THE NONLINEAR DISCRIMINANT AND STEPWISE NONLINEAR DISCRIMINANT ANALYSES

Yao Dirong (姚棣荣)

Department of Geography, Hangzhou University, Hangzhou

Received March 20, 1986

ABSTRACT

The nonlinear discriminant function, when covariance matrixes of each population are not equal to each other, is discussed on the basis of Bayes' criterion, and by using the stepwise discriminant method, a method for calculating the nonlinear discriminant function is provided, which is called "stepwise nonlinear discriminant analysis". In addition, an appropriate discriminant analysis model is selected by testing whether the covariance matrixes of each population are equal, which was proposed by Box. The calculations show that, the discriminant effects of this method are superior not only to linear discriminant analysis, but also to nonlinear discriminant analysis in which the stepwise discriminant algorithm is not used when covariance matrixes of each population are not equal to each other. Satisfactory results have been obtained in applying this method. This is an important improvement on the linear discriminant analysis used in the weather typing prediction at present.

I. INTRODUCTION

In the weather typing prediction at present, discriminant analysis has widely been used (Wang and Li et al., 1974; Li and Yao, 1977) mainly for linear problems. However, we often meet problems which tend to be nonlinear and, if we still use linear discriminant analysis, the effects of forecast would be very bad. Thus, we must take into account the nonlinear discriminant (Clark et al., 1975; Lorenz, 1980). Many works (Gal-Mowcorterchin et al., 1971; Yao and Liu, 1983) have shown that the discriminant effects are, to some extent, better than those of linear discriminant analysis in some problems, and their results conform more to reality when nonlinear discriminant analysis is used. Nonlinear discriminant analysis is rational and advantageous.

But the above works are only within the limits of two-category discriminant. Moreover, the term numbers in the nonlinear discriminant function increase greatly with the predictors' increment because of the occurrence of nonlinear terms in the nonlinear discriminant function. Thus, not only the calculation amounts increase, but also the discriminant effects of nonlinear discriminant analysis cannot compare with those of linear discriminant analysis. Recently, Yao and Liu (1985) have provided a stepwise sieve method of predictors in the nonlinear discriminant function, and have obtained good results. This paper makes a systematic exposition of nonlinear discriminant analysis and sets up the so-called "stepwise nonlinear discriminant analysis" of the multiple-category nonlinear discriminant function (including the two-category one) making use of the algorithm of the stepwise discriminant analysis (Li and Yao, 1977). At the same time, the model selection of the discriminant analysis is discussed.

II. THE MATHEMATICAL MODEL

Assume that each individual composed of p variables $x_1, x_2, \ldots, x_p$ comes from G populations $A_1, A_2, \ldots, A_G$, respectively. For a given individual $X=(x_1, x_2, \ldots, x_p)$, we need to discriminate which population it belongs to in G populations. Bayes' criterion for this problem is that the mean losses of classification errors on each population reach a minimum, i.e. the posterior probability

$$p(g|X)=\frac{Q_g f_g(X)}{\sum_{h=1}^{G} Q_h f_h(X)}, \quad g=1,2,\cdots G \tag{1}$$

reaches a maximum.

If the probability density function $f_g(X)$ and the prior probability Q_g of each population A_g $(g=1,2,\ldots,G)$ are known and the mean losses of classification errors on each population are equal, the discriminant function can be established as follows:

$$Y_g(X)=Q_g f_g(X), g=1,2,\cdots,G. \tag{2}$$

If

$$Y_{g^*}(X)=\operatorname*{Max}_{1\leqslant g\leqslant G}\{Y_g(X)\}, \tag{3}$$

then the individual X belongs to the population A_{g^*}.

