

HIGH-DIMENSIONAL DIAGONAL CLASSIFIERS

OKWONU FRIDAY ZINZENDOFF

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE, DELTA STATE UNIVERSITY, P.M.B.1
okwonufz@delsu.edu.ng.; fokwonu@gmail.com

ABSTRACT. The advancement in technology has necessitated the collection of massive high-dimensional data in bioinformatics, medical sciences, business, climatology, and magnetic resonance images. These collections have given rise to high-dimensional (p) and small sample(n) (HDSS) problems in group classifications. The classification coefficient of the HDSS data cannot be constructed due to the ill-conditioning or singularity of the covariance matrix. The curse of dimensionality suffered by the classical methods has rendered the formulation of the coefficient impossible as such different dimensionality reduction methods have been proposed to avert the curse of dimensionality and ill conditioning. Dimensionality reduction methods such as the principal component analysis have been applied as a prelude to reducing the dimension of the HDSS data to enhance constructing the coefficients of any classification methods. Some of these dimension reduction methods with infused constants as a benchmark selection, may increase or decrease computational time and improve classification accuracy at the cost of information loss. In this paper, we proposed two methods that can solve the ill-conditioning and singularity, or dimensionality problems associated to HDSS data directly. These methods are diagonal FLCM (DFLCM) and point extraction method (PEM). We investigated the robustness of these methods based on the optimal probability benchmark (OPB). The computational time (CPU time) of all the methods is also investigated. We applied three real data sets to determine the comparative performance, and the classification results revealed that the proposed DFLCM is very robust more than the Naïve Bayes classifier in terms of the probability of correct classification and CPU time. Therefore, this study affirmed that the OPB is suitable to determine the robustness of HDSS classification problems.

MSC (2020): 62H30, 62H12, 62R07.

Key words: High-dimensional data, Classification, Probability, CPU time, Robustness.

Introduction

Technological advancement has generated massive high-dimensional data set originating from health sciences, genetics, business, astronomy, magnetic resonance images, epidemiology, climatology, seismological study, chromatography, ecology [1, 2, 3 4], herbal medicine [5], bioinformatics [6] finance [7] tumour classification [8] and functional magnetic resonance images [9]. The availability and applications of these categories of data have rendered the classical classification methods unsuitable for computational and analytical purposes [10, 11]. As such, many of these classical methods have been modified and new classification methods have emerged to cope with the increasing data dimension (p) and small sample (n) otherwise referred to as HDSS ($p > n$). One of such method that could not cope with $p > n$ is the Fisher linear classification (FLCM) [3, 12]. Thus, problems are often encountered when the data dimension is greater than the sample size as such these classification techniques break down [13, 14]. For the HDSS problem, the numerical values of some of the variables do not contribute meaningfully to improve classification results, such variables decrease the proportion of correct classification of any methods. To improve the classification accuracy for the HDSS problems, researchers tend to find data reduction techniques to extract useful information from the data set thereby reducing the dimension of the data.

The principal component analysis [15, 16, 17, 18] is an established data dimension reduction technique that has been applied to extract useful information from HDSS problems. Variants of principal component analysis (PCA) for dimension reduction have been discussed [19, 20]. Several other regularized discriminant analyses methods have been proposed to solve HDSS for two groups classification problems [21, 22, 23, 24, 25, 26, 27, 28, 29, 30]. The nearest shrunken centroids [31, 32] sparse partial least square discriminant analysis [33] are dimension reducing techniques to solve HDSS problems.

Other alternative methods such as the independent classification rule (ICR), diagonal linear classification (DLC), and diagonal quadratic classification (DQC) methods were proposed such that the diagonal of the covariance matrix is applied to construct the coefficient of the classification methods [3, 34]. These methods avert the curse of singularity or

dimensionality [6, 35] and the ill conditioning of the covariance matrix. Feature annealed independence rule (FAIR) was proposed as an alternative version of ICR [13, 14]. Besides the FAIR, the sure independence screening (SIS) was also proposed for dimensionality reduction as a step to applying other well-known lower-dimensional classification techniques to determine class separation [8].

