STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **4**, June 2016, pp 118–131.
Published online in International Academic Press (www.IAPress.org)

The Best Model of the Swiss Banknote Data - Validation by the 95% CI of error rates and discriminant coefficients

Shuichi Shinmura *

Faculty of Economics, Seikei University, Japan

(Received: 11 October 2015; Accepted: 5 January 2016)

Abstract The discriminant analysis is not the inferential statistics since there are no equations for standard error (SE) of error rate and discriminant coefficient based on the normal distribution. In this paper, we proposed the "k-fold cross validation for small sample" and can obtain the 95% confidence interval (CI) of error rates and discriminant coefficients. This method is the computer-intensive approach by statistical and mathematical programming (MP) software such as JMP and LINGO. By the proposed approach, we can choose the best model with the minimum mean of error rate in the validation samples (**Minimum M2 Standard**). In this research, we examine the sixteen linear separable models of Swiss banknote data by eight linear discriminant functions (LDFs). M2 of the best model of Revised IP-OLDF is the smallest value of all models. We find all coefficients of six Revised IP-OLDF among sixteen models rejected by the 95% CI of discriminant coefficients (**Discriminant coefficient standard**). We compare t-values of the discriminant scores. The t-value of the best model has the maximum values among sixteen models (**Maximum t-value Standard**). Moreover, we can conclude that all standards support the best model of Revised IP-OLDF.

Keywords K-fold Cross Validation for Small Sample, Best Model, Fisher's Linear Discriminant Function (Fisher's LDF), Logistic Regression, Support Vector Machine (SVM), Revised IP-OLDF

AMS 2010 subject classifications 62G09, 62H30, 68R05, 90C05, 90C08, 90C11, 90C20

DOI: 10.19139/soic.v4i2.178

1. Introduction

There are four problems of the discriminant analysis [23, 27]. We developed an optimal linear discriminant function (Optimal LDF, Revised IP-OLDF) based on a minimum number of misclassification (minimum NM, MNM) criterion by the integer programming (IP) [15]. All LDFs except for Revised IP-OLDF cannot discriminate the cases on the discriminant hyperplane theoretically (**the problem 1**) [11, 12, 13, 14, 15]. Another LDFs except for a hard-margin SVM (H-SVM) [32] and Revised IP-OLDF cannot discriminate linear separable data theoretically (**the problem 2**). Because there is the defect of the generalized inverse matrix technique , a quadratic discriminant function (QDF) and regularized discriminant analysis (RDA) [5] cause the problem in the particular case (**the problem 3**). Revised IP-OLDF solves the problem1 and problem2. The discriminant analysis is not the inferential statistics because there are no equations for standard error (SE) of error rate and discriminant coefficient based on the normal distribution (**the problem 4**). Therefore, we proposed the "k-fold cross validation for small sample" method [18] and can obtain the 95% confidence interval (CI) of error rates and discriminant coefficients [25, 26]. We had better considered this method is not the traditional inferential statistics because it is the computer-intensive approach by statistical and mathematical programming (MP) software such as JMP [8] and LINGO [10]. By this

*Correspondence to: Faculty of Economics, Seikei University, 3-3-1 Kichijoji-kitamachi Musashino Tokyo, 180-8633 Japan. Email: sshinmura@gmail.com

ISSN 2310-5070 (online) ISSN 2311-004X (print)

break-through, we can choose the best model with the minimum mean of error rate in the validation samples (**Minimum M2 Standard**) instead of the leave-one-out (LOO) method [7]. We compare two statistical LDFs and six MP-based LDFs by this new model selection procedure. JMP script codes the theory of Fisher's LDF and logistic regression [1]. LINGO codes the theory of H-SVM, two soft-margin SVM (SVM4 for penalty c=10000, SVM1 for c=1), Revised IP-OLDF, Revised LP-OLDF by the linear programming (LP) and Revised IPLP- OLDF [19, 24]. Although we get the remarkable results by many small samples, we cannot explain the meaning of the 95% CI of discriminant coefficients well [18]. After many trials, we fix the constant of eight LDFs to 1. Seven LDFs except for Fisher's LDF become almost the same as a trivial LDF [28] for the pass/fail determination of exam scores [20]. In this research, we examine the sixteen linear separable models of Swiss banknote data [4] by eight LDFs. M2 of the best model of Revised IP-OLDF is the smallest value among eight LDFs. We find all coefficients of six Revised IP-OLDF among sixteen models rejected by the 95% CI of discriminant coefficients (**Discriminant coefficient standard**). We compare t-values of the discriminant scores. The t-value of the best model has the maximum values among sixteen models (**Maximum t-value Standard**). Therefore, we can conclude both standards support the best model of Revised IP-OLDF.

2. Method

In this research, we discriminate the Swiss banknote data and its re-sampling samples by eight LDFs. We focus on two means of error rates such as "M1 and M2" in the training and validation samples and propose the model with minimum M2 is the best model because M1 decreases monotonously as same as MNM. We compare eight M2s of the best model of eight LDFs and decide the final best model among eight LDFs. Moreover, we discuss the discriminant coefficients by fixing the constant of eight LDFs to 1. We confirm the validity of the best model by both the "discriminant coefficient and maximum t-value" standards.

2.1. Eight LDFs

In this research, we compare two statistical LDFs by JMP and six MP-based LDFs by LINGO. Two statistical LDFs are Fisher's LDF and a logistic regression [1]. Fisher proposed Fisher's LDF based on the variance-covariance matrices and found the discriminant analysis [2]. From 1971 to 1974, we developed the diagnostic logic between normal and abnormal symptoms of an electrocardiogram (ECG) data by Fisher's LDF and QDF. Our research was inferior to the decision tree logic developed by the medical doctor. After this experience, we concluded these discriminant functions are fragile for the discrimination of the normal and abnormal diseases because of two main reasons.

(1)There are many cases nearby the discriminant hyperplane. All LDFs except for Revised IP-OLDF cannot discriminate the cases on the discriminant hyperplane correctly (**the problem 1**). The problem1 means that NM of these LDFs may not be correct.

(2)If the value of some variables increases or decrease, the probability belonging to abnormal disease increases from 0 to 1. The discriminant functions based on the variance-covariance matrices assume the typical abnormal patients are the average of the abnormal classes. However, the typical patients are far from the normal patients. Taguchi and Jugulum proposed the Mahalanobis-Taguchi (MT) method using Mahalanobis distance based on the variance-covariance matrix [31]. They claim that the cases belonging to abnormal states are far from the normal state. Their claim is the same perception as our claim. Therefore, most Japanese researchers in the medical diagnosis use logistic regression in the equation (1) If some independent variable increases or decreases, the probability 'p' belonging to class1 (abnormal symptom) increases from 0 (class2) to 1 (class1). Although the maximum likelihood calculates the SE of logistic regression and the SE becomes the large value for the linear separable model [3], we judge it can discriminate the linear separable data correctly if NM of logistic regression and MNM of Revised IP-OLDF are zero.

$$Log(p/(1-p)) = f(\mathbf{x}) \tag{1}$$

Where
p: the probability belongs to class1;
x: the independent variables;
$f(\mathbf{x})$: the linear regression model.

