A Trivial Linear Discriminant Function

Shuichi Shinmura *

Faculty of Economics, Seikei University, Japan

Abstract   In this paper, we focus on the new model selection procedure of the discriminant analysis. Combining re-sampling technique with k-fold cross validation, we develop a k-fold cross validation for small sample method. By this breakthrough, we obtain the mean error rate in the validation samples (M2) and the 95% confidence interval (CI) of discriminant coefficient. Moreover, we propose the model selection procedure in which the model having a minimum M2 was chosen to the best model. We apply this new method and procedure to the pass/ fail determination of exam scores. In this case, we fix the constant =1 for seven linear discriminant functions (LDFs) and several good results were obtained as follows: 1) M2 of Fisher's LDF are over 4.6% worse than Revised IP-OLDF. 2) A soft-margin SVM for penalty c=1 (SVM1) is worse than another mathematical programming (MP) based LDFs and logistic regression . 3) The 95% CI of the best discriminant coefficients was obtained. Seven LDFs except for Fisher's LDF are almost the same as a trivial LDF for the linear separable model. Furthermore, if we choose the median of the coefficient of seven LDFs except for Fisher's LDF, those are almost the same as the trivial LDF for the linear separable model.

Keywords   Trivial Linear Discriminant Function (LDF), Fisher's LDF, Logistic Regression, Support Vector Machine (SVM), Revised IP-OLDF.

1. Introduction

In this paper, we propose new model selection procedure of the discriminant analysis by the "k-fold cross-validation for small sample" method [14, 18]. Fisher [1] described the linear discriminant function (Fisher's **LDF**) and founded the discriminant theory. He never formulated the standard error (**SE**) of Fisher's LDF. Therefore, there was no sophisticated model selection procedure of the discriminant analysis. Although Lachenbruch et al. [4] had proposed a leave-one-out (**LOO**) method for model selection of the discriminant analysis, they could not achieve the new method because of lack of computer power. If we fix "k=100", we can obtain 100 LDFs and 100 error rates in the training and validation samples. From the 100 error rates, we calculate two means of error rates such **M1** and **M2** in the training and validation samples. We consider the model with minimum M2 among all possible combination models [3] is the best model. We apply this new method and procedure for three data sets of the pass/fail determinations and obtain good results. We had better distinguished these computer-intensive approaches from the traditional inferential statistics with the SE. Genuine statisticians without computer power had established the inferential statistics by their intellectual brain. Now, we can utilize the computer power of statistical and MP solver. The researchers who wish to discriminate their research data can decide the best model. We obtain the 95% confidence interval (CI) of the error rate and discriminant coefficients [14, 17, 18, 21, 22]. The new method and procedure give us precise and deterministic judgment about the model selection of the discriminant analysis.

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

Moreover, we obtain the good results of the discriminant coefficients by fixing the constant=1. Although we could obtain the 95% CI of the discriminant coefficients until now, we could not explain the meaning of those clearly [22]. By fixing the constant=1, most LDFs except for Fisher's LDF are almost the same as trivial LDFs.

2. Method

In this research, we discriminate three original data and those 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 M2s is the best model. We compare eight M2s of the best model of eight LDFs and decide the best model among eight LDFs. Moreover, we discuss the discriminant coefficients by the 95% CI for discriminant coefficients by fixing the constant=1.

2.1. Eight LDFs

In this research, we compare two statistical LDFs and six MP-based LDFs [21]. Two statistical LDFs are Fisher's LDF and a logistic regression. Fisher proposed Fisher's LDF based on the variance-covariance matrices and found the discriminant analysis. After Fisher's LDF, we can use a quadratic discriminant function (QDF) and the multi-class discrimination based on the variance-covariance matrices. 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 first problem of the discriminant analysis, **the problem1**). The problem1 means that the number of misclassifications (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 [25]. 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, the Framing study developed the logistic regression in the equation (1). If some independent variable increases or decreases, the probability 'p' belonging to class1 increases from 0 (class2) to 1 (class1). Most statisticians respect Fisher who opened a new frontier of much statistical theory by his intellectual consideration without computer power. Therefore, we think no researchers discuss this point seriously. However we can use a powerful computer power and software such as statistical software JMP [5] and MP solver LINGO [7]. We are in the next generation blessed unlike Fisher era. We should develop analytical techniques tailored to the characteristics of the individual data without the normal distribution.

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

where
p: the probability belongs to class1;
$\mathbf{x}$: the independent variables.

We can obtain the maximum/minimum value of the function by MP, regardless of the presence or absence of constraints. Therefore, Schrage [6] introduced several definitions of regression models. Quadratic Programming (QP) defines the ordinal least square method. Linear Programming (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 [24]. However, statistical users rarely used these discriminant functions because there was no evaluation of the real data.

Vapnik [26] proposed three different SVM models. H-SVM in equation (2) indicates the discrimination of linearly separable data clearly if we fix $c = 0$ and drop 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' correctly. In this research, two S- SVMs such as SVM4 ($c = 10^4$) and SVM1 ($c = 1$) are examined. 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.

$$\text{MIN} = \|\mathbf{b}\|^2/2 + c * \Sigma e_i; \quad y_i * (^t\mathbf{x_i}\mathbf{b} + b_0) >= 1 - e_i, \tag{2}$$

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

On the other hand, Shinmura [8] developed an Optimal LDF by Integer Programming (**IP-OLDF)** based on the MNM criterion. We found several new facts about the discriminant theory [9, 10]. However, we found IP-OLDF cannot find true MNM if data does not satisfy the general position [11, 12, 13]. Therefore, we developed the **Revised IP-OLDF** in equation (3). Only Revised IP-OLDF is free from the **problem1**. Other LDFs must count the number of cases on the discriminant hyperplane. Moreover, only H-SVM and Revised IP-OLDF can recognize a linear separable model theoretically. Another LDFs cannot recognize a linear separable data and cannot judge the data is overlap or not (the **problem 2**).