Often the case, the above dimension reduction methods may result in information loss as such the classification accuracy of these methods may lack the exact information the data set would naturally portray [36]. The application of the HDSS data on the classical classification methods to perform classification task is impossible due to the difficulty in constructing the classification coefficient resulting from the ill conditioning of the covariance matrix. In this paper, we propose DFLCM and point extraction method (PEM), the DFLCM is a diagonal modification of the FLCM while the PEM extract vital data point from each sample variables based on group probabilities. The objective of the proposed methods is to avert the curse of singularity and ill conditioning of the HDSS data in other to construct the classification coefficient. The PEM shows the proportion of sparsity and density of the data matrix, as such, we can determine the overall proportion of the data points associated with the variables selected from each sample to construct the model. The structure of PEM allows us to determine the group probabilities and the extraction probabilities within each group. The PEM can be described as a method that extracts vital information based on within-group probability from each sample based on the numerical strength of the corresponding variables. This method adopts and utilizes vital data points from the entire data set to construct the classification coefficient via probability benchmark comparison. The PEM does not rely on the information from the covariance matrix to construct the classification coefficient for the HDSS problem. The PEM modulus operandis is described as follows; different samples with corresponding variables have different numerical strengths in the data matrix, as such deleting the entire data points associated with a variable corresponding to the samples may generate numerical differential in the data matrix. Therefore, PEM considers each variable data point as relevant until the benchmark isolates the irrelevant data points from the data matrix. Thus, the relevant data points are applied to construct the model. This can be expressed as; data points in group 1 may provide vital information(relevant) in variables 2,6,8,11,21, and group 2 may contain vital information in variables 1,4, 5,16,27,32,45 and so on for $p > n$ problem, this is illustrated in Table 1. We also considered the modification of the independent classification rule. The modified ICR is called the adjusted ICR (AICR). To give credence to the different methods, we applied real data set to investigate the comparative performance of the proposed methods, the ICR and the AICR, respectively. To determine the robustness of these methods, we applied the optimal probability benchmark to analyze their comparative performance.

This paper is structured as follows. In Section 2, we discussed the diagonal classifier (DFLCM), independence classification rule (ICR), adjusted independence classification (AICR), and finally in this section, we described the proposed point extraction method (PEM). Real data application, results, and discussion are presented in Section 3 followed by a concluding remark in Section 4.

Method

For the HDSS problem, developing the coefficient of the classical methods such as the FLCM becomes impractical to perform classification tasks due to the curse of singularity, dimensionality, and ill conditioning of the covariance matrix[37, 38]. In this section, we will navigate different procedures to perform classification for $p > n$ problems.

Diagonal FLCM (DFLCM)

Let X_{ij} and Y_{ij} such that $j > i$ are independent random samples from two different groups where n and p denote sample size and dimension. We assumed that the two groups have equal sample sizes, that is $n_k, k = 1, 2, i = 1, 2, \dots, n_k, n = \sum_{k=1}^2 n_k$. Let $p_k = \frac{n_k}{N}$, such that $\sum_{k=1}^2 p_k = 1$ be the priori probability from the respective groups.

$$\text{Let } \bar{X}_k = \frac{1}{n_j} \sum_{j=1}^p X_j, j = 1, \dots, p$$

$$S_k^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_k} (X_{ij} - \bar{X}_k)^2$$

$$S_p = \sum_{c=1}^{n_k} \frac{1}{(n_c - k)} (n_k - 1) S_k^2, c = k = 1, 2 \quad (1)$$

be the group mean vectors (X_k), covariance matrices (S_k^2) and pooled sampled covariance matrix (S_p), respectively. Computing the inverse of Equation (1) is difficult since $p > n$. To construct the model coefficient, we must use the diagonal of Equation (1), that is,

$$\hat{S} = \text{diag}(S_p) \quad (2)$$

Equation (2) allows us to solve the problem of dimensionality. Hence the model coefficient can be constructed as

$$\Delta_k = (\bar{X}_1 - \bar{X}_2) \hat{S}^{-1} X_{k,ij} \quad (3)$$

Equation (3) mimics the independence classification rule discussed in [13, 14]. To perform the classification task, we adopt a similar classification procedure discussed in [39]. Therefore, the decision rule is defined as; assign an object

to group one (G_1) if $\Delta_1 \geq \Delta_2$ otherwise, assign the object to group two $\Delta_1 < \Delta_2$. This process can be summarized as follow

$$\gamma = \begin{cases} X_{1j} \in G_1, \text{ if } \Delta_1 \geq \Delta_2 \\ \text{Otherwise } X_{1j} \in G_1, \text{ if } \Delta_1, \Delta_2 \end{cases}$$

The DFLCM is a blended concept of the FLCM, ICR, and predictive classifiers models.