We can obtain the maximum/minimum value of the function by MP, regardless of the presence or absence of constraints. Therefore, Schrage [9] introduced several definitions of regression models. Quadratic Programming (QP) defines the ordinal least square method. LP defines the "Least Absolute Values (LAV) Regression". Nonlinear Programming (NLP) represents several Lp-norm regression. However, there were few pieces of research about the regression analysis. On the other hands, there were many pieces of research about MP-based discriminant models [30]. However, statistical users rarely used these discriminant functions because there was no evaluation of the real data.

Vapnik [32] proposed three different SVM models. H-SVM in equation (2) indicates the discrimination of linearly separable data clearly if we delete two terms such as '$c * \Sigma e_i$' and '$-e_i$'. Real data are rarely linearly separable. For this reason, **S-SVM** has been defined in equation (2) with two objects. These two objects are combined by defining some "**penalty c**." However, S-SVM does not have the rule to determine 'c' uniquely. In this research, we evaluate two S-SVMs such as SVM4 (c=10^4) and SVM1 (c=1). We know the "M1 & M2" of SVM4 are almost better than SVM1. Some researchers misunderstand S-SVM can discriminate the linearly separable data exactly and prefer to choose the small penalty c without the examination of real data.

$$\min = ||\mathbf{b}||^2/2 + c * \Sigma e_i; \quad y_i * ({}^t\mathbf{x}_i\mathbf{b} + b_0) >= 1 - e_i; \tag{2}$$

Where
$y_i = 1/-1$ for $\mathbf{x}_i \in$class1/class2;
$\mathbf{x}_i$: p-independent variables (p-variables);
b: p-discriminant coefficients; b_0: the constant and free variable;
c: penalty c;
e_i: non-negative decision variable.

On the other hand, Shinmura developed **IP-OLDF** based on the MNM criterion [11, 16, 17]. We found two important facts about the discriminant theory as follows: 1) We defines IP-OLDF on both data and discriminant coefficient spaces before Revised IP-OLDF. N-linear hyperplanes made by the values of n-cases divide the discriminant coefficient space into finite convex polyhedron (CV). LDF corresponding to the interior point of CV has unique NM and misclassifies the same cases. Therefore, there is/are the optimal convex polyhedron with MNM (OCP). 2) Let us MNM_k be MNM of k-variables model and $\text{MNM}_{(k+1)}$ be MNM of (k+1)-variables model added one variable to the former model. MNM decreases monotonously ($\text{MNM}_k >= \text{MNM}_{(k+1)}$) because the (k+1)-dimensional space include k-dimensional space. If MNM_k = 0, all MNMs including these k variables are zero. IP-OLDF finds Swiss banknote data having six independent variables is linear separable by the model (X4, X6). Although there are sixty-three models (=2^6-1=63), we know MNMs of sixteen models including (X4, X6) are zero and forty-seven models are not linearly separable. By this fact, we are successful in feature selection of microarray data that composes several small gene sub-spaces with MNM=0 [29]. However, we found IP-OLDF cannot find true MNM if data does not satisfy the general position [15]. Therefore, we developed the **Revised IP-OLDF** in equation (3) that looks for the interior point of the OCP directly. If some LDF cannot look for the interior point of the convex polyhedron, it cannot solve **the problem 1**. Therefore, only Revised IP-OLDF is free from the problem1 and its NM equals to MNM. Other LDFs must count the number of cases on the discriminant hyperplane. If there are k cases on the discriminant hyperplane, real NM may increase k because we cannot discriminate these k-cases correctly. Moreover, only H-SVM and Revised IP-OLDF can recognize a linear separable model theoretically. Another LDFs cannot recognize a linear separable data/models and cannot judge the data is overlap or not because the status "not to overlap" equal to "MNM=0" (**the problem 2**).

$$\min = \Sigma e_i; y_i * ({}^t\mathbf{x}_i\mathbf{b} + b_0) >= 1 - M * e_i; \tag{3}$$

Where
e_i: 0/1 integer decision variable;

M: big M constant (M=10000);
b_0: free decision variables.

If e_i is a real non-negative variable in equation (3), we utilize **Revised LP-OLDF,** which is an L1-norm LDF. Revised IPLP-OLDF is a mixed model of Revised LP-OLDF and Revised IP-OLDF. The CPU time of Revised IPLP-OLDF was very faster than Revised IP-OLDF before 2012 [19]. However, it is slower than Revised IP-OLDF after 2012 because LINGO IP solver improves the computation time tremendously [24]. We expect Revised IPLP-OLDF can be supposed to obtain an estimate of MNM faster than Revised IP-OLDF for large samples in near future. Although we intend to use for the microarray data, Revised IP-OLDF can discriminate the microarray data very easy because the data is linearly separable [29].

In this research, we compare Revised IP-OLDF with seven LDFs by the "k-fold cross validation for small sample" method and evaluate the best model of Revised IP-OLDF among eight LDFs. Moreover, it established the Matroska feature selection method for the microarray data and revealed the structure of the microarray data.

2.2. The Model Selection Procedure by K-fold Cross Validation for Small Sample Method

We examined the effectiveness of Revised IP-OLDF by several small samples. It was difficult for us to compare Revised IP-OLDF with seven LDFs because we had no validation samples. Therefore, we proposed the "k-fold cross validation for small sample" method that is a combination of k-fold cross validation and resampling technique. We can evaluate eight LDFs by two means of error rates such as M1 and M2 in the training and validation samples [21, 22]. Although Fisher developed Fisher's LDF, he never formulated the equation of SEs of error rates and discriminant coefficients. Moreover, there were no good model selection procedures instead of the LOO method in the discriminant analysis [7].

In this research, we propose the new method and model selection procedure as follows:

1)We discriminate an original data by eight LDFs and two discriminant functions such as QDF and RDA. In principal, we discriminated all possible models. Goodnight established this technique in the regression analysis by the sweep operator [6]. By this technique, we can overlook the whole picture of the study.

2)We discriminate re-sampling samples by the new method. In this research, we fix k=100 to obtain two means of error rates and 95% CI of error rates and discriminant coefficients.

3)We consider the best model with the minimum values of M2 (**Minimum M2 Standard**) and compare eight M2s of eight best models. Although Vapnik defined the generalization ability, we claim the best model has good generalization ability. Moreover, we discuss the 95% CI of discriminant coefficients by fixing the constant of eight LDFs to 1 [28]. We expect all coefficient of the best model rejected at 5% level by the 95% CI of discriminant coefficient and t-value of the best model has a maximum value (**Maximum t-value Standard**). We compare t-value instead of p-value because the range of p-value is in [0, 1] that is more narrow than t-value.

3. Swiss Banknote Data

3.1. Outlook of Data

The Swiss banknote data consists of two kinds of bills such as 100 genuine ($y_i = 1$) and 100 counterfeit bills ($y_i = -1$). There are six independent variables (6-variables) such as X1 is the length of the bill; X2 and X3 are the widths of the left and right edges; X4 and X5 are the bottoms and top margin widths; X6 is the length of the image diagonal. **Table 1** shows the full model by the regression analysis. We judge only a coefficient of X1 accepted at 5% level. Forward stepwise selects the variable as follows: X6, X4, X5, X3, X2, and X1. We choose the models with minimum values of AIC, BIC, and $|Cp - (p + 1)|$. AIC chooses 5-variables model (X2-X6), BIC chooses 3-variables model (X4-X6), and Cp chooses the full mode. Because t-test, AIC, BIC and Cp statistics choose different three models, we cannot decide the appropriate model uniquely. On the other hand, IP-OLDF finds two variables model (X4, X6) was linear separable by the examination of all possible combination of six independent variables. Because MNM decreases monotonously ($\text{MNM}_k >= \text{MNM}_{(k+1)}$), we know sixteen models including (X4, X6) are linearly separable. On the other hand, other forty-seven models are not linearly separable. Therefore, we can choose the best model among these sixteen linear separable models.