Assume that the population $A_g(g=1,2,\ldots,G)$ is distributed according to a multivariate Gaussian density function with mean vector $\boldsymbol{\mu}_g$ and covariance matrix $\boldsymbol{\Sigma}_g$, i.e. $f_g(X)\sim N(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, then, the density function of the population A_g is

$$f_g(X)=\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_g|^{1/2}}\exp\left[-\frac{1}{2}(X-\boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(X-\boldsymbol{\mu}_g)\right], \tag{4}$$
$$g=1,2,\cdots,G$$

where the mean vector of the population $g(g=1,2,\ldots,G)$ is

$$\boldsymbol{\mu}_g=(\mu_{1g}, \mu_{2g}, \cdots\cdots \mu_{pg})',$$

while the covariance matrix of the population $g(g=1,2,\ldots,G)$ is

$$\boldsymbol{\Sigma}_g=(\sigma_{ij})_g=\begin{pmatrix}\sigma_{11} & \sigma_{12}\cdots\sigma_{1p}\\ \sigma_{21} & \sigma_{22}\cdots\sigma_{2p}\\ \cdots & \cdots\cdots\cdots\\ \sigma_{p1} & \sigma_{p2}\cdots\sigma_{pp}\end{pmatrix}_g$$

Substituting Eq. (4) into (2), carrying out logarithmic computations for $Q_g f_g(\mathbf{X})$, and omitting the terms which bear no relation to g when the covariance matrixes of each population are not equal to each other, we obtain the following nonlinear discriminant function:

$$Y_g(X)=-\frac{1}{2}X'\boldsymbol{\Sigma}_g^{-1}X+\boldsymbol{\mu}_g'\boldsymbol{\Sigma}_g^{-1}X-\frac{1}{2}\boldsymbol{\mu}_g'\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\mu}_g$$
$$-\frac{1}{2}\ln|\boldsymbol{\Sigma}_g|+\ln Q_g, \quad g=1,2,\cdots,G. \tag{5}$$

The parameters $\boldsymbol{\mu}_g$ and $\boldsymbol{\Sigma}_g$ $(g=1,2,\ldots,G)$ of the population are often unknown in fact. These parameters must be estimated by using the sample data, i.e.

$$\boldsymbol{\mu}_g\approx\bar{X}_g=(\bar{x}_{1g}, \bar{x}_{2g}, \cdots, \bar{x}_{pg})', \quad g=1,2\cdots,G. \tag{6}$$
$$\boldsymbol{\Sigma}_g\approx \boldsymbol{S}_g=(s_{gij}), \quad i,j=1,2,\cdots,p, \quad g=1,2,\cdots,G, \tag{7}$$

where

$$\bar{x}_{ig}=\frac{1}{n_g}\sum_{k=1}^{n_g}x_{igk},$$

$$S_{gij}=\frac{1}{n_g-1}\sum_{k=1}^{n_g}(x_{igk}-\bar{x}_{ig})(x_{jgk}-\bar{x}_{jg}) \tag{8}$$

$$i,j=1,2,\cdots,p,\qquad g=1,2,\cdots,G$$

are the mean values of the variable i for population g and the covariance of population g, respectively. In Eq. (8), x_{igk} is the kth observation value of variable i for population g, and n_g is the number of observations in population g.

Note that if N is the total number of observations, we have

$$N=\sum_{g=1}^{G}n_g \tag{9}$$

By using the sample data, the prior probability Q_g can be estimated from

$$Q_g\approx q_g=n_g/N. \tag{10}$$

Then, Eq. (5) will become

$$Y_g(\boldsymbol{X})=-\frac{1}{2}\boldsymbol{X}'\boldsymbol{S}_g^{-1}\boldsymbol{X}+\bar{\boldsymbol{X}}_g'\boldsymbol{S}_g^{-1}\boldsymbol{X}-\frac{1}{2}\bar{\boldsymbol{X}}_g'\boldsymbol{S}_g^{-1}\bar{\boldsymbol{X}}_g-\frac{1}{2}\ln|\boldsymbol{S}_g|+\ln q_g, \tag{11}$$

$$g=1,2,\cdots,G.$$

Let

$$\begin{pmatrix} c_{g11} & c_{g12}\cdots & c_{g1p}\\ c_{g21} & c_{g22}\cdots & c_{g2p}\\ \cdots\cdots & \cdots\cdots & \cdots\cdots\\ c_{gp1} & c_{gp2} & c_{gpp}\end{pmatrix}=-\frac{1}{2}\boldsymbol{S}_g^{-1},\qquad g=1,2,\cdots,G, \tag{12}$$

$$(c_{g1},c_{g2},\cdots c_{gp})=\bar{\boldsymbol{X}}_g'\boldsymbol{S}_g^{-1},\qquad q=1,2,\cdots,G, \tag{13}$$

$$c_{g0}=-\frac{1}{2}\bar{\boldsymbol{X}}_g'\boldsymbol{S}_g^{-1}\bar{\boldsymbol{X}}_g-\frac{1}{2}\ln|\boldsymbol{S}_g|+\ln q_g, g=1,2,\cdots,G. \tag{14}$$