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

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 combined 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 [15]. However, it is slower than Revised IP-OLDF after 2012 because LINGO IP solver improves the computation time tremendously [20]. 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.

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.

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

We examined Revised IP-OLDF by several small samples. It was difficult for us to compare Revised IP- OLDF with seven LDFs because we could not validate the effectiveness of Revised IP-OLDF. Therefore, we proposed the "k-fold cross-validation for small sample" method and can evaluate eight LDFs by two means of error rates such as M1 and M2 in the training and validation samples. Although Fisher developed Fisher's LDF, he never formulated

Table 1. NMs of four discriminant functions by forward stepwise in 2012 midterm exams.

p	10%					50%					90%				
	Var.	RIP	Logi	LDF[*1]	QDF[*1]	Var.	RIP	Logi	LDF[*1]	QDF[*1]	Var.	RIP	Logi	LDF[*1]	QDF[*1]
1	4	4	8	6	6	4	12	12	14	14	3	8	30	12	12
2	2	**0**	**0**	**1**	**1**	1	6	5	9	9	1	5	12	9	9
3	1	0	0	1	1	2	3	3	8	8	4	3	3	10	10
4	3	0	0	1	1	3	**0**	**0**	7	7	**2**	**0**	**0**	**10**	**10**

*1: We obtained some NMs of Fisher's LDF and QDF by two options of "prior probability=Same" or "prior probability= Proportional" using JMP until now. After this paper, we fix the option as the "prior probability = Proportional".

the equation of SEs of error rates and discriminant coefficients. Therefore, there were no good model selection procedures instead of the LOO method.

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 a Regularized Discriminant Analysis (RDA) [2]. In principal, we discriminated all possible models. Goodnight established this technique in the regression analysis by the sweep operator [3]. 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. [22]

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=1.

2.3. Original Data and Re-sampling Sample

After 2010, we teach the preliminary statistical course for approximately 130 freshmen. Midterm and final exams consisted of 100 questions with ten choices. We consider discrimination using four testlet scores as independent variables [16]. If the pass mark is 50 points, we can easily obtain **a trivial LDF** ($f = T1 + T2 + T3 + T4 - 50$), NM of which is zero. If $f \geq 0$ or $f < 0$, the students pass or fail the exam, respectively. Usually, all LDFs except for Revised IP-OLDF are not free from the problem1. Because we can define the discriminant rule by exam scores (or independent variables), we can obtain above trivial LDF that is free from the problem1. We claim the pass/fail determination using exam scores gives us deep knowledge about the discrimination and offers good test data for linear separable data. We had many papers about the medical diagnosis. However, we have no connection with medical doctors now. Therefore, we use these data instead of the medical data.

Table 1 shows the discrimination of four testlet scores for 10%, 50% and 90% levels of the midterm exams in 2012. One hundred twenty-four students attended the exam. 'p' denotes the number of independent variables selected by the forward stepwise technique. In 10% level, the 2-variables model (T4, T2) was linearly separable. There are 15 discriminant models by all combination of four variables, only four LDFs of which are linearly separable. The pass mark is 36 points, and ten students fail the exam. In 50% level, the only full model is linearly separable. The pass mark is 63 points, and fifty-seven students fail the exam. In 90% levels, the only full model is linearly separable. The pass mark is 78 points, and one hundred twelve students fail the exam. "RIP and Logi" are Revised IP-OLDF and logistic regression. Both LDFs can discriminate a linear separable model correctly. However, logistic regression cannot sometimes discriminate the linear separable data correctly. On the other hand, Fisher's LDF and QDF cannot discriminate all linear separable models in these data.

Fig.1 are three scatter plots. X-axis is a first principal component. Y-axes correspond to second, third and fourth principal components from the left to right. Three 95% probability ellipses correspond to three groups such as $SCORE <= 35$, $36 <= SCORE <= 77$ and $78 <= SCORE$. Three groups consist of 10, 102 and 12 students. First and third groups are almost symmetry because two ellipses and cases are almost the same. If we check both
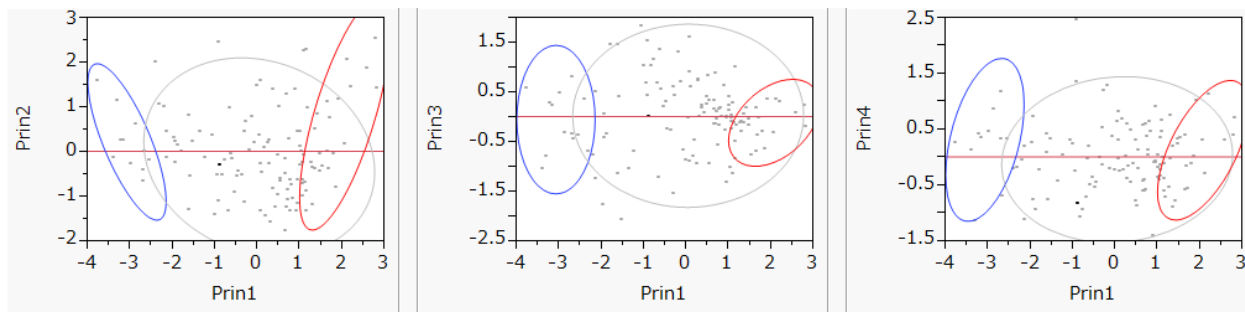
Figure 1. Scatter Plots of three groups.

