A Classification Study in High-Dimensional Data of Linear Discriminant Analysis and Regularized Discriminant Analysis

AUTCHA ARAVEEPORN, SOMSRI BANDITVILAI
Department of Statistics, School of Science,
King Mongkut's Institute of Technology Ladkrabang
10520, Bangkok,
THAILAND

Abstract: - The objective of this work is to compare linear discriminant analysis (LDA) and regularized discriminant analysis (RDA) for classification in high-dimensional data. This dataset consists of the response variable as a binary or dichotomous variable and the explanatory as a continuous variable. The LDA and RDA methods are well-known in statistical and probabilistic learning classification. The LDA has created the decision boundary as a linear function where the covariance of two classes is equal. Then the RDA is extended from the LDA to resolve the estimated covariance when the number of observations exceeds the explanatory variables, or called high-dimensional data. The explanatory dataset is generated from the normal distribution, contaminated normal distribution, and uniform distribution. The binary of the response variables is computed from the logit function depending on the explanatory variable. The highest average accuracy percentage evaluates to propose the performance of the classification methods in several situations. Through simulation results, the LDA was successful when using large sample sizes, but the RDA performed when using the most sample sizes.

Key-Words: - high-dimensional data, linear discriminant analysis, regularized discriminant analysis

Received: September 11, 2022. Revised: April 2, 2023. Accepted: April 26, 2023. Published: May 10, 2023.

1 Introduction

The discriminant analysis is a statistical technique that is helped the researcher to separate response variables in terms of categorical data depending on the explanatory variable. This method comprises a discriminant function or decision function in the form of a linear or quadratic function to divide two or more classes of the response variable. [1], illustrated the discriminant analysis to challenge the classifying data. This paper demonstrated that the discriminant analysis had good predictive accuracy in the normal distribution. [2], applied the cosine similarity measure based on decision rue in the discriminant analysis.

Linear discriminant analysis is a well-known technique for dimensionality reduction problems. Pre-processing step is a machine learning and pattern classification application, [3]. This technique comes from the assumption of a standard covariance matrix based on the multivariate normal distribution. The decision boundary function is created for computing the population. The maximization of the likelihood function is to evaluate the observation and the proportion of each population. [4], applied linear discriminant analysis for small sample sizes in the classification of face recognition, bioinformatics, and text recognition. [5], developed the linear discriminant analysis to neighborhood linear discriminant analysis. Then, the scatter matrices are defined on a neighborhood consisting of reverse nearest neighbors.

When the assumption of the covariance matrix has an individual for each group, this leads to so-called quadratic discriminant analysis. The linear discriminant analysis is straightforward, where the number of observations is greater than the number of the explanatory variable. However, it becomes a severe problem where the number of the explanatory variable is greater than the number of observations, or it defines the high-dimensional data. The quadratic discriminant analysis cannot be inverted for computation because the sample covariance matrix is singular. To overcome these problems, the linear discriminant analysis makes some adaptations to a new method as regularized discriminant analysis, [6]. [7], improved the covariance in regularized discriminant analysis on the high-dimensional low-sample size data for the ill-posed inverse problem. [8], conducted a large dimensional experiment of regularized discriminant analysis classifiers with its two popular methods, known as regularized LDA and QDA.

The LDA has extended to flexible discriminant analysis (FDA), [9], a valuable multigroup classification tool. FDA obtained nonparametric versions of discriminant analysis by replacing linear regression with any nonparametric regression method, and this technique can improve its classification performant results. [10], considered the high-dimensional data for the within-class covariance singular matrix, called penalized LDA, that evaluated the performance of the resulting methods in the simulation study. [11], described a penalized version of LDA designed for highly correlated independent variables. [12], fitted the Gaussian mixture to each class to facilitate effective classification in non-normal settings.

This article aims to study the binary classification of high-dimensional data by comparing LDA and QDA. Through simulation data, we generate explanatory variables from the normal distribution, contaminated normal distribution, and uniform distributions, while response variables are obtained from the logit function. The maximum average accuracy percentage investigates the performance of two methods.

