นิพนธ์ต้นฉบับ

วิธีโปรแกรมเชิงเส้นสำหรับปัญหาการแบ่งกลุ่มในการวิเคราะห์การจำแนก

Linear Programming Approaches to the Classification Problem in Discriminant Analysis

วิโรจน์ ลิ่วคงสถาพร, วท.ม. (สถิติประยุกต์)
กองแผนงาน สถาบันบัณฑิตพัฒนบริหารศาสตร์ กรุงเทพมหานคร

Wiroj Liewkongsthaporn, M.S. (Applied Statistics)
Planning Division, National Institute of Development Administration, Bangkok.

Abstract

The classification problem in discriminant analysis involves the construction a rule for the assignment of an observation into one of several groups. Various discriminant analysis techniques have been proposed in the fields of statistics, management science and artificial intelligence. This paper introduces two most widely used linear programming (LP) formulations to the classification problem. The results of experimental comparisons of LP and the statistical approaches are reported.

บทคัดย่อ

ปัญหาการแบ่งกลุ่มในการวิเคราะห์การจำแนกเป็นปัญหาที่เกี่ยวกับการสร้างกฎเกณฑ์ในการจัดให้ค่าสังเกตอยู่ในกลุ่มใดกลุ่มหนึ่ง วิธีการวิเคราะห์การจำแนกแบบต่างๆ ได้ถูกพัฒนาขึ้นทั้งในสาขาสถิติ วิทยาการจัดการ และปัญญาประดิษฐ์ บทความนี้ได้เสนอตัวแบบโปรแกรมเชิงเส้นที่ใช้กันมากที่สุดในปัญหาการแบ่งกลุ่ม จำนวน 2 ตัวแบบ และได้รายงานผลการทดลองเชิงเปรียบเทียบระหว่างวิธีโปรแกรมเชิงเส้นและวิธีทางสถิติ

Introduction

The goal of classification problem in discriminant analysis is to derive a rule for the assignment of an observation into one of several groups. The original discriminant analysis techniques were statistical approaches. They have been joined recently by methods in the areas of management science and artificial intelligence.

Perhaps, Fisher's (1936) linear discriminant function (FLDF) is the most widely used discriminant analysis method. The primary purpose of FLDF is to separate populations. However, it can also be used to classify. Under the assumption of multivariate normality with common covariance matrix and known parameters, FLDF is optimal in minimizing the expected cost of misclassifications. When the covariance matrices are not equal, Smith's (1947) quadratic discriminant function (QDF) and Anderson-Bahadur (1962) linear discriminant function were proposed. The alternative statistical method of classification is logistic regression, or logistic discriminant (LGD) model. Unlike discriminant functions, the LGD approach treats the distribution for each population as a multinomial. Under nonnormality, Press and Wilson (1978) recommended the LGD model with maximum likelihood estimators for solving the classification problem.

In recent years, neural networks (NNs) have been applied to the classification problem in discriminant analysis. Several authors compared NNs with statistical models (Patuwo et. al.,1993; Subramanian et. al., 1993; Tam and Kiang, 1992; Yoon et. al., 1993). Markham and Ragsdale (1995) combined NNs and statistical methods to provide more accurate composite predictions for the classification problem. NNs are widely used in discriminant analysis as well as statistical approach.

For the last two decades, various mathematical programming (MP) approaches have been developed for solving the classification problem in discriminant analysis. MP-based methods include LP, goal programming and mixed integer programming models. Only LP approaches are discussed here. Freed and Glover (1981a, 1981b) and Hand (1981) introduced LP approaches to the classification problem. Glorfeld and Gaither (1982) and Freed and Glover (1982) debated the relative complexity of the LP approaches and FLDF. Several authors have been proposed various normalization constraints to solve unacceptable and trivial solution from the LP approaches (Freed and Glover, 1986b; Glen, 1999; Glover, 1990; Koehler, 1989; Markowski and Markowski, 1985; Ragsdale and Stam, 1991). Unlike NNs and statistical models, LP approaches to the classification problem are not widely used although LP is the well-known model.

In this paper, the two most widely used LP formulations to the classification problem in discriminant analysis are introduced. (For other LP models, see Freed and Glover, 1986a; Gochet et. al., 1997; Lee and Ord, 1990.) Comparisons of the LP models and the classical discriminant analysis are presented.

Linear Programming Formulations to the Classification Problem

Let x_{ij} be the observation on variable i (i = 1, 2, ..., p) of case j (j = 1, 2, ..., n).