Table 1. The full model by the regression analysis (Left) and four statistics by the forward stepwise technique (Right)*.

Var.	Coefficient	SE	t	p	R2	AIC	BIC	Cp
c	24.09	6.55	3.68	0.00				
X6	-0.21	0.02	-13.90	0.00	0.81	-34.16	-24.39	292.02
X4	0.15	0.01	14.77	0.00	0.88	-128.72	-115.74	107.00
X5	0.16	0.02	9.22	0.00	0.92	-205.74	-189.56	10.66
X3	0.11	0.04	2.77	0.01	0.92	-205.99	-186.64	10.26
X2	-0.12	0.04	-2.69	0.01	0.92	-210.90	-188.40	5.32
X1	0.02	0.03	0.56	0.57	0.92	-209.06	-183.43	7.00

*: We compute this table in Sept. 2015 again. Some values are different from the same old table.

Figure 1 is a scatter plot by (X4, X6). Genuine and Counterfeit bills are represented by the markers '◯' and '×', respectively. Two circles are 99 % confidence probability ellipses those are expected to include 99% bills in each ellipse if two classes are supposed to be normal distributions. We understand that counterfeit bills are not well controlled because those variance is larger than the genuine bills. There may be following reasons why no one finds this data is linearly separable. 1) Until now, only H-SVM can recognize linear separable models theoretically. However, it can adopt only for linear separable models. Therefore, no one tries to use H-SVM for the discrimination. In addition to this, researchers of SVM are interested in kernel SVM because its idea is attractive. 2) Nobody considers the importance of linear separable models. Many researchers claim the purpose of the discriminant analysis is to discriminate the overlapping data, not the linearly separable data. However, all LDFs except for H-SVM and Revised IP-OLDF cannot discriminate the overlapping data correctly because "MNM=0" defines the data is not overlap. Moreover, although many statisticians are interested in the analysis of the microarray data, it is linearly separable. 3) All possible combination of regression models gives us the clear and deterministic perception of the data. Therefore, we discriminate all possible combination of discriminant models as possible as we do. If some researchers try to check all combination of scatter plots the same as **Figure 1,** they may be suspicious the model (X4, X6) is linearly separable. IP-OLDF found this data was linear separable by the examination of sixty-three discriminations.

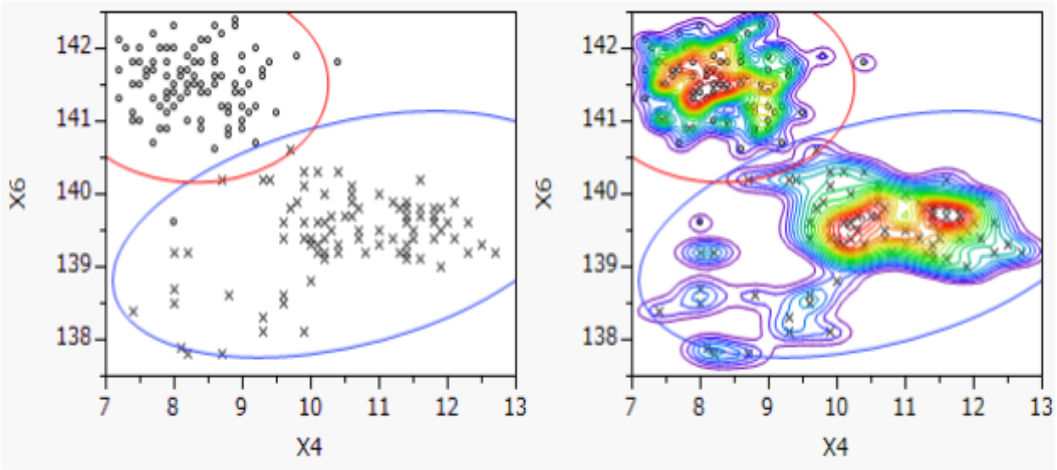

Figure 1. Swiss Banknote Data ($\mathrm{MNM}_{(X4,X6)} = 0.$)

Table 2. MNM and nine NMs of 16 linear separable models.

SN	p	var.	RIP	SVM1	LDF	QDF	RDA
1	6	1-6	0	1	1	1	1
2	5	2-6	0	1	1	1	1
3	5	1,3-6	0	1	1	1	1
4	5	1,2,4-6	0	1	1	1	1
5	5	1-4,8	0	2	1	1	1
8	4	3-6	0	1	1	1	1
9	4	2,4-6	0	1	1	1	1
10	4	1,4-6	0	1	1	1	1
11	4	2-4,6	0	2	1	1	1
12	4	1,3,4,6	0	2	1	1	1
13	4	1,2,4,6	0	2	2	1	1
23	3	4-6	0	1	1	1	1
24	3	3,4,6	0	2	1	1	1
25	3	1,4,6	0	2	2	2	1
26	3	2,4,6	0	2	1	1	1
43	2	4,6	0	2	3	1	1

In this research, we investigate a total of sixty-three discriminant models. Sixteen models including two variables (X4, X6) are linearly separable. Other forty-seven models are not linearly separable. This data is adequate whether or not eight LDFs can discriminate linear separable models correctly. We focus on the linearly separable models because evaluation is very explicit. **Table 2** shows the results of 16 linearly separable models. 'SN' is the sequential number of the discriminant model. 'p' is the number of variables and 'var.' is a suffix of the variable. '1-6' means 6-variables model (X1, X2, X3, X4, X5, X6). RIP column is MNM of Revised IP-OLDF. NMs of H-SVM, SVM4 for penalty c=10^4, Revised LP-OLDF (LP), Revised IPLP-OLDF (IPLP) and logistic regression (logistic) are zero and omitted from the table. "SVM1, LDF, QDF and RDA" columns show NMs of S-SVM (SVM1) for penalty c=1, Fisher's LDF (LDF), QDF and RDA. These four discriminant functions cannot recognize linear separable models. We observe this fact in other data. Therefore, we can conclude that S-SVM with small penalty c and discriminant functions based on the variance-covariance matrices are feeble for the discrimination of linearly separable models.

Table 3 shows the results of eleven models that exclude X6. We can understand MNMs are the minimum NMs of nine discriminant functions except for H-SVM. Revised IPLP-OLDF is the second best because it obtains the estimate of MNM. Next, we observed logistic regression is often better than other six LDFs.