This study is divided into four sections: the first section discusses the importance and background of linear discriminant analysis and regularized discriminant analysis. Section 2, the general definitions related to discriminant analysis, proposes the theorems of these methods. Section 3 presents the simulation study and results used to construct the response and explanatory variables in the high-dimensional data. A discussion of our simulation results is shown in section 4. Finally, the conclusion and recommendations are provided in Section 5.

2 Discriminant Analysis

The explanation of LDA and RDA relates to the Bayes theory concept based on a multivariate normal distribution. The two classes have a normal distribution in the real world, then

$$x \sim \begin{cases} N(\mu_1, \sigma_1^2), & \text{if } x \in C_1 \\ N(\mu_2, \sigma_2^2), & \text{if } x \in C_2 \end{cases},$$

where C_1 and C_2 denote the first and the second class. The definition of the probability distribution is

$$P(x \in C_1) = f_1(x) = \pi_1,$$
$$\text{and } P(x \in C_2) = f_2(x) = \pi_2,$$

where the prior distributions denote $f_1(x)$ and $f_2(x)$ by π_1 and π_2. According to the Bayes theorem, the posterior distribution is written by

$$P(x \in C_1 \mid X = x) = \frac{P(X = x \mid x \in C_1) P(x \in C_1)}{P(X = x)} = \frac{f_1(x)\ \pi_1}{\sum_{k=1}^{|C|} P(X = x \mid x \in C_k)\pi_k}, \quad (1)$$

where $|C|$ is the number of class. The likelihood and the prior functions of class are $f_1(x)$ and π_1. Therefore, the posterior distribution in (1) becomes

$$\frac{f_1(x)\ \pi_1}{\sum_{k=1}^{|C|} P(X = x \mid x \in C_k)\pi_k} = \frac{f_1(x)\ \pi_2}{\sum_{k=1}^{|C|} P(X = x \mid x \in C_k)\pi_k},$$

then $$f_1(x)\pi_1 = f_2(x)\pi_2. \quad (2)$$

Now, the thinking of a multivariate dataset of discriminant analysis is $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ with n observations where $x_i = (x_{i1}, x_{i2}, ..., x_{ip})^{\mathrm{T}}, i = 1, 2, ..., n$ in p variables. This dataset focuses on the multivariate normal distribution called $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \Sigma)$. The probability distribution function for $\boldsymbol{x}$ is

$$f(\boldsymbol{x} \mid \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}{2} \right), \quad (3)$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_p)$ denotes the mean of the dataset, Σ denotes the covariance matrix, and Σ^{-1} denotes the inverse of the covariance matrix.

Therefore, the two classes of multivariate normal distribution in (2) and (3) become

$$\frac{1}{\sqrt{(2\pi)^p |\Sigma_1|}} exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_1)}{2} \right) \pi_1 = \frac{1}{\sqrt{(2\pi)^p |\Sigma_2|}} exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{\mu}_2)^T \Sigma_2^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_2)}{2} \right) \pi_2, \quad (4)$$

2.1 Linear Discriminant Analysis

The linear discriminant analysis mentions the equal covariance matrix on two classes $\Sigma_1 = \Sigma_2 = \Sigma$, [13]. Therefore, the probability distribution function in (3) becomes:

$$\begin{aligned}&\frac{1}{\sqrt{(2\pi)^p|\Sigma|}}\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}{2}\right)\pi_1\\&=\frac{1}{\sqrt{(2\pi)^p|\Sigma|}}\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_2)^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)}{2}\right)\pi_2,\end{aligned} \quad (5)$$

where π_1 and π_2 are the probability of two classes, and $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the mean of two classes.