The maximize minimum deviation (MMD) model

Freed and Glover (1981a) firstly proposed the MMD model for the two-group problem. This model determines the coefficients, w_i (i = 1, 2, ..., p), that maximize the minimum deviation, d, between a case score, $w_1 x_{1j} + w_2 x_{2j} + \ldots + w_p x_{pj}$, and the cutoff score, c. The form of the MMD formulation is:

$$\text{Maximize } d$$

Subject to

$$w_1 x_{1j} + w_2 x_{2j} + \ldots + w_p x_{pj} - d \geq c \qquad \text{for all observations in group 1}$$
$$w_1 x_{1j} + w_2 x_{2j} + \ldots + w_p x_{pj} + d \leq c \qquad \text{for all observations in group 2}$$
$$d,\ w_i \text{ unrestricted in sign, } i = 1, 2, \ldots, p$$

The maximize sum of deviation (MSD) model

The MSD model maximizes the sum of the distances from the case scores to the cutoff (Freed and Glover, 1981a). Let dj , j = 1, 2, ..., n, represent the distance of group member j's score from the cutoff c. The basic form of the MSD formulation is:

$$\text{Maximize } \sum_{j=1}^{n} d_j$$

Subject to

$$w_1 x_{1j} + w_2 x_{2j} + \ldots + w_p x_{pj} - d_j \geq c \qquad \text{for all observations in group 1}$$
$$w_1 x_{1j} + w_2 x_{2j} + \ldots + w_p x_{pj} + d_j \leq c \qquad \text{for all observations in group 2}$$
$$d_j,\ w_i \text{ unrestricted in sign, } i = 1, 2, \ldots, p\ ;\ j = 1, 2, \ldots, n$$

Comparisons

To illustrate for comparison of the MSD model and FLDF, the data of a bankruptcy problem are used and shown in Table 1. Glen (1999) used the MSD model with normalization constant ± 1, that is c = 1, to generate the discriminant function

$$1.291\, x_1 - 2.689\, x_2 + 0.541\, x_3 - 0.249\, x_4 = 1$$

where x_1 is the cash flow to total debt ratio, x_2 is the net income to total assests ratio, x_3 is the current assests to current liabilities ratio, and x_4 is the current assests to net sales ratio. The FLDF are produced by running the SPSS discriminant program. The misclassification rates of the MSD and FLDF procedures are 0.15 and 0.09 respectively.

A number of studies have compared the performance of LP and statistical approaches to the classification problem. Experimental results indicated that the MMD formulation outperforms FLDF when overlap between groups is lowest and the covariance matrices are most unequal (Bajgier and Hill, 1982) and that the MSD formulation generally is the most reliable predictor of group membership (Freed and Glover, 1986a). Joachimsthaler and Stam (1988) suggested that the MSD model is a useful alternative to LGD. However, Nath et. al. (1992) noted that statistical methods are generally superior to LP based approaches. Further work may combine the statitstical and LP approaches to create composite predictions that are more accurate than either of the individual techniques used in isolation.

Table 1 Data for an example (Glen, 1999)

Case	Group	X_1	X_2	X_3	X_4
1	1	-0.4485	-0.4106	1.0865	0.4526
2	1	-0.5633	-0.3114	1.5134	0.1642
3	1	0.0643	0.0156	1.0077	0.3978
4	1	-0.0721	-0.0930	1.4544	0.2589
5	1	-0.1002	-0.0917	1.5644	0.6683
6	1	-0.1421	-0.0651	0.7066	0.2794
7	1	0.0351	0.0147	1.5046	0.7080
8	1	-0.0653	-0.0566	1.3737	0.4032
9	1	0.0724	-0.0076	1.3723	0.3361
10	1	-0.1353	-0.1433	1.4196	0.4347
11	1	-0.2298	-0.2961	0.3310	0.1824
12	1	0.0713	0.0205	1.3124	0.2497
13	1	0.0109	0.0011	2.1495	0.6969
14	1	-0.2777	-0.2316	1.1918	0.6601
15	1	0.1454	0.0500	1.8762	0.2723
16	1	0.3703	0.1098	1.9941	0.3828
17	1	-0.0757	-0.0821	1.5077	0.4215
18	1	0.0451	0.0263	1.6756	0.9494
19	1	0.0115	-0.0032	1.2602	0.6038
20	1	0.1227	0.1055	1.1434	0.1655
21	1	-0.2843	-0.2703	1.2722	0.5128

Table 1 (continued)