Because we had no validation samples before the new method, we evaluated different LDFs in the training samples using simple regression as follows [13]. We could evaluate all LDFs by MNM in the training samples because MNM is the minimum NMs among all LDFs. In this research, we evaluate QDF and RDA by MNM also. These are not LDFs. **Table 4** shows the results of regression analysis such as "each NMs = c+b*MNM" using 63 NMs including sixteen linear separable models. QDF is the worst result because the constant of the simple regression is 1.59, and the discriminant coefficient is 0.99, which is almost 1. Simple regression line (QDF = 1.59 + 0.99*MNM) can predicts good NMs of QDF by MNMs because R-square is 0.991. This result implies us that NMs of QDF are almost 1.59 higher than MNMs. The error rate of QDF is 0.8% (=1.59/200) greater than that of Revised IP-OLDF, also. We conclude that NMs of QDF, RDA, LDF and SVM1 are at least 1.28 higher than MNMs. On the contrary, NMs of SVM4, Revised LP-OLDF, Revised IPLP-OLDF and logistic regression are half of NMs of these functions. Revised IPLP-OLDF is expected to be good estimates of MNM. However, Revised LP-OLDF and logistic regression are better than Revised IPLP-OLDF for 63 models. We agree everybody do not accept this explanation about the superiority of Revised IP-OLDF.

Table 3. MNM and eight NMs of 11 non-linear separable models.

Var.	RIP	SVM4	SVM1	LP	IPLP	logistic	LDF	QDF	RDA
1-5	2	3	3	3	2	2	7	6	6
1,3-5	2	3	3	3	2	2	7	6	6
2-5	2	3	5	3	2	2	6	6	5
1,2,4,5	2	3	5	3	2	2	8	6	8
1-4	12	17	17	17	12	15	19	14	14
3-5	2	3	4	3	2	2	6	6	5
2,4,5	2	2	5	2	2	2	8	6	5
1,4,5	2	4	5	4	2	2	9	6	6
1,3,4	13	17	17	17	14	15	19	16	16
2-4	13	17	17	17	13	17	19	15	15
1,2,4	13	19	19	19	13	16	22	18	18

Table 4. Comparison of eight discriminant functions [13].

	c b	$_R2$	MNM=1	MNM=40
SVM4	0.72 1.04	0.988	1.76	42.32
SVM1	1.57 1.02	0.989	2.59	42.37
LP	-0.03 0.97	0.998	0.94	38.77
IPLP	0.68 1.03	0.987	1.71	41.88
logistic	0.32 1	0.995	1.32	40.32
LDF	1.28 1.11	0.974	2.38	45.68
QDF	1.59 0.99	0.991	2.58	41.19
RDA	1.56 0.99	0.992	2.55	41.16

3.2. The 100-fold Cross Validation for Small Sample Method

We generate re-sampling sample from the Swiss banknote data and evaluate eight LDFs by our new method as follows.

1)Let n (=200) be the number of cases and p (=7) be the number of variables including the constant y_i ($y_i = 1$ for class1; $y_i = -1$ for class2). We copy the original data (n cases by p-variables) 100 times and generate pseudo-population sample (100*n cases by p-variables).

2)We add the random number to this sample (100*n cases by (p+1)-variables) and sort it in ascending order by the random number. We divide this sample by 100 sub-samples and add the sub-group number from 1 to 100.

3)We use 100 sub-samples as the training samples (n cases by (p+1)-variables) and the pseudo-sample as the validation sample (100*n cases by (p+1)-variables). This operation implies us that we re-sample 100 sub-samples from the pseudo-population. If we consider one sub-sample is the training sample, and other 99 sub-samples is the validation sample, we cannot estimate results uniformly because 100 validation samples are different. Moreover if we fix the validation sample uniquely we can control the training samples and validation sample very easy. For example, we can validate the validation sample is generated by the original data whether both samples is the same distribution. In this research, we discuss eight LDFs by our method.

Table 5 shows sixteen linearly separable models and eleven models corresponding to **Table 3**. "M1 & M2" are the mean error rates in the training and validation samples. All sixteen M1s of Revised IP-OLDF, H-SVM, SVM4, LP, IPLP and logistic regression are zero. SVM1 and Fisher's LDF cannot recognize all linear separable models. In other data, we observed SVM4, LP, IPLP and logistic regression cannot recognize all linear separable models. Only Revised IP-OLDF choose the third model as the best model, M2 of which is 0.26%. HSVM, SVM4, Revised

IPLP-OLDF, and Revised LP-OLDF chooses eighth model as the best model, M2 of those are 0.38, 0.37, 0.41 and 0.27%, respectively. SVM1 and logistic regression choose a twelveth model, M2 of those are 0.52 and 0.41%, respectively. Only Fisher's LDF chooses a seventh model, M2 of which is 0.54%. The best model of Revised IP-OLDF has the minimum value of M2 among eight LDFs. Seven 'M2Diff' of third models are 0.21, 0.21, 0.28, 0.23, 0.01, 0.26, and 0.29%, respectively. Next, we focus eleven models those are not linearly separable and choose the twenty-third model of Revised IPLP-OLDF as the best model among eleven models. Because six M2Diffs are 0.08, 0.84, -0.03, 0.12, 0.33 and 1.75%, Revised IPLP-OLDF and Revised IP-OLDF are better than another five LDFs for eleven models.

3.3. The 95% CI of Discriminant Coefficient

3.3.1. ***Consideration of twenty-seven models.*** We can obtain the 95% CI of the coefficient by our new method. We use the median as eight LDFs. **Table 6** shows the result of the median represented by the symbol. 'SN' is the sequential number of twenty-seven models. First sixteen models are a linear separable model that include (X4, X6). We select another eleven models without X6. Therefore, these models are not linearly separable. 'Model' shows the suffix of variables. Each model has two expressions by the symbol. The lower row is the original coefficient, and upper row is the modified coefficient by fixing the constant of LDF to 1. The rule of symbols are as follows:

1)Symbols "$+ -$ and 0" show the coefficients are positive (lower limit of the 95% $CI > 0$), negative (upper limit of the 95% $CI < 0$) and zero (the 95% CI include 0) at 5% significant level, respectively. If the model has the symbol '0', we should not choose this model by the "discriminant coefficient standard".

2)Symbol 'd' means the variable dropped from the model.

3)Symbol 'Z' means a hundred coefficients are 0s. If the model has the symbol 'Z', we consider this model is redundant. Moreover this model is the same as the model dropped the variables with the symbol 'd'.

4)Symbol '1' means a hundred constants are 1s. Symbol '*' means the constant is 1/0 because the same original constants are '0'.

3.3.2. ***Revised IP-OLDF***. The equation (4) is the full model of Revised IP-OLDF. We represent the full model as the symbol '+ZZ++- Z' in Table 6. The symbols of X2, X3, and the constant are 'Z'. This fact tells us we can drop these three variables from the full model, and the six variables model is redundant. If the symbol of the constant is 'Z' in the second row, the first row is the same as the second row because we need not divide by the original constant. Equation (5) shows the third model (X1, X3, X4, X5, X6). We choose this model as the best model for the minimum M2 standard. The symbol is '+dZ++-Z'. Equation (6) shows the fourth model (X1, X2, X4, X5, X6). The symbol is '+Zd++-Z'. Equation (7) displays the eighth model (X1, X4, X5, X6). The symbol is '+dd++-Z'. Because we think symbols of "Z & d" have the same effect at first, these four models are equivalent. However, four M2s of four models are different such as 0.30%, 0.26%, 0.30%, and 0.27% because the 95% CI of each coefficient are different. We can judge X2 and X3 are less meaningful among six variables.

$$\begin{aligned} \mathrm{SN}=1:\mathrm{RIP} &= 1.037*X1+Z*X2+Z*X3+2.197*X4+2.285*X5-1.812*X6+Z \qquad (4)\\ &= 1.037*X1+2.197*X4+2.285*X5-1.812*X6.\\ &\quad [0.147,1.455]\ [0.878,3.729]\ [0.539,4.278]\ [-2.438,-0.556] \end{aligned}$$

$$\begin{aligned} \mathrm{SN}=3:\mathrm{RIP} &= 1.037*X1+d*X2+Z*X3+2.197*X4+2.292*X5-1.908*X6+Z \qquad (5)\\ &= 1.037*X1+2.197*X4+2.292*X5-1.908*X6.\\ &\quad [0.231,1.353]\ [0.878,3.124]\ [0.539,4.049]\ [-2.317,-0.612] \end{aligned}$$

Table 5. 100-fold cross validation for small sample method.