Take the natural logarithm in (5) two sides, and the simplified term shows that

$$\begin{aligned}&-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}-\frac{1}{2}\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{\mu}_1+\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{x}+\ln(\pi_1)\\&=-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}-\frac{1}{2}\boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{\mu}_2+\boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{x}+\ln(\pi_2).\end{aligned} \quad (6)$$

where (6) is $\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{x}$, and multiply two sides by two, and we have:

$$\begin{aligned}&2\left(\Sigma^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)\right)^T\boldsymbol{x}+\left((\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^T\Sigma^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)\right)\\&+2\ln\left(\frac{\pi_2}{\pi_1}\right)=0.\end{aligned} \quad (7)$$

For obtaining (7), this equation can be seen in the form of a linear function $\boldsymbol{A}^T\boldsymbol{x}+\boldsymbol{b}=\boldsymbol{0}$ which is called the LDA. The decision boundary to discriminate the two classes is

$$\begin{aligned}\delta(\boldsymbol{x}) = &2\left(\hat{\Sigma}^{-1}(\hat{\boldsymbol{\mu}}_2-\hat{\boldsymbol{\mu}}_1)\right)^T\boldsymbol{x}+\left((\hat{\boldsymbol{\mu}}_2-\hat{\boldsymbol{\mu}}_1)^T\Sigma^{-1}(\hat{\boldsymbol{\mu}}_2-\hat{\boldsymbol{\mu}}_1)\right)\\&+2\ln\left(\frac{\hat{\pi}_2}{\hat{\pi}_1}\right).\end{aligned} \quad (8)$$

The classification corresponds to assigning two classes as

$$\delta(x)=\begin{cases}1\ ,\ if\ \ \delta(\boldsymbol{x})<0\\2\ ,\ if\ \ \delta(\boldsymbol{x})>0\end{cases}. \quad (9)$$

The parameters associated with (9) are approximated from the multivariate dataset as the mean and covariance matrices following:

$$\hat{\boldsymbol{\mu}}_k=\frac{\sum_{i=1}^{n_k}\boldsymbol{x}_i}{n_k}\ ,\ k=1,2$$

$$\hat{\Sigma}\ =\frac{(n_1-1)\hat{\Sigma}_1+(n_2-1)\hat{\Sigma}_2}{n-2},\ \ n=n_1+n_2,$$

$$\hat{\Sigma}_k=\frac{1}{n_k-1}\sum_{i=1}^{n_k}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_k)(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_k)^T$$

and $\hat{\pi}_k=\frac{n_k}{n}$,

where $\hat{\Sigma}$ is called the pooled covariance matrix.

2.2 Regularized Discriminant Analysis

In high-dimensional data, the performance of linear discriminant analysis is far from optimal since the lack of observation is unstable data. Therefore, the regularized discriminant analysis is proposed to resolve the singularity problem. [14], proposed the regularization in a covariance matrix (Σ) by defining

$$\tilde{\Sigma}\ =\alpha\hat{\Sigma}\ +(1-\alpha)I_p, \quad (10)$$

where α is defined as the regularized parameter on values $0\le\alpha\le1$. Then, the regularization probably is adjusted by the sample correlation matrix $\hat{R}=\hat{D}^{1/2}\hat{\Sigma}\hat{D}^{1/2}$ in the same way,

$$\tilde{R}\ =\alpha\hat{R}\ +(1-\alpha)I_p, \quad (11)$$

where $\hat{D}$ is the diagonal matrix of the pooled covariance matrix ($\hat{\Sigma}$). Then, the regularized covariance matrix is modified by (10) and (11) as

$$\tilde{\Sigma}\ =\hat{D}^{1/2}\tilde{R}\hat{D}^{1/2}. \quad (12)$$

Now, the decision boundary depends on regularized

covariance matrix that can define the corresponding linear discriminant analysis as,

$$\delta(x) = 2\left(\tilde{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1)\right)^T x + \left((\hat{\mu}_2 - \hat{\mu}_1)^T \tilde{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1)\right) + 2\ln\left(\frac{\hat{\pi}_2}{\hat{\pi}_1}\right), \quad (13)$$

where the $\tilde{\Sigma}$ can be from (12) and the classify criterion is the same as (9).

3 Simulation Study and Results

The simulation study will classify the binary response variables (y) based on an explanatory variable ($\boldsymbol{x}$) by using linear discriminant analysis and regularized discriminant analysis. The explanatory variables are generated on the normal distribution, contaminated normal distribution, and uniform distribution.