Independence classification rule (ICR)

Since the coefficient of the FLCM could not be constructed for $p > n$, a unique distance classification measure called the independence classification rule (ICR) was proposed [14] whereby the diagonal of the covariance matrix was used to construct the coefficient of ICR. Due to the diagonal use, it is expected that the ICR would avert the curse of singularity. This method relies on the group mean vectors and the diagonal of the covariance matrix to construct the coefficient of the model. This is described as follows

$$\begin{aligned} \bar{X}_c &= \frac{1}{n_c} \sum_{i=1}^{n_c} X_{ci}, & \bar{\bar{X}} &= \frac{1}{k} \sum_{c=1}^k \bar{X}_c \\ S_c^2 &= \frac{1}{n_c-1} \sum_{i=1}^{n_c} (X_i - \bar{X}_c)^2 \\ S_p &= \frac{1}{\sum_{c=1}^k (n_c - k)} (n_c - 1) S_c^2, c = 1, \dots, k, i = 1, \dots, n_c \\ \hat{S} &= \text{diag}(S_p) \end{aligned} \quad (4)$$

Therefore, the ICR [3, 14] for discrimination and classification is given as

$$W_L = (\bar{X}_1 - \bar{X}_2) \hat{S}^{-1} (X_j - \bar{\bar{X}}) = \gamma d \quad (5)$$

where $\gamma = \bar{X}_2 \hat{S}^{-1}$, and $d = (X_j - \bar{\bar{X}})$.

By this rule, an object is assigned to G_1 if $W_L \geq 0$ otherwise assign to G_2 if $W_L < 0$. The ICR is very flexible, and its performance depends on the group data deviation (d) from the ground mean vector $\bar{\bar{X}}$. The independent rule method has a similar model coefficient (γ) with the Fisher linear classification method. Taking the absolute value of d may lead to a higher misclassification rate. Therefore, the strength of the independence classification rule depends on the signs of d . The ICR is also called the Naïve Baye classifier.

Adjusted ICR (AICR)

For the AICR, we require the definition of the group probabilities for the classification benchmark. Let $\pi = \sum_{c=1}^k \pi_c = 1$, be the overall probability and π_c be the group probabilities. The adjusted ICR is a direct derivation of the ICR with slight modification to the classification benchmark. The AICR is replicated from Equation (5) as

$$\hat{W}_L = (\bar{X}_1 - \bar{X}_2) \hat{S}^{-1} (X_j - \bar{\bar{X}}) = \gamma d \quad (6)$$

Therefore, the classification rule for the adjusted ICR is defined as assign X_1 to G_1 if $\hat{W}_L \geq \log(\pi)!$ otherwise, assign X_1 to G_2 if $\hat{W}_L < \log(\pi)!$ where $!$ denotes the factorial symbol.

Point extraction method (PEM)

This method applies within-group probability to extract relevant data points with a strong numerical value from each variable in each group to construct the classification coefficients. Based on the extraction, we can determine the sparsity and density of the data matrix. This allows us to determine the relevant data points used to build the model. The PEM can be described in detail as follow.

Let X_{ij} and Y_{ij} be $p > n$ be as defined above, $M = p$, let

$$P(X_i) = \frac{X_{ij}}{\sum_{i=1}^n X_{ij}} \quad (7)$$

$$P(Y_i) = \frac{Y_{ij}}{\sum_{i=1}^n Y_{ij}} \quad (8)$$

denote the first phase probability transformation for the two groups. Then the second phase is given as

$$C_X = \frac{P(X_i)}{M} \quad (9)$$

$$C_Y = \frac{P(Y_i)}{M} \quad (10)$$

Based on Equations (7-10), we can isolate the weight of the irrelevant data points from the data matrix for the two groups as follows.

$$W_X = \begin{cases} 1, C_X > P(X_i) \\ 0, C_X < P(X_i) \end{cases} \quad (11)$$

$$W_Y = \begin{cases} 1, C_Y > P(Y_i) \\ 0, C_Y < P(Y_i) \end{cases} \quad (12)$$

