

Nonlinear Discriminant Analysis Method Based on the Class Cover and Its Application

Liwen HUANG

*College of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou 362000,
China*

E-mail: Livern@126.com

Abstract This paper introduces the related concepts of the hybrid spherical-shaped dataset and proposes a new discriminant analysis method based on the spherical-shaped dataset (SDAM), then SDAM is further improved by the idea of the class cover and presents the nonlinear discriminant analysis method (NDAM). To demonstrate the effectiveness of these two methods, this work constructs seven hybrid spherical-shaped datasets and uses nine UCI datasets. Numerical experiments on these examples indicate that SDAM can preferably solve the discriminant problem for the hybrid spherical-shaped dataset, but this method does not always work well for real datasets; NDAM overcomes the drawbacks of SDAM and better solves the discriminative problem of real datasets. Besides, it has better stability.

Keywords nonlinear discriminant analysis; discriminant rule; class cover; hybrid spherical-shaped dataset

1 Introduction

The discriminant analysis method (DAM) is a statistical method used to distinguish the samples from different classes, and it was introduced by Fisher^[1] for two-class problems in the 1930s and was extended to more than two classes by Rao^[2]. Over the past decades, DAM has been widely improved, and many new theories and discriminant rules have been proposed, such as fuzzy linear discriminant analysis^[3], discriminant analysis of feature selection and error probability^[4], fast and robust discriminant analysis^[5], weighted maximum margin discriminant analysis with kernels^[6]. These methods have been successfully applied to pattern recognition, geology, remote sensing, medicine and other fields^[7–9], and achieved good results.

The traditional discriminant analysis is not very efficient when a dataset has certain types of complexity. Bayes discriminant analysis requires that each class follows a normal distribution, and the covariance matrices of all classes are equal^[10]. Fisher linear discriminant analysis (FLDA) performs well when the datasets are the large between-class difference and the small within-class difference^[11]. Linear discriminant analysis (LDA) achieves a good effect on the classification of many low-dimensional sample data^[12,13], since it is simple and effective in use, it is still one of the widely used methods. However, it may not be suitable for the discrimination

Received May 25, 2022, accepted December 11, 2022

Supported by the National Natural Science Foundation of China (52070119) and the Natural Science Foundation of Fujian Province (2021J01970)

and classification of complex data, especially for the discrimination of high-dimensional sample data^[14]. To better handle the discriminant classification problem of complex data, many improved discriminant analysis methods have been proposed. For the discriminant problem the class surrounded by the other class, Yang^[15] proposed a new discriminant method based on volcano mouth data and verified its performance, Huang and Su^[16] have modified this method by using the idea of linear and nonlinear. Aiming at the discriminant problem of complex data, Pacheco, et al.^[17], Brusco, et al.^[18] and Huang^[19] discussed the discriminant method based on selecting optimal variables. Recently, many methods, including [20–22], have been developed to process the discriminant problem of high-dimensional data. In order to test the validity of the classification methods, the optimal classification method was found to be different through a large number of datasets^[23]. In fact, various assumptions are often made when constructing classification methods. Due to the complexity of the data, actual data may not always conform to these assumptions. Moreover, many classification methods involve parameter setting problems, including the idea of the kernel-based method^[24–26], and the results of classification are often associated with the selection of kernel function or the parameters. For exploring a simple and effective method, this paper presents a new nonlinear discriminant analysis method, and further improves this method for wide application in real datasets. The purpose is to enhance the stability and the correct ratio of the discriminant model.

In the following sections, this paper will discuss the discriminant analysis method based on spherical-shaped data (SDAM), nonlinear discriminant analysis method (NDAM) based on the class cover, numerical experiments the paper has been conducted, and the conclusions.

2 Discriminant Analysis Method Based on Spherical-Shaped Data

In practical problems, the shape of the data is diverse, including spherical-shaped data, bar-shaped data, irregularly shaped data, etc. Spherical data are common in astronomy, biology, geology, medicine and meteorology¹. The classification of spherical data that are naturally generated in many data mining and knowledge discovery applications such as text classification, visual scenes categorization and gene expression analysis^[28]. In the past few decades, a large literature on the statistical analysis of spherical data has developed, see [29], [30], [31], etc. Some scholars refer to the set of points on a sphere as spherical shell data^[32,33] or spherical data^[34,35], while others refer to the set of points on and within a sphere as spherical data^[36,37]. In order to better distinguish spherical shell data and spherical data, and discuss the classification of such data, the following concepts of spherical shell-shaped class and spherical class are introduced.

Definition 1 *If set S satisfies the following conditions: 1) The shape of the data elements in S is spherical. 2) For any element $P \in S$, $\exists r > 0$ such that the distance $d(P, P_0) \leq r$, where P_0 is a fixed element of set S , and r is a finite positive number. Then S is called a spherical-shaped class.*

Definition 2 *If S_1 and S_2 are two spherical classes, and S_2 is inside of S_1 such that $S = S_1 - S_2$, then S is called a spherical shell-shaped class.*

For example, in the two-dimensional space, from Definition 1 and Definition 2, G_p and G_q

¹More details and thorough discussions about the statistics of spherical data in particular and directional data in general can be found in [27].

are the spherical-shaped classes in Figure 1(a). However, G_p is the spherical-shaped class, and G_q is the spherical shell-shaped class in Figure 1(b).

Unless otherwise specified, we use the following notation in this paper. Let G_i be a spherical-shaped or a spherical shell-shaped class, $i = 1, 2, \dots, k, k > 0$. $x_{(\alpha)}^{(i)}$ is the α -th sample of G_i , $\alpha = 1, 2, \dots, n_i$, $\bar{x}^{(i)} = (\bar{x}_1^{(i)}, \bar{x}_2^{(i)}, \dots, \bar{x}_m^{(i)})$ is the average of G_i , and n_i is the sample size of the i -th class. $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m)$ is the average of the samples of G_1, G_2, \dots, G_k , and $x = (x_1, x_2, \dots, x_m)$ is a sample of the class to be determined.