The normal distribution is the common data with parameter μ and variance σ^2 in the following function:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0.$$

The simulation data is generated from a normal distribution with a mean of zero and a variance of twenty-five or called $N(\mu,\sigma^2) = N(0,25)$ and the probability density is shown in Fig. 1.

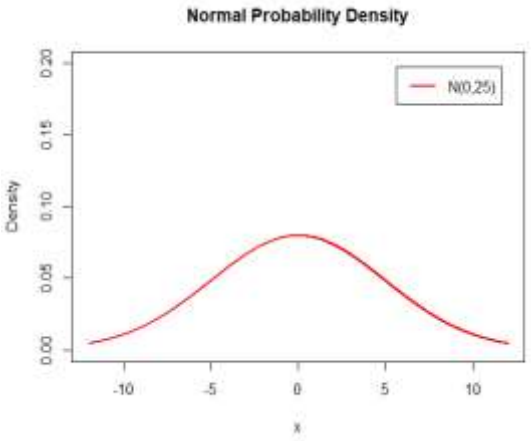

Fig. 1: The normal probability density with mean zero and variance twenty-five.

The contaminated normal distribution is a mixture of two normal distributions with a mixing probability of contaminated data p and $1-p$, where $0 < p \le 0.1$. Then the contaminated normal probability density is

$$f\left(x;\mu,\sigma^2\right) = (1-p)N(\mu,\sigma^2) + pN(\mu,c^2\sigma^2),$$

where c is a parameter that determines the wider standard deviation. In this case, we used the ten percent of contaminated data ($p = 0.1$) and $c = 5$. The mean and variance are defined as normal distribution, and the histogram of the contaminated normal distribution is shown in Fig. 2.

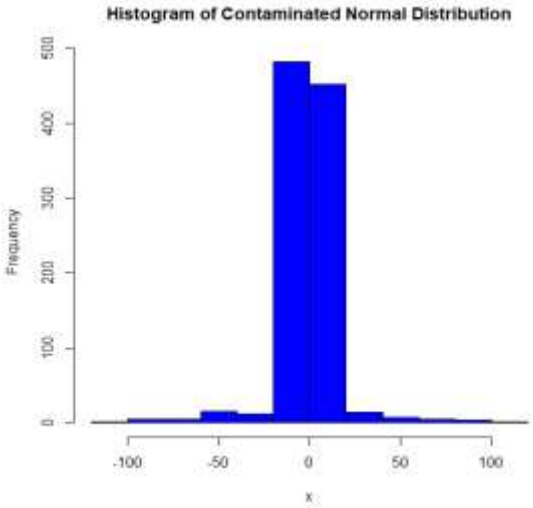

Fig. 2: The histogram of contaminated normal distribution with mean zero, variance twenty-five, $p = 0.1$, and $c = 5$.

Finally, the uniform distribution is the symmetric distribution with parameters a and b, which are the minimum and maximum values. The uniform probability density is written by

$$f(x) = \frac{1}{b-a}, \quad a < x < b,$$

where the mean is $E(X) = \frac{b+a}{2}$, and variance is $Var(X) = \frac{(b-a)^2}{12}$. This explanatory variable is simulated in the range of -2 to 2 with a mean zero and a variance of 1.333. The probability density is exhibited in Fig. 3.

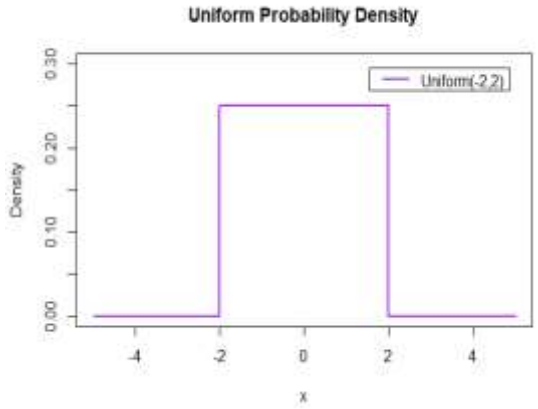

Fig. 3: The uniform probability density in the range of -2 to 2.