Case	Group	X_1	X_2	X_3	X_4
1	2	0.5135	0.1001	2.4871	0.5368
2	2	0.0769	0.0195	2.0069	0.5304
3	2	0.3776	0.1075	3.2651	0.3548
4	2	0.1933	0.0473	2.2506	0.3309
5	2	0.3248	0.0718	4.2401	0.6279
6	2	0.3132	0.0511	4.4500	0.6852
7	2	0.1184	0.0499	2.5210	0.6925
8	2	-0.0173	0.0233	2.0528	0.3484
9	2	0.2169	0.0779	2.3489	0.3970
10	2	0.1703	0.0695	1.7973	0.5174
11	2	0.1460	0.0518	2.1692	0.5500
12	2	-0.0985	-0.0123	2.5029	0.5778
13	2	0.1398	-0.0312	0.4611	0.2643
14	2	0.1379	0.0728	2.6123	0.5151
15	2	0.1486	0.0564	2.2347	0.5563
16	2	0.1633	0.0486	2.3080	0.1978
17	2	0.2907	0.0597	1.8381	0.3786
18	2	0.5383	0.1064	2.3293	0.4835
19	2	-0.3330	-0.0854	3.0124	0.4730
20	2	0.4785	0.0910	1.2444	0.1847
21	2	0.5603	0.1112	4.2918	0.4443
22	2	0.2029	0.0792	1.9939	0.3018
23	2	0.4746	0.1380	2.9166	0.4487
24	2	0.1661	0.0351	2.4527	0.1370
25	2	0.5808	0.0371	5.0594	0.1268

References

Anderson, T. W. and R. R. Bahadur, 1962, Classification into two multivariate normal distributions with different covariance matrices, **Annals of Mathematical Statistics**, 33: 420-431.

Bajgier, S. M. and A. V. Hill, 1982, An experimental comparison of statistical and linear programming approaches to the discriminant problem, **Decision Sciences**, 13: 604-618.

Fisher, R. A. 1936, The use of multiple measurements in taxonomic problems, **Annals of Eugenics**, 7: 179-188.

Freed, N. and F. Glover, 1981a, A linear programming approach to the discriminant problem, **Decision Sciences**, 12: 68-74.

Freed, N. and F. Glover, 1981b, Simple but powerful goal programming formulations for the discriminant problem, **European Journal of Operational Research**, 7: 44-60.

Freed, N. and F. Glover, 1982, Linear programming and statistical discrimination - The LP side, **Decision Sciences**, 13: 172-175.

Freed, N. and F. Glover, 1986a, Evaluating alternative linear programming models to solve the two-group discriminant problem, **Decision Sciences**, 17: 151-162.

Freed, N. and F. Glover, 1986b, Resolving certain difficulties and improving the classification power of LP discriminant analysis formulations, **Decision Sciences**, 17: 589-595.

Glen, J. J. 1999, Integer programming methods for normalisation and variable selection in mathematical programming discriminant analysis models, **Journal of the Operational Research Society**, 50: 1043-1053.

Glorfeld, L. W. and N. Gaither, 1982, On using linear programming in discriminant problems, **Decision Sciences**, 13: 167-171.

Glover, F. 1990, Improved linear programming models for discriminant analysis, **Decision Sciences**, 21: 771-785.

Gochet, W., A. Stam, V. Srinivasan, and S. Chen, 1997, Multigroup discriminant analysis using linear programming, **Operations Research**, 45: 213-225.

Hand, D. J. 1981, **Discrimination and Classification**, Wiley, New York.

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

Koehler, G. J. 1989, Characterization of unacceptable solutions in LP discriminant analysis, **Decision Sciences**, 20:239-257.

Lee, C. K. and J. K. Ord, 1990, Discriminant analysis using least absolute deviations, **Decision Sciences**, 21: 86-96.

Markham, I. S. and C. T. Ragsdale, 1995, Combining neural networks and statistical predictions to solve the classification problem in discriminant analysis, **Decision Sciences**, 26: 229-242.

Markowski, E. P. and C. A. Markowski, 1985, Some difficulties and improvements in applying linear programming formulations to the discriminant problem, **Decision Sciences**, 16: 237-247.

Nath, R., W.M. Jackson, and T. W. Jones, 1992, A comparison of the classical and the linear programming approaches to the classification problem in discriminant analysis, **Journal of Statistical Computation and Simulation**, 41: 73-93.

Patuwo, E., M. Y. Hu, and M. S. Hung, 1993, Two-group classification using neural networks, **Decision Sciences**, 24:825-845.

Press, S. J. and S. Wilson, 1978, Choosing between logistic regression and discriminant analysis, **Journal of the American Statistical Association**, 73: 699-705.

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

Smith, C. A. B. 1947, Some examples of discrimination, **Annals of Eugenics**, 13: 272-282.

Subramanian, V., M. S. Hung, and M. Y. Hu, 1993, An experimental evaluation of neural networks for classification, **Computers and Operations Research**, 20: 769-782.

Tam, K. Y. and M. Y. Kiang, 1992, Managerial applications of neural networks: The case of bank failure prediction, **Management Science**, 38: 926-947.

Yoon, Y., G. Swales, and T. M. Margavio, 1993, A comparison of discriminant analysis versus artificial neural networks, **Journal of the Operational Research Society**, 44: 51-60.