|_2 \le \|x - \xi_i(x)\|_2 \right\}$$

where

$$\xi_i(x) = \arg\min \|x - z\|_2$$

$$z \in X$$

$$\psi(z) \ne \psi(x)$$

$$z \ne \xi_j(x), j = 1, ..., i - 1 \tag{15}$$

To consider general overview of around the sample x , the sets of $I(x)$ which contains the arbitrary number k_R ($k_R = 3$) from the nearest enemy $\xi_i(x)$ that we define by Eq. (6) .To avoid overlap of the enemies of x pattern towards each other so, the nearest next enemy is selected in a way that the angle between the line connecting the pattern of the new enemy pattern of x and connecting line of former enemies to pattern of to be placed at an angle - more than an arbitrary angle Φ ($\Phi = 20$).

$$I(x) = \arg\min \sum_{i \in J} \|x - \xi_i(x)\|_2$$

$$J \subseteq \{1, ..., n\}$$

$$|J| = k_R$$

$$\Phi(\xi_i(x) - x, x - \xi_i(x)) \ge \Phi_R, \forall_{i,j \in J, i \ne j} \tag{6}$$

$\Phi(.,.)$ The definer of the angle is between the two vectors. Now, in order to find the way of samples' dispersion in the space, the Cosines simulation is used. So, Cosine simulation between x and y with regard to the enemy is obtained through formula of 7:

$$C_{i,x}(y) = \frac{\langle y - x, \xi_i(x) - x \rangle}{\|y - x\|_2 \cdot \|\xi_i(x) - x\|_2} \tag{7}$$

If the friendly sample of y to be near connection between x and $\xi_i(x)$ it means Cosine similarity will be positive and x will be within non - pattern border. Then, we must calculate $C_{i,x}(x)$ for all samples y within each $R_i(x)$, and all the enemies of $I(x)$ will be calculated like the 8 relationship.

$$S_i(x) = \left\{C_{i,x}(y), \forall y \in R_i(x)\right\}$$
$$S(x) = \bigcup_{i \in I(x)} S_i(x) \tag{8}$$

Because all the samples of $y \epsilon R_i(x)$ their position within a sphere passing through s(x) $\xi_i(x)$ and centered at x. so the distribution of casinos' values in s(x) specifies the ratio of scattering friendly instances of y around the x. If the values to be positive in s(x) the more y samples in the most common part of the circle and conical whose summit is in x position and its vector is placed in $x - \xi_i(x)$ (the width of conical entrance by the value of T is controlled). from the other side, the large negative value in s(x) shows the sample of y is out of the cone and the criteria of distinctions between the Boundary and non-Boundary sets of X is expressed as Eq. (9) that the median(s(x)) stating the average of s(x).

$$X_B = \left\{x \in X : midian(S(x)) < -T \vee |R_i(x) \le 2|\right\}$$
$$X_{NB} = X - X_B \tag{9}$$

3.3   Making new scattering matrices:

Now, we can make scattering matrices based on the pattern of Boundary and non – Boundary patterns Eq. (7).

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

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

$n^{(B)}$ The number of the patterns of Boundary set and $n^{(NB)}$, is the number of models of non-Boundary set. As it is shown, in designing scattering matrix of between-class of $\mathbf{S}^{(b)}$, out of the difference between the pattern of the Boundary pattern and the class means and the design of within-class scatter matrix $\mathbf{S}^{(w)}$ uses the difference between non–Boundary patterns and the class means is used [53-55].

3.4 Discussing about the effectiveness of the proposing measure

Scattering matrix between-classes S$^{(b)}$ which it's prove is mentioned in [8].