Then Eq. (11) may be written as

$$Y_g(\boldsymbol{X})=\sum_{i=1}^{p}\sum_{j=1}^{p}c_{gij}x_ix_j+\sum_{i=1}^{p}c_{gi}x_i-c_{g0},\ g=1,2,\cdots,G. \tag{15}$$

For a given individual $\boldsymbol{X}_0$, the values $Y_g(\boldsymbol{X}_0)$, $g=1,2,\ldots,G$ of G nonlinear discriminant functions may be obtained from Eq. (15). If

$$Y_g*(\boldsymbol{X}_0)=\operatorname*{Max}_{1\leqslant g\leqslant G}\{Y_g(\boldsymbol{X}_0)\}, \tag{16}$$

then the individual $\boldsymbol{X}_0$ belongs to the population A_g*.

If we assume that the covariance matrix of each population is equal to each other, then we obtain the following linear discriminant function:

$$Y_g(X)=\bar{\boldsymbol{X}}_g'\boldsymbol{S}_g^{-1}\boldsymbol{X}-\frac{1}{2}\bar{\boldsymbol{X}}_g'\boldsymbol{S}^{-1}\bar{\boldsymbol{X}}_g+\ln\ q_g, g=1,2,\cdots,G, \tag{17}$$

where

$$\boldsymbol{S}=\mathbf{W}/N-G,$$

while $\mathbf{W}=(w_{ij})$, $i,j=1,2,\ldots,p$, is a within cross-product matrix. Its element is

$$w_{ij}=\sum_{g=1}^{G}\sum_{k=1}^{n_g}(x_{igk}-\bar{x}_{ig})(x_{jgk}-\bar{x}_{jg}),\quad i,j=1,2,\cdots,p. \tag{18}$$

Considering

$$(c_{g1},c_{g2},\cdots c_{gp})=\bar{\boldsymbol{X}}_g'\boldsymbol{S}^{-1}, g=1,2,\cdots,G, \tag{19}$$

$$c_{g0} = -\frac{1}{2}\bar{X}'_g S^{-1} \bar{X}_g + \ln q_g, \qquad g=1,2,\cdots,G. \tag{20}$$

Eq. (17) may be written as

$$Y_g(X) = \sum_{i=1}^{p} c_{gi} x_i + c_{g0}, \qquad g=1,2,\cdots,G. \tag{21}$$

The discriminant rules corresponding to Eq. (21) are the same as the above.

III. THE STEPWISE NONLINEAR DISCRIMINANT

Previous discussions show that when the number of variables is p, the number of terms is $p+1$ in the linear discriminant function, but $2p+c_p^2+1$ in the nonlinear discriminant function. Thus, the number of terms in Eq. (15) is considerably increased when there are a large number of variables. This makes computation more difficult and does not prove that the nonlinear discriminant function has an advantage over the linear one, even though the presence of the nonlinear term makes the number of terms increase rapidly and discriminant effects rise obviously in the nonlinear discriminant function. In fact, in the discriminant function the ability of the division of each term on the various kinds of groups is different. In linear discriminant analysis the algorithm of the stepwise discriminant has been widely used to select important variables (Li, Yao and Yang, 1977), and has obtained good effects. We believe that this algorithm may be used to do nonlinear discriminant analysis as well. Retained are *the variables that play an important role in dividing various kinds of groups in the nonlinear* discriminant function.

It can be seen from Eq. (15) that, besides the linear terms with p variables, there exist additional $p+c_p^2$ nonlinear terms in the nonlinear discriminant function with p variables. Therefore, if we carry out a transformation of variables for those nonlinear terms, then together with p original variables, all told $2p+c_p^2$ variables are obtained. Thus, nonlinear discriminant analysis with p variables may become linear discriminant analysis with $2p+c_p^2$ variables. Therefore, the stepwise nonlinear discriminant analysis proposed by the author is divided into two steps:

First, a transformation of variables is made to give p variables, and $2p+c_p^2$ variables are obtained, i.e. the original p variables x_i $(i=1,2,\ldots,p)$ plus $p+c_p^2$ new variables $x_i x_j$ $(i,j=1,2,\ldots,p;\ i\leqslant j)$.