Therefore, Equation (11-12) contains the weight of the relevant data points. Equations (13-14) consist of the weighted relevant data points associated with each group and variables,

$$e_X = W_X P(X_i) \quad (13)$$

$$e_Y = W_Y P(Y_i) \quad (14)$$

Table 1: Point extraction matrix (Equation (13-14)) ($p > n$)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7		X_p
1	w	w	0	w	0	0	w		w
2	0	0	w	0	w	w	0		0
3	w	0	w	0	w	0	w		w
:	:	:	:	:	:	:	:		:
n	0	w	0	w	w	w	w		w

The PEM retains the original dimension of the data matrix due to the zigzag relevant numerical strength associated with each sample and variables. It preserves all the variables based on the sample size information. The PEM cannot be described as a data reduction method rather a data point extraction method. Table 1 described Equations (13-14), respectively. Equations (13-14) is transformed to $n \times p$, from the above we have $n \times n$ transformed matrix form, that is

$$E_{n \times n} = e_X e_Y$$

$$\delta = \sum E_{n \times n}$$

$$\alpha = \frac{E_{n \times n}}{\delta}$$

$$\Delta = \frac{(E_{n \times n} - \alpha)(E_{n \times n} - \alpha)^T}{\delta}$$

$$\nabla_W = \frac{e_X}{p} - \frac{e_Y}{p}$$

$$W = \left(\frac{e_X}{p} - \frac{e_Y}{p} \right) \frac{(E_{n \times n} - \alpha)(E_{n \times n} - \alpha)^T}{\delta} \quad (15)$$

Equation (15) is the coefficient of the PEM, where W is $1 \times p$. The benchmark is given as

$$\beta = \frac{\phi}{k'} \quad (16)$$

$$\text{where } \phi = \frac{e_X}{p} + \frac{e_Y}{p}, \gamma_a = W e_X, \gamma_b = W e_Y.$$

Therefore, the decision rule is to assign an object to group one (G_1) if $\gamma_a \geq \beta$ otherwise assign an object to group two (G_2) if $\gamma_a < \beta$. Therefore, the classification summary is

$$\theta = \begin{cases} \gamma_a \geq \beta, & \text{if } \gamma_a \in G_1 \\ \gamma_b < \beta, & \text{if } \gamma_b \in G_2 \end{cases}$$

The optimal performance benchmark for high dimensional data can be determined by using the following formula. The probability of misclassification ($\mathfrak{M}$) is defined as

$$\mathfrak{M} = \left[\frac{(1-\tau)}{2 \times \tau} \right] \times \tau$$

where

$$\tau = \left(\frac{\sum_{i=1}^{n_1} \gamma_{ai} \in \Delta_1 + \sum_{i=1}^{n_2} \gamma_{bi} \in \Delta_2}{n_1 + n_2} \right) \quad (17)$$

Equation (17) explain the probability of the group membership being correctly classified and the proportion of membership prediction accuracy (ϕ) is defined as

$$\phi = \cap - \mathfrak{M} \quad (18)$$

where $\cap = 1$ denotes the optimal performance benchmark. The robustness of the high dimensional data can be evaluated using Equation (18).

Results and discussion

For this section, all data sets are uniformly reshuffled and divided into training 60% and testing 40% and seeded 123456. The results reported are based on the average classification accuracy over 1000 runs. Three real data set are considered viz: Musk version 2 data set [40,41 42, 43] sonar rock signal [44] and ozone level data set [45]. Equation (18) is applied to determine the robustness of any of these methods.

Applications

Musk version 2 data set.

The musk and non-musk main data set consist of $n=6598, p=166$, [43]. For $p > n$ problem we took $n=102$, and $p=166$. Previous studies [40, 41, 42] used this data set for $n > p$ problems. The neural network method achieved the highest classification accuracy of 91% [40] and the lowest classification accuracy was reported in [40] with 58.8% based on C4.5 prune. The objective is to determine whether the new molecules will be classified as musk and non-musk. The proportion of membership prediction accuracy (ϕ), and CPU time is reported in Table 2. The result reported in Table 2 showed that the DFLCM, ICR, and AICR performed comparably. Fig. 1 shows the CPU time with the AICR having the smallest CPU time followed by DFLCM, respectively.

