ISSN 2229-5518

Classifying Students 2nd Level at National Crypto Institute with Vertex Discriminant Analysis and Kernel Discriminant Analysis

Yeni Farida, Muhammad Nur Aidi, I Made Sumertajaya

Abstract— Discriminant analysis is a multivariate technique to classify the observations into the known groups and one or more new observations are classified into one of the known groups based on the discriminant function. Classical discriminant analysis, also known as Fisher discriminant analysis, assumes that the data from each group have homogeneous variance-covariance matrices. If the data does not meet the assumption, we can use Vertex Discriminant Analysis (VDA) or Kernel Discriminant Analysis (KDA) to perform discriminant analysis. In this paper, we compare the performance of both methods in classifying "data". In classifying students based "Grade Point Achievent (GPA)", the data that we was use are final scores of subject in first semester and scores of psikotest. With the training data about 70%, the results show that 71.7% of the data was correctly classified using VDA and 37.7% of the data was correctly classified using KDA. For testing data, VDA have 43.5% data was correctly and KDA have 52.2% data was correctly, so VDA is better than KDA for validations and KDA is better than VDA for predictions in this case.

Index Terms— classifying students, kernel discriminant, national crypto institute, vertex discriminant

1 Introduction

Discriminant analysis is a multivariate technique in which the objective is to separate observations in a particular group and classify new observations into previously defined groups based on the discriminant function [3]. Discriminant analysis was introduced by Ronald A. Fisher in 1936, known as the Fisher discriminant analysis. The analysis requires data assuming that they have the same a variance-covariance matrices group. If the assumption is met, it will generate the best discriminant function that is the function that gives the minimum misclassified opportunity and vice versa.

As the science developed, discriminant analysis method is also growing, such as quadratic discriminant analysis, canonical discriminant analysis, kernel discriminant analysis, and vertex discriminant analysis. Other factors such as data distribution and variance-covariance matrices need to be considered in the selection of analytical methods to solve the problem of data classifying.

National Crypto Institute (NCI) is an official school whose students are garrisoned. Learning process at NCI is using the package system and applying the drop-out system. The use of the drop-out system lead to student achievement requires to be monitored. Student achievement consistency in learning activities can be seen from their Grade Point Achievent (GPA) value. Richardson et al. [2] states that the psychological aspect has a correlation to learning achievement. Therefore, classification of the best students in the 4th semester based on GPA in the 2nd and 3rd semester using academic scores and the scores of psychological tests is conducted in this research. By doing this classification, student's ability in the learning process can be seen; the student's potential can be detected so that if there are deviations can be repaired immediately. The accuracy and the exactness of the classification can be verified using discriminant analysis with a group of outstanding students as a response variable (Y) and the final score of nine subjects at the 1st half and the score of psychological tests (X). Discriminant analysis used in this research is the Vertex Discriminant Analysis (VDA) and Kernel Discriminant Analysis (KDA). The goal of this research is to compare the accuracy of VDA method classification and KDA classification.

2 Research Method

2.1 Data

The data used in this research is the empirical data in the form of NCI students 2nd level Year 2015/2016; consisting of the final score of the subjects in the first semester and the scores of psychological tests as explanatory variables X to measure the level of achievement (response variable Y). There were 21 explanatory variables used which was described in table 1.

- *Yeni Farida is currently pursuing masters degree program in applied statistics in Bogor Agricultural University, Indonesia, PH +6281298480822 E-mail: faridajunaedi@gmail.com*
- *Muhammad Nur Aidi is Lecturer, Departement of Statistics, Bogor Agricultural University, Bogor, Indonesia*
- *I Made Sumertajaya is Lecturer, Departement of Statistics, Bogor Agricultural University, Bogor,Indonesia*

