SECOND-ORDER TERMS IN CLASSIFICATION MODELS

ประดิษฐ์ วรรณรัตน์[1]

Abstract

Much research has been devoted to evaluating the performance of a variety of mathematical programming models relative to the standard parametric procedures. While many of these models have displayed excellent classificatory performance when compared to Fisher's method, these mathematical programming models cannot compete with Smith's quadratic discriminant method on certain data sets. In this paper, the appropriateness of including second-order terms in mathematical programming models is investigated. A simulation study is conducted to assess the relative performance of two mathematical programming models with higher-order terms in comparison to the performance of the parametric procedures. The two mathematical programming models have been proved to remain rotationally and translationally invariant when all the second-order terms are included in the models.

Introduction

Mathematical programming approaches for solving the statistical classification problem have been given considerable attention since their introduction by Freed and Glover (1981) and Hand (1981). These mathematical programming models provide alternate approaches to the standard parametric procedures. Fisher's (1936) linear discriminant function (LDF) procedure and

---

[1] คณะบริหารธุรกิจ สถาบันบัณฑิตพัฒนบริหารศาสตร์

Smith's (1947) quadratic discriminant function (QDF) procedure are the standard procedures. In many research papers involving mathematical programming models for discriminant analysis, the objective is to find a discriminant rule which is either optimal or competitive to the parametric approaches in correctly classifying observations from a set of new observations or a representative holdout sample. These new approaches would be of benefit to the practitioner interested in applying an appropriate model to a set of data which does not meet the assumptions necessary for optimality by the parametric procedure. Applications of discriminant analysis can be found in numerous disciplines including business, artificial intelligence, economics, and biology. Johnson and Wichern (1992) provides a list of applications. Gehrlein (1986) and Mahmood and Lawrence (1987) use mathematical programming models on real data to construct viable classification rules.

Some papers have focused on the undesirable problems associated with mathematical programming models. Koehler (1989a, 1989b), Markowski and Markowski (1987), Rubin (1989, 1991), and Glover (1990) have investigated problems that plagued certain mathematical programming models. These problems included obtaining unbounded solutions, trivial solutions, and solutions which were not invariant under data translation or rotation. These problems have inspired numerous variations of mathematical programming models. Normalization constraints, such as those discussed in Glover, Keene and Duea (1988) and Glover (1990) were introduced to overcome the undesirable problems with early mathematical programming formulations.

Each of the linear programming based models obtains a classification rule by optimizing an objective function which is a surrogate for minimizing the number of misclassifications. The Bayesian parametric methods actually optimize the expected misclassification costs. To directly minimize the number of misclassifications in the training sample, mixed integer programming (MIP) models have been proposed. However, it does not always follow that these models will perform optimally in the holdout sample. Because of the computationally intensive nature of these models, several researchers have proposed heuristic algorithms to make the MIP approach more computationally efficient. Koehler and Erenguc (1990), Banks and Abad (1991), and Rubin (1990a) have investigated heuristic algorithms that appear to yield good, albeit suboptimal, solutions to the MIP model.

In this paper, the appropriateness of including second-order terms in mathematical programming models is investigated. Rubin (1990b) found that Smith's quadratic procedure was superior to some 15 linear programming models when the data follow a multivariate normal distribution with various parameter values for the means, variances, and correlations. This result is not totally surprising since the quadratic method allows for a nonlinear classification function. A Monte Carlo simulation study in which the MSD (minimize the sum of the external deviations) model and MIP model are evaluated with various sets of normal data with and without outliers was performed. The second-order terms and first-order terms are included in these models to determine the increase or decrease in classificatory

performance due to the presence of these terms. The importance of including the crossproduct terms is also assessed. A theorem establishing the invariance of the MSD and MIP models to rotations or translations of the data when all of the first-order and second-order terms are included in the mathematical programming models was developed in this study.

Silva and Stam (1993) also conducted a simulation using second-order terms in the hybrid and MSD models. The hybrid model is presented in Glover (1990). Silva and Stam's study differs from this study in that they only considered large training samples of size 100 from exponentially distributed random variables. Furthermore, they did not consider the MIP model in their study. For the highly nonnormal data generated in their study, the hybrid model and the MSD model greatly benefitted from the second-order terms.