NMs of the full model in three levels of Table 1, those of Fisher's LDF and QDF are 1s, 7s, and 10s, respectively. We cannot explain whether the increasing trend is common or not for another data.

3. Pass/Fail Determination by Exam Scores (50% level in 2012)

In this chapter, we discuss the discrimination at the 50% level. The pass mark is 63 points. **Table 1** tells us only full model is a linear separable model. We know a trivial LDF is $f = T1 + T2 + T3 + T4 - 63$. In this research, we omit four 1-variable models because there is no meaning about the discrimination.

3.1. NMs of Original Data

Table 2 shows the MNM and nine 'Diff2' of seven LDFs and two discriminant functions. Seven LDFs are as follows: H-SVM, SVM4, SVM1, Revised LP-OLDF (LP), Revised IPLP-OLDF (IPLP), logistic regression (Logistic) and Fisher's LDF (LDF). Two discriminant functions are QDF and RDA. Nine 'Diff2' are the differences of (nine NMs - MNM). The first column is the sequential number (SN) of eleven models from 4-variables to 2-variables that correspond to the second column 'Model'. We check the number of cases on the discriminant hyperplane of seven LDFs. We show it in the parenthesized numbers. Only five 2- variables model of Revised LP-OLDF cannot avoid the problem1. We cannot check the number of cases on the discriminant hyperplane of four statistical discriminant functions analyzed by JMP. SVM1, LDF, QDF, and RDA cannot recognize the linear separable model (the problem2). 'Diff2.' tells us the following facts.

1)We can roughly evaluate nine discriminant functions as follows. Revised IPLP-OLDF is the best result because ten NMs of IPLP are the same as MNM. Logistic regression is the second best because four NMs of 'Logistic' are the same as MNM. Three statistical discriminant functions except for logistic regression are the worst result.

2)Two models of LP have minus values because there are several cases on the discriminant hyperplane of these two models. We cannot evaluate the problem1 of other three models (SN = 9, 10, 11). Until now, we discuss the problem1 for the models having minus values of 'Diff2'. However, we cannot solve the problem1 for five models of Revised LP-OLDF clearly. Moreover, we cannot find the problem1 of four statistical discriminant functions because we cannot know the number of cases on the discriminant hyperplane. We expect statistical software output this number for users.

3.2. The Mean of Error Rates (M1 and M2)

Table 3 shows the results by the new method. We omit QDF and RDA because those are not LDFs. We examine 11 discriminant models of seven LDFs, and only full model of H-SVM. First 11 rows are eleven models of Revised IP-OLDF (RIP). M1s and M2s are the mean of error rates in the training and validation samples. We confirm the full models of eight LDFs are the best models because the values of M2 are the minimum values among 11

Table 2. MNM and nine Diff2 (50% level).

SN	MODEL	RIP	H-SVM	SVM4	SVM1	LP	IPLP	Logistic	LDF	QDF	RDA
1	1-4	0	0	0	3	0	0	0	7	5	4
2	1,2,4	3		1	1	1	0	0	5	5	3
3	2-4	5		3	2	3	0	0	5	5	3
4	1,3,4	3		0	0	0	0	1	6	2	2
5	1-3	13		4	5	4	0	1	4	3	2
6	2,4	6		1	1	-2(6)	0	1	7	4	6
7	1,4	5		2	2	2	0	0	4	3	3
8	3,4	10		1	2	-2(6)	0	1	1	1	0
9	1,2	16		2	3	0(6)	0	1	4	5	5
10	2,3	26		6	6	2(11)	3	1	7	7	7
11	1,3	17		2	1	0(3)	0	1	2	1	1

models. Eight M2s of the best models are 1, 1.09, 1.09, 1.79, 1.02, 1.06, 0.97, and 5.58%, respectively. "M1Diff2. & M2Diff." are the differences of (M1 and M2 of seven LDFs - those of RIP). Minimum 'M2Diff.' of seven LDFs are 0.09, 0.09, 0.79, 0.02, 0.06, -0.03 and 4.58%, respectively. M2 of Fisher's LDF is 4.579% greater than that of Revised IP-OLDF. Although we cannot evaluate the influence of the problem1, the best model of Logistic is 0.03% less than that of Revised IP-OLDF. Other five LDFs are not so bad because the 'M2Diff.' are within 0.79% higher than that of Revised IP-OLDF. The column 'Diff1' is the difference between of (M2 - M1). Some statisticians claim the model with a minimum value of 'Diff1' has good generalization ability. If this claim is correct, the sixth model (T1, T2, T3) of Fisher's LDF has good generalization ability. "M1 and M2" of this model are very high, 9.03% and 9.15%. Therefore, we cannot permit this claim. If we look at the second best model among ten models that are not linearly separable, only Fisher's LDF chooses the fourth model, M2 of which is 5.60%. On the other hand, other seven LDFs accept the second model. Second 'M2Diff.' of six LDFs are 0.41, 0.56, 0, 0.35, 0.27 and 3.03%, respectively. This fact may imply us Revised IP-OLDF is superior to other LDFs for the non-linear separable models, not for linear separable models.