Table 2: Performance analysis of the different methods

Methods	ϕ	CPU time
DFLCM	1.000	6.93
ICR	0.999	7.20
AICR	0.999	6.89
PEM	0.941	8.59

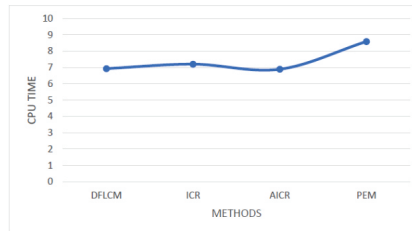


Fig 1: Comparative analysis of CPU time based on methods

Mines vs rock sonar signal

The objective is to classify sonar signal as bounced off a metal cylinder (M) and bounced off a roughly cylindrical rock (R) [44]. In Gorman and Sejnowski (1988), they used neural network techniques and nearest neighbor classifier and it was applied for $n > p$ problems with classification accuracy 90.4% and 82.7% based on the training set ($n=104$). For the $p > n$ problem, we used $p = 60, n = 50$. Table 3 contain the proportion of membership prediction accuracy of the different methods over 1000 runs. The DFLCM appeared to be the best classifier which is much better than the 90.4% [44] although with different study conditions. The ICR and AICR are more robust than 82.7%. For this data set, the DFLCM has better performance for both accuracy and time. Fig.2 depicts the comparative CPU time between DFLCM and AICR for the different methods.

Table 3: Performance analysis of the different methods

Methods	ϕ	CPU time
DFLCM	0.956	0.68
ICR	0.907	0.72
AICR	0.911	0.67
PEM	0.735	0.81

Ozone level detection

This data was previously used to forecast ozone days. It was reported that the data sets contain many irrelevant variables and sample bias[45]. This led us to apply this data set for $p > n$ problem. The original data set [45] contains $n = 2536, p = 72$ with group one ozone day and group two normal day. For $p > n$ problem, we elect to retain the dimension (p) and selected $n = 66$. Based on this, we applied these methods to classify ozone days and non-ozone days. Table 4 contains the classification results of the different methods with DFLCM as the best method followed by the ICR and AICR method. Fig.3 contains the CPU time of the different methods, the analysis showed that DFLCM appeared to have the fastest computational time followed by the ICR method.

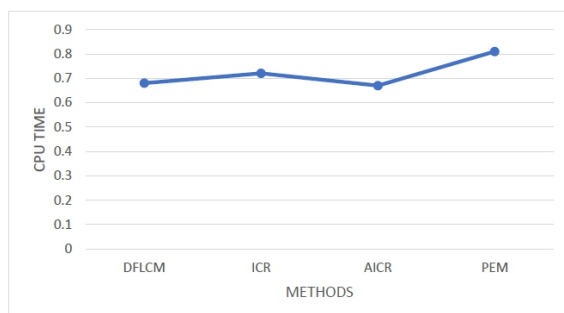


Fig 2: Comparative analysis of CPU time based on methods

Table 4: Performance analysis of the different methods

Methods	ϕ	CPU time
DFLCM	0.993	0.98
ICR	0.973	1.04
AICR	0.973	1.07
PEM	0.921	1.25

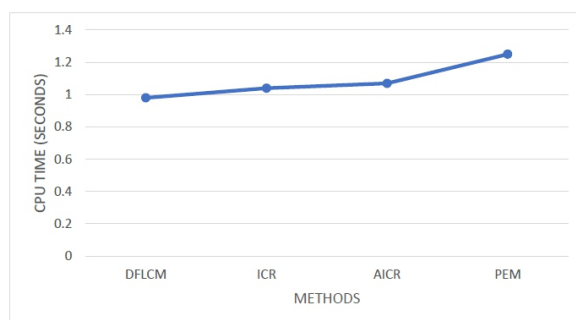


Fig 3: Comparative analysis of CPU time based on methods

It is vital to state that the performance of any classification method strictly relies on the nature of the dataset [6]. This implies that a particular model may result in minimum misclassification error due to certain dataset, also this model may result in high misclassification error for another data set. Based on this conviction, it is important for a researcher to consider a suitable model and data set to avoid the generalization of the under performance of a particular classification model. In this regard, model under performance should be investigated vigorously with different data set before generalization, or the results should be accompanied by detailed discussions. In general, no classification model performs excellently with any available data set hence model performance is data dependent. It is observed that the problems associated with $p > n$ are high computational cost and high misclassification error [8, 46, 47] due to irrelevant variables. In comparative terms, it implies that a low dimensional data set may result in a minimum computational cost and a better classification rate otherwise. In this study, we have observed that the PEM method is comparably suitable to do comparative performance analysis with the diagonal classifiers. Table 5 show the number of relevant (bold) and irrelevant data extracted for all the data set. The number in italics indicates the total data points in the data matrix. In future work, we will compare the PEM methods with dimension reduction techniques and variable selection techniques to determine further performance analysis.