TABLE 1
LIST OF VARIABLES

No	Variables	Explanation
1	X_1	Cryptanalysis Transposition Alphabetic
2	X_2	Introduction to Cryptography
3	X_3	Basic Math I
4	X_4	Electrical Magnet Physics
5	X_5	Electrical Magnet Physics Practice
6	X_6	Introduction to the Information Technology
7	X_7	Introduction to Statistics
8	X_8	Introduction to Public Administration
9	X_9	Information Technology Practice
10	X_{10}	Exact Technique Test
11	X_{11}	SatzerganzngValue – sentence completion
12	X_{12}	Wortausuahl Value – searching different word
13	X_{13}	Analogien Value - searching word association
14	X_{14}	Gmeinsamkeiten Value - searching for words that include two senses
15	X_{15}	Brands Eutgaben Value - remembering words
16	X_{16}	RechenAufgaben Value – simple math
17	X_{17}	ZahlenReihen Value – sequences of numbers
18	X_{18}	Form Ausuahl Value – form arrangement
19	X_{19}	WurfalAufgaben Value - cubes
20	X_{20}	*Intelligent Quotient*
21	X_{21}	Figural Ability Test

2.2 Methods of Data Analysis

The stages of data analysis in this research are as follow:

1. Define response variable Y
   Make a group of outstanding students based on the average GPA 2nd semester and 3rd.
2. Descriptive Analysis
   Descriptive analysis was performed to explore the general description of data patten that aimed to get the appropriate next analysis.
3. Box's M Test
   Box's M Test was performed to know the type variance-covariance matrices that were used.
4. Split the data into two, namely 70% of the training data and testing data is 30%.
5. Perform VDA analysis methods.
   VDA method discriminant function establishment of training data is carry out in stages as follows: [1]
   a. Determining initial iteration m = 0 where **A**(0)=0 and **b**(0)=0.
   b. Defining $y_j = \upsilon_j$ and determine the node value as a group indicator of each group by the following equation:
   $$\upsilon_j = \begin{cases} (k-1)^{-\frac{1}{2}}1 & \text{jika } j = 1 \\ c1 + de_{j-1} & \text{jika } 2 \le j \le k \end{cases}, \quad c = -\frac{1+\sqrt{k}}{(k-1)^{\frac{3}{2}}}, \ d = \sqrt{\frac{k}{k-1}},$$
   e_{j-1} is the vector standard unit with value 1 when all j and value 0 for the other in $\mathbb{R}^{k-1}$
   c. Majorize the regularized loss function in equation: $R(A,b) \le \frac{1}{n}\sum_{i=1}^{n_j} w_i \|r_i - s_i\|^2 + \lambda \sum_{j=1}^{k-1} \|a_j\|^2 + d$ with ith current residual $r_i^{(m)} = y_i - A^{(m)}z_i - b^{(m)}$ and d is a constant that depends on the remnant r_i at iteration m.
   d. Minimize the surrogate function and determine $\mathbf{A}^{(m+1)}$ dan $\mathbf{b}^{(m+1)}$ by solving k-1 set of linear equations with k are the number of groups.
   e. If $\|A^{(m+1)} - A^{(m)}\| < \gamma$ dan $|R(A^{(m+1)}, b^{(m+1)}) - R(Am,bm)<\gamma$ both hold for $\gamma=10-4$ then stop; otherwise repeat steps 3 through 5.
   f. Discriminant function is formed with parameters obtained and classification of observations i-this carried by the formula:
   $\hat{y} = \text{argmin}_{j=1,..,k} \|v_{j(i)} - \hat{A}z_{j(i)} - \hat{b}\|$.
   g. Evaluate the precision and accuracy of classification.