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

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

$n(y_u)$ The number of samples belonging to the class y_u This change of formulation based on the distance of weight is between the pair of sample data so, in the formula of (8) , the values of non- zero weights ,to scattering between non-bordering patterns , by similar label is specified and the values of zero – weight represents the difference between non-bordering patterns [44, 46-52]. As it is observed, the patterns which are in the non - Boundary region do not have any effect on are in calculating between classes , classes , as well as they separating between the classes because they have been well-separated therefore, the other point is that this reality that based on the criteria of Boundary and non- Boundary patterns , the Boundary patterns between two class that in original space have been well–separated , is classified as non- Boundary pattern and does not have any effect in the estimate of scattering matrix between the classes . The goal designing scattering matrix (7) is to find W direction that by using it in Chernoff's criteria we can maximize Chernoff's j_c (3) optimizer and improve the number of specified patterns. Therefore, it is evident that in the obtained mapping space, the classes' overlapping is decreases considerably then based on this we replace new scattering matrices in Chernoff's criterion:

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

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

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

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

$$S_c W = \lambda S_w W$$

$$A = [W_1, W_2, ..., W_d] \quad (10)$$

By forming analyzing matrix (10), we can calculate W vector equal to Eigenvector which is the biggest Eigenvalue of analysis.

3.5 implementation Algorithm of CBPHDA:

Input: train data matrix $X = \{X_i \mid X_i \in R^d\}$ and $\{y_i \mid y_i \in \{1,2,...,l\}$ is the class label of $X_i\}$

Output: transformation matrix W_c

1. Obtain $X^{(NB)}$ and $X^{(B)}$ based on algorithm RPS

2. for each class $i = 1,...C$

a. $X_{(i)} \leftarrow \{X_j\}_{y_j=i}$

c. $S^{(b)} \leftarrow S^{(b)} + \left(X^B - m(i)(1_{n^{(B)}})^t\right)\left(X^B - m(i)(1_{n^{(B)}})^t\right)^t$ where

$1_{n^{(B)}}$ denote the $n^{(B)}$-dimensional vector with ones.

d. $S^{(w)} \leftarrow S^{(w)} + \left(X^{NB} - m(i)(1_{n^{(NB)}})^t\right)\left(X^{NB} - m(i)(1_{n^{(NB)}})^t\right)^t$ where

$1_{n^{(NB)}}$ denote the $n^{(NB)}$-dimensional vector with ones.

e. $S_C \leftarrow \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} P_i P_j tr \left((S_w)^{-1} \times S_w^{1/2} \left[\left(S_w^{-1/2} S_{ij}^{(w)} S_w^{-1/2} \right)^{-1/2} \times S_w^{-1/2} S_{ij}^{(b)} S_w^{-1/2} \left(S_w^{-1/2} S_{ij}^{(w)} S_w^{-1/2} \right)^{-1/2} + \frac{1}{\pi_i \pi_j} (\log(S_w^{-1/2} S_{ij}^{(w)} S_w^{-1/2}) - \pi_i \log(S_w^{-1/2} S_i^{(w)} S_w^{-1/2}) - \pi_j \log(S_w^{-1/2} S_j^{(w)} S_w^{-1/2})) \right] S_w^{1/2} \right).$

3. Obtain the transformation matrix

a. $S_c \Psi = \lambda S_w \Psi$ where $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_d$.

b. $A = [w_1, w_2,....,w_d]$

4 The experiments

In this part, some tests has been introduced to prove the effectiveness of proposing method. The tests have been carried out on data based which has been downloaded from UCI Machine Learning [17]. The list of these Datasets have been mentioned in table (1) and most of these datasets have been used in the articles of [2, 1, 12, 13]. Column D is the equivalent of the best chosen characteristic for the act of classification. To compare proposing method 4 other method has been used then, all of these methods, by using techniques have tried to solve the issue of data heterogeneity for increasing the rate of classification in LDA , therefore, based on this, for comparison, these methods have been used. Available unclear values in the Datasets have been replaced by average value of related features. The Output of proposing method (CBPHDA) has been compared with (FDA) Fisher Discriminant Algorithm (HAD) Heteroscedastic discriminant analysis (BHDA) Boundary Heteroscedastic discriminant analysis (SCDA) Super-class discriminant analysis) [34] other methods, which indicates that the output CBPHDA better than other methods.