Through simulation, the explanatory variables are greater than the observed data (n) standing on the high-dimensional data. The number of explanatory variables is defined as
30 $(n=15, 20, 25)$,
60 $(n=20, 30, 40, 50, 55)$,
and 100 $(n=20, 30, 40, 50, 70, 95)$. The response variable is obtained from the logit function $p(\boldsymbol{x}_i) = \dfrac{e^{\boldsymbol{x}_i \underset{\sim}{\beta}}}{1+e^{\boldsymbol{x}_i \underset{\sim}{\beta}}}$, where $\boldsymbol{x}$ are the explanatory and $\underset{\sim}{\beta}$ are the parameter of correlation coefficients. If $p(\boldsymbol{x}_i) \geq 0.5$, the response variables are denoted as $y_i = 1$, and $y_i = 0$, when $p(\boldsymbol{x}_i) < 0.5$.

The R program was conducted to simulate data and approximated the decision boundary to classify the response variable. The confusion matrix was created to decide the performance of these classification methods. The predicted data were evaluated to compare with the real data using the accuracy percentage (Table 1).

Table 1. The confusion matrix of real data (y_i) and predicted data ($\hat{y}_i$).

Predicted data	Real data	
	$y_i = 1$	$y_i = 0$
$\hat{y}_i = 1$	True Positive (TP)	False Positive (FP)
$\hat{y}_i = 0$	False Negative (FN)	True Negative (TN)

$$\text{Accuracy Percentage} = \frac{TP+TN}{TP+TN+FP+FN} \times 100.$$

The average accuracy percentage for the classification of the linear discriminant analysis and regularized discriminant analysis are shown in Table 2, Table 3, and Table 4. Then Fig. 4, Fig. 5, and Fig. 6 show the average accuracy percentage trend when sample sizes are increased.

Table 2. The average accuracy percentage of linear discriminant analysis (LDA) and regularized discriminant analysis (RDA) under 30 independent variables.

Sample Sizes (n)	Normal		Contaminated Normal		Uniform	
	LDA	RDA	LDA	RDA	LDA	RDA
15	85.14	**99.60**	84.21	**97.13**	85.69	**99.72**
20	92.71	**99.63**	90.64	**97.12**	93.15	**99.54**
25	98.44	**99.36**	96.30	**96.52**	98.44	**99.27**

In Table 2, the RDA employs the highest average accuracy percentage in all cases. It can see that the increased sample size of RDA does not affect classification except for LDA. When the sample sizes increase, the average accuracy percentage of LDA is increased, as shown in Fig. 4.

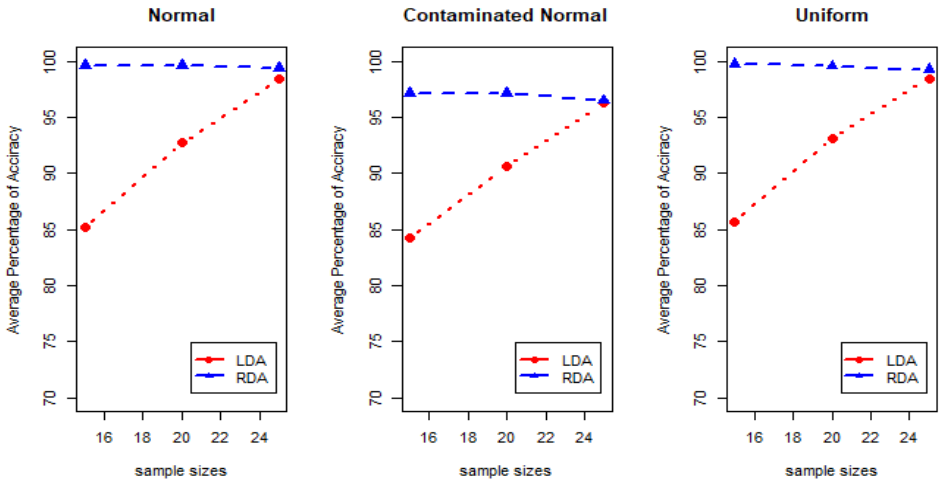

Fig. 4: The trend of the average accuracy percentage of linear discriminant analysis (LDA) and regularized discriminant analysis (RDA) under 30 independent variables.