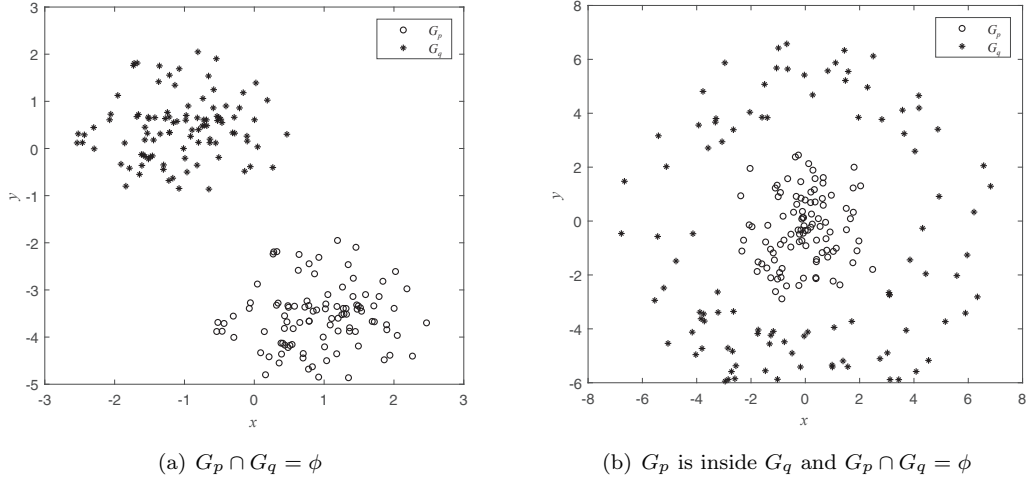


Figure 1 The relation between G_p and G_q

2.1 Theoretical Basis

Definition 3 Let G_i be a spherical-shaped or a spherical shell-shaped class, $i = 1, 2, \dots, k$. For any given two classes G_p and G_q , if $G_p \cap G_q = \phi$, $1 \leq p \neq q \leq k$, then the dataset consisting of G_1, G_2, \dots, G_k is called a hybrid spherical shell-shaped dataset.

$G_p \cap G_q = \phi$ has two implications in the above definition. For example, in the two-dimensional space, the first implication is that G_p and G_q are disjoint, as shown in Figure 1(a), denoted by $G_p \tilde{\cap} G_q = \phi$; the second implication is that G_p is surrounded by G_q , but G_p and G_q do not overlap, as shown in Figure 1(b), denoted by $G_p \tilde{\subset} G_q = \phi$ or $G_q \tilde{\supset} G_p = \phi$.

Definition 4 If $x \in G_p$, $\bar{x}^{(p)}$ is the average of G_p , and \bar{x} is the average of the samples of G_1, G_2, \dots, G_k , then the distance between the sample x and G_p is defined as follows:

$$d(x, G_p) = \sqrt{\sum_{j=1}^m \left(\frac{x_j - \bar{x}_j^{(p)}}{\bar{x}_j} \right)^2}, \quad \text{where } \bar{x}_j \neq 0. \quad (1)$$

Obviously, $d(x, G_p)$ is dimensionless.

If the radius of G_p is denoted by $d^{(p)}$, then $d^{(p)}$ can be defined as follows.

Definition 5 If set S' represents the edge samples of G_p , and

$$d^{(p)} = \sup_{x \in S'} d(x, G_p), \quad (2)$$

then $d^{(p)}$ is called the radius of G_p .

Generally, it is hard to get the exact value of $d^{(p)}$. In practice, $d^{(p)}$ can be usually replaced by the following form:

$$d^{(p)} = \max_{1 \leq \alpha \leq n_p} d(x_{(\alpha)}^{(p)}, G_p). \quad (3)$$

To better distinguish the samples between two classes, the following two theorems are introduced.

Theorem 1 Let G_i be a spherical-shaped or a spherical shell-shaped class, $d^{(i)}$ be the radius of G_i , $i = p, q$, and $G_p \tilde{\cap} G_q = \phi$. For any given sample x ,

- 1) If $x \in G_p$, then $\frac{d(x, G_p)}{d^{(p)}} < 1$ and $\frac{d(x, G_q)}{d^{(q)}} < 1$.
- 2) If $x \in G_q$, and $x \notin G_p$, then $\frac{d(x, G_p)}{d^{(p)}} > 1$ and $\frac{d(x, G_q)}{d^{(q)}} < 1$.
- 3) If x is a sample of the class to be determined, then

$$\frac{d(x, G_p)}{d^{(p)}} = \begin{cases} < 1, & x \text{ is inside of } G_p, \\ > 1, & x \text{ is outside of } G_p, \end{cases} \quad (4)$$

and

$$\frac{d(x, G_q)}{d^{(q)}} = \begin{cases} < 1, & x \text{ is inside of } G_q, \\ > 1, & x \text{ is outside of } G_q. \end{cases} \quad (5)$$

Theorem 2 Let G_i be a spherical-shaped or a spherical shell-shaped class, $d^{(i)}$ be the radius of G_i , $i = p, q$, and $G_p \tilde{\cap} G_q = \phi$. For any given sample x ,

- 1) if $x \in G_p$, then $\frac{d(x, G_p)}{d^{(p)}} < 1$ and $\frac{d(x, G_q)}{d^{(q)}} > 1$.
- 2) if $x \in G_q$, then $\frac{d(x, G_p)}{d^{(p)}} > 1$ and $\frac{d(x, G_q)}{d^{(q)}} < 1$.
- 3) if x is a sample of the class to be determined, then

$$\frac{d(x, G_p)}{d^{(p)}} = \begin{cases} < 1, & x \text{ is inside of } G_p, \\ > 1, & x \text{ is outside of } G_p, \end{cases} \quad (6)$$

and

$$\frac{d(x, G_q)}{d^{(q)}} = \begin{cases} < 1, & x \text{ is inside of } G_q, \\ > 1, & x \text{ is outside of } G_q. \end{cases} \quad (7)$$

Theorem 1 and Theorem 2 reveal the relationship between internal or external points and the classes. When there are more than two classes, the relationship between the samples and the classes becomes more complicated, so it is necessary to analyse the relationships among the classes. The following three definitions are about the relationship among the classes.