RIP		M1	M2	t	Diff.	Model	
53m42s	1	0	0.30	453	0.30	1-6	
	2	0	0.77	307	0.77	2-6	
	3	0	0.26	456	0.26	1,3-6	
	4	0	0.30	453	0.30	1,2,4-6	
	5	0	0.70	243	0.70	1-4,6	
	6	0	0.74	409	0.74	3-6	
	7	0	0.75	419	0.75	2,4-6	
	8	0	0.27	454	0.27	1,4-6	
	9	0	0.77	362	0.77	2-4,6	
	10	0	0.63	379	0.63	1,3,4,6	
	11	0	0.62	379	0.62	1,2,4,6	
	12	0	0.69	402	0.69	4-6	
	13	0	0.67	353	0.67	3,4,6	
	14	0	0.60	379	0.60	1,4,6	
	15	0	0.66	366	0.66	2,4,6	
	16	0	0.47	359	0.47	4,6	
	23	0.84	1.69	315	0.85	2,4,5	
HSVM		M1	M2	t	Diff1	M1Diff	M2Diff
35m6s	1	0	0.53	-147	0.53	0.00	0.23
	2	0	0.46	182	0.46	0.00	-0.30
	3	0	0.46	-163	0.46	0.00	0.21
	4	0	0.45	-158	0.45	0.00	0.15
	5	0	0.72	141	0.72	0.00	0.02
	6	0	0.46	192	0.46	0.00	-0.28
	7	0	0.43	-185	0.43	0.00	-0.32
	8	0	0.38	-164	0.38	0.00	0.11
	9	0	0.70	149	0.70	0.00	-0.06
	10	0	0.66	147	0.66	0.00	0.03
	11	0	0.65	143	0.65	0.00	0.03
	12	0	0.39	184	0.39	0.00	-0.30
	13	0	0.63	147	0.63	0.00	-0.04
	14	0	0.60	142	0.60	0.00	-0.01
	15	0	0.59	142	0.59	0.00	-0.07
	16	0	0.46	140	0.46	0.00	-0.01
SVM4		M1	M2		Diff1	M1Diff	M2Diff
44m46s	3	0	0.464		0.46	0.00	0.21
	8	0	0.374		0.37	0.00	0.10
	23	1.21	1.764		0.56	0.37	0.08
SVM1		M1	M2		Diff1	M1Diff	M2Diff
46m17s	3	0.26	0.54		0.28	0.26	0.28
	12	0.32	0.52		0.21	0.32	-0.17
	23	1.94	2.52		0.58	1.11	0.84
IPLP		M1	M2		Diff1	M1Diff	M2Diff
47m31s	3	0	0.49		0.49	0.00	0.23
	8	0	0.41		0.41	0.00	0.14
	23	0.84	1.66		0.82	0.00	-0.03
LP		M1	M2		Diff1	M1Diff	M2Diff
19m58s	3	0.00	0.27		0.27	0.00	0.01
	8	0.00	0.27		0.27	0.00	0.00
	23	1.22	1.81		0.59	0.38	0.12
Logistic		M1	M2		Diff1	M1Diff	M2Diff
46m	3	0.00	0.52		0.52	0.00	0.26
	12	0.00	0.41		0.41	0.00	-0.27
	23	1.51	2.02		0.51	0.67	0.33
LDF		M1	M2		Diff1	M1Diff	M2Diff
55m	3	0.53	0.55		0.02	0.53	0.29
	7	0.51	0.54		0.03	0.51	-0.20
	23	3.10	3.43		0.33	2.27	1.75

Table 6. The 95% CI of the twenty-seven models.

SN	Model	RIP	LP	IPLP	HSVM	SVM4	SVM1	LDF	Logistic
1	1-6	+ZZ++-Z	0Z0−−+Z	000000*	0000001	0000001	0000001	1	0
		+ZZ++-Z	000−−+0	0Z0−−+Z	000++-0	000++-0	000++-0	0+−−-+-	0000000
2	2-6	d00000*	d00000*	d00000*	d000001	d00000o	d000001	1	0
		d00++-0	d00−−+0	d00−−+0	d00++-*	d00++-0	d00++-0	d+−−-+-	d000000
3	1,3-6	+dZ++-Z	-d0−−+Z	0d0000*	0d00001	0d00001	0d00001	1	0
		+dZ++-Z	0d0−−+0	0d0−−+0	0d0++-0	0d0++-0	0d0++-0	0d0++-;	0d00000
4	1,2,4-6	+Zd++-Z	00d−−+*	00d000*	00d0001	00d0001	00d0001	1	0
		+Zd++-Z	00d−−+0	00d−−++	00d++-0	00d++-0	00d++-0	00d++-+	00d0000
5	1-4,6	000+d-*	0000d0*	0000d0*	0000d01	0000d01	0000d01	1	0
		000+d-+	000-d+0	0000d+0	000+d-0	000+d-0	000+d-0	0-++d-+	0000d00
6	3-6	dd0000*	dd0000*	dd0000*	dd0++-1	dd0++-1	dd0++-1	1	0
		dd0++-0	dd0−−+0	dd0−−+0	dd0++-+	dd0++-0	dd0++-0	dd0−−+-	dd00000
7	2,4-6	d0d000*	d0d000*	d0d000*	d0d++-1	d0d0001	d0d++-1	1	0
		d0d++-0	d0d−−+0	d0d−−+0	d0d++-+	d0d++-0	d0d++-0	d0d++-;	d0d0000
8	1,4-6	+dd++-Z	0dd−−+*	0dd000*	0dd0001	0dd0001	0dd0001	1	0
		+dd++-Z	0dd−−+0	0dd−−++	0dd++-0	0dd++-0	0dd++-0	0dd++-+	0dd0000
9	2-4,6	d000d0*	d000d0*	d000d01	d00+d-1	d000d-1	d00+d-1	1	0
		d000d-0	d00-d+-	d000d+0	d00+d-+	d000d-+	d000d-+	d-++d-+	d000d00
10	1,3,4,6	0d00d0*	0d00d0*	0d00d0*	0d00d01	0d00d01	0d00d01	1	0
		0d0+d-0	0d0-d+0	0d00d+0	0d0+d-0	0d0+d-0	0d0+d-0	-d++d-+	0d00d00
11	1,2,4,6	00d+d-*	00d0d0*	00d0d0*	00d0d01	00d0d01	00d0d01	2	0
		00d+d-0	00d-d+0	00d-d+0	00d+d-0	00d+d-0	00d+d-0	00d+d-+	00d0d00
12	4-6	ddd++-*	ddd000*	ddd+00*	ddd++-1	ddd++-1	ddd++-1	1	0
		ddd++-0	ddd−−+-	ddd−−+0	ddd++-+	ddd++-+	ddd++-+	ddd++-+	ddd0000
13	3,4,6	dd00d-*	dd00d0*	dd00d01	dd0+d-1	dd00d-1	dd0+d-1	1	0
		dd0+d-+	dd0-d+-	dd00d+-	dd0+d-+	dd00d-+	dd00d-+	dd++d-+	dd00d00
14	1,4,6	0dd+d-*	0dd0d0*	0dd0d0*	0dd+d-1	0dd0d01	0dd+d-1	2	0
		0dd+d-+	0dd-d+-	0dd-d+0	0dd+d-+	0dd+d-0	0dd+d-0	0dd+d-+	0dd0d00
15	2,4,6	d0d+d-1	d0d0d01	d0d0d01	d0d+d-1	d0d+d-1	d0d+d-1	1	0
		d0d+d-+	d0d-d+-	d0d-d+0	d0d+d-+	d0d+d-+	d0d+d-+	d0d+d-+	d0d0d00
16	4,6	ddd+d-1	ddd+d-1	ddd+d-1	ddd+d-1	ddd+d-1	ddd+d-1	3	0
		ddd+d-+	ddd-d+-	ddd-d+-	ddd+d-+	ddd+d-+	ddd+d-+	ddd-d+-	ddd0d00
22	3-5	dd000d1	dd000d1	dd000d1	dd−−-d1	dd−−-d1	dd−−-d1	6	2
		dd0++d0	dd−-d+	dd0−−d0	dd+++d-	dd0++d0	dd+++d-	dd0++d-	dd0−−d+
25	1,3,4	0d00dd1	0d00dd*	0d00dd1	0d00dd1	0d00dd1	0d00dd1	19	15
		0d0+dd0	+d−-dd0	+d−−dd0	-d++dd0	-d++dd0	-d++dd0	-d++dd-	dd0−−d+
27	1,2,4	00d0dd1	00d0dd*	00d0dd1	00d0dd1	00d0dd1	00d0dd1	22	17
		00d+dd0	+-d-dd0	+-d-dd0	-+d+dd0	-+d+dd0	-+d+dd0	-+d+dd-	+-d-dd+