6. Perform KDA analysis methods
   Stages in the KDA method on training data use a normal distribution (based on event of explanatory variable), i.e:
   a. Determine the value of the kernel constant,
   $A(k_j) = \frac{4}{2p+1}$.
   b. Calculates optimum bandwidth of each group,
   $$h = \left(\frac{\left(A(k_j)\right)}{n_j}\right)^{\frac{1}{p+4}}$$
   c. Calculate the variance-covariance matrices (S_j) of each group, S_j^{-1}, $|S_j|$ and prior probability of each group.
   d. Calculating the kernel density function estimation of each observation with the equation[4]:
   $$\hat{f}_j(\boldsymbol{x}) = \frac{1}{n_j} \frac{1}{(2\pi)^{\frac{p}{2}} h^p |\boldsymbol{S_j}|^{\frac{1}{2}}} exp \sum_{i=1}^{n_j} \left(-\frac{(\boldsymbol{x}-\boldsymbol{y_i})^{\boldsymbol{t}} \boldsymbol{S}_j^{-1} (\boldsymbol{x}-\boldsymbol{y_i})}{2h^2}\right)$$
   e. Calculate the posterior probability of each observation with Bayes rule: $p(\pi_j|\boldsymbol{x}) = \frac{p_j . \hat{f}_j(\boldsymbol{x})}{\sum_{j=1}^{k} p_j . \hat{f}_j(\boldsymbol{x})}$
   f. Classification of x observation into groups j,
   $\hat{y} = \text{argmax}_{j=1,..,k}\, p(\pi_j|\boldsymbol{x})$
   g. Perform data testing classification with the function established from the data training.
   h. Evaluate the precision and accuracy of classification.

7. Comparing the results of the precision and accuracy of the classification of the two methods.

3 RESULT AND DISCUSSION

3.1 Descriptive Analysis

In this research, students' classification is using GPA and the division of discriminant group is using data distribution of the average GPA. The group division can be seen in Table 2.

TABLE 2
DESCRIPTION OF GPA FOR EACH GROUP

Group	n	Average	SD	Max	Min
1	22	2.695	0.190	2.89	2.24
2	30	3.062	0.076	3.19	2.90
3	24	3.365	0.128	3.62	3.20
Total	76				

TABLE 3
DESCRIPTION OF OUTSTANDING STUDENT'S DATA IN NCI FOR EACH VARIABLE IN EACH GROUP

Var	Group 1 Average	Group 1 SD	Group 2 Average	Group 2 SD	Group 3 Average	Group 3 SD	*p-value* Wilks' Lambda
X_1	76.09	9.36	82.3	9.77	85.31	7.37	0.003
X_2	68.78	8.04	73.68	6.45	76.54	9.35	0.005
X_3	70.14	7.18	74.37	8.13	80.53	8.02	0
X_4	64.35	5.68	66.33	5.93	70.01	7.15	0.01
X_5	76.67	3.68	77.81	2.71	76.86	3.21	0.369
X_6	68.53	6.96	73.41	7.6	75.23	9.7	0.02
X_7	66.11	6.22	70.4	8.57	75.72	9.28	0.001
X_8	70.92	5.27	77.19	6.61	77.39	8.53	0.002
X_9	75.92	10.2	79.74	8.49	83.28	7.89	0.023
X_{10}	76.64	10.47	84.03	8.63	84.75	9.3	0.007
X_{11}	99.73	5.84	101.4	5.9	101.33	7.66	0.608
X_{12}	113.5	6.84	115.1	5.64	113.04	7.33	0.48
X_{13}	104.86	5.72	105.87	5.38	104.96	5.99	0.773
X_{14}	120.73	5.92	118.77	8.27	119.25	6.52	0.608
X_{15}	112.91	9.42	113.43	7.79	115.25	6.28	0.563
X_{16}	104.32	9.92	104.33	7.46	105.88	12.72	0.824
X_{17}	111.27	8.35	113.57	7.95	112.5	9.46	0.635
X_{18}	106.32	7.56	106.63	8.76	108.25	6.87	0.661
X_{19}	112.91	9.92	111.63	9.01	109.17	12.98	0.479
X_{20}	119.46	6.94	121.57	8.5	120.96	9.23	0.66
X_{21}	100	11.68	104.93	12.9	108.75	16.58	0.108

Table 3 illustrates that students in group 3 had the highest scores average in almost every variable except variable X5,X12, X13, X14, X17. Based on Wilk's Lambda Test results from Table 3, it is discovered that only 9 variables were significant (between variables have different average) i.e variables with p-value less than 0.05. Based on these results the next variable to be used only 9 variables, i.e.X1, X2, X3, X4, X6, X7, X8, X9, X10.