Table 3. The average accuracy percentage of linear discriminant analysis (LDA) and regularized discriminant analysis (RDA) under 60 independent variables.

Sample Sizes (n)	Normal		Contaminated Normal		Uniform	
	LDA	RDA	LDA	RDA	LDA	RDA
20	77.71	**99.86**	79.62	**98.64**	77.04	**99.82**
30	85.61	**99.67**	87.19	**98.28**	85.64	**99.65**
40	94.00	**99.31**	93.05	**98.15**	94.19	**99.52**
50	**99.27**	99.23	**97.94**	97.78	**99.28**	99.16
55	**99.94**	99.06	**99.50**	97.66	**99.95**	99.14

From the average accuracy percentage in Table 2, the RDA is appropriate for the small sample sizes, but LDA outperforms the large sample sizes. The average accuracy percentage of LDA is increased when the sample sizes increase, as shown in Fig. 5.

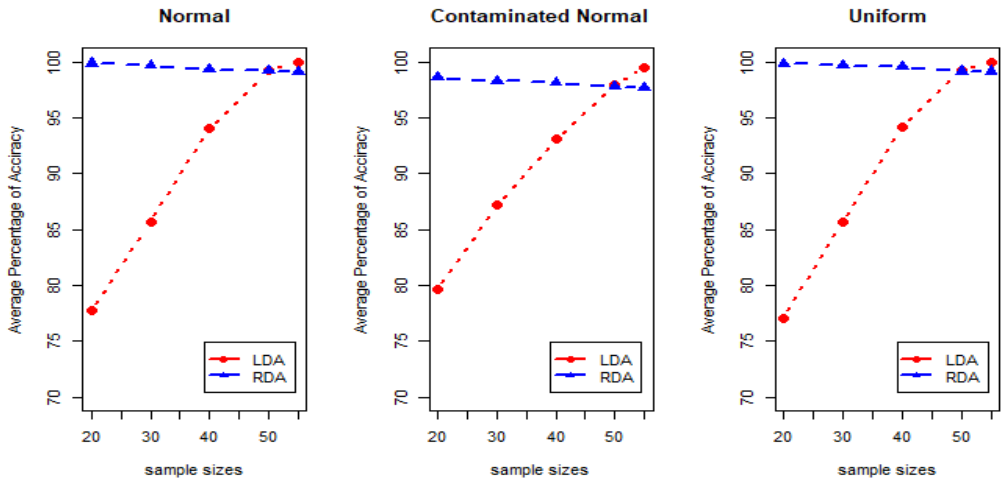

Fig. 5: The trend of the average accuracy percentage of linear discriminant analysis (LDA) and regularized discriminant analysis (RDA) under 60 independent variables.

Table 4. The average accuracy percentage of linear discriminant analysis (LDA) and regularized discriminant analysis (RDA) under 100 independent variables.

Sample Sizes (n)	Normal		Contaminated Normal		Uniform	
	LDA	RDA	LDA	RDA	LDA	RDA
20	70.74	**99.96**	73.61	**99.45**	70.70	**99.96**
30	75.60	**99.87**	78.69	**99.14**	75.33	**99.89**
40	80.75	**99.80**	83.32	**99.09**	80.82	**99.79**
50	85.92	**99.54**	87.51	**98.83**	86.16	**99.62**
70	96.14	**99.33**	94.93	**98.58**	96.29	**99.42**
95	**99.99**	98.99	**99.94**	98.19	**99.99**	98.89

According to the results in Table 4, the RDA performs well in most cases, but the LDA is a perfect classification in the largest sample sizes. The average LDA accuracy percentage increases when the sample sizes increase, as shown in Fig. 6.