Table 1. The UCI dataset used for the experiments [17]

Dataset name	Number data	Number class	Number feature(number dimension)
Haberman	306	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
Teaching Assistant Evaluation	151	3	5
Iris	150	3	5
Soybean	47	4	36
Breast cancer Wisconsin	699	2	11
Hayes Roth	132	3	5
25PDB	1674	4	64

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

Dataset name	FDA	D	HDA	D	BHDA	D	SCDA	D	CBPHDA	D
Haberman	61.8056	1	55.6999	1	74.9247	1	64.5161	1	99.8623	1
Australian credit	67.6713	1	66.9161	1	74.4988	1	81.3846	1	81.4815	1
German credit	63.2000	1	58.4000	1	65.3000	1	69.5000	1	87.9315	1
Primary tumor	61.3387	2	63.9947	1	65.9777	1	78.2902	1	93.7500	1
Banknote authentication	97.4447	1	95.2618	1	97.3348	1	96.7201	1	98.6631	1
Vote	75.7558	2	94.2653	1	95.4979	1	94.2318	1	97.108	1
hepatitis	64.9451	2	64.1758	2	65.3956	1	78.5714	1	78.7121	1
Liver	57.4118	2	61.1513	1	67.7059	1	77.1429	1	86.9565	1
Zoo	88.2117	2	89.4038	1	96.3027	1	100.000	1	100.000	1
Wine	85.8443	1	87.8657	1	94.5050	1	94.9346	1	95.9583	1
new-thyroid	42.4026	1	45.9870	1	79.2338	1	99.5238	1	100.000	1
Teaching Assistant Evaluation	62.9583	1	52.4167	1	66.3333	1	79.4168	4	90.1667	2
Iris	87.8711	1	72.6417	1	98.7222	1	97.3333	1	100.000	1
Soybean	65.1720	1	70.7988	1	76.5000	1	75.5000	1	80.2288	2
Breast cancer Wisconsin	88.5300	2	87.2298	1	96.5714	1	96.4161	1	97.7391	1
Hayes Roth	82.4176	1	80.4505	1	86.8706	1	85.4326	1	87.9524	1
25PDB	64.2314	1	66.2146	1	83.3326	1	83.3333	1	93.9314	1

4.1 the steps of doing the experiment

To evaluate the effectiveness of proposing method from Fisher classifier from cross-validation strategies: 1) leave-one-out (LOO) and 2) 10-fold cross validation has been used. in the strategy of leave-one-out(LOO) cross-validation the input data is used for teaching and the rest of the data for recognition and experiment although leave-one-out(LOO) cross-validation is a good method to evaluate effectiveness [18][19], it has largely been criticized by the researchers. Hence, in this paper 10-fold cross-validation is used to do the act of classification. In order to avoid the singularity problem of within-class scattering matrix which is a current way of using linear classification analysis, a credit method introduced in [20] was used. First, the ranking of scattering matrix has been said in (7) so, if its ranking was not complete, *aI* has been added to them which a= 0001 and I is like unit matrix. To avoid problems with log and square root of matrix A inverse, of mentioned method in [11] has been used, therefore; to calculate F function, of specific A matrix has been used. A matrix is analyzed as specific analysis (VDV^{-1}) 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 $f(A) = Vf(D)V^{-1}$ 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 [9]. In order to do this, a positive small fixed amount has been added to specific matrix D either negative or zero. Also in the Fig. 4. And Fig. 5. the proportion of the used methods in the experiments with CBPHDA on the using datasets has been shown.
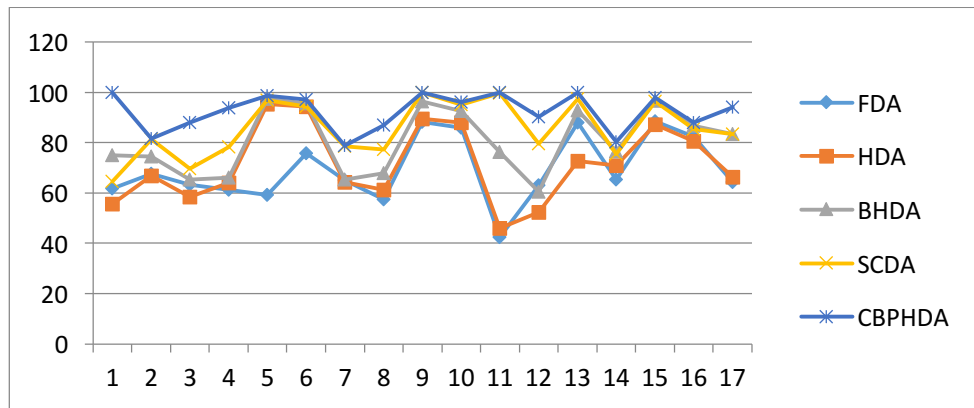