Second, the algorithm of the general stepwise discriminant is adopted for $2p+c_p^2$ variables and the nonlinear discriminant function is set up.

If the number of variables is great enough or the computation is limited by the capacity of computer, then the two-step sieve method (Li and Yao, 1976) can be adopted, i.e. the Wilks' values which possess the ability of single factor discriminant are first calculated to $2p+c_p^2$ variables as follows:

$$U_i = w_{ii}/t_{ii}, \qquad i=1,2,\cdots,2p+c_p^2, \tag{22}$$

where

$$w_{ii} = \sum_{g=1}^{G} \sum_{k=1}^{n_g} (x_{igk}-\bar{x}_{ig})^2, \qquad i=1,2,\cdots,2p+c_p^2, \tag{23}$$

are the square sum within cross-product, while

$$t_{ii} = \sum_{g=1}^{G} \sum_{k=1}^{n_g} (x_{igk}-\bar{x}_i)^2, \qquad i=1,2,\cdots,2p+c_p^2, \tag{24}$$

are the square sum of total cross-product, where $\bar{x}_{ig}$ is the mean of variable x_i for group g (see Eq. (8)). $\bar{x}_i$ the total mean of variable x_i, i.e.

$$\bar{x}_i=\frac{1}{N}\sum_{g=1}^{G}\sum_{k=1}^{n_g}x_{igk}, i=1,2,\cdots,2p+c_p^2. \tag{25}$$

At the same time, by using the statistical values for Wilks values U_i,

$$F_i(G-1,N-G)=\frac{1-U_i}{U_i}\cdot\frac{N-G}{G-1}, \quad i=1,2,\cdots,2p+c_p^2, \tag{26}$$

F-tests are made. Several variables of less U_i (i.e. bigger F_i) are screened primarily, and then a few variables are screened with rigorous algorithm of the stepwise discriminant from the variables obtained by primary selection and the better equations of the nonlinear discriminant function are set up.

The basic principle and computational procedures of the stepwise discriminant analysis are the same as those of Li and Yao (1977) and Li et al. (1977).

IV. SELECTION OF THE DISCRIMINANT MODEL

Whether the linear or the nonlinear discriminant function is adopted in the discriminant analysis depends mainly on whether the covariance matrixes of each normal population are equal to each other or not. Therefore, in selecting an exact model of the discriminant, first, it must be tested whether the covariance matrixes of each normal population are equal to each other or not. We use the following verification method proposed by Box (see Gal-Mowcorterchin et al., 1971).

Assuming that S_1, S_2, ..., S_G are the estimate values of the covariance matrixes for the sample of G populations, respectively. The null hypothesis Ho states as follows: S_1, S_2, ..., S_G are all from normal population with the same covariance matrix tested.

Assuming that N_1, N_2, ..., N_G are the sample size of each population, respectively. Let $n_g=N_g-1(g=1,2,\dots,G)$. The statistic Tc conforms a chi-square distribution with the degree of freedom $(G-1)\ p(p-1)/2$ when the null hypothesis Ho is founded. Here

$$T=\left(\sum_{g=1}^{G}n_g\right)\ln(de+s)-\sum_{g=1}^{G}n_g\ \ln(de+s_g), \tag{27}$$

$$c=1-\frac{2p^2+3p-1}{\sigma(p+1)(G-1)}\left[\sum_{g=1}^{G}\frac{1}{n_g}-\frac{1}{\sum_{g=1}^{G}n_g}\right], \tag{28}$$

while the pooled covariance matrix S of the sample has the form

$$S=\sum_{g=1}^{G}n_gS_g\Big/\sum_{g=1}^{G}n_g. \tag{29}$$

Giving the level of significance α, if $Tc>X_\alpha^2$, then the null hypothesis Ho is rejected, i.e. the covariance matrix of each population is not equal to each other. Conversely, if $Tc<X_\alpha^2$, then the null hypothesis Ho is accepted, i.e. the covariance matrix of each population is equal to each other.

Thus, we may confirm which discriminant model should be selected.

V. SEVERAL EXAMPLES

In order to verify whether the above scheme is feasible or not, we conduct comparisons and explanations to the following several examples.