3.3. The 95% CI of Discriminant Coefficients

We examine the 95% CI of the best models. **Table 4** shows the median and 95% CI of six MP-based LDFs by fixing the constant=1. It is our mistake Fisher's LDF and logistic regression by JMP script designed by us do not output 100 discriminant coefficients. If the 95% CI include zero, we can judge the pseudo-population coefficient is zero. If the value of 2.5% is greater than 0 or the value of 97.5% is less than 0, we estimate the pseudo-population coefficient a positive or negative value. Following this judgment, only five coefficients of T2 except for SVM1 are zero, and other coefficients are negative. Four medians of the full model are almost -0.016. This fact implies us these LDFs are the same as the trivial LDF in the equation (4):

$$F = -0.016 * T1 - 0.016 * T2 - 0.016 * T3 - 0.016 * T4 + 1 = -T1 - T2 - T3 - T4 + 62.5. \quad (4)$$

We can find this surprising result by fixing the constant=1. When we did not fix the constant, we could not find this result [22]. We discriminate the pseudo-population sample having 12,400 cases by the Fisher's LDF and logistic regression. The equation (5) is a logistic regression. If we divide five coefficients by 2178, we can obtain the same trivial LDF. Because the numbers in parentheses are SEs, all coefficients are zero.

$$\begin{aligned} \text{Logist1234} &= -34.33(5062) * T1 - 35.07(4631) * T2 - 35.27(4899) * T3 - 35.14(4607) * T4 + 2178(277555) \\ &= -0.016 * T1 - 0.016 * T2 - 0.016 * T3 - 0.016 * T4 + 1. \end{aligned} \quad (5)$$

On the other hand, we get Fisher's LDF in equations (6). We obtain these coefficients by the regression analysis. If we divide the coefficients by -3.22, we know Fisher's LDF is not the same as the trivial LDF because the third coefficient becomes -0.008. This fact tells us that Fisher's LDF does not follow the real data. It assumes the data is

Table 3. M1s and M2s of eight LDFs (50%level).

RIP	M1	M2	Diff1.	Model	
1	0	1.00	1.00	1,2,3,4	
2	2	3.12	1.45	1,2,4	
3	3	5.66	2.46	1,3,4	
4	1	3.36	1.86	2,4	
5	9.13	12.57	3.44	1,3,4	
6	4.51	5.57	1.07	1,2,3	
7	2.98	4.40	1.41	1,4	
8	6.24	9.16	2.92	3,4	
9	11.75	14.35	2.60	1,2	
10	19.42	22.16	2.74	2,3	
11	12.18	14.58	2.40	1,3	
H-SVM	M1	M2	Diff1.	M1Diff.	M2Diff.
1	0.00	1.09	1.09	0.00	0.09
SVM4	M1	M2	Diff1.	M1Diff.	M2Diff.
1	0.00	1.09	1.09	0.00	0.09
2	2.61	3.53	0.92	0.94	0.41
9	14.57	15.21	0.65	2.82	0.86
SVM1	M1	M2	Diff1.	M1Diff.	M2Diff.
1	0.47	1.79	1.32	0.47	0.79
2	2.82	3.68	0.86	1.15	0.56
7	4.38	5.06	0.68	1.40	0.66
IPLP	M1	M2	Diff1.	M1Diff.	M2Diff.
1	0.00	1.02	1.02	0.00	0.02
2	1.67	3.12	1.45	0.00	0.00
LP	M1	M2	Diff1.	M1Diff.	M2Diff.
1	0.00	1.06	1.06	0.00	0.06
2	2.59	3.47	0.88	0.92	0.35
9	13.91	14.56	0.65	2.16	0.21
Logistic	M1	M2	Diff1.	M1Diff.	M2Diff.
1	0.00	0.97	0.97	0.00	-0.03
2	2.60	3.39	0.79	0.93	0.27
9	14.31	14.60	0.30	2.56	0.26
LDF	M1	M2	Diff1.	M1Diff.	M2Diff.
1	5.15	5.58	0.43	5.15	4.58
2	5.65	6.15	0.49	3.98	3.03
4	5.40	5.60	0.21	3.90	2.25
6	9.03	9.15	0.12	4.52	3.58

the normal distribution. The numbers in parentheses are SEs, and all coefficients are rejected at 5%.

$$
\begin{aligned}
LDF1234 &= 0.059(0.001^*)*T1 + 0.056(0.003^*)*T2 + 0.026(0.002^*)*T3 + 0.056(0.0007^*)*T4 - 3.22(0.026^*) \\
&= -0.018*T1 - 0.017*T2 - 0.008*T2 - 0.017*T4 + 1. \tag{6}
\end{aligned}
$$

Table 4. The 95% C.I. of six LDFs.

RIP	T1	T2	T3	T4	c
97.5 %	-0.0075	0.0082	-0.0069	-0.0117	1
Median	-0.0160	-0.0160	-0.0161	-0.0162	1
2.5 %	-0.0204	-0.0243	-0.0262	-0.0246	1
H-SVM	T1	T2	T3	T4	c
97.5 %	-0.0075	0	-0.0081	-0.0120	1
Median	-0.0160	-0.0160	-0.0160	-0.0160	1
2.5 %	-0.0196	-0.0217	-0.0258	-0.0232	1
SVM4	T1	T2	T3	T4	c
97.5 %	-0.0075	0	-0.0081	-0.012	1
Median	-0.0160	-0.0160	-0.0160	-0.0160	1
2.5 %	-0.0196	-0.0217	-0.0258	-0.0232	1
SVM1	T1	T2	T3	T4	c
97.5 %	-0.0106	-0.0010	-0.0085	-0.0121	1
Median	-0.0154	-0.0161	-0.0164	-0.0171	1
2.5 %	-0.0202	-0.0230	-0.0250	-0.0208	1
IPLP	T1	T2	T3	T4	c
97.5 %	-0.0080	0.0071	-0.0069	-0.0118	1
Median	-0.0160	-0.0160	-0.0160	-0.0160	1
2.5 %	-0.0204	-0.0239	-0.0272	-0.0243	1
LP	T1	T2	T3	T4	c
97.5 %	-0.0095	0.0071	-0.0082	-0.0117	1
Median	-0.0160	-0.0160	-0.0160	-0.0160	1
2.5 %	-0.0204	-0.0243	-0.0264	-0.0234	1