$$\begin{aligned} \text{SN} = 4 : \text{RIP} &= 1.037 * X1 + Z * X2 + d * X3 + 2.20 * X4 + 2.30 * X5 - 1.84 * X6 + Z \qquad (6)\\ &= 1.037 * X1 + 2.200 * X4 + 2.300 * X5 - 1.840 * X6. \\ &\quad [0.147, 1.455]\ \ [0.878, 3.729]\ \ [0.539, 4.278]\ \ [-2.438, -0.556] \end{aligned}$$

$$\begin{aligned} \text{SN} = 8 : \text{RIP} &= 1.037 * X1 + d * X2 + d * X3 + 2.197 * X4 + 2.3 * X5 - 1.84 * X6 + Z. \qquad (7)\\ &= 1.037 * X1 + 2.197 * X4 + 2.3 * X5 - 1.84 * X6. \\ &\quad [0.231, 1.297]\ \ [0.878, 3.192]\ \ [0.659, 4.242]\ \ [-2.314, -0.612] \end{aligned}$$

The equation (8) is the original twelveth model, the symbol of which is 'ddd++-0'. Shinmura [28] shows the remarkable result if we fix the constant of LDFs to 1 by dividing the original coefficients. We divide each original

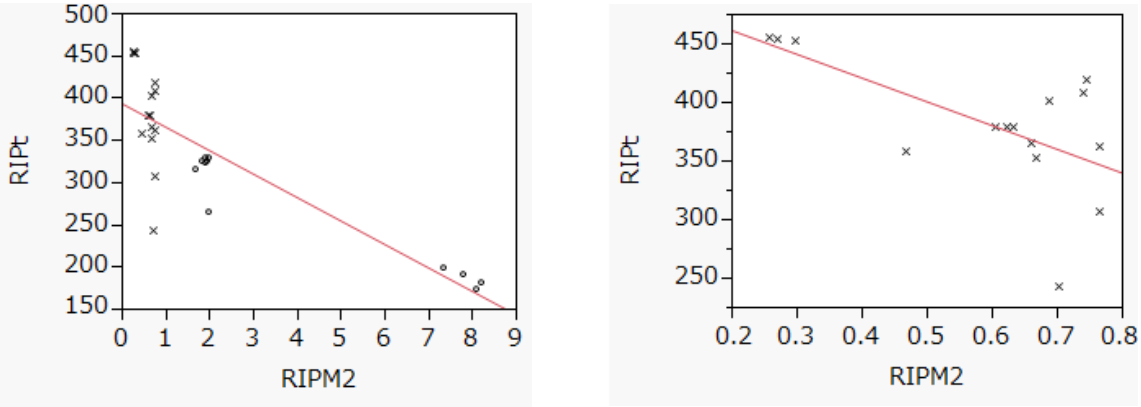

Figure 2. The scatter plots of M2 vs. t-value (Left: r= -0.8449, Right: r=-0.6678)

coefficient by (the constant (143.36) + 0.000001); it becomes 'ddd++-*' in equation (9). Because fewer twenty-five constants are zero, the constants have 0/1 values. We denote this status as the symbol '*'. We know the values of coefficients are quite different with the equation (4) (5) (6) (7). The equation(10) is the sixteenth model, the symbol of which is 'ddd+d-+'. If we divide the coefficients by the original constant, we obtain the equation (11) denoted by 'ddd+d-1'. The symbol '1' means all constant are one. Until now, we cannot understand the useful meaning of the 95% CI of discriminant coefficients. When we survey the full model of the pass/fail determination of exam scores, we find seven LDFs except for Fisher's LDF become the trivial LDFs. In this research, we investigate sixteen linear separable models and find all coefficients of six models have positive or negative values, not zero. The result of the "minimum M2 standard" and the "discriminant coefficient standard" matches well only for six linear separable models of Revised IP-OLDF. We claim the "discriminant coefficient standard" supports the best model.

$$\begin{aligned} \mathrm{SN}=12:\mathrm{RIP} &= 4.346*X4+5.432*X5-1.498*X6+143.356. \qquad (8)\\ &\quad [0.795,7.712]\ \ [0.884,11.873]\ \ [-2.598,-0.437]\ \ [0,250.687]\\ &= 0.023*X4+0.033*X5-0.012*X6+1/0 \qquad (9)\\ &\quad [0.0006,7E6]\ \ [0.005,1E7]\ \ [-1E6,-0.008]* \end{aligned}$$

$$\begin{aligned} \mathrm{SN}=16:\mathrm{RIP} &= 3.846*X4-5.321*X6+699.395. \qquad (10)\\ &\quad [0.364,44]\ \ [-48,-2.544]\ \ [345,6348]\\ &= 0.007*X4-0.008*X6+1. \qquad (11)\\ &\quad [0.001,0.009]\ \ [-0.008,-0.007](1) \end{aligned}$$

We calculate the discriminant scores by Revised IP-OLDF and calculate 't-value' of two classes in Table5. Although t-test assumes two distribution of discriminant scores belongs to the normal distribution, we ignore this assumption and use t-value as an index of evaluation for the discriminant scores. T-value of the best model is 456 and is the maximum value among twenty-four discriminant scores. **Figure 2** shows the scatter plots. The x-axis is M2s (RIPM2), and the y-axis is t-values (RIPt). Left plot is twenty-seven points, and the correlation is r = -0.84. Symbole '$\times$' are sixteen linear separable models and symbol '.' are eleven models. Right plot is only sixteen linear separable models and the correlation is r = -0.67. Although the best model has the maximum t-value (Maximum t-value Standard), we are afraid t-test always support the best model because the t-value of the best model may not the highest value of all models. However, this standard supports the best model for this data.