3.2 Variance-covariance Matrices Test

Examination of variance-covariance matrices homogeneous is using M Box's test. The Box's test results shows that the p-value = 0609. P-value is greater than α = 0:05 which show that the test results are significant. Thus the variance-covariance matrices between groups are alike (homogeneous).

3.3 VDA Methods

Before establishment discriminant functions, the data is splited into two; 53 students as training data (70%) and 23 students as testing data (30%).

The VDA method used is based on linear, because the data is not normally double spread, thus the data is transformed into the standard normal form (z). The results of the package VDA analysis issued coefficients which form discriminant function as follows:

D_{1i}= -0.008 - 0.031z_{1i} + 0.031z_{2i} - 0.098z_{3i} - 0.011z_{4i} - 0.059z_{6i} - 0.063z_{7i} - 0.070z_{8i} - 0.023z_{9i} - 0.087z_{10i}

D_{2i}= -0.085 - 0.072z_{1i} - 0.001z_{2i} + 0.038z_{3i} + 0.021z_{4i} - 0.084z_{6i} - 0.013z_{7i} - 0.098z_{8i} + 0.052z_{9i} - 0.043z_{10i}

The value of D1i and D2i is a coordinate in Euclidean spacewhich determines the classification of observation to-i. In this research, VDA using three vertex as the center point of groups indicator; group 1 (0.707; 0.707), group 2 (0259; -0966) and group 3 (-0966; 0259). If the observations have value D1i dan D2i which are both positive, such observations will tend to be classified into group 1. The smaller the coefficient value then the value D1i dan D2i will be smaller, and vice versa.

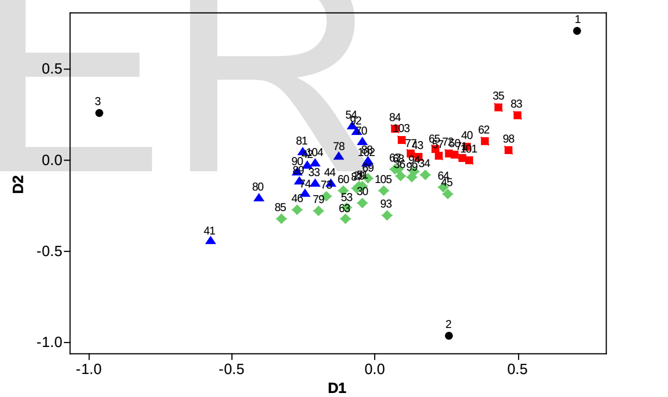

Fig.1. D_{1i} dan D_{2i} plot value as well as the central point of the class indicator

Figure 1 illustrates the classification of observations from this research. Observations with a red symbol will tend to be classified into group 1, green symbol into group 2 and blue symbol into groups 3. The classification accuracy of the VDA method training data can be seen in Table 4.

TABLE 4
THE CLASSIFICATION ACCURACY OF VDA

	Model Classification				
	Group	1	2	3	n
Actual Classification	1	12	2	1	15
	2	2	15	5	22
	3	1	4	11	16
	n	15	21	17	53

3.4 KDA Methods

The KDA method is non-parametric discriminant analysis that does not produce parameter as well as VDA method. Observations classificationare performed by Bayes' rule by taking advantage of posterior probability. The results of training data classification accuracy in KDA methods are presented in Table 5.

TABLE 5
THE CLASSIFICATION ACCURACY OF KDA

	Model Classification				
	Group	1	2	3	n
Actual Classification	1	5	3	7	15
	2	4	8	9	21
	3	4	6	7	17
	n	13	17	23	53

Based on Table 4 and Table 5, the observations of the training data are classified correctly by VDA method as much as 38 observations (71.7%) and by KDA method as much as 20 observations (37.7%).