Table 5: Number of selected data points (bold) by PEM

Data 1	$2 \times (p \times n)$	Relevant point by PEM	irrelevant point by PEM	Δ_1 relevant	Δ_2 relevant	Δ_1 irrelevant	Δ_2 irrelevant
Training (60%)	20,318.4	10,204	10,114.4	5,297	4,907	4,862.2	5,252.2
Testing (40%)	13,545.6	8,056	5,489.6	4,189	3,867	2,583.8	2,905.8
Total	33,864	18,260	15,604	9,486	8,774	7,446	8,158
Data 2	$2 \times (p \times n)$	Relevant point by PEM	irrelevant point by PEM	Δ_1 relevant	Δ_2 relevant	Δ_1 irrelevant	Δ_2 irrelevant
Training (60%)	3,600	2,010	1,590	1,012	998	788	802
Testing (40%)	2,400	1,611	789	820	791	380	409
Total	6,000	3,621	2,379	1,832	1,789	1,168	1,211
Data 3	$2 \times (p \times n)$	Relevant point by PEM	irrelevant point by PEM	Δ_1 relevant	Δ_2 relevant	Δ_1 irrelevant	Δ_2 irrelevant
Training (60%)	5,328	5,032	296	2,516	2,516	148	148
Testing (40%)	4,176	3,944	232	1,972	1,972	116	116
Total	9,504	8,976	528	4,488	4,488	264	264

$\Delta_i, i = 1, 2$ denotes groups.

Conclusions

Three diagonal classifiers for HDSS problems were considered for this paper. The point extraction method was also discussed. Three real data applications taken from the UCI machine learning repository were used with sample size modifications to suit HDSS problems. The results have shown that the proposed DFLCM methods appeared to be the best in terms of proportion of membership prediction accuracy and computational time (CPU time). The modified AICR method showed comparative performance followed by the conventional ICR and PEM in that order for these data sets. The analysis has demonstrated that the DFLCM is very capable to perform HDSS classification tasks based on these data set. It also indicated fast CPU computational time compared to the other method. We remark that the DFLCM and the other methods can be applied to investigate any HDSS problems with high-speed computers. This study concludes that these methods are suitable and comparable to carry out classification analysis based on the application of HDSS data.

Acknowledgement

We deeply acknowledge the donors of the relevant data used in this research especially Thomas G. Dietterich, Lathrop and Lozano-Perez, 1997, Jain et al, 1994, Dietterich, T. G., et al 1994, Dua D., and Graff, C. (2019) and K. Zhang and Fan 2008. All the data set was retrieved from <https://www.openml.org/search> and verified in UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>].

REFERENCES

1. El Anbari, M.; Bensmail, H. Computational and statistical challenges with high dimensionality: A new method and efficient algorithm for feature selection in knowledge discovery. *In Qatar Foundation Annual Research Forum Proceedings* (Vol. 2012), No. 1. Bloomsbury Qatar Foundation Journals.
2. Fan, J.; Li R. Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *In Proceedings of the International Congress of Mathematicians Madrid, August 22-30, 2006* (pp. 595-622). European Mathematical Society-EMS-Publishing House GmbH.
3. Herbert, P.; Tiejun, T. Recent advances in discriminant analysis for high-dimensional data classification. *J. Biom. Biostat* **3** 2, 1-2.