MATHEMATICAL PROGRAMMING MODELS

There is a plethora of variations on the MSD model. The model presented in Ragsdale and Stam (1991) was selected. This model is very similar to the original model suggested by Hand (1981). This model does not require any normalization constraint such as that proposed by Freed and Glover (1986), Glover, Keene, and Duea (1988), and Glover (1990). Some of these normalization constraints have undesirable side effects as illustrated by Koehler (1989a, 1989b). The MSD model by Ragsdale and Stam (1991), however, does include a gap which separates the hyperplanes used for classification. Hand (1981) referred to the

gap as a "safety margin." Koehler (1989a) showed that Hand's model does not have the undesirable side effects displayed by some other mathematical programming models.

The MSD model of Ragsdale and Stam (1991) is presented below. The training sample consists of n_i ($i = 1, 2$) observations from each of two groups for a total of $n = n_1 + n_2$ observations. The notation G_1 and G_2 will denote the sets of observations from groups 1 and 2 respectively.

Notation:

d_i denotes the external (undesirable) deviation of a misclassified observation's discriminant score from 0 or ε. For a correctly classified observation, d_i is equal to zero.

a_{ij} denotes the jth attribute value for observation i.

x_j denotes the weight for attribute j.

x_0 denotes the constant term in the discriminant function.

ε denotes the minimum gap size separating the discriminant scores between the two groups.

p denotes the number of predictor variables (attributes).

MSD Formulation:

$$\text{Minimize} \sum_{i \in G_1} d_i + \sum_{i \in G_2} d_i$$

subject to

$$x_0 + \sum_{j=1}^{p} a_{ij}x_j - d_i \leq 0 \qquad i \in G_1$$

$$x_0 + \sum_{j=1}^{p} a_{ij}x_j + d_i \geq \varepsilon \qquad i \in G_2$$

x_j is a sign-unrestricted variable ($j = 0, 1, \ldots, p$).
d_i is a nonnegative variable ($i = 1, 2, \ldots, n$).
ε is a small positive constant.

This approach does have a drawback in that observations may be classified into the gap. To correct this problem, simply dividing the gap in half and using all discriminant scores below (above) $\varepsilon/2$ for assigning values to group 1 (group 2) appears to provide reasonably good results. However, to further refine the model, additional analysis could be performed on those observations with scores between 0 and ε as illustrated in Stam and Ragsdale (1992).

Another two-group mathematical programming formulation that was used in this study is the MIP model which is similar to that presented in Koehler and Erenguc (1990). By replacing the d_i's in the MSD model with binary variables I_i and multiplying the I_i's by a large constant M in the constraints, it is easy to construct the MIP model. Using the same notation as in the MSD model, the MIP formulation is expressed below.

MIP formulation:

$$\text{Minimize} \sum_{i \in G_1} I_i + \sum_{i \in G_2} I_i$$

subject to

$$x_0 + \sum_{j=1}^{p} a_{ij}x_j - MI_i \leq 0 \qquad i \in G_1$$

$$x_0 + \sum_{j=1}^{p} a_{ij}x_j + MI_i \geq \varepsilon \qquad i \in G_2$$

x_j is a sign-unrestricted variable ($j = 0, 1, \ldots, p$).
I_i is a binary variable ($i = 1, 2, \ldots, n$).
ε is a small positive constant.
M is a large positive constant.

To form second-order mathematical programming formulations, the squared attribute values and the crossproduct values of all attributes need to be included as additional predictor variables.

Note that the second-order terms in either the MSD or MIP formulation still have constraints that are linear in the x parameters. The constraints are obviously nonlinear in the attribute values. The second-order mathematical programming models have all of the terms present in Smith's quadratic discriminant function. Thus the second-order mathematical programming formulations have the potential of being competitive with the quadratic method in problems requiring a nonlinear classification function.

The following lemma and theorem are presented to establish that the MSD and MIP models with all second-order terms are translationally and rotationally invariant. Furthermore, the MIP model will not have more misclassifications on the training sample than the MSD or QDF methods if all second-order terms are included in the model.

Lemma

Any linear combination of second and first-order terms of $a_i = (a_{i1}, a_{i2}, \ldots, a_{ip})^T$ can be expressed as $a_i^T W a_i + a_i^T x$, where $W = (w_{hk})$ is a symmetric matrix and $x = (x_1, x_2, \ldots, x_p)^T$. The coefficients of the square terms are w_{hh}, the coefficients of the crossproduct terms are $2w_{hk}$, and the coefficients of the first-order terms are x_j.