Definition 6 If G_i is a spherical shell-shaped class, $i = 1, 2, \dots, s-1$, G_s is a spherical shaped or a spherical shell-shaped class, and $G_s \tilde{\cap} G_{s-1} \tilde{\cap} \dots \tilde{\cap} G_1$, then the relation of these classes is defined as the s -th layer inclusion relation, and G_i ($1 \leq i \leq s$) is the class of the i -th layer inclusion relation.

Definition 7 If G_i is a spherical shell-shaped class, $i = 1, 2, \dots, s-1$, G_s and G'_s are spherical-shaped or spherical shell-shaped classes, $G_s \tilde{\subset} G_{s-1} \tilde{\subset} \dots \tilde{\subset} G_1$, $G'_s \tilde{\subset} G_{s-1} \tilde{\subset} \dots \tilde{\subset} G_1$, and $G_s \tilde{\cap} G'_s = \phi$, then the relationship between G_s and G'_s is defined as the s -th layer coordinative relation, denoted by $G_s \tilde{\cup} G'_s$.

Definition 7 can be generalized to the case when there are more than two classes.

Definition 8 If G_i is a spherical shell-shaped class, $i = 1, 2, \dots, s-1$, G_{s_j} is a spherical-shaped or a spherical shell-shaped class, $G_{s_j} \tilde{\subset} G_{s-1} \tilde{\subset} \dots \tilde{\subset} G_1$, and $G_{s_j} \tilde{\cap} G_{s_l} = \phi$, $1 \leq j \neq l \leq t$, then the relationship among these classes ($G_{s_1}, G_{s_2}, \dots, G_{s_t}$) is defined as the s -th layer coordinative relation, denoted by $G_{s_1} \tilde{\cup} G_{s_2} \tilde{\cup} \dots \tilde{\cup} G_{s_t}$.

If G_1, G_2, \dots, G_s are coordinative relation, by Definition 8, then the relationship can be denoted by $G_1 \tilde{\cup} G_2 \tilde{\cup} \dots \tilde{\cup} G_s$. Especially, when these classes are the first layer coordination relation, they can also be denoted by $G_1 \tilde{\cap} G_2 \tilde{\cap} \dots \tilde{\cap} G_s$.

Theorem 3 Let G_i be a spherical shell-shaped class, $i = 1, 2, \dots, s-1$, G_s be a spherical-shaped or a spherical shell-shaped class, $G_s \tilde{\subset} G_{s-1} \tilde{\subset} \dots \tilde{\subset} G_1$, and $d^{(i)}$ be the radius of G_i , $i = 1, 2, \dots, s$. For any given sample x , if $x \in G_t$ ($1 \leq t \leq s$), then

$$\frac{d(x, G_i)}{d^{(i)}} = \begin{cases} < 1, & 1 \leq i \leq t, \\ > 1, & t < i \leq s. \end{cases} \quad (8)$$

Proof From Theorem 1, for any given sample x , if $x \in G_t$ ($1 \leq t \leq s$), then $\frac{d(x, G_t)}{d^{(t)}} < 1$. Given a class G_i , when $G_t \tilde{\subset} G_i$ ($1 \leq i < t$), $\frac{d(x, G_i)}{d^{(i)}} < 1$; and when $G_t \tilde{\supset} G_i$ ($t < i \leq s$), $\frac{d(x, G_i)}{d^{(i)}} > 1$. Thus, if $x \in G_t$ ($1 \leq i \leq t$), then

$$\frac{d(x, G_i)}{d^{(i)}} = \begin{cases} < 1, & 1 \leq i \leq t, \\ > 1, & t < i \leq s. \end{cases} \quad (9)$$

Theorem 4 Let G_i be a spherical-shaped or a spherical shell-shaped class, $d^{(i)}$ be the radius of G_i , $i = 1, 2, \dots, s$, and $G_1 \tilde{\cup} G_2 \tilde{\cup} \dots \tilde{\cup} G_s$. For any given sample x , if $x \in G_t$ ($1 \leq t \leq s$), then

$$\frac{d(x, G_i)}{d^{(i)}} = \begin{cases} < 1, & i = t, \\ > 1, & i \neq t. \end{cases} \quad (10)$$

Proof From Theorem 2, for any given sample x , if $x \in G_t$ ($1 \leq t \leq s$), then $\frac{d(x, G_t)}{d^{(t)}} < 1$. Given a class G_i , when $i \neq t$, $\frac{d(x, G_i)}{d^{(i)}} > 1$. Thus, if $x \in G_t$ ($1 \leq t \leq s$), then

$$\frac{d(x, G_i)}{d^{(i)}} = \begin{cases} < 1, & i = t, \\ > 1, & i \neq t. \end{cases} \quad (11)$$

Theorem 3 and Theorem 4 reveal the relationship between internal or external points and the classes.

For a hybrid spherical shell-shaped dataset, the relationship between a sample and the classes can be revealed by the following theorem.

Theorem 5 Let G_1, G_2, \dots, G_s be a hybrid spherical shell-shaped dataset, and these classes have l ($1 \leq l \leq s$) layer inclusion relation. For any given sample x , $Q(x, G_i)$ and N are denoted by the following form:

$$Q(x, G_i) = \frac{d(x, G_i)}{d^{(i)}}, N_i(x) = \begin{cases} 1, & Q(x, G_i) < 1, \\ 0, & \text{other,} \end{cases} i = 1, 2, \dots, s, N(x) = \sum_{i=1}^s N_i(x). \quad (12)$$

If $N(x) = j$ ($1 \leq j \leq l$), $G_t^{(j)}$ is a class of the j -th layer coordinative relation, and $Q(x, G_t^{(j)}) < 1$ ($1 \leq t \leq l$), then $x \in G_t^{(j)}$.

Proof Suppose G_1, G_2, \dots, G_s have l ($1 \leq l \leq s$) layer inclusion relation, their relations are expressed as $G'_l \tilde{\subset} G'_{l-1} \tilde{\subset} \dots \tilde{\subset} G'_1$, the remaining classes and each layer class have coordinative relation respectively, and G'_j ($1 \leq t \leq l$) is the j -th layer class. From Theorem 3, for any given sample x , if $x \in G'_j$, then

$$Q(x, G'_i) = \begin{cases} < 1, & 1 \leq i \leq j, \\ > 1, & j < i \leq l. \end{cases} \quad (13)$$

Hence, if $x \in G'_j$, then $N(x) = j$.