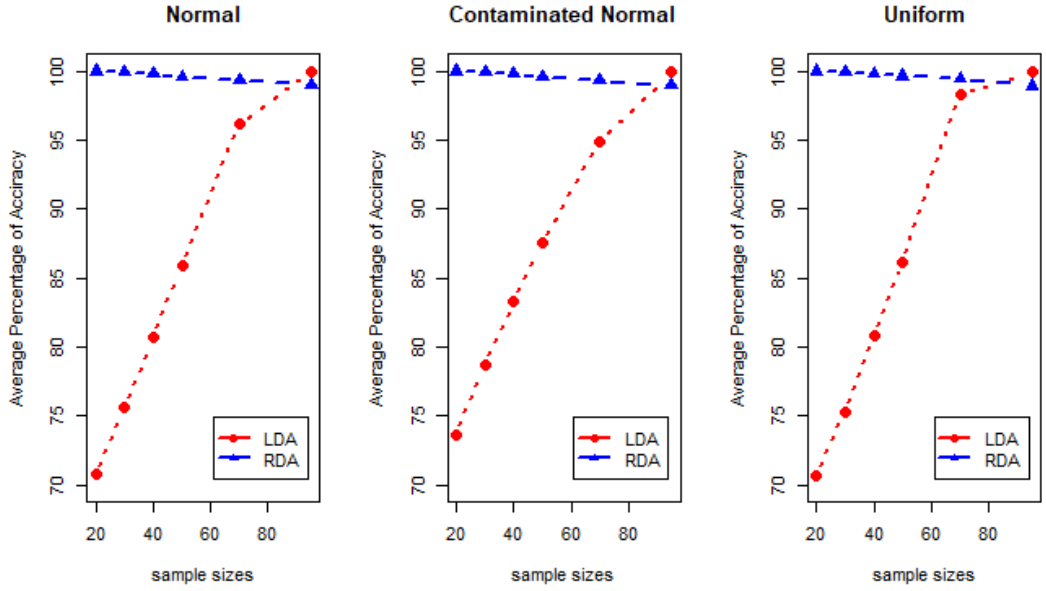

Fig. 6: The trend of the average accuracy percentage of linear discriminant analysis (LDA) and regularized discriminant analysis (RDA) under 100 independent variables.

4 Discussion

The classification performance for the binary response variable depended on the explanatory variables via the normal, contaminated normal, and uniform distributions shown in Table 2, Table 3, and Table 4. Starting with the first table, the average accuracy percentage in RDA for small explanatory variables is more significant than LDA for all sample sizes. Moreover, when the explanatory variables are increased to the moderate and high range, the average accuracy percentage in RDA is more significant than LDA in most sample sizes, as shown in Table 3 and Table 4. Meanwhile, in the largest sample sizes, the average accuracy percentage in LDA is greater than RDA. The average accuracy percentage increases when the sample sizes are increased, as shown in Fig. 4, Fig. 5, and Fig. 6. The several distributions give the same performance methods, but the normal and uniform distributions present the highest average accuracy percentage. The choice of data distribution plays a vital role in good classification accuracy, [15].

5 Conclusion

This paper provided a binary classification by applying the high-dimensional data for linear discriminant analysis (LDA) and regularized discriminant analysis (RDA). We explained the benefit of explanatory variables on several distributions for predicting binary response

variables. Through a simulation study, the RDA outperformed more than the LDA in most sample sizes. However, the LDA was reasonable working with the largest sample sizes.

When considering the distribution, the average accuracy percentage of the normal and uniform distributions was slightly different because of the symmetric distribution. In the case of outlier data, the RDA performed well for classification. These results explained that the RDA was adequate for a classification based on high-dimensional data in most cases. Therefore, we concluded that the RDA could classify the situation of the sizeable explanatory variable and the sample sizes. Furthermore, the RDA was recommended for small sample sizes, [16], and large dimensional data, [17]. For future work, the RDA can apply the classification of psychological tasks, [18].

The simulation data is mainly used in this research. For future work, the real dataset in high-dimensional distribution, especially medical data such as large-scale gene expression data for classification disease in small patients. This research focuses the discriminant classification. Then the machine learning method can apply in this case.

Acknowledgments:
This research is supported by King Mongkut's Institute of Technology Lad-krabang.

References:

[1] T. Ramayah, N.H. Ahmad, H. A. Halim, S, R. M. Zanai, M. H. Lo, Discriminant analysis: An illustrate example, *African Journal of Business Management*, Vol.4, No.9, 2010, pp. 1654-1667.