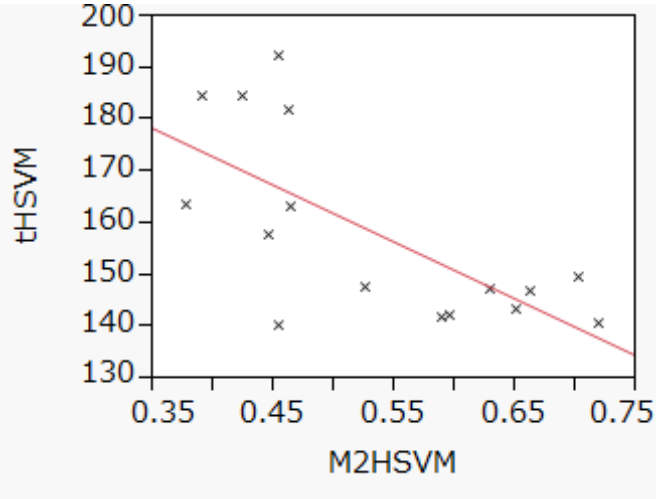

Figure 3. The scatter plots of M2 vs. t-value (r = -0.6871).

We can confirm the coefficients of all eleven models include '0.' Although we discuss the second best model among these eleven models in Table 5, we need not consider these models by the "discriminant coefficient standard".

3.3.3. ***Hard Margin SVM (HSVM) and Other LDFs.*** The equation (12) is the full model of H-SVM. The symbol is '000++-0'. To avoid zero divides, we divide the coefficients by (the original constant (-102.38) + 0.000001). Following this notation rule, we can represent the equation (13) as the symbol '0000001'. In **Table 6**, this equation is the first row, and original symbol is the second row because we think the model with the constant=1 is better than the initial coefficient. The sign of twelveth and sixteenth models are 'ddd++-1' and 'ddd+d-1', respectively. Although the twelveth model of HSVM is the best model among sixteen models, its sign is 'ddd++-1'. Although the discriminant coefficient standard supports the best model of HSVM, the maximum t value standard is not compatible with this model. We forecast HSVM is superior to Revised IP-OLDF for linear separable data because the support vector is efficient for the linear separable data. On the other hand, we forecast Revised IP-OLDF may overestimate for the linear separable data because it may look for one of the interior points of OCP arbitrary. From this research, we suppose it looks for the gravity of OCP that causes the results in stability. This forecast is the future work. **Figure 3** shows the scatter plot of sixteen models. The correlation is r = - 0.69.

$$\begin{aligned} \text{HSVM} = \quad & 0.993 * X1 + 0.567 * X2 + 0.276 * X3 + 1.645 * X4 + 1.391 * X5 - 1.635 * X6 - 102.38. \quad (12) \\ & [-0.57, 1.74] \ [-0.54, 1.41] \ [-1.06, 0.83] \ [0.37, 2.54] \ [0.49, 3.02] \ [-2.17, -0.59] \ [-456, 285] \\ = \quad & -0.004 * X1 - 0.003 * X2 + 0.0002 * X3 - 0.004 * X4 - 0.004 * X5 + 0.01 * X6 + 1. \quad (13) \\ & [-0.03, 0.13] \ [-0.02, 0.05] \ [-0.03, 0.07] \ [-0.1, 0.3] \ [-0.11, 0.24] \ [-0.27, 0.09] (1) \end{aligned}$$

The coefficients of 12^{th}-, 16^{th}- and twenty-second models of SVM4 and SVM1 do not include the symbol '0'. We need not discuss these models because the values of M2s are bigger than those of Revised IP-OLDF. Although Revised LP-OLDF and Revised IPLP-OLDF choose the sixteenth model, we need not discuss these models for the same reason. Because Fisher's LDF and logistic regression by JMP script do not output 100 discriminant coefficients, we discriminate the original data by the regression analysis and logistic regression. First rows of Fisher's LDF and logistic regression show the NMs by the original Swiss banknote data in **Table 6**. If we divide this number by two, it is the error rate because the sample size is 200 cases. The full models of Fisher's LDF and

logistic regression are the equation (14) and (15). Bracketed numbers are the standard errors. Coefficients right shoulder * is variable rejected at 5% level. The symbol of Fisher's LDF and logistic regression are '0++++++' and '0000000', respectively. **Table 6** shows the 95% CI supports the eighth model of Fisher's LDF. However, Fisher's LDF never discriminates sixteen linear separable models correctly. The logistic regression calculates the SE from Hessian matrix. The values of SEs are enormous, and all CI include zero. Therefore, JMP outputs a warning message such the "estimation is unstable" for the linearly separable model [3]. However, if we find "NM=0" on the ROC by JMP output and "MNM=0" by Revised IP-OLDF, we judge it is the linearly separable model. In general, we recommend using an exact logistic regression supported by SAS for avoiding this complex work. Although we cannot accept sixteen linear separable models because all coefficient accepts at 5% significant level, it can discriminate the linear separable models correctly. Although we accept only two models among eleven models, there is no meaning for the discrimination.

$$\begin{array}{rcl} \text{LDF} & = & -0.03X1 + 0.23^{*}X2 - 0.22^{*}X3 - 0.30^{*}X4 - 0.31^{*}X5 + 0.42^{*}X6 - 47.18^{*}. \\ & & \quad (0.06) \quad (0.09) \quad (0.08) \quad (0.02) \quad (12.18) \quad (0.03) \quad (13.10) \end{array} \tag{14}$$

$$\begin{array}{rcl} \text{Logistic} & = & 30.33X1 - 3.36X2 + 4.86X3 + 36.69X4 + 50.72X5 - 28.63X6 - 3594.33. \\ & & \quad (15) \quad (28411) \quad (35162) \quad (48244) \quad (8772) \quad (18142) \quad (8954) \quad (8608558) \end{array} \tag{15}$$

4. Conclusion

Table 7 shows the comparison of six models chosen by Revised IP-OLDF. Because SVM1 and Fisher's LDF cannot discriminate the linear separable models correctly, there is no need to compare among eight LDFs. We omit Revised IPLP-OLDF and Revised LP-OLDF because these LDFs are inferior to Revised IP- OLDF. Column 'M2' shows the value of M2. The number before the colon indicate the rank of useful models. The third model of Revised IP-OLDF has the most minimum value 0.26 among all models by eight LDFS and the maximum value 456 of the t-test. If we compare the third model of Revised IP-OLDF with HSVM, SVM4, and logistic regression, we can confirm the third model of Revised IP-OLDF is the best model. Both the coefficient and t-test standards support the "minimum M2 standard". Until now, we cannot decide the best model uniquely. Even if Fisher developed the SE of error rate and discriminant coefficients, we could not choose the best model uniquely. Because we obtain the powerful computer power and user-friendly solvers such as LINGO and JMP, we had better developed the new theory by the computer-intensive approach. Most statisticians respect Fisher. He opened a new frontier of much statistical theory by his thoughtful consideration without computer power. Therefore, we think no researchers discuss our claim seriously. However, we can use a powerful computer power and software such as statistical software JMP and MP solver LINGO. We are in the next generation blessed, unlike Fisher era. We should develop a new theory of discriminant analysis and discriminate every kind of data sets without restriction of the normal distribution.