It can be seen from this if $N(x) = j$, then the sample x belongs to a class of the j -th layer coordinative relation.

Suppose there are r ($1 \leq r \leq s - l$) classes of the j -th layer coordinative relation, denoted by $G_1^{(j)}, G_2^{(j)}, \dots, G_r^{(j)}$, namely $G_1^{(j)} \tilde{\cup} G_2^{(j)} \tilde{\cup} \dots \tilde{\cup} G_r^{(j)}$. From Theorem 4, if $x \in G_t^{(j)}$, then

$$Q(x, G_i^{(j)}) = \begin{cases} < 1, & i = t, \\ > 1, & i \neq t. \end{cases} \quad (14)$$

Thus, if $N(x) = j$, and in the j -th layer inclusion relation, $\exists G_t^{(j)}$ such that $Q(x, G_t^{(j)}) < 1$, then $x \in G_t^{(j)}$.

As mentioned above, suppose the dataset consisting of G_1, G_2, \dots, G_k are a hybrid spherical shell-shaped dataset, these k classes have s ($1 \leq s \leq k$) layer inclusion relation or coordinative relation, then the discriminant rule of SDAM can be established as follows:

Rule 1 Let l ($1 \leq l \leq s$) be the number of layers in the inclusion relation of G_1, G_2, \dots, G_s , for any given sample x , define

$$Q(x, G_i) = \frac{d(x, G_i)}{d^{(i)}}, N_i(x) = \begin{cases} 1, & Q(x, G_i) < 1, \\ 0, & \text{other,} \end{cases} i = 1, 2, \dots, s, N(x) = \sum_{i=1}^s N_i(x). \quad (15)$$

1) If $N(x) = j$ ($1 \leq j \leq l$), and in the class of the j -th inclusion relation, $\exists G_t$ such that $Q(x, G_t) < 1$ ($1 \leq t \leq l$), then $x \in G_t$.

2) If $N(x) = 0$, and $\exists G_t$ such that $Q(x, G_t) = \min_{1 \leq i \leq s} Q(x, G_i) > 1$, then $x \in G_t$.

3) If $N(x) = l + 1$ and $Q(x, G_t) = \min_{1 \leq i \leq s} Q(x, G_i) < 1$, then $x \in G_t$.

2.2 Judgment of the Hybrid Spherical Shell-shaped Dataset

Let G_i be a spherical-shaped or a spherical shell-shaped class, $i = p, q$, then the relationship between G_p and G_q has three forms, namely $G_p \tilde{\cap} G_q = \phi$, $G_p \tilde{\subset} G_q = \phi$ or $G_q \tilde{\supset} G_p = \phi$, and $G_p \tilde{\cap} G_q \neq \phi$.

Therefore, in order to distinguish the relationship between G_p and G_q , we introduce the function as follows.

$$f(G_p, G_q) = \begin{cases} 0, & G_p = G_q, \\ 1, & G_p \tilde{\cap} G_q = \phi, \\ 2, & G_p \tilde{\subset} G_q = \phi, \\ -1, & G_p \tilde{\cap} G_q \neq \phi, \end{cases} \quad 1 \leq p, q \leq k. \quad (16)$$

From (16), the relationship between G_p and G_q can be described by the upper triangular matrix below, $1 \leq p \neq q \leq k$.

$$\begin{bmatrix} 0 & f(G_1, G_2) & f(G_1, G_3) & \cdots & f(G_1, G_k) \\ & 0 & f(G_2, G_3) & \cdots & f(G_2, G_k) \\ & & \ddots & \vdots & \vdots \\ & & & 0 & f(G_{k-1}, G_k) \\ & & & & 0 \end{bmatrix}. \quad (17)$$

According to (17), the relationship between any two classes in G_1, G_2, \dots, G_k can be analysed.

If the dataset consisting of G_1, G_2, \dots, G_k is a hybrid spherical shell-shaped dataset, then the relationship between G_p and G_q has two forms, one is $G_p \tilde{\cap} G_q = \phi$, the other is $G_p \tilde{\subset} G_q = \phi$. In either case, their common feature is that the two classes do not overlap. In other words, these two classes have no intersection area. However, if the dataset consisting of G_1, G_2, \dots, G_k is not of hybrid spherical-shaped type, then these classes may have an intersection area. In general, the more samples the dataset has in the intersection area, the less likely it fits the hybrid spherical shell-shaped dataset. Therefore, by the number of samples in the intersection area among the classes, a dataset can be judged whether it is a hybrid spherical shell-shaped dataset.

Suppose n'_{pq} represents the sample size in the intersection area between G_p and G_q , $1 \leq p, q \leq k$. According to (16), if $f(G_p, G_q) = 1$, see Figure 1(a), or $f(G_p, G_q) = 2$, see Figure 1(b), then $n'_{pq} = 0$. But if $f(G_p, G_q) = -1$, then $n'_{pq} \neq 0$.

From Theorem 1, if $G_p \tilde{\subset} G_q \neq \phi$, then

$$n'_{pq} = \sum_{x \in G_p} N_p(x).$$

From Theorem 2, if $G_p \tilde{\cap} G_q \neq \phi$, then

$$n'_{pq} = u_{pq} + v_{pq},$$

where

$$u_{pq} = \sum_{x \in G_p} N_p^q(x), \quad v_{pq} = \sum_{x \in G_q} N_q^p(x).$$

$$N_i^j(x) = \begin{cases} 1, & Q(x, G_i) < Q(x, G_j) < 1, \\ 0, & \text{other}, \end{cases} \quad i \neq j = p, q.$$

Once n'_{pq} has been determined, the fitting degree of the hybrid spherical shell-shaped dataset is defined as follows.

$$R = \frac{\sum_{p=1}^{k-1} \sum_{q=1}^k n'_{pq}}{n}. \quad (18)$$

It is easy to know that $0 \leq R < 1$. If $R = 0$, then the dataset consisting of G_1, G_2, \dots, G_k is a hybrid spherical shell-shaped dataset, and the closer the value of R is to zero, the better it fits this type of dataset.