Proof:

A linear combination of the second and first-order terms of $a_i = (a_{i1}, a_{i2}, \ldots, a_{ip})^T$ can be written as

$$\sum_{j=1}^{p} a_{ij}^2 x_{jj} + \sum_{j=1}^{p} a_{ij}x_j + \sum_{h>k} \sum a_{ih}a_{ik}x_{hk} = \sum_{h\geq k} \sum a_{ih}a_{ik}x_{hk} + \sum_{j=1}^{p} a_{ij}x_j$$

$$= \sum_{h=1}^{p} \sum_{k=1}^{p} a_{ih}a_{ik}w_{hk} + \sum_{j=1}^{p} a_{ij}x_j$$

$$= a_i^T W a_i + a_i^T x$$

where $W = (w_{hk})$ and

$$w_{hk} = \begin{cases} x_{hk}/2 & \text{if } h > k \\ x_{kh}/2 & \text{if } h < k \\ x_{hk} & \text{if } h = k \end{cases}$$

Theorem

If all of the first order-terms and second-order terms are included in the MIP formulation and in the MSD formulation then

1. The MIP method will not have more misclassifications than the QDF method or the MSD method on the training sample.
2. The MSD and MIP methods are rotationally and translationally invariant.

Proof:

The first statement follows since the MIP procedure directly minimizes the total number of misclassifications on the training sample and since each of the MIP, MSD, and QDF procedures are assumed in this theorem to contain all second-order terms. To show that the second statement holds, let P be an orthogonal matrix and let c be a p x 1 vector of constants. By Lemma 1, the discriminant score of observations for either the MSD or MIP formulation can be written as

$$x_0 + a_i^T W a_i + a_i^T x$$

Now consider an orthogonal rotation P and a translation c of the a_i vector. We have

$$
\begin{aligned}
&x_0 + [P(a_i + c)]^T W [P(a_i + c)] + [P(a_i + c)]^T x \\
&\quad = x_0 + a_i^T (P^T W P) a_i + (Pc)^T W P a_i + a_i^T (P^T W P) c \\
&\qquad + c^T (P^T W P) c + a_i^T P^T x + c^T P^T x \\
&\quad = \tilde{x}_0 + a_i^T \tilde{W} a_i + a_i^T \tilde{x}
\end{aligned}
$$

where
$$\tilde{x}_0 = x_0 + c^T (P^T W P) c + c^T P^T x$$
$$\tilde{W} = P^T W P$$
$$\tilde{x} = P^T x + 2 P^T W P c$$

We can see that $\tilde{x}_0 + a_i^T \tilde{W} a_i + a_i^T \tilde{x}$ is still a linear combination of the second-order terms of the values of the vector a_i.

Note that if some of the second-order terms are missing, such as the crossproduct (interaction) terms, then it is possible that the QDF procedure may produce fewer misclassifications than the MIP procedure on the training sample.

EXPERIMENTAL STUDY

A Monte Carlo simulation study was performed to assess the performance of both the mathematical programming formulations and the standard parametric procedures. The notation MIPQ and MSDQ were used to denote the use of the MIP and MSD procedures, respectively, with all of the second-order terms in the models. For some configurations, the MIPQ and MSDQ without the crossproduct terms were also considered and were denoted by indicating "no X_1X_2" in parenthesis after the names of the models.

Five different data configurations were examined in the simulation study. Each of the population distributions was normally distributed except for one. These data configurations are described in Table 1. Configuration A is the only configuration in which a first-order (linear) classification rule would be optimal. For the other configurations, it is expected that a second-order (nonlinear) classification rule would be the classification rule of choice. Configuration B was selected because it is a configuration in which the QDF model can easily perform well. Configuration C was selected because the means of the two populations are equal. Configuration D is the only configuration to contain nonnormal data. The second population of Configuration D consists of a normal population with mean vector $(2, 2)^T$ and 2 independent variables with variances equal to one, but this population also contains a 15% contamination from a set of normally distributed outliers. The outlier group has mean vector $(-10, -10)^T$ and 2 independent variables with variances equal to 9. Configuration E was selected to determine the importance of the crossproduct term when correlation was present in the populations.

For each configuration in this simulation study, training sample sizes of $n_i = 25$, $i = 1, 2$, and $n_i = 50$, $i = 1, 2$ were used for each of the two groups. Validation sample sizes of 500 were used for each group, for a total of 1000 observations for each validation sample. In each simulation experiment, two attribute values were generated for each observation. The simulation study was performed using the SAS statistical package (version 6.07) on

a Solbourne 5E/902 computer operating under UNIX. All experimental conditions were replicated 50 times.