Table 7. Comparison of Six Models

		RIP			HSVM			SVM4		Logistic
SN	Model	M2	coeff.	t	M2	coeff.	t	M2	coeff.	M2
1	1-6	4: 0.30	+ZZ++-Z	4: 453	6: 0.53	0000001	5-147	6:0.52	0000001	6: 0.55
3	1,3-6	1: 0.26	+dZ++-Z	1: 456	5: 0.46	0d00001	3:-163	4: 0.46	0d00001	4: 0.52
4	1,2,4-6	3: 0.30	+Zd++-Z	3: 453	3: 0.45	00d0001	4:-158	3: 0.45	00d0001	5: 0.55
8	1,4-6	2: 0.27	+dd++-Z	2: 454	1: 0.38	0dd0001	2:-164	1: 0.37	0dd0001	2: 0.46
12	4-6	6: 0.69	ddd++-*	5: 402	2: 0.39	ddd++-1	1: 184	2: 0.39	ddd++-1	1: 0.41
16	4,6	5: 0.47	ddd+d-1	6: 359	4: 0.46	ddd+d-1	6: 140	5 0.46	ddd+d-1	3: 0.47

Acknowledgments

This research started in 1997 and finished in 2015 by this paper. It was achieved by "What's Best! and LINGO of LINDO Systems Inc." and "SAS and JMP of SAS Institute Inc." The problem 4 in this paper is our last goal after 2012. Moreover, we can propose the Matroska feature selection method for the microarray data by our new theory of the discriminant analysis.

REFERENCES

1. Cox, DR., (1958). The regression analysis of binary sequences (with discussion). J Roy Stat Soc B 20: 215-242.
2. Fisher, R. A., (1936). The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, **7**, 179-188.
3. Firth, D., (1993). Bias reduction of maximum likelihood estimates. Biometrika, **80**, 27-39.
4. Flury, B., Riedel, H., (1988). Multivariate Statistics: A Practical Approach. Cambridge University Press.
5. Friedman, J. H., (1989). Regularized Discriminant Analysis. Journal of the American Statistical Association **84/**405, 165-175.
6. Goodnight J.H. (1981) A tutorial on the SWEEP Operator. The American Statistician 33, 149-158.
7. Lachenbruch, P. A., Mickey, M. R., (1968). Estimation of error rates in discriminant analysis. Technometrics, **10**, 1-11.
8. Sall, J. P., Creighton, L., Lehman, A., (2004). JMP Start Statistics, Third Edition. SAS Institute Inc.
9. Schrage, L., (1991). LINDO -An Optimization Modeling System (Fourth Edition). The Scientific Press.
10. Schrage, L., (2006). Optimization Modeling with LINGO. LINDO Systems Inc.
11. Shinmura, S., (1998). Optimal Linear Discriminant Functions using Mathematical Programming. Journal of the Japanese Society of Computer Statistics, **11 /** 2, 89-101.
12. Shinmura, S., (2000a). A new algorithm of the linear discriminant function using integer programming. New Trends in Probability and Statistics, **5**, 133-142.
13. Shinmura, S., (2000b). Optimal Linear Discriminant Function using Mathematical Programming. Dissertation, March 200, 1-101, Okayama Univ.
14. Shinmura, S., (2004). New Algorithm of Discriminant Analysis using Integer Programming. IPSI 2004 Pescara VIP Conference CD-ROM, 1-18.
15. Shinmura, S., (2007a). Comparison of Revised IP-OLDF and SVM. ISI2009, 1-4.
16. Shinmura, S., (2007b). Overviews of Discriminant Function by Mathematical Programming. Journal of the Japanese Society of Computer Statistics, **20/**1-2, 59-94.
17. Shinmura, S., (2009). Practical discriminant analysis by IP-OLDF and IPLP-OLDF. IPSI 2009 Belgrade VIPSI Conference CD-ROM, 1-17.
18. Shinmura, S., (2010a). The optimal linear discriminant function. Union of Japanese Scientist and Engineer Publishing.
19. Shinmura, S., (2010b). Improvement of CPU time of Revised IP-OLDF using Linear Programming. Journal of the Japanese Society of Computer Statistics, **22/**1, 39-57.
20. Shinmura, S., (2011a)Problems of Discriminant Analysis by Mark Sense Test Data. Japanese Society of Applied Statistics, **4/**3, 157-172.
21. Shinmura, S., (2011b). Beyond Fisher's Linear Discriminant Analysis - New World of Discriminant Analysis. ISI2011 CD-ROM, 1-6.
22. Shinmura, S., (2013). Evaluation of Optimal Linear Discriminant Function by 100-fold Cross-validation. 2013 ISI CD-ROM, 1-6.
23. Shinmura, S., (2014a). End of Discriminant Functions based on Variance-Covariance Matrices. ICORES, 5-14, 2014.
24. Shinmura, S., (2014b). Improvement of CPU time of Linear Discriminant Functions based on MNM criterion by IP. Statistics, Optimization and Information Computing, **2**, 14-129.
25. Shinmura, S., (2014c). Comparison of Linear Discriminant Function by K-fold Cross-validation. Data Analytic 2014, 1-6.
26. Shinmura, S., (2015a). The 95% confidence intervals of error rates and discriminant coefficients. Statistics, Optimization and Information Computing, **3**, 66-78.
27. Shinmura, S., (2015b). Four Serious Problems and New Facts of the Discriminant Analysis. In Pinson, E., Valente, F., Vitoriano, B., (Eds.), Operations Research and Enterprise Systems, 15-30, Springer (ISSN: 1865-0929, ISBN: 978-3-319-17508-9, DOI: 10.1007/978-3-319-17509-6).
28. Shinmura, S., (2015c). A Trivial Linear Discriminant Function. Statistics, Optimization and Information Computing, 322-335.
29. Shinmura, S., (2015d). Matroska Feature Selection Method for Microarray Data. Free paper (16) on Research Gate, 1-6.
30. Stam, A., (1997). Nontraditional approaches to statistical classification: Some perspectives on lp-Norm methods. Annals of Operations Research, **74**, 1-36.
31. Taguchi, G., Jugulum, R., (2002). The Mahalanobis-Taguchi Strategy - A Pattern Technology System. John Wiley & Sons.
32. Vapnik, V., (1995). The Nature of Statistical Learning Theory. SpringerVerlag.