2.3 The Method of SDAM

Suppose there are k classes G_1, G_2, \dots, G_k , from Subsection 2.1 and Subsection 2.2, SDAM can be briefly described in the following four steps:

1) Analyse the relations among the classes according to the training sample, and judge the type of dataset. If $R = 0$ or R is close to zero, then proceed to 2); otherwise, this method does not work well and stop.

2) Compute the center point of \bar{x} and $\bar{x}^{(p)}$, where $\bar{x} = \frac{1}{n} \sum_{\alpha=1}^n x_{(\alpha)}$, and $\bar{x}^{(p)} = \frac{1}{n_p} \sum_{\alpha=1}^{n_p} x_{(\alpha)}^{(p)}$, $p = 1, 2, \dots, k$.

3) Compute the radius of G_p , that is, $d^{(p)} = \max_{1 \leq \alpha \leq n_p} d(x_{(\alpha)}^{(p)}, G_p)$, $p = 1, 2, \dots, k$.

4) For any given sample x , compute the value of $Q(x, G_i)$, $i = 1, 2, \dots, k$, and count the number of $N(x)$, then classify its class according to Rule 1.

2.4 Simulation Examples

To demonstrate the improvements of SDAM, we construct a dataset with 4 classes, each class has 100 samples, and the scatter diagram is shown in Figure 2. Figure 2 shows that G_1 is the spherical shell-shaped class, G_2, G_3, G_4 are the spherical-shaped classes, and the dataset consisting of G_1, G_2, G_3, G_4 can be regarded as the hybrid spherical shell-shaped dataset.

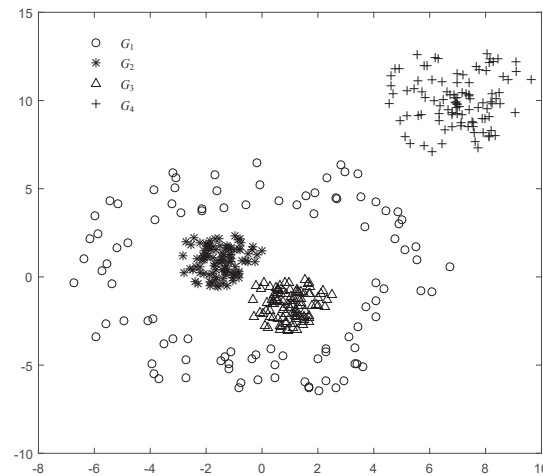


Figure 2 A realization of dataset from simulation setting

We design three experimental schemes, take the top 70, 80 and 90 samples of each class as training samples, and use the remaining samples of each class as test samples. According to the three experimental schemes, FLDA, Ensembles, SVM-Kernel, and SDAM are used to establish the discriminant model. The related results of SDAM are shown in Table 1 and Table 2 is the classification results of different discriminant methods.

Table 1 Three experimental schemes and related results

| No. | Sample number per class | | Class | Average of each class | | $d^{(i)}$ |
|--------|-------------------------|-------------|-------|-----------------------|---------|-----------|
| | Training sample | Test sample | | | | |
| Task 1 | Before 70 | After 30 | G_1 | -0.0865 | -0.0010 | 4.3727 |
| | | | G_2 | -0.9683 | 0.4059 | 0.8452 |
| | | | G_3 | 0.6233 | -0.6648 | 0.8452 |
| | | | G_4 | 4.4315 | 4.2599 | 1.5028 |
| Task 2 | Before 80 | After 20 | G_1 | -0.1440 | -0.1724 | 4.5584 |
| | | | G_2 | -1.0136 | 0.3938 | 0.8869 |
| | | | G_3 | 0.6256 | -0.7424 | 0.8869 |
| | | | G_4 | 4.5320 | 4.5165 | 1.5157 |
| Task 3 | Before 90 | After 10 | G_1 | -0.1064 | -0.1140 | 4.4230 |
| | | | G_2 | -0.9746 | 0.4042 | 0.9840 |
| | | | G_3 | 0.6283 | -0.7041 | 0.9840 |
| | | | G_4 | 4.4527 | 4.4140 | 1.7156 |

Table 2 Classification results of different discriminant methods

| No. | Correct ratio (%) | | | |
|--------|--|--|--|---|
| | FLDA | Ensembles | SVM-Kernel | SDAM |
| Task 1 | 51.79 ^a /53.33 ^b | 85.00 ^a /84.17 ^b | 95.00 ^a /95.00 ^b | 97.86 ^a /98.33 ^b |
| Task 2 | 49.38 ^a /50.00 ^b | 74.38 ^a /73.75 ^b | 96.25 ^a /93.75 ^b | 98.13 ^a /97.50 ^b |
| Task 3 | 50.83 ^a /52.50 ^b | 84.16 ^a /82.50 ^b | 96.11 ^a /92.50 ^b | 99.44 ^a /100.00 ^b |

^a Correct ratio of training sample; ^b Correct ratio of test sample.

Note: SVM-Kernel method derives from SVM-KM-matlab, and learning parameters are as follows: $c = 1000$, $\lambda = 1e-7$, $\text{kerneloption} = 2$, $\text{kernel} = \text{'poly'}$, $\text{verbose} = 1$.

Table 2 shows that the correct ratios of SVM-Kernel, Ensembles, and SDAM are superior to those of FLDA for both training samples and test samples. With the increase of training samples and the decrease of test samples, the correct ratio of Ensembles becomes unstable. For SVM-Kernel, the correct ratio of training samples is more than 95%, and the correct ratio of test samples is more than 92%. This shows SVM-Kernel method has a better classification

effect on the hybrid spherical shell-shaped datasets under the set parameters. Regardless of training samples, or test samples, SDAM proposed in this paper has achieved a good effect.

To examine the effect of datasets with higher dimensions, more classes, and more layers, six different types of datasets are built through data simulation, and SDAM is used to run each dataset 1000 times. Table 3 shows the average results of each dataset.