The simulation study did not compare the performance of the models with higher sample sizes. This is partly due to the computational intensiveness of the MIP procedures at higher sample sizes, particularly if the degree of overlap in the groups is large. Simulation studies such as Joachimsthaler and Stam (1988) and Rubin (1990), and Stam and Jones (1990) also limited the size of training samples to 50.

TABLE 1

Data Configurations for Simulation Study

	First Population		Second Population	
Configuration	Mean Vector	Covariance Matrix	Mean Vector	Covariance Matrix
A	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
B	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$
C	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 25 & 0 \\ 0 & 25 \end{bmatrix}$
D	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
			15% of observations from	
			$\begin{bmatrix} -10 \\ -10 \end{bmatrix}$	$\begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix}$
E	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}$	$\begin{bmatrix} -2 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 4 & 2.4 \\ 2.4 & 4 \end{bmatrix}$

RESULTS

The results of the simulation study are summarized in Tables 2 through 6. Table 2 displays the results for Configuration A. For this configuration, the LDF procedure had the lowest average misclassification rate on the validation samples for both training sample sizes of 25 and 50. The QDF procedure performed nearly as well as the LDF procedure. However, for the case of 25 observations per training group, the LDF procedure had the highest average misclassification rate on the training sample. For training sample sizes of 50, the parametric methods were somewhat close in performance to the MSD and MIP procedures which did not include the second-order terms.

Table 2

Percentage of Misclassification for Configuration A

Method	Training Sample Size = 25				Training Sample Size = 50			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
LDF	7.52	2.82	8.45	1.18	7.10	2.21	8.20	0.90
QDF	7.32	2.98	8.64	1.22	7.12	2.31	8.29	0.95
MSD	7.08	3.34	8.88	1.32	7.16	2.54	8.36	0.90
MIP	4.44	2.30	9.76	2.18	5.28	1.84	8.87	1.06
MSDQ	5.84	3.20	11.73	2.46	6.58	2.19	9.38	1.43
MIPQ	3.36	2.16	14.68	4.58	4.28	1.78	11.07	2.28
MSDQ (no x_1x_2)	6.00	3.30	10.50	2.23	6.70	2.28	9.00	1.28
MIPQ (no x_1x_2)	3.68	2.26	12.89	2.57	4.68	1.85	10.52	1.84

Table 2 shows that the addition of second-order terms in the MSD and MIP models decreased the performance of these procedures in the validation samples for Configuration A, particularly for the case of 25 observations per group in the training sample. The omission of the crossproduct terms in the MSDQ and MIPQ procedures improved the performance of these models in the validation samples. However, the use of the higher-order terms in the MSDQ and MIPQ procedures showed marked improvement in the classificatory performance in the training sample over the MSD and MIP procedure with 2 variables. The standard deviations of the misclassification rates for the MSDQ and MIPQ models showed a substantial increase when compared to the MSD and MIP models. As expected, a linear first-order model is the most appropriate model for this configuration.

For Configuration B, it is expected that the QDF procedure would be the optimal. Indeed, on the validation sample, the average misclassification rate for the QDF is smaller than all of the other models as displayed in Table 3. For training samples of size 25 per group, it is interesting to see that the addition of higher-order terms markedly increased the holdout misclassification rates of the MSD and MIP procedures. However for samples of size 50, the MSDQ's holdout classificatory performance is slightly better than the first-order MSD model's performance.

Interestingly, the omission of the crossproduct term showed a marked improvement in the MSDQ and MIPQ models. Note that the

MSD's and MSDQ's misclassification rates on the validation samples were always lower than that of the LDF procedure except when all second-order terms were used with training samples of size 25 per group. For this configuration, the MSD and MSDQ procedures appear to outperformed the MIP and MIPQ procedures on the validation samples, particularly for small sample size.

Table 3

Percentage of Misclassification for Configuration B

Method	Training Sample Size = 25				Training Sample Size = 50			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
LDF	7.22	2.96	8.45	1.19	7.74	2.03	8.32	1.13
QDF	5.34	3.18	7.13	1.29	5.68	1.84	6.66	1.12
MSD	6.32	3.74	8.32	1.14	6.44	2.32	7.88	0.94
MIP	4.16	2.44	9.10	2.13	4.64	1.89	8.43	1.39
MSDQ	3.96	3.50	10.00	2.96	4.90	1.88	7.62	1.37
MIPQ	2.32	1.90	13.10	3.84	3.08	1.40	9.19	2.01
MSDQ (no x_1x_2)	4.08	3.42	9.39	2.88	4.98	2.05	7.26	1.15
MIPQ (no x_1x_2)	2.56	2.24	10.99	3.50	3.36	1.47	8.52	1.69