4. Pass/Fail Determination by Exam Scores (90% level in 2012)

In this section, we discuss the discrimination at the 90% level. The pass mark is 78 points. **Table 1** tells us only full model is a linear separable model. We know a trivial linear separable LDF is $f = T1 + T2 + T3 + T4 - 78$ (or 77.5).

4.1. NMs of Original Data

Table 5 shows the "MNM and nine 'Diff2.' of eight LDFs and two discriminant functions. We check the number of cases on the discriminant hyperplane of six MP-based LDFs. Three models of Revised LP-OLDF cannot avoid the problem1. We cannot check the number of cases on the discriminant hyperplane of four statistical discriminant functions. Three statistical discriminant functions and SVM1 cannot recognize the linear separable model (the problem2). 'Diff2' tells us the following facts.

1)We can roughly evaluate nine discriminant functions as follows. Revised IPLP-OLDF is the best because ten NMs of IPLP are the same as MNM.

2)SVM1, LDF, RDA, and QDF cannot discriminate the linearly separable model exactly (the problem2).

3)Although logistic regression is the second best at the 50% level, ten NMs of non-linear separable models are greater than MNM. Moreover, logistic regression is worse than SVM4, SVM1, LP, and IPLP. This result is crucial because logistic regression is as same as IPLP and is better than SVM4, SVM1, and LP in other data.

4)Three statistical discriminant functions are worse than RIP and IPLP.

5)We find the problem1 of Revised LP-OLDF for three models (SN = 3, 8,10).

Table 5. MNM and nine Diff2 (90% level).

SN	Model	RIP	H-SVM	SVM4	SVM1	LP	IPLP	Logistic	LDF	QDF	RDA
1	1-4	0	0	0	1	0	0	0	10	1	10
2	1,2,4	3		1	3	1	1	3	9	6	9
3	2-4	5		1	0	0(1)	0	2	4	3	4
4	1,3,4	2		1	1	1	0	1	8	5	8
5	1-3	5		2	2	2	0	2	5	2	5
6	2,4	5		2	2	2	0	10	7	2	7
7	1,4	5		3	3	3	0	5	7	1	7
8	3,4	6		4	5	*4(1)*	0	5	3	5	3
9	1,2	11		1	1	1	0	27	1	6	1
10	2,3	8		0	2	0(2)	0	32	2	3	2
11	1,3	7		2	2	2	0	5	2	2	2

4.2. The Mean of Error Rates (M1 and M2)

Table 6 shows the results by the new method. SVM1 and Fisher's LDF cannot recognize the full model is linearly separable. Especially, M1 of Fisher's LDF is 11.613%. Only Fisher's LDF chooses the fourth model as the best model. Other LDFs choose the full model as the best model. Seven 'M2Diff.' of full model are - 0.03, -0.03, 0.16, -0.02, -0.05, 0.08 and 11.53%, respectively. Although we cannot evaluate the influence of the problem1, four best models of H-SVM, SVM4, IPLP, and LP are better than Revised IP-OLDF within 0.05%. This fact implies us M2 of Revised IP-OLDF may be wrong for the linearly separable models. On the other hand, the best model of Fisher's LDF is 11.525% worse than RIP. If we look at the best model among ten models that are not linearly separable, all LDFs choose a fourth model. Six 'M2Diff.' of fourth model are 0.29, 0.78, 0.04, 0.23, 0.33 and 10.25%, respectively. Although M2 of Revised IP-OLDF is less than six LDFs, five LDFs except for Fisher's LDF are not so bad. Only Fisher's LDF is 10.25% worse than Revised IP-OLDF. This fact may imply us that Revised IP-OLDF is superior to other LDFs for non-linear separable models, not for linear separable models. Sixth 'Diff1' of Fisher's LDF is the minimum value - 0.152%. This fact means that sixth M2 (26.72%) is 0.15% less than M1 (26.87%). Some statisticians claim Fisher's LDF has good generalization ability in this model. Our research shows this claim is entirely wrong.

4.3. The 95% CI of Discriminant Coefficients

Table 7 shows the median and 95% CI of six MP-based LDFs. Three coefficients of T2 of RIP, IPLP and LP are zero, and other coefficients are negative. If four medians of the full model are -0.0128, this LDF is the same as a trivial LDF such as f = T1+T2+T3+T4-78 (or 77.5). All MP-based LDFs are almost equal to the trivial LDF in the equation (7).

$$F = -0.0128 * T1 - 0.0128 * T2 - 0.0128 * T3 - 0.0128 * T4 + 1 = -T1 - T2 - T3 - T4 + 78. \quad (7)$$

The equation (8) shows logistic regression is almost the same trivial LDF. On the other hand, Fisher's LDF in equations (9) is different from the trivial LDF.

$$
\begin{aligned}
\text{Logist1234} &= -22.8(1123) * T1 - 26.84(2945) * T2 - 27.46(1352) * T3 - 23.94(1167) * T4 + 1873.27(89580) \\
&= -0.0122 * T1 - 0.0143 * T2 - 0.0147 * T3 - 0.0128 * T4 + 1.
\end{aligned} \quad (8)
$$

$$
\begin{aligned}
\text{LDF1234} &= 0.026(0.013^*) * T1 - 0.006(0.026) * T2 + 0.080(0.017^*) * T3 + 0.009(0.0067) * T4 - 1.788(0.24^*) \\
&= -0.015 * T1 + 0.003 * T2 - 0.045 * T3 - 0.005 * T4 + 1
\end{aligned} \quad (9)
$$

Table 6. M1s and M2s of eight LDFs (90%).