Table 3 Classification results of SDAM

| Dataset | Dimension | Class | Layer | R | Correct ratio |
|---------|-----------|-------|-------|--------|---------------|
| Data1 | 4 | 5 | 2 | 0.0374 | 96.33% |
| Data2 | 5 | 7 | 4 | 0.0169 | 98.36% |
| Data3 | 6 | 8 | 5 | 0.0826 | 91.72% |
| Data4 | 7 | 7 | 4 | 0.0615 | 93.89% |
| Data5 | 8 | 7 | 5 | 0.0201 | 98.00% |
| Data6 | 9 | 6 | 5 | 0.0383 | 96.22% |

From Table 3, each R is close to 0, so these datasets can be approximated as the hybrid spherical shell-shaped dataset. From the discriminant results of these datasets, SDAM has achieved good results, which shows this method is suitable for this kind of dataset and has good stability.

3 Nonlinear Discriminant Analysis Method Based on the Class Cover

Although DAM is an effective method to predict the classification of new observations, it is not sufficient to deal with the discriminant problem of the hybrid spherical shell-shaped dataset. SDAM overcomes this problem. Compared with some common discriminant analysis methods, SDAM has two potential benefits: It does not need to set any parameters, and it has achieved a good effect on the discriminant problem of the hybrid spherical shell-shaped dataset.

However, in real datasets, many datasets may not be of the hybrid spherical shell-shaped dataset, which limits the application of SDAM. To enhance the performance of SDAM, the class cover problem is introduced into statistical pattern classification^[38–40], and NDAM based on the class cover is proposed. The main idea of NDAM is to approximately treat a non-spherical-shaped class as a class consisting of a finite number of spherical-shaped classes.

For a given non-spherical-shaped class B_p ($1 \leq p \leq k$), the goal is to design a pre-classifier that automatically identifies the spherical-shaped classes in B_p , and these spherical-shaped classes can cover all samples or most of samples of B_p .

Let B_p be the union of k_p spherical-shaped classes, denoted by $G_1^{(p)}, G_2^{(p)}, \dots, G_{k_p}^{(p)}$, where $1 \leq k_p < n_p$, and $c_i^{(p)}$ and $d_i^{(p)}$ are the center point and the radius of $G_i^{(p)}$, where $G_i^{(p)} \cap G_j^{(p)} = \emptyset, 1 \leq i \neq j \leq k_p$.

It is crucial to determine the center point $c_i^{(p)}$ and the radius $d_i^{(p)}$ of $G_i^{(p)}$, and a new adaptive strategy of pre-classifier is proposed for choosing the center points and the radii of B_p . The steps are expressed as follows:

$$1) B = \bigcup_{1 \leq i \leq k, i \neq p} B_i.$$

2) Search the nearest two samples of B_p , and let the initial center point $c_1^{(p)}$ of $G_1^{(p)}$ be the average of these two samples.

3) The sample in B_p closest to the center point $c_1^{(p)}$ is added to $G_1^{(p)}$, and subsequently update the center point $c_1^{(p)}$ until $G_1^{(p)} \cap B \neq \phi$. Once $c_1^{(p)}$ has been determined, then compute $d_1^{(p)}$ by the following form:

$$d_1^{(p)} = \max_{x \in G_1^{(p)}} d(x, G_1^{(p)}).$$

4) $B_p = B_p \setminus G_1(p)$, $B = B \cup G_1^{(p)}$.

5) Repeat 2)~4), until $B_p = \phi$ or $G \cap B \neq \phi$, here G is the set of the spherical-shaped class in B_p .

Thus, for any given sample x , the value of $Q(x, B_p)$ can be defined by the following form:

$$Q(x, B_p) = \min_{1 \leq i \leq k_p} \left\{ \frac{d(x, G_i^{(p)})}{d_i^{(p)}} \right\} = \begin{cases} 0 \sim 1, & \text{if } x \in \bigcup_{1 \leq i \leq k_p} G_i^{(p)}, \\ > 1, & \text{otherwise.} \end{cases} \quad (19)$$

From the above-mentioned definition of $Q(x, B_p)$, the discriminant rule can be established as follows:

Rule 2 For any given sample x , compute the value of $Q(x, B_i)$, $i = 1, 2, \dots, k$. If $Q(x, B_t) = \min_{1 \leq i \leq k} Q(x, B_i)$, then $x \in B_t$.

Consequently, NDAM can be briefly described in the following steps:

1) $p = 1$.

2) For the training sample of B_p , determine the center points and the radii according to the adaptive strategy of the pre-classifier.

3) $p = p + 1$, if $p < k$, then return to 2); otherwise proceed to 4).

4) For any given sample x , classify its class according to Rule 2.

4 Numerical Examples

4.1 Simulation Examples

For the sake of convenience, the simulation dataset in Subsection 2.4 is used as an example to test the performance of NDAM. Following the steps in Section 3, the discriminant results of NDAM are shown in Table 4.

The results in Table 4 indicate NDAM is also appropriate for the discriminant problem of the hybrid spherical shell-shaped dataset, and its classification performance is superior to those of Esembles, SVM-Kernel, and SDAM for both training samples and test samples. However, the discriminant effect of boundary points needs to be further improved.

Table 4 Classification results of three tasks

| Method | Correct ratio (%) | | |
|--------|--|--|--|
| | Task 1 | Task 2 | Task 3 |
| NDAM | 100.0 ^a /99.17 ^b | 100.0 ^a /98.75 ^b | 100.0 ^a /100.0 ^b |

^a Correct ratio of training sample;

^b Correct ratio of test sample.

4.2 Real Data Examples

To examine the performance of NDAM in real datasets, nine datasets are selected from UCI Machine Learning Repository^[41]; these datasets are Banknote Dataset, Breast Dataset, Cryotherapy Dataset^[42], Ecoli Dataset, Haberman's Survival Dataset, Iris Dataset, Wine Dataset, Verbetra2c Dataset and Verbetra3c Dataset, respectively, and their basic information and the value of R are shown in Table 5.

Table 5 shows the iris dataset can be approximately regarded as the hybrid spherical shell-shaped dataset, while other datasets are not. For the above nine datasets, the last 5% samples of each class are used as test samples, and the remaining 95% samples are used as training samples. Then, FLDA, Ensembles, SVM-Kernel, SDAM, and NDAM are adopted to build the discriminant models, and their results are shown in Table 6.