[2] C. Liu, Discriminant analysis and similarity measure, *Pattern Recognition*, Vol.47, No.1, 2014, pp.359 -367.

[3] A. Tharwat, T. Gaber, A. Ibrahim, A. E. Hassanien, Linear discriminant analysis: detailed tutorial, *Ai Communications*, Vol.30, No.2, 2017, pp.169-190.

[4] A. Sharma, K. K. Paliwal, Linear discriminant analysis for the small sample size problem: an overview, *International Journal of Machine Learning and Cybernetics*, Vol.6, 2017, pp.443-454.

[5] F. Zhu, J. Gao, J. Yang, N. Ye, Neighborhood linear discriminant analysis, *Pattern Recognition,* Vol.123, No.2, 2022, Article no. 108422.

[6] F. H. Friedman, Regularized discriminant analysis, *Journal of the American Statistical Association*, Vol.84, 1889, pp.165-175.

[7] S. Yang, H. Xiong, K. Xu, L. Wang, J. Bian, Z. Sun, Improving covariance-regularized discriminant analysis HER-based predictive analytics of diseases, *Applied Intelligence*, Vol.51, 2021, pp.377-395.

[8] K. Elkhalil, A. Kammoun, R. Couillet, T. Y. Al-Naffouri, M. S. Alouini, A large dimensional study of regularized discriminant analysis, *IEEE Transections on Signal Processing*, Vol.68, 2020, pp.2464-2479.

[9] H. Trevor, R. Tibshirani, A. Buja. Flexible Discriminant Analysis by Optimal Scoring, *Journal of the American Statistical Association*, Vol.89, No. 428, 1994, pp.1255-1270.

[10] D.M. Witten, R. Tibshirani, Penalized Classification using Fisher's Linear Discriminant, *The Journal of the Royal Statistical Society, Series B,* Vol.75, No.5, 2011, pp.753-772.

[11] T. Hastie, A. Buja, R. Tibshirani, Penalized Discriminant Analysis, *The Annals of Statistics*, Vol. 23, No. 1, pp. 73-102.

[12] T. Hastie, R. Tibshirani, Discriminant Analysis by Gaussian Mixtures, *The Journal of the Royal Statistical Society, Series B,* Vol.58, No.1, 1966, pp.155-176.

[13] B. Ghojogh, M. Crowley, *Linear and quadratic discriminant analysis: Tutorial*, 2019, Available at http:// arXiv preprint arXiv:1906.02590.

[14] Y.Guo, T. Hastie, R. Tibshirani, Regularized Discriminant Analysis and Its Application in Microarrays, *Biostatistics*, Vol. 1, No.1,2005, pp. 1-8.

[15] K. Ksushbu, P. Nishad, V. Kasyap, I. Gupta. A Classification and Distribution Model for Data Leakage Prevention and Detection, *International Research Journal of Modernization in Engineering Technology and Science*, Vol. 3, No.2, 2021, pp. 348-354.

[16] J. Ye, T. Wang, Regularized discriminant analysis for high dimensional, low sample size data, *Proceeding of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining.* Philadelphia, Pennsylvania, USA, 2006, pp.454-463.

[17] X. Yang, K. Elkhalil, A. Kammoun, T. Y. Al-Naffouri, M. S. Alouini, Regularized Discriminant Analysis: A large Dimensional Study, *2018 IEEE International Symposium on Information Theory (ISIT)*. Vali, Colorado,

USA, 2018, pp. 536-540.
[18] R. Fu, M. Han, Y. Tian, P. Shi, Improvement motor imagery EEG classification based on sparse common spatial pattern and regularized discriminant analysis, *Journal od Neuroscience.* Vol. 343, 2020, Article no. 108833.

Contribution of Individual Authors to the Creation of a Scientific Article:

-Autcha Araveeporn has conceptualized the research and organized the simulation process to the discussion.

-Somsri Banditvilai has derived the results and made the conclusion.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself:

This research is supported by King Mongkut's Institute of Technology Ladkrabang.

Conflict of Interest

The authors have no conflict of interest to declare.