RIP		M1	M2	Diff1.	Model	
14m8s	1	0	1.06	1.06	1,2,3,4	
	4	0.61	2.32	1.70	1,3,4	
	6	3.96	4.69	0.73	2,4	
H-SVM		M1	M2	Diff1.	M1Diff.	M2Diff.
5m23s	1	0	1.02	1.02	0	-0.03
SVM4		M1	M2	Diff1.	M1Diff.	M2Diff.
9m23s	1	0	1.02	1.02	0	-0.03
	4	1.07	2.61	1.53	0.46	0.29
	11	6.28	6.95	0.67	1.90	0.46
SVM1		M1	M2	Diff1.	M1Diff.	M2Diff.
8m53s	1	0.20	1.22	1.02	0.20	0.16
	4	1.70	3.10	1.40	1.09	0.78
	10	7.83	8.45	0.62	1.44	1.21
IPLP		M1	M2	Diff1.	M1Diff.	M2Diff.
13m16s	1	0	1.04	1.04	0	-0.02
	4	0.61	2.35	1.74	0	0.04
	6	4.19	5.10	0.90	0.23	0.41
LP		M1	M2	Diff1.	M1Diff.	M2Diff.
4m23s	1	0	1.01	1.01	0	-0.05
	4	1.02	2.54	1.52	0.41	0.23
	6	5.43	6.07	0.65	1.47	1.39
Logistic		M1	M2	Diff1.	M1Diff.	M2Diff.
12m	1	0	1.14	1.14	0	0.08
	4	1.28	2.64	1.36	0.669	0.33
	10	7.40	7.51	0.11	1.008	0.27
LDF		M1	M2	Diff1.	M1Diff.	M2Diff.
15m	1	11.61	12.58	0.97	11.62	11.53
	4	11.96	12.56	0.60	11.35	10.25
	6	26.87	26.72	-0.15	22.911	22.03

5. Pass/Fail Determination by Exam Scores (10% level in 2012)

In this section, we discuss the discrimination at the 10% level. The pass mark is 36 points. **Table 1** tells us four models are linearly separable models. A trivial LDF is $f = T1 + T2 + T3 + T4 - 36$ (35.5).

5.1. NMs of Original Data

Table 8 shows the MNM and nine 'Diff2' of seven LDFs and two discriminant functions. There is no model having the cases on the discriminant hyperplane. SVM1, LDF, QDF, and RDA cannot recognize twelve linear separable model among sixteen linear separable models. 'Diff2' tells us the following facts.

1)Revised IPLP-OLDF is the same as MNM.

2)H-SVM, SVM4, Revised LP-OLDF, Revised IPLP-OLDF and logistic regression can recognize four linear separable models. However, most models of seven non-linear separable models are worse than RIP.

5.2. The Mean of Error Rates (M1 and M2)

Table 9 shows the results by the new method. We obtained the following outcomes.

Table 7. The 95% CI of six LDFs.

RIP	T1	T2	T3	T4	c
97.5 %	-0.0075	0.0045	-0.0063	-0.0048	1
Median	-0.0128	-0.0141	-0.0150	-0.0128	1
2.5 %	-0.0190	-0.0357	-0.0255	-0.0147	1
H-SVM	T1	T2	T3	T4	c
97.5 %	-0.0080	-0.0020	-0.0039	-0.0079	1
Median	-0.0130	-0.0130	-0.0152	-0.0128	1
2.5 %	-0.0190	-0.0292	-0.0228	-0.0171	1
SVM4	T1	T2	T3	T4	c
97.5 %	-0.0080	-0.0018	-0.0039	-0.0079	1
Median	-0.0129	-0.0130	-0.0142	-0.0128	1
2.5 %	-0.0190	-0.0292	-0.0228	-0.0171	1
SVM1	T1	T2	T3	T4	c
97.5 %	-0.0081	-0.0022	-0.0039	-0.0102	1
Median	-0.0130	-0.0130	-0.0128	-0.0130	1
2.5 %	-0.0188	-0.0253	-0.0210	-0.0171	1
IPLP	T1	T2	T3	T4	c
97.5 %	-0.0080	0.0058	-0.0062	-0.004	1
Median	-0.0130	-0.0141	-0.0150	-0.0128	1
2.5 %	-0.0204	-0.0358	-0.0225	-0.0146	1
LP	T1	T2	T3	T4	c
97.5 %	-0.0080	0.0058	-0.0072	-0.0040	1
Median	-0.0130	-0.0141	-0.0150	-0.0128	1
2.5 %	-0.0190	-0.0358	-0.0255	-0.0146	1

Table 8. MNM and nine Diff2 (10% level).