Example 1, from Chang and Fang (1982):

In this example, $G=2$, $p=7$, $N=35$, $N_1=12$, $N_2=23$.

From Eqs. (27) and (28), we obtain $Tc=61.1058$, while $X^2_{0.05}(21)=32.671$, i.e. $Tc>X^2_{0.05}(21)$. It shows that the covariance matrixes of two populations are not equal in this example. The nonlinear discriminant functions should be adopted.

Using the algorithm of the stepwise discriminant, we obtain the results as follows:

The equations of the linear discriminant function ($p=7$) are

$$Y_1=-5.4755+0.3710x_5+17.6037x_6-0.0026x_7,$$
$$Y_2=-5.5328-0.1507x_5+38.4613x_6+0.0411x_7.$$

The equations of the nonlinear discriminant function ($p=35$) are

$$Y_1=-11.6857-2.3471x_5-0.0229x_7-0.0469x_5^2+0.5958x_3x_4-16.9000x_3x_6,$$
$$Y_2=-4.8183+0.9929x_5+0.0309x_7-0.0269x_5^2+0.1327x_3x_4+0.4430x_3x_5.$$

Table 1. Comparisons between Two Discriminant Models

Model	F-level	Numbers of Entry Variables	U	X^2	X^2_α	Numbers of Classification Errors
Linear Discriminant	2.0	3	0.3967	29.121	$X^2_{0.05}(3)$ 7.315 $X^2_{0.01}(3)$ 11.345	3
Nonlinear Discriminant	2.0	5	0.2852	38.270	$X^2_{0.05}(5)$ 11.070 $X^2_{0.01}(5)$ 15.068	1

The equations of the linear discriminant function with 7 variables have been set up by Chang and Fang (1982), and the numbers of the classification error are also 3. It may be seen that from either the verification of the classification effects or the cases of the classification error, the equations of the nonlinear discriminant function have an advantage over the linear one.

Example 2, from Speciality of Mathematics of Computation, Department of Mathematics, Nanjing University (1979):

In this example, $G=3$, $p=5$, $N=23$, $N_1=11$, $N_2=7$, $N_3=5$.

From Eqs. (27) and (28), we obtain $Tc=41.9477$, while $X^2_{0.05}(20)=30.410$, i.e. $Tc>X^2_{0.05}(20)$. It shows that the covariance matrixes of three populations are not equal to each other in this example. The nonlinear discriminant functions should also be adopted.

Using the algorithm of the stepwise discriminant, we obtain the results as follows:

The equations of the linear discriminant function ($p=5$) are

$$Y_1=-10.0339+0.0308x_2+1.4669x_5,$$
$$Y_2=-17.6363+0.0694x_2+0.9853x_5,$$
$$Y_3=-18.3113+0.0328x_2+2.2307x_5.$$

The equations of the nonlinear discriminant function ($p=20$) are

$$Y_1=-8.3655+0.0609x_2+0.0067x_3x_5,$$
$$Y_2=-16.8695+0.0922x_2-0.0019x_3x_5,$$
$$Y_3=-15.7770-0.0661x_2+0.0407x_3x_5.$$

Table 2. Comparisons between Two Discriminant Models

Model	F-level	Numbers of Entry Variables	U	X^2	X^2_a	Numbers of Classification Errors
Linear Discriminant	2.0	2	0.4422	15.9128	$X^2_{0.05}$ (4) 9.488 $X^2_{0.01}$ (4) 13.277	5
Nonlinear Discriminant	2.0	2	0.4007	17.8314	$X^2_{0.05}$ (4) 9.488 $X^2_{0.01}$ (4) 13.277	4

Two variables are screened out to both linear and nonlinear discriminant models in this example for the same F-level. But U, the ability of the classification of nonlinear discriminant model is greater than that of the linear one. In the test of significance, the former is more remarkable than that of the latter and the numbers of the classification error decrease by one when the nonlinear discriminant function is used.

Example 3, from Yao and Liu (1985):

In this example, $G=2$, $p=3$, $N=16$, $N_1=8$, $N_2=8$.

From Eqs. (27) and (28), we obtain $Tc=10.2733$, while $X^2_{0.05}(3)=7.815$, i.e. $Tc>X^2_{0.05}(3)$. The covariance matrixes of two populations are not equal either in this example. The nonlinear discriminant analysis should be adopted.