For Configuration C, it is expected that the LDF procedure would have very poor discriminatory power since the group means are equal. Table 4 shows that linear first-order models perform poorly in this setting. The high overlap of the two groups made the first-order MIP model impractical to compute for training

samples of size 50. This was the only experimental situation in which a model was not assessed on 50 replications of the data. As expected, the second-order terms helped the MSDQ and MIPQ models to show dramatic improvement over the first-order MSD and MIP models. The MSDQ and MIPQ showed substantial improvement with larger training sample size. However, the QDF procedure easily outperformed the MSDQ and MIPQ procedures on the validation samples despite the better classificatory performance on the training samples of the MSDQ and MIPQ models.

Table 4

Percentage of Misclassification for Configuration C

Method	Training Sample Size = 25				Training Sample Size = 50			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
LDF	37.64	6.90	42.93	3.90	40.92	5.37	43.98	3.49
QDF	6.72	3.62	9.18	1.35	7.12	2.40	8.49	0.94
MSD	32.20	5.04	36.03	2.54	34.54	3.92	37.36	2.50
MIP	23.52	2.84	34.68	1.76	*	*	*	*
MSDQ	4.56	3.94	11.61	2.85	6.02	2.30	9.54	1.31
MIPQ	2.92	2.40	13.01	3.04	3.84	1.47	10.85	2.28

* Computationally too intensive to complete runs for this model.

In Configuration D, the second population contained 15% of its observations as outliers. It was expected that the nonnormality of this data set would weaken the performance of the QDF procedure. The results in Table 5 revealed that the first-

order MSD and MIP procedures could not match the performance of the QDF procedure which benefited from its nonlinear structure. The LDF model clearly showed little discriminatory ability for this configuration. The MSD model was more easily affected than the MIP model by the outliers and showed poor discriminatory ability as well.

In Configuration D, the second-order terms in the MSDQ and MIPQ models made these models capable of outperforming the QDF model for both samples of size 25 and 50. While the MIP model outperformed the MSD model on the validation sample, the MSDQ model outperformed the MIPQ model on the validation sample for this configuration.

Table 5

Percentage of Misclassification for Configuration D

Method	Training Sample Size = 25				Training Sample Size = 50			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
LDF	34.80	15.14	41.33	14.20	37.06	10.41	40.43	10.10
QDF	9.20	4.08	13.14	3.12	10.10	3.65	12.07	2.28
MSD	24.40	13.50	29.52	13.49	24.18	13.09	27.44	13.35
MIP	11.00	4.80	16.54	2.13	11.74	2.54	16.15	1.25
MSDQ	5.24	3.12	11.22	2.96	5.48	2.19	9.20	1.57
MIPQ	3.20	2.46	14.34	4.69	3.62	1.55	10.93	2.25

Configuration E is the only configuration in which correlation is present in the predictor variables. As expected, the QDF model was the optimal procedure for this configuration.

Since the purpose of using this configuration was to determine the importance of including the crossproduct term in the MSDQ and MIPQ models, only the second-order models and the LDF model were evaluated with this configuration. Table 6 shows that omitting the crossproduct term actually decreased the misclassification rates of the MSDQ and MIPQ models on the validation sample, although it does not on the training sample. Thus having correlation present among the attributes values does not always necessitate including the crossproduct term to enhance the classificatory performance of the mathematical programming model on the validation sample.

Table 6

Percentage of Misclassification for Configuration E

Method	Training Sample Size = 25				Training Sample Size = 50			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
LDF	6.76	3.58	7.68	1.28	7.48	2.33	7.42	0.96
QDF	5.04	3.38	6.27	0.83	5.28	2.47	5.87	0.76
MSDQ	3.24	3.30	9.75	2.64	4.36	2.36	7.27	1.70
MIPQ	1.96	1.92	13.02	4.36	2.50	1.26	9.02	2.21
MSDQ (no x_1x_2)	3.28	3.12	8.45	2.48	4.66	2.45	6.63	1.11
MIPQ (no x_1x_2)	2.20	2.06	11.22	4.11	2.88	1.53	7.70	1.75

CONCLUSION

In the paper, the appropriateness of including second-order terms in the MSD and MIP procedures was investigated. The use of second-order terms in these mathematical programming models allows for greater flexibility in choosing an appropriate mathematical programming based discriminant procedure. These second-order mathematical programming models can be easily implemented on any standard linear programming software package.