Table 5 Information on different datasets

| UCI Dataset | Number | Variable | Class | R |
|-------------|--------|----------|-------|--------|
| Banknote | 1372 | 4 | 2 | 0.1399 |
| Breast | 106 | 9 | 6 | 0.4245 |
| Cryotherapy | 90 | 6 | 2 | 0.2889 |
| Ecoli | 336 | 7 | 8 | 0.2470 |
| Haberman | 306 | 3 | 2 | 0.2614 |
| Iris | 150 | 4 | 3 | 0.0400 |
| Verbetra2c | 310 | 6 | 2 | 0.3516 |
| Verbetra3c | 310 | 6 | 3 | 0.2903 |
| Wine | 178 | 13 | 3 | 0.1011 |

Table 6 Classification results of different discriminant methods

| UCI Dataset | Correct ratio (%) | | | | |
|-------------|---|--|--|---|---|
| | FLDA | HDAM | SVM-Kernel | SDAM | NDAM |
| Banknote | 97.70 ^a /97.06 ^b | 100.00 ^a /98.53 ^b | 100.00 ^a /100.00 ^b | 86.27 ^a /80.88 ^b | 99.69 ^a /97.06 ^b |
| Breast | 52.89 ^a /50.00 ^b | 56.73 ^a /100.00 ^b | 32.69 ^a /100.00 ^b | 32.69 ^a /100.00 ^b | 90.39 ^a /100.00 ^b |
| Cryotherapy | 89.54 ^a /75.00 ^b | 100.00 ^a /100.00 ^b | 56.98 ^a /75.00 ^b | 76.74 ^a /75.00 ^b | 93.02 ^a /75.00 ^b |
| Ecoli | 58.70 ^a /64.29 ^b | 77.64 ^a /85.71 ^b | 90.99 ^a /100.00 ^b | 82.61 ^a /71.43 ^b | 86.03 ^a /92.86 ^b |
| Haberman | 74.91 ^a /66.67 ^b | 76.29 ^a /73.33 ^b | 72.85 ^a /73.33 ^b | 73.20 ^a /73.33 ^b | 85.57 ^a /73.33 ^b |
| Iris | 97.92 ^a /100.00 ^b | 96.53 ^a /100.00 ^b | 100.00 ^a /83.33 ^b | 95.83 ^a /100.00 ^b | 97.92 ^a /100.00 ^b |
| Verbetra2c | 80.34 ^a /93.33 ^b | 91.86 ^a /73.33 ^b | 67.80 ^a /66.67 ^b | 64.07 ^a /86.67 ^b | 91.53 ^a /80.00 ^b |
| Verbetra3c | 69.49 ^a /60.00 ^b | 77.97 ^a /80.00 ^b | 48.48 ^a /46.67 ^b | 75.93 ^a /66.67 ^b | 91.86 ^a /73.33 ^b |
| Wine | 94.15 ^a /100.00 ^b | 100.00 ^a /100.00 ^b | 33.33 ^a /28.57 ^b | 86.55 ^a /100.00 ^b | 98.25 ^a /100.00 ^b |

^a Correct ratio of training sample; ^b Correct ratio of test sample.

From Subsection 2.4, Table 2 and Table 3 indicate that SDAM achieves better results for the hybrid spherical shell-shaped dataset. However, the results given in Table 6 indicate the discriminant performances of partial real datasets do not work well; the reason is that real datasets may not be of the hybrid spherical type, which is consistent with the values of R. NDAM overcomes this drawback through the adaptive strategy of pre-classifier, and each non-spherical-shaped class is divided into several spherical-shaped classes. Through this strategy, this method improves the discriminant performance and stability. Therefore, whether it is a hybrid spherical shell-shaped dataset or a non-hybrid spherical shell-shaped dataset, the experiments show that NDAM has better results and stability than SDAM.

In real datasets, as can be seen from Table 6, none of the methods are universally superior to other methods, each method works well in some datasets, but not so well in others. Overall, the discriminant effect of NDAM is better than that of FLDA, Ensembles, and SVM-Kernel for training samples and test samples. This shows NDAM has better stability in practical applications.

5 Conclusion

This paper has proposed the SDAM, and this method has the advantage of not needing to set parameters. The numerical experiment results indicate that SDAM can solve the discriminant problem for hybrid spherical shell-shaped datasets and has a better classification performance. However, it does not perform well for some no-hybrid spherical shell-shaped datasets.

To overcome the shortcomings of SDAM, NDAM has been proposed. This method utilizes the adaptive strategy of pre-classifier, which aims to transform the discriminant problem of non-spherical-shaped classes into a discriminant problem of spherical-shaped classes. The numerical experiment results indicate that NDAM can preferably solve the discriminant problem in real datasets and has better classification performance and stability.

However, NDAM has some problems that need to be solved further, such as the radius of each class and the classification of boundary points. In addition, as the variables increase, how to make NDAM achieve better results is also worthy of further discussion. Thus, the next stage will continue to study these problems and the classification performance of high dimensional dataset in order to improve the correct ratio of the discriminant model.

References

- [1] Fisher R A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 1936, 7(2): 179–188.
- [2] Rao C R. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1948, 10(2): 159–203.
- [3] Chen Z P, Jiang J H, Li Y, et al. Fuzzy linear discriminant analysis for chemical data sets. *Chemometrics and Intelligent Laboratory Systems*, 1999, 45(1–2): 295–302.
- [4] Pavlenko T. On feature selection, curse-of-dimensionality and error probability in discriminant analysis. *Journal of Statistical Planning and Inference*, 2003, 115(2): 565–584.
- [5] Huberta M, Van Driessen K. Fast and robust discriminant analysis. *Computational Statistics and Data Analysis*, 2004, 45(2): 301–320.
- [6] Zheng W M, Zou C R, Zhao L. Weighted maximum margin discriminant analysis with kernels. *Neurocomputing*, 2005, 67(8): 357–362.