4. Li, C. Q.; Xu, Q. S. High-dimensional spectral data classification with nonparametric feature screening. *Journal of Chemometrics* **34** (2020), 3, e3199.
5. Gad, H. A.; El-Ahmady, S. H.; Abou-Shoer, M. I.; Al-Azizi, M. M. Application of chemometrics in authentication of herbal medicines: a review. *Phytochemical Analysis* **24** (2013) 1, 1-24.
6. Bommert, A.; Sun, X.; Bischl, B.; Rahnenfehrer, J.; Lang, M. Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational Statistics & Data Analysis* **143** (2020), 106839.
7. Harikumar, S.; AS, A.; Kaimal, R. A depth-based nearest neighbor algorithm for high-dimensional data classification. *Turkish Journal of Electrical Engineering and Computer Sciences* **27** (2019), 6, 4082-4101.
8. Fan, J.; Lv, J. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **70** (2008), 5, 849-911.
9. Mourao-Miranda, J.; Bokde, A. L.; Born, C.; Hampel, H.; Stetter, M. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *NeuroImage* **28** (2005), 4, 980-995.
10. Pappu, V.; Pardalos, P. M. High-dimensional data classification. In *Clusters, Orders, and Trees: Methods and Applications: In Honor of Boris Mirkin's 70th Birthday* (2014), pp. 119-150. New York, NY: Springer New York.
11. Johnstone, I. M.; Titterton, D. M. Statistical challenges of high-dimensional data. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* **367** (2009), 1906, 4237-4253.
12. Okwonu, F. Z.; Othman, A. R. A model classification technique for linear discriminant analysis for two groups. *International journal of computer science issues (IJCSI)* **9** (2012), 3 no. 2, 125-128.
13. Fan, J.; Fan, Y. High dimensional classification using features annealed independence rules. *Annals of statistics* **36** (2008), 6, 2605-2644.
14. Bickel, P. J.; Levina, E. Some theory for Fisher's linear discriminant function, naive Bayes', and some alternatives when there are many more variables than observations. *Bernoulli* **10** (2004), 6, 989-1010.
15. Ghosh, D. Singular value decomposition regression models for classification of tumors from microarray experiments. In *Biocomputing 2002* (2002), pp. 18-29.
16. Bair, E.; Hastie, T.; Paul, D.; Tibshirani, R. Prediction by supervised principal components. *Journal of the American Statistical Association* **101** (2006), 473, 119-137.
17. Ma, J.; Yuan, Y. Dimension reduction of image deep feature using PCA. *Journal of Visual Communication and Image Representation* **63** (2019), 102578.
18. Monakhova, Y. B.; Godelmann, R.; Kuballa, T.; Mushtakova, S. P.; Rutledge, D. N. Independent components analysis to increase efficiency of discriminant analysis methods (FDA and LDA): Application to NMR fingerprinting of wine. *Talanta* **141** (2015), 60-65.
19. Seghouane, A. K.; Shokouhi, N.; Koch, I. Sparse principal component analysis with preserved sparsity pattern. *IEEE Transactions on Image Processing* **28** (2019), 7, 3274-3285.
20. Velliangiri, S.; Alagumuthukrishnan, S. J. P. C. S. A review of dimensionality reduction techniques for efficient computation. *Procedia computer science* **165** (2019), 104-111.
21. Ye, B.; Liu, P. Classification of High-Dimensional Data: A Random-Matrix Regularized Discriminant Analysis Approach. *Inter. Journal of Innov. Comp., Info. and Control* **15** (2019), 3, 955-967.
22. Wang, C.; Jiang, B. On the dimension effect of regularized linear discriminant analysis. *Electronic Journal of Statistics* **12** (2018) 2709-2742.
23. Guo, Y.; Hastie, T.; Tibshirani, R. Regularized linear discriminant analysis and its application in microarrays. *Biostatistics* **8** (2007), 1, 86-100.
24. Fu, R.; Tian, Y.; Bao, T.; Meng, Z.; Shi, P. Improvement motor imagery EEG classification based on regularized linear discriminant analysis. *Journal of medical system* **43** (2019), 6, 169-182.
25. Wu, W.; Mallet, Y.; Walczak, B.; Penninckx, W.; Massart, D. L.; Heuerding, S.; Erni, F. Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data. *Analytica Chimica Acta* **329** (1996), (3), 257-265.
26. Elkhailil, K.; Kammoun, A.; Couillet, R.; Al-Naffouri, T. Y.; Alouini, M. S. A large dimensional study of regularized discriminant analysis. *IEEE Transactions on Signal Processing* **68** (2020), 2464-2479.
27. Zhang, Z.; Dai, G.; Xu, C.; Jordan, M. I. Regularized discriminant analysis, ridge regression and beyond. *The Journal of Machine Learning Research* **11** (2010), 2199-2228.