Fig. 4. The output exhibition of the results of experiment on the using datasets.
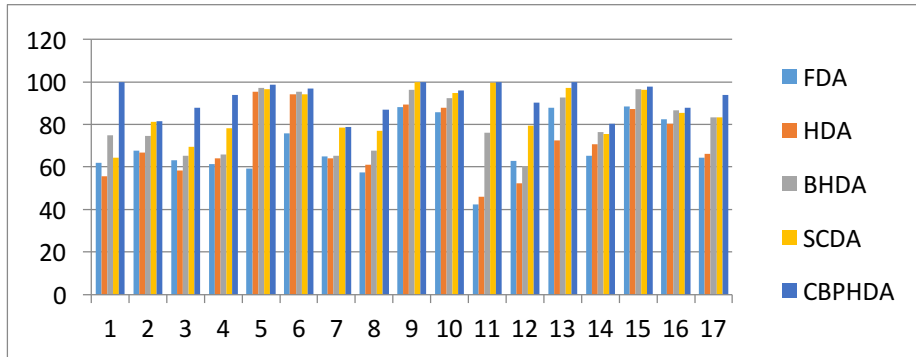
Fig. 5. The column display of the experiment results on the using datasets

4.2 discussion on experimental method's outputs

This section is devoted to discuss the observations resulted from the conducted experiments on the datasets however; the experiments showed that in the database of Iris between class of 1-2 and also 1-3 there is no bordering sample but , between class of 2&3 there are 20 bordering sample thus, we can conclude that the distance of class 1 in proportion to 2&3 is so far and, as result, the scattering matrix between the class of 1&2 and other class has no influence on designing Chernoff's criterion so, it is better in this state instead of using the estimate of distance between class pair Fisher's linear classification which is based on the mean intervals is being used because the best method for classifying the distanced classes are Fisher's classifier.

5   Conclusion

In this paper by using a new Algorithm, we could specify Boundary and non- Boundary patterns and draw scattering matrices based on the pattern's Boundary as well as their non- Boundary however; we theoretically showed that using these matrices cause the increasing number of extracted features and also removing the limitation the number of extracted features by Chernoff's criterion.

6   References

[1]. Fukunaga, K. (2013). Introduction to statistical pattern recognition. Academic press.
[2]. Duin, R. P. W., & Loog, M. (2004). Linear dimensionality reduction via a heteroscedastic extension of LDA: the Chernoff criterion. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 26(6), 732-739.
[3]. Sugiyama, M. (2007). Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. The Journal of Machine Learning Research, 8, 1027-1061.

[4]. ReinholdHaeb-Umbach, M. (2000). MULTI-CLASS LINEAR DIMENSION REDUCTION BY GENERALIZED FISHER CRITERIA. In The Proceedings of the 6~ (th) International Conference on Spoken Language Processing (Volume II).

[5]. Na, J. H., Park, M. S., & Choi, J. Y. (2010). Linear boundary discriminant analysis. Pattern Recognition, 43(3)

[6]. Kim, H., Drake, B. L., & Park, H. (2007). Multiclass classifiers based on dimension reduction with generalized LDA. Pattern recognition, 40(11), 2939-2945.

[7]. Salvi, G. (2003, August). Accent clustering in Swedish using the Bhattacharyya distance. In 15th International Congress of Phonetic Science (pp. 1149-1152).

[8]. Sugiyama, M. (2007). Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. The Journal of Machine Learning Research, 8, 1027-1061.

[9]. Rueda, L., Oommen, B. J., & Henríquez, C. (2010). Multi-class pairwise linear dimensionality reduction using heteroscedastic schemes. Pattern Recognition, 43(7), 2456-2465.

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

[11]. Masip, D., Kuncheva, L. I., & Vitrià, J. (2005). An ensemble-based method for linear feature extraction for two-class problems. Pattern Analysis and Applications, 8(3), 227-237.