- [7] Phinyomark A, Hu H, Phukpattaranont P, et al. Application of linear discriminant analysis in dimensionality reduction for hand motion classification. *Measurement Science Review*, 2012, 12(3): 82–89.
- [8] Garibotto V, Montandon M L, Viaud C T, et al. Regions of interest-based discriminant analysis of DaTSCAN SPECT and FDG-PET for the classification of dementia. *Clinical Nuclear Medicine*, 2013, 38(3): e112–e117.
- [9] Li H, Ye Z J, Xiao G R. Hyperspectral image classification using spectral-spatial composite kernels discriminant analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(6): 2341–2350.
- [10] Guo F X. Multi-statistics data analysis. Fuzhou: Fujian Science and Technology Press, 1990.
- [11] Chen S C, Li D H. Modified linear discriminant analysis. *Pattern Recognition*, 2005, 38(3): 441–443.
- [12] Hand D J. Classifier technology and the illusion of progress. *Statistical Science*, 2006, 21(1): 1–15.
- [13] Tayaoka A, Tayaoka E, Hirajima T, et al. Image processing system for air classification using linear discriminant analysis. *Computational Water, Energy, and Environmental Engineering*, 2017, 6(2): 192–204.
- [14] Mai Q. A review of discriminant analysis in high dimensions. *WIREs Computational Statistics*, 2013, 5(3): 190–197.
- [15] Yang Z Q. Discriminant analysis of one class surrounded by the other class. *Mathematical Statistics and Applied Probability*, 1989, 4(3): 292–296.
- [16] Huang L W, Su L T. Hierarchical discriminant analysis and its application. *Communications in Statistics — Theory and Methods*, 2013, 42(11): 1951–1957.
- [17] Pacheco J, Casado S, Porras S. Exact methods for variable selection in principal component analysis: Guide functions and pre-selection. *Computational Statistics and Data Analysis*, 2013, 57(1): 95–111.
- [18] Brusco M J, Voorhees C M, Calantone R J, et al. Integrating linear discriminant analysis, polynomial basis expansion, and genetic search for two-group classification. *Communications in Statistics — Simulation and Computation*, 2019, 48(6): 1623–1636.
- [19] Huang L W. Fisher stepwise discriminant analysis method based on selecting optimal variables. *Journal of Systems Science and Mathematical Sciences*, 2021, 41(8): 2338–2348.
- [20] Mai Q, Zou H, Yuan M. A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika*, 2012, 99(1): 29–42.
- [21] Pepler P T, Uys D W, Nel D G. Discriminant analysis under the common principal components model. *Communications in Statistics — Simulation and Computation*, 2017, 46(6): 4812–4827.
- [22] Wang Y, Wang X X. Classification using semiparametric mixtures. *Journal of Applied Statistics*, 2019, 46(11): 2056–2074.
- [23] Fernández-Delgado M, Cernadas E, Barro S, et al. Do we need hundreds of classifiers to solve real world classification problems. *Journal of Machine Learning Research*, 2014, 15(1): 3133–3181.
- [24] Camps-Valls G, Bruzzone L. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2005, 43(6): 1351–1362.
- [25] Wang P, Wang Z Z, Ji Y L, et al. Identification of volcanic rock based on kernel fisher discriminant analysis. *Well Logging Technology*, 2015, 39(3): 390–394.
- [26] Liu Y F, Pi D C, Cheng Q Y. Ensemble kernel method: SVM classification based on game theory. *Journal of Systems Engineering and Electronics*, 2016, 27(1): 251–259.
- [27] Mardia K V. Statistics of directional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1975, 37(3): 349–371.
- [28] Amayri O, Bouguila N. Beyond hybrid generative discriminative learning: Spherical data classification. *Pattern Analysis and Applications*, 2015, 18(1): 113–133.
- [29] Fisher N I, Lewis T, Embleton B J J. Statistical analysis of spherical data. Cambridge: Cambridge University Press, 1993.
- [30] Amayri O. Statistical analysis of spherical data: clustering, feature selection and applications. PhD Thesis, Concordia University, 2014.
- [31] Amayri O, Bouguila N. On online high-dimensional spherical data clustering and feature selection. *Engineering Applications of Artificial Intelligence*, 2013, 26(4): 1386–1398.
- [32] Li M K, Jiang T, Guan J. Radius/center constraint fuzzy c-spherical shell clustering algorithm. *Opto-Electronic Engineering*, 2011, 38(4): 41–47.
- [33] Wei L M, Xie W X. Fuzzy c-spherical shell clustering algorithm based on Euclidean distance. *Systems*

- Engineering and Electronics, 1999, 21(5): 43–47.
- [34] Fan W T, Bouguila N. Spherical data clustering and feature selection through nonparametric Bayesian mixture models with von Mises distributions. *Engineering Applications of Artificial Intelligence*, 2020, 94: 103781.
- [35] Leong P, Carlile S. Methods for spherical data analysis and visualization. *Journal of Neuroscience Methods*, 1998, 80(2): 191–200.
- [36] Chen Y S, Ma S L, Chen X, et al. Hyperspectral data clustering based on density analysis ensemble. *Remote Sensing Letters*, 2017, 8(2): 194–203.
- [37] Han H. Determination of center and radius of clusters with approaching method. *Journal of Jiangnan University (Natural Science Edition)*, 2013, 41(5): 62–64.
- [38] DeVinney J G, Priebe C E, Marchette D J, et al. Random walks and catch digraphs in classification. *Proceedings of the 34th Symposium on the Interface: Computing Science and Statistics*, 2002, 34: 1–10.
- [39] Marchette D J, Wegman E J, Priebe C E. Fast algorithms for classification using class cover catch digraphs. *Handbook of Statistics*, 2005, 24: 331–358.
- [40] Marchette D. Class cover catch digraphs. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010, 2(2): 171–177.
- [41] Dua D, Taniskidou E K. UCI machine learning repository, Irvine, CA: University of California, School of Information and Computer Science, <http://archive.ics.uci.edu/ml>, 2017.
- [42] Khozeimeh F, Alizadehsani R, Roshanzamir M, et al. An expert system for selecting wart treatment method. *Computers in Biology and Medicine*, 2017, 81: 167–175.