Using the algorithm of the stepwise discriminant, we obtain the results as follows:

The equations of the linear discriminant function ($p=3$) are

$$Y_1=-5.3558+0.5292x_1+0.1638x_2,$$

$$Y_2=-5.8202+1.4738x_1+0.0580x_2.$$

The equations of the nonlinear discriminant function ($p=9$) are

$$Y_1=-7.9471+0.8943x_1+0.6410x_2-0.0073x_2^2,$$

$$Y_2=-14.7206+2.1498x_1+0.9416x_2-0.0135x_2^2.$$

Table 3. Comparisons between Two Discriminant Models

Model	F-level	Numbers of Entry Variables	U	X^2	X^2_a	Numbers of Classification Errors
Linear Discriminant	2.0	2	0.4986	9.0472	$X^2_{0.05}$ (2) 5.991 $X^2_{0.01}$ (2) 9.210	3
Nonlinear Discriminant	2.0	3	0.3245	14.0682	$X^2_{0.05}$ (3) 7.815 $X^2_{0.01}$ (3) 11.345	1

The equations of the linear discriminant function with 3 variables have been set up by Yao and Liu (1985). The numbers of the classification error are 4 for the above equations. In addition, we have tested two discriminant models using the independent sample data from 1973 to 1982. The results of calculation show that the accuracy of prediction of the linear discriminant function is 70%, while that of the nonlinear one is 90%. It is thus clear that the discriminant effect of the nonlinear discriminant function is very good in this example,

VI. CONCLUSIONS

We have obtained several results by the previous analysis as follows:

(1) The results of the nonlinear discriminant function are more practical than that of the linear one in the weather typing prediction and the effects of the classification of the former is, to some extent, better than that of the latter. It shows that the use of the nonlinear discriminant function is important in the weather classification prediction.

(2) The scheme of the stepwise nonlinear discriminant analysis proposed by the author is feasible, and has important significance to the practical use.

(3) In order to select the exact discriminant model, we should test whether the covariance matrixes of the sample data of each population are equal to each other or not, in terms of the Box criterion in the practical use.

REFERENCES

Chang Yaoting and Fang Kaitai (1982), An Introduction to Multivariate Statistical Analysis, Science Press, Beijing, China, 247 pp. (in Chinese).

Clark, T.I. et al. (1975), Distinguishing between CAT and non-CAT areas by use of discriminant function analysis, *Mon. Wea. Rev.*, **103**:514–520.

Gal-Mowcorterchin, M.K. et al. (1971), Using discriminant analysis to the prediction of verglas, The Collection of Papers of Institute of Hydrometeorology USSR, No. 90, 3–39 (in Russian).

Li Maicun and Yao Dirong (1967), On two step screens of predictor, The References of Meteorological Science and Technology, No. 7, 12–16 (in Chinese).

Li Maicun and Yao Dirong (1977), Application of multiple-category discriminant analysis to the prediction of typhoon tracks, *Journal of Hangzhou University* (Natural Science Edition), **4**:51–72 (in Chinese).

Li Maicun, Yao Dirong and Yang Ziqian (1977), The screening of predictors in multiple-category discriminant analysis, Acta Applied Mathematics, No. 4, 58–73 (in Chinese).

Lorenz, E.N. (1980), Nonlinear statistical weather prediction, The Collection of Papers Presented at the WMO Symposium on Probabilistic and Statistical Methods in Weather Forecasting, 3–9.

Speciality of Mathematics of Computation, Department of Mathematics, Nanjing University (1979), The Basis and Methods of Probability Statistics, Science Press, Beijing, China, 305 pp. (in Chinese).

Wang Zonghao and Li Maicun et al. (1974), The Methods to Probability Statistics in Weather Forecasting, Science Press, Beijing, China, pp. 102–125 (in Chinese).

Yao Dirong and Liu Yuezheng (1983), The nonlinear discriminant function in weather prediction, *Journal of Hangzhou University* (Natural Science Edition), **10**:127–138 (in Chinese with English abstract).

Yao Dirong and Liu Yuezheng (1985), The screening of predictors in nonlinear discriminant function, *Scientia Atmospherica Sinica*, **9**:54–61 (in Chinese with English abstract).