3.5 The Evaluation of VDA and KDA Methods

The evaluation of analysis method accuracy is performed by classifying the testing data. As an illustration of the evaluation process, the student with ID number 61 (ID 61) is used as the first observation of testing for both methods.

VDA methods evaluation towards observation ID 61 with the discriminant function which is formed from the training data obtained value D1 = 0.1845 and D2 = 0.0588. Estimation classification of observations ID 61 presented in Table 6.

TABLE 6
THE CENTER POINT OF THE GROUP, THE DISTANCE BETWEEN THE CENTER POINTS OF OBSERVATION GROUP

Group		v_j	$\lVert v_j - \hat{A}z_i - \hat{b} \rVert$
1	v_1	0.7071 0.7071	0.8327
2	v_2	0.2588 -0.9659	1.0274
3	v_3	-0.9659 0.2588	1.1677

Based on Table 6, observation 61 is classified into group 1 because it has the closest distance to the center point of the group 1. The results of all testing data classification can be seen in Table 7.

TABLE 7
THE TESTING DATA CLASSIFICATION WITH VDA METHOD

	Model Classification				
	Group	1	2	3	n
Actual Classification	1	3	2	0	5
	2	3	4	4	11
	3	1	3	3	7
	n	7	9	7	23

It can be seen from Table 7, the VDA method appropriately classifies the testing data as many as 10 observations (43.5%). The results of testing data classifications are worse than the training data.

The evaluation for KDA methods against observations ID 61. Classification illustration for observation ID 61 are presented in Table 8. The results of the overall classification of the testing data can be seen in Table 9.

TABLE 8
ESTIMATE VALUE OF KERNEL DENSITY FUNCTIONS, PRIOR PROBABILITY AND POSTERIOR PROBABILITY

Group	$\hat{f}_t(x)$	p_t	$p(\pi_t \mid x)$
1	7.3969E-16	0.283	0.92471
2	5.9589E-17	0.396	0.07449
3	6.4049E-19	0.321	0.00080

Based on Table 8, observations with ID 61 are classified into group 1 for posterior odds of group 1 have the largest value.

TABLE 9
THE TESTING DATA CLASSIFICATION WITH KDA METHOD

	Model Classification				
	Group	1	2	3	n
Actual Classification	1	4	2	1	7
	2	1	4	4	9
	3	0	3	4	7
	n	5	9	9	23

It can be seen from Table 9, the KDA method properly classify the testing data as much as 12 observations (52.2%). Based on the results of testing data classification, it can be concluded that the KDA method provides better classification accuracy than VDA for the research.

4 CONCLUSION

Comparison of VDA and KDA methods on classification case studies of outstanding students 2nd level at NCI which data characteristics data is variance-covariance matrices of each group are homogeneous, yield the result that:

1. VDA method has the classification capability quite well with the percentage of classification accuracy of 71.7% for the training data and predictive capability of 43.5%.
2. KDA method has the ability classification with classification accuracy percentage of 37.7% for the training data and predictive capability of 52.2%.

It can be concluded that for the validation VDAmethodhas better capabilities while for the prediction KDA method has better capability in the case of the classification of the outstanding student.

ISSN 2229-5518

REFERENCE

[1] K Lange , TT Wu. 2008. An MM Encryption ForMulticategory Vertex Discriminant Analysis. J Comput Graph Stat. 17 (3): 527-544. (Journal or magazine citation)

[2] M Richardson,C Abraham,R Bond. 2012. Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis. *Psychological Bulletin.* 138(2):353-387. . (Journal or magazine citation)

[3] RA Johnson, DW Wichern. 2007. *Applied Multivariate Statistical Analysis*. New Jersey (US): Pearson Prentice Hall. 6th ed. (Book style)

[4] X Zhang, ML King, RJ Hyndman. 2004. Bandwidth Selection for Multivariate Density Estimation Using MCMC. (Working Paper)