Previous research with mathematical programming models have primarily investigated linear discriminant functions. It is easy to find data for which these functions are not the optimal classification rule. This simulation study demonstrates that second-order mathematical programming models can be competitive with the quadratic discriminant procedure for some nonnormal set of data. While the second-order terms substantially improved the performance of the MSD and MIP procedures for some configurations, there were configurations in which the inclusion of these terms in the MSD and MIP procedures decreased the classificatory performance, particularly for the situation in which the LDF procedure was the optimal classification rule. The study also proved that the MSDQ and MIPQ models will be both translationally invariant and rotationally invariant, all of the second-order terms must be included in the model. This study demonstrated that second-order terms in mathematical programming models may be very appropriate and effective for certain data configurations. Further research is needed to demonstrate the

appropriateness of other nonlinear terms in mathematical programming models.

REFERENCES

[1] Banks, W. and Abad, P. (1991). An efficient optimal solution algorithm for the classification problem. *Decision Sciences* 22, 5, 1008-1023.

[2] Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7, 179-188.

[3] Freed, N. and Glover, F. (1981). A linear programming approach to the discriminant problem. *Decision Sciences* 12, 1, 68-74.

[4] Freed, N. and Glover, F. (1986). Resolving certain difficulties and improving the classification power of LP discriminant analysis formulations. *Decision Sciences* 17, 4, 589-595.

[5] Gehrlein, W. (1986). General mathematical programming formulations for the statistical classification problem. *Operations Research Letters* 5, 6, 299-304.

[6] Glover, F. (1990). Improved linear programming models for discriminant analysis. *Decision Sciences* 21, 4, 771-785.

[7] Glover, F., Keene, S., and Duea, R. (1988). A new class of models for the discriminant problem. *Decision Sciences* 19,2, 269-280.

[8] Hand, D. (1981). *Discrimination and Classification*. New York: John Wiley and Sons.

[9] Joachimsthaler, E. and Stam, A. (1988). Four approaches to the classification problem in discriminant analysis: An experimental study. *Decision Sciences* 19, 2, 322-333.

[10] Johnson, R. and Wichern, D. (1992). *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, New Jersey.

[11] Koehler, G. (1989a). Characterization of unacceptable solutions in LP discriminant analysis. *Decision Sciences* 20, 2, 239-257.

[12] Koehler, G. (1989b). Unacceptable solutions and the hybrid discriminant model. *Decision Sciences* 20, 4, 844-848.

[13] Koehler, G. and Erenguc, S. (1990). Minimizing misclassifications in linear discriminant analysis. *Decision Sciences* 21, 1, 63-85.

[14] Mahmood, M. and Lawrence, E. (1987). A performance analysis of parametric and nonparametric discriminant approaches to business decision making. *Decision Sciences* 18, 2, 308-326.

[15] Markowski, E. and Markowski, C. (1985). Some difficulties and improvements in applying linear programming formulations to the discriminant problem. *Decision Sciences* 16, 3, 237-247.

[16] Ragsdale, C. and Stam, A. (1991). Mathematical programming formulations for the discriminant problem: An old dog does new tricks. *Decision Sciences* 22, 2, 296-307.

[17] Rubin, P. (1989). Evaluating the maximize minimum distance formulation of the linear discriminant problem. *European Journal of Operational Research* 41, 2, 240-248.

[18] Rubin, P. (1990a). Heuristic solution procedures for a mixed-integer programming discriminant model. *Managerial and Decision Economics* 11, 4, 255-266.

[19] Rubin, P. (1990b). A comparison of linear programming and parametric approaches to the two group discriminant problem. *Decision Sciences* 21, 2, 373-386.

[20] Rubin, P. (1991). Separation failure in linear programming discriminant models. *Decision Sciences* 22, 3, 519-535.

[21] Silva, A. and Stam, A. (1993). Second order mathematical programming formulations for discriminant analysis. Forthcoming in *European Journal of Operational Research*.

[22] Smith, C. (1947). Some examples of discrimination. *Annals of Eugenics* 13, 272-282.

[23] Stam, A. and Jones, D. (1990). Classification performance of mathematical programming techniques in discriminant analysis: Results for small and medium sample sizes. *Managerial and Decision Economics* 11, 4, 243-253.

[24] Stam, A. and Ragsdale, C. (1992). On the classification gap in MP-based approaches to the discriminant problem. *Naval Research Logistics* 39, 545-559.