SN	Model	RIP	H-SVM	SVM4	SVM1	LP	IPLP	Logistic	LDF	QDF	RDA
1	1-4	0	0	0	0	0	0	0	1	0	1
2	1,2,4	0	0	0	2	0	0	0	1	1	2
3	2-4	0	0	0	0	0	0	0	0	1	2
4	1,3,4	0	0	0	3	0	0	0	1	1	1
5	1-3	2		1	2	2	0	0	4	0	1
6	2,4	4		2	2	2	0	2	6	2	4
7	1,4	3		3	3	3	0	3	3	3	4
8	3,4	3		1	1	1	0	0	2	2	2
9	1,2	7		0	0	0	0	5	3	0	0
10	2,3	7		4	4	4	0	4	4	4	4
11	1,3	8		2	2	2	0	7	2	3	2

1)Revised IP-OLDF, Revised IPLP-OLDF, Revised LP-OLDF and Fisher's LDF chosen fourth model as the best model. However, H-SVM, SVM4, SVM1 and logistic regression chosen the full model as the best model. Because the full model of logistic regression has the minimum M2 among all models, we choose this model as the best model. Seven 'M2Diff' are -0.07, -0.07, -0.07, 0.12, 0.08, -0.11 and 9.66, respectively. M2 of Revised IP-OLDF is within 0.11% bigger than four M2s of H-SVM, SVM4, SVM1 and logistic regression. Fisher's LDF is 9.66% bigger than Revised IP-OLDF.

Table 9. M1s and M2s of eight LDFs (10%).

RIP		M1	M2	Diff.	Model	
14m8s	1	0	0.88	0.88	1,2,3,4	
	4	0	0.86	0.86	2,4	
	5	0.73	2.44	1.71	1,3,4	
H-SVM		M1	M2	Diff.	M1Diff.	M2Diff.
5m23s	1	0.00	0.81	0.81	0	-0.07
	4	0.00	0.90	0.90	0	0.03
SVM4		M1	M2	Diff.	M1Diff.	M2Diff.
9m23s	1	0	0.81	0.81	0	-0.07
	4	0	0.90	0.90	0	0.04
	5	1.36	2.84	1.48	0.629	0.40
SVM1		M1	M2	Diff.	M1Diff.	M2Diff.
8m53s	1	0	0.81	0.81	0	-0.07
	4	0.76	1.71	0.95	0.758	0.85
	5	1.65	3.15	1.50	0.919	0.71
IPLP		M1	M2	Diff.	M1Diff.	M2Diff.
13m16s	1	0	1.00	1.00	0	0.12
	4	0.00	0.85	0.85	0	-0.02
	5	0.73	2.44	1.70	0	0
LP		M1	M2	Diff.	M1Diff.	M2Diff.
4m23s	1	0	0.96	0.96	0	0.08
	4	0	0.80	0.80	0	-0.07
	5	1.36	2.85	1.48	0.629	0.40
Logistic		M1	M2	Diff.	M1Diff.	M2Diff.
12m	1	0	0.77	0.77	0	-0.11
	4	0	0.91	0.91	0	0.05
	5	1.59	2.83	1.24	0.855	0.39
LDF		M1	M2	Diff.	M1Diff.	M2Diff.
15m	1	9.64	10.54	0.90	9.637	9.66
	4	9.54	9.91	0.37	9.54	9.05
	5	11.44	12.04	0.60	10.71	9.60

2)Among seven models that were not linearly separable, seven LDFs choose fifth models. Moreover, M2 of Revised IP-OLDF was the minimum value. Six 'M2Diff' are 0.4, 0.71, 0, 0.4, 0.39 and 9.6%, respectively. Revised IPLP-OLDF and Revise IP-OLDF are better than other five LDFs. Fisher's LDF is 9.6% bigger than Revised IP-OLDF. This fact may imply Revised IP-OLDF is superior to other LDFs for non- linear separable models.

5.3. The 95% CI of Discriminant Coefficients

We examine the coefficients of the full model as the best model. A trivial LDF is

$$f = T1 + T2 + T3 + T4 - 35.5 = -0.028 * T1 - 0.028 * T2 - 0.028 * T3 - 0.028 * T4 + 1.$$

The equation (10) is Revised IP-OLDF. The equation (11) is a logistic regression. The equation (10) is Fisher's LDF. Seven LDFs are not the same as the trivial LDF. We cannot explain the reason that the full models of all LDFs are not similar to the trivial LDF. This fact is new complicated research theme in near future.

$$\text{Revised IP-OLDF} = -0.008 * T1 - 0.061 * T2 - 0.032 * T3 - 0.024 * T4 + 1 \qquad (10)$$

$$
\begin{aligned}
\text{Logi1234} &= -2.6(725) * T1 - 21(2489) * T2 - 6.6(1248) * T3 - 6.97(724) * T4 + 296(26825) \\
&= -0.009 * T1 - 0.071 * T2 - 0.022 * T3 - 0.024 * T4 + 1.
\end{aligned} \tag{11}
$$

$$
\begin{aligned}
\text{LDF1234} &= 0.006(0.011) * T1 + 0.075(0.022^*) * T2 + 0.007(0.015) * T3 + 0.023(0.006^*) * T4 - 1.683(0.201^*) \\
&= -0.004 * T1 - 0.045 * T2 - 0.004 * T3 - 0.014 * T4 + 1.
\end{aligned} \tag{12}
$$