28. Fu, R.; Wang, H.; Bao, T.; Han, M. EEG intentions recognition in dynamic complex object control task by functional brain networks and regularized discriminant analysis. *Biomedical Signal Processing and Control* **61** (2020), 101998, 1-12.
29. Krzeszowski, T.; Wiktorowicz, K. . Combined regularized discriminant analysis and swarm intelligence techniques for gait recognition. *Sensors* **20** (2020), 23, 6794-67808.
30. Friedman, J. H.. Regularized discriminant analysis. *Journal of the American statistical association* **84** (1989), 405, 165-175.
31. Xu, J.; Xu, Q.; Yi, L.; Chan, C. O.; Mok, D. K. W. Correlation-assisted nearest shrunken centroid classifier with applications for high dimensional spectral data. *Journal of Chemometrics* **30** (2016), 1, 37-45.
32. Tibshirani, R.; Hastie, T.; Narasimhan, B.; Chu, G. Class prediction by nearest shrunken centroids, with applications to DNA microarrays. *Statistical Science* **18** (2003), 1, 104-117.
33. Chung, D.; Keles, S. . Sparse partial least squares classification for high dimensional data. *Statistical applications in genetics and molecular biology* **9** (2010), (1), 17, 1-30
34. Dudoit, S.; Fridlyand, J.; Speed, T. P. Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American statistical association* **97** (2002), 457, 77-87.
35. Li, Y.; Chai, Y.; Zhou, H.; Yin, H. A novel dimension reduction and dictionary learning framework for high-dimensional data classification. *Pattern Recognition* **112** (2021), 107793.
36. Chen, W.; Xu, Y.; Yu, Z.; Cao, W.; Chen, C. P.; Han, G. . Hybrid dimensionality reduction forest with pruning for high-dimensional data classification. *IEEE Access* **8** (2020), 40138-40150.
37. Xiao, X.; Fu, D.; Shi, Y.; Wen, J. (2020). Optimized Mahalanobis-Taguchi System for High-Dimensional Small Sample Data Classification. *Computational Intelligence and Neuroscience* **2020** 1, 4609423, 1-15.
38. Okwonu, F. Z.; Othman, A. R.. Heteroscedastic variancecovariance matrices for unbiased two groups linear classification methods. *Applied Mathematical Sciences* **7** (2013), 138, 6855-6865.
39. Huberty, C. J.; Holmes, S. E. Two-group comparisons and univariate classification. *Educational and Psychological Measurement*, 43 (1983), 1) 15-26.
40. Dietterich, T.; Jain, A.; Lathrop, R.; Lozano-Perez, T. A comparison of dynamic reposing and tangent distance for drug activity prediction. *Advances in neural information processing systems* **6** (1993), 216-223.
41. Jain, A. N.; Dietterich, T. G.; Lathrop, R. H.; Chapman, D.; Critchlow Jr, R. E.; Bauer, B. E.; Teresa A. W; Lozano-Perez, T. Compass: A shape-based machine learning tool for drug design. *Journal of Computer-Aided Molecular Design* **8** (1994), 6, 635-652.
42. Dietterich, T. G.; Lathrop, R. H.; Lozano-Perez, T. Solving the multiple instance problem with axis-parallel rectangles. *Artificial intelligence* **89** (1997), no. 1-2, 31-71.
43. Dua, D.; Graff, C. UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science, zuletzt abgerufen am: 14.09. 2019. Google Sch (2019).
44. Gorman, R. P.; Sejnowski, T. J. Analysis of hidden units in a layered network trained to classify sonar targets. *Neural networks* **1** (1988), 1, 75-89.
45. Zhang, K.; Fan, W. Forecasting skewed biased stochastic ozone days: analyses, solutions and beyond. *Knowledge and Information Systems* **14** (2008), 3, 299-326.
46. Zanin Zambom, A.; Matthews, G. J. Sure Independence Screening in the Presence of Data That is Missing at Random. *Statistical Papers from Loyola University Chicago* (2019), pg 1-27.
47. Nandy, D.; Chiaromonte, F.; Li, R. Covariate information number for feature screening in ultrahigh-dimensional supervised problems. *Journal of the American Statistical Association* **117** (2022), 539, 1516-1529.

[Cite this article](#)

Okwonu Z. F. *High-Dimensional Diagonal Classifiers*, **Acta NUUZ**, 2026, 2/3, pp. 169-177.