[12]. Devijver, P. A., & Kittler, J. (1982). Pattern recognition: A statistical approach (Vol. 761). London: Prentice-Hall.

[13]. Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems.Annals of eugenics, 7(2), 179-188.

[14]. Fisher, Ronald A. "The use of multiple measurements in taxonomic problems."Annals of eugenics 7.2 (1936): 179-188.

[15]. Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).

[16] Friedman, J. H. (1989). Regularized discriminant analysis. Journal of the American statistical association, 84(405), 165-175.

[17]. D.J. Newman, S. Hettich, C.L. Blake, C.J. Merz, UCI repository of machine learning databases, 1998 <http://archive.ics.uci.edu/ml>.

[18]. Vapnik, V., & Chapelle, O. (2000). Bounds on error expectation for support vector machines. Neural computation, 12(9), 2013-2036.

[19]. Chapelle, O., Vapnik, V., Bousquet, O., & Mukherjee, S. (2002). Choosing multiple parameters for support vector machines. Machine learning, 46(1-3), 131-159.

[20]. Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).

[21]. Yaghoubi, A., & Ghaffari, H. R. Improved LDA by using Distributing Distances and Boundary Patterns.

[22]. Wang, Z., Ruan, Q., Liu, S., & Guo, S. (2011). Regularized neighborhood boundary discriminant analysis for facial expression recognition.

[23]E. Marchiori, Hit miss networks with applications to instance selection, Journal of Machine Learning Research 9 (2008) 997–1017.

[24]. E. Marchiori, Graph-based discrete differential geometry for critical instance filtering, in: Joint European Conference on Machine Learning (ECML)/ European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), Bled, Slovenia, 2009.

[25]. E. Marchiori, Class conditional nearest neighbor for large margin instance selection, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 364–370.

[26]. Nikolaidis, K., Goulermas, J. Y., & Wu, Q. H. (2011). A class boundary preserving algorithm for data condensation. Pattern Recognition, 44(3), 704-715.

[27]. H. Brighton, C. Mellish, Advances in instance selection for instance-based learning algorithms, Data Mining Knowledge Discovery 6 (2002) 153–172.

[28]. E. Marchiori, Hit miss networks with applications to instance selection, Journal of Machine Learning Research 9 (2008) 997–1017.

[29]. B. Smyth, M.T. Keane, Remembering to forget, in: Proceeding of the 14th International Conference on Artificial Intelligence, 1995, pp. 377–382.

[30]. D.R. Wilson, T.R. Martinez, Reduction techniques for instance-based learning algorithms, Mach. Learning 38 (2000) 257–286.

[31]. Y. Cheng, Mean shift, mode seeking, and clustering, IEEE Trans. Pattern Anal. Mach. Intell. 17 (1995) 790–799.

[32]. B.V. Dasarathy, Minimal consistent set (MCS) identification for optimal nearest neighbor decision systems design, IEEE Trans. Systems Man Cybernet. 24 (3) (1994) 511–517.

[33]. ARIF, M., WANG, G., & BALAS, V. E. (2018). Secure VANETs: Trusted Communication Scheme between Vehicles and Infrastructure Based on Fog Computing. Studies in Informatics and Control, 27(2), 235-246.

[34]. Zhu, X. (2013). Super-class Discriminant Analysis: A novel solution for heteroscedasticity. Pattern Recognition Letters, 34(5), 545-551.

[35] Gheisari, Mehdi, and Mehdi Esnaashari. "Data Storages in Wireless Sensor Networks to Deal With Disaster Management." Emergency and Disaster Management: Concepts, Methodologies, Tools, and Applications. IGI Global, 2019. 655-682.