6. Conclusion

In this research, we discuss the new method and model selection procedure of the discriminant analysis. We discriminate the pass/fail determinations at 10%, 50%, and 90% levels. We select the best models of eight LDFs by the "minimum M2 standard" method. Two studies by 50% and 90% choose the same best models because only full models are linearly separable. We get surprising results about the best models of all MP- based LDFs and logistic regression those are almost the same as trivial LDFs. Both Fisher's LDFs are quite different from trivial LDFs. We can obtain these results by fixing the constant=1. Absolute values of 'M2Diff.' of six LDFs except for Fisher's LDF are within 0.08% and 0.16%, respectively. However, those of Fisher's LDF are 4.58% and 11.53%, respectively. Next, we select the second best model among ten non- linear separable models by six LDFs. Analysis of 50% selects the second 3-variables model and all 'M2Diff.' are greater than zero. Especially, that of Fisher's LDF is 3.026%. Analysis of 90% selects the fourth 3-variables model and all 'M2Diff.' are greater than 0.04%. Especially, that of Fisher's LDF is 10.247%. On the other hand, there are four linear separable models at 10% level. Four LDFs such as RIP, IPLP, LP and Fisher's LDF select the fourth 2-variables model, and four LDFs such as H-SVM, SVM4, SVM1 and logistic regression select the full model. We choose the full model as the best model. Moreover, all LDFs are not the same as trivial LDFs.

Base on the above results at 10%, 50%, and 90% level, we could summarize as follows.

1)Three M2s of Fisher's best LDF are 9.66% 4.58% and 11.53% worse than Revised IP-OLDF. Only Fisher's LDFs are fragile for the pass/fail determination by exams scores. Therefore, we are worried to obtain the same results about the medical diagnosis.

2)Two best models of Revised IP-OLDF and logistic regression are the same as trivial LDFs at 50% and 90% level. However all LDFs are not the same as trivial LDFs at 10% level. We cannot explain the reason theoretically.

3)If we select the second best LDFs for non-linear separable models, all LDFs select the same models and M2 of Revised IP-OLDFs have the minimum values. This fact may imply us Revised IP-OLDF is superior to other LDFs for the non-linear separable models, although only Revised IPOLDF and H-SVM can recognize the linear separable models.

7. Acknowledgments

REFERENCES

1. Fisher, R. A., (1936). The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics,**7**, 179-188.
2. Friedman, J. H., (1989). Regularized Discriminant Analysis. Journal of the American Statistical Association **84**/405, 165-175.
3. Goodnight, J.H., (1981) A tutorial on the SWEEP Operator. The American Statistician 33, 149-158.
4. Lachenbruch, P. A., Mickey, M. R., (1968). Estimation of error rates in discriminant analysis. Technometrics, **10**, 1-11.
5. Sall, J. P., Creighton, L., Lehman, A., (2004). JMP Start Statistics, Third Edition. SAS Institute Inc.
6. Schrage, L., (1991). LINDO -An Optimization Modeling System (Fourth Edition)-. The Scientific Press.

7. Schrage, L., (2006). Optimization Modeling with LINGO. LINDO Systems Inc.
8. Shinmura, S., (1998). Optimal Linear Discriminant Functions using Mathematical Programming. Journal of the Japanese Society of Computer Statistics, **11**/2, 89-101.
9. Shinmura, S., (2000). A new algorithm of the linear discriminant function using integer programming. New Trends in Probability and Statistics, **5**, 133-142.
10. Shinmura, S., (2004). New Algorithm of Discriminant Analysis using Integer Programming. IPSI 2004 Pescara VIP Conference CD-ROM, 1-18.
11. Shinmura, S., (2007a). Comparison of Revised IP-OLDF and SVM. ISI2009, 1-4.
12. Shinmura, S., (2007b). Overviews of Discriminant Function by Mathematical Programming. Journal of the Japanese Society of Computer Statistics, **20**/1-2, 59-94.
13. Shinmura, S., (2009). Practical discriminant analysis by IP-OLDF and IPLP-OLDF. IPSI 2009 Belgrade VIPSI Conference CD-ROM, 1-17.
14. Shinmura, S., (2010a). The optimal linear discriminant function. Union of Japanese Scientist and Engineer Publishing.
15. 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.
16. Shinmura, S., (2011a). Problems of Discriminant Analysis by Mark Sense Test Data. Japanese Society of Applied Statistics, **40**/3, 157-172.
17. Shinmura, S., (2011b). Beyond Fisher's Linear Discriminant Analysis - New World of Discriminant Analysis -. ISI2011 CD-ROM, 1-6.
18. Shinmura, S., (2013). Evaluation of Optimal Linear Discriminant Function by 100-fold Cross-validation. 2013 ISI CD-ROM, 1-6.
19. Shinmura, S., (2014a). End of Discriminant Functions based on Variance-Covariance Matrices. ICORES, 5-14, 2014.
20. Shinmura, S., (2014b). Improvement of CPU time of Linear Discriminant Functions based on MNM criterion by IP. Statistics, Optimization and Information Computing, **2**, 114-129.
21. Shinmura, S., (2014c). Comparison of Linear Discriminant Function by K-fold Cross-validation. Data Analytic 2014, 1-6.
22. Shinmura, S., (2015a). The 95% confidence intervals of error rates and discriminant coefficients. Statistics, Optimization and Information Computing, **3**, 66-78.
23. 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).
24. Stam, A., (1997). Nontraditional approaches to statistical classification: Some perspectives on lp-norm methods. Annals of Operations Research, **74**, 1-36.
25. Taguchi, G., Jugulum, R., (2002). The Mahalanobis-Taguchi Strategy -A Pattern Technology System. John Wiley & Sons.
26. Vapnik, V., (1995). The Nature of Statistical Learning Theory. Springer-Verlag.