[36]Mehdi Gheisari, Hamed Baloochi, M.G., Khajehyouse_, M.: An evaluation of two proposed systems of sensor datas storage in total data parameter. International Geoinformatics Research and Development Journal (March 2012)

[37] Porkar, P., Gheisari, M., Bazyari, G.H., Kaviyanjahromi, Z.: A comparison with two sensor data storages in energy. In: ICCCI. ASME Press (2011)

[38] Rezaeiye, P.P., Gheisari, M.: Performance analysis of two sensor data storages. In: Proceedings of 2nd International Conference on Circuits, Systems, Communica- tions & Computers (CSCC). pp. 133{136 (2011)

[39] Rezaeiye, P.P., Rezaeiye, P.P., Karbalayi, E., Gheisari, M.: Statistical method used for doing better corneal junction operation. In: Material and Manufacturing Tech- nology III. Advanced Materials Research, vol. 548, pp. 762{766. Trans Tech Pub lications (9 2012)

[40] Rezaeiye, P.P., Rezaeiye, P.P., Beig, E.F.G.M., Mohseni, H., Kaviani, R., Gheisari, M., Golzar, M.: Agent programming with object oriented (c++). In: (ICECCT). pp. 1{10. IEEE (2017)

[41] GHEISARI, M.. Design, Implementation and Evaluation of SemHD: a New Semantic Hierarchical Sensor Data Storage. Indian Journal of Innovations and Developments, [S.l.], p. 115-120, mar. 2012. ISSN 2277 – 5390

[42] Gheisari, M., Esnaashari, M. (2017). A survey to face recognition algorithms: advantageous and disadvantageous. Journal Modern Technology & Engineering, 2(1), 57-65.

[43] M. Gheisari *et al*., "NSSSD: A new semantic hierarchical storage for sensor data," *2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, Nanchang, 2016, pp. 174-179.

[44] M. Gheisari, G. Wang and M. Z. A. Bhuiyan, "A Survey on Deep Learning in Big Data," *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, Guangzhou, 2017, pp. 173-180.

[45] M. Jafari, J. Wang, Y. Qin, M. Gheisari, A. S. Shahabi and X. Tao, "Automatic text summarization using fuzzy inference," *2016 22nd International Conference on Automation and Computing (ICAC)*, Colchester, 2016, pp. 256-260.

[46] Gheisari, Mehdi. "The Effectiveness of Schema Therapy Integrated with Neurological Rehabilitation Methods to Improve Executive Functions in Patients with Chronic Depression." Health Science Journal 10.4 (2016).

[47] Gheisari, Mehdi, et al. "MAPP: A Modular Arithmetic Algorithm for Privacy Preserving in IoT." Ubiquitous Computing and Communications (ISPA/IUCC), 2017 IEEE International

Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on. IEEE, 2017.

[48] Ashourian, Mohsen, Mehdi Gheisari, and Ali Hashemi. "An Improved Node Scheduling Scheme for Resilient Packet Ring Network." Majlesi Journal of Electrical Engineering 9.2 (2015): 43.

[49] M. Gheisari, G. Wang and S. Chen, " IoT-SDNPP: A Method for Privacy-Preserving in IoT-Based Smart City with Software Defined Networking"18th International Conference on Algorithms and Architectures for Parallel Processing Guangzhou, China | 15-17 Nov 2018

[50] Wenyin Yang, Guojun Wang, Kim-Kwang Raymond Chood, and Shuhong Chen, "HEPart: A balanced hypergraph partitioning algorithm for big data applications," Future Generation Computer Systems, 83: 250-268, June 2018.

[51] Yinglong Dai and Guojun Wang*, "Analyzing Tongue Images Using a Conceptual Alignment Deep Autoencoder," IEEE ACCESS, 6: 5962-5972, March 2018.

[52] Sancheng Peng, Guojun Wang*, Yongmei Zhou, Cong Wan, Cong Wang, Shui Yu, and Jianwei Niu, "An Immunization Framework for Social Networks through Big Data Based Influence Modeling," IEEE Transactions on Dependable and Secure Computing, PP(99): 1-1, July 2017.

[53] A. Javadpour and H. Memarzadeh-Tehran and F. Saghafi, "A temperature monitoring system incorporating an array of precision wireless thermometers", 2015 International Conference on Smart Sensors and Application (ICSSA)

[54] A. Javadpour and H. Memarzadeh-Tehran, "A wearable medical sensor for provisional healthcare", 2015 2nd International Symposium on Physics and Technology of Sensors (ISPTS)

[55] A. Javadpour and S. Kazemi Abharian and G. Wang," Feature Selection and Intrusion Detection in Cloud Environment Based on Machine Learning Algorithms", 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)