Б. З. ДОКТОРОВ, асс.
*Ленинградская Высшая партийная школа*

# ДВЕ ЗАДАЧИ ИЗ ТЕОРИИ ДИСКРИМИНАНТНОГО АНАЛИЗА

1. Пусть рассматриваются два $n$-мерных нормальных распределения с одинаковыми ковариационными матрицами В и разностью средних $d = \mu_1 - \mu_2$. Один из способов получения коэффициентов линейной дискриминантной функции (л. д. ф.)

$$z = b^T x \tag{1}$$

связан с решением экстремальной задачи

$$\max_b \{(b^T d)^2 / b^T \mathrm{B} b\}. \tag{2}$$

Решением (2) является вектор

$$b^T_{\text{опт}} = d^T \mathrm{B}^{-1} = (\mu_1 - \mu_2)^T \mathrm{B}^{-1}, \tag{3}$$

определяющий коэффициенты л. д. ф. Настоящий метод получения коэффициентов (1) восходит к Р. Фишеру [1].

Максимальное значение (2) определяет расстояние Махаланобиса между совокупностями

$$\Delta_m = b^T_{\text{опт}} d = d^T \mathrm{B}^{-1} d. \tag{4}$$

Если при решении задачи классификации наблюдений используется не оптимальное правило (3), а некоторый вектор

$$b_1^T = d_1^T \mathrm{B}_1^{-1}, \tag{5}$$

то значение дроби в (2) равно

$$\Delta = \frac{\{b_1^T d\}^2}{b_1^T \mathrm{B}^{-1} b_1} = \frac{\{d_1 \mathrm{B}_1^{-1} d\}^2}{d_1^T \mathrm{B}_1^{-1} \mathrm{B} \mathrm{B}_1^{-1} d_1}. \tag{6}$$

Поскольку (3) оптимально, то

$$\alpha = \Delta_m - \Delta \geqslant 0 \tag{7}$$

указывает на величину отличия правила (5) от оптимального.

Будем говорить, что линейные решающие правила $z = b^T x$ и $z_1 = b_1^T x$ линейно зависимы, если $b = f b_1$, где $f$ — скаляр. Из (2) и (7) можно получить, что $\alpha = 0$ тогда и только тогда, когда решающие правила линейно зависимы.

Прежде чем переходить к исследованию свойств л. д. ф., сформулируем некоторые положения теории коммутирующих матриц:

а) если $AB = BA$ и $B^{-1}$ существует, то $B^{-1}A = AB^{-1}$;

б) если $AB = BA$ и существуют $A^{-1}$ и $B^{-1}$, то $A^{-1}B^{-1} = B^{-1}A^{-1}$;

в) справедлива теорема [2]: если дано конечное множество попарно коммутирующих нормальных матриц, то все эти матрицы одним и тем же унитарным преобразованием могут быть приведены к диагональному виду.

**Теорема 1.** Решающие правила $z = d^T B^{-1} x$ и $z_1 = d^T B_1^{-1} x$ линейно зависимы тогда и только тогда, когда $D = dd^T$, $B$ и $B_1$ попарно коммутативны.

Если выполнено

$$DB^{-1} = B^{-1}D,\ BB_1 = B_1B \text{ и } DB_1^{-1} = B_1^{-1}D, \tag{8}$$

то непосредственной проверкой убеждаемся, что $\alpha = 0$. Вторую часть доказательства проводим от противного.

Пусть $\alpha = 0$, т. е.

$$B^{-1}DB_1^{-1} = B_1^{-1}DB^{-1}, \tag{9}$$

но одно из соотношений (8) не выполнено, например

$$B^{-1}D \neq DB^{-1}. \tag{10}$$

Умножая (10) справа на $B_1^{-1}$, имеем

$$B^{-1}DB_1^{-1} \neq DB^{-1}B_1^{-1} \neq DB_1^{-1}B^{-1} \neq B_1^{-1}DB^{-1},$$

что противоречит (9).

Можно проверить, что если $z$ и $z_1$ линейно зависимы, то

$$z = \frac{d^T B^{-1} d}{d^T B_1^{-1} d} z_1. \tag{11}$$

**Теорема 2.** Решающие правила $z = d^T B^{-1} x$ и $z_1 = d_1^T B^{-1} x$ линейно зависимы тогда и только тогда, когда $d = d_1$.

**Теорема 3.** Решающие правила $z = d^T B^{-1} x$ и $z_1 = d_1^T B_1^{-1} x$ линейно зависимы тогда и только тогда, когда выполнены условия теорем 1 и 2.

*Следствие.* Решающие правила $z = d^T B^{-1} x$ и $z = d^T x$ линейно зависимы тогда и только тогда, когда $D$ и $B$ коммутативны.

**Теорема 4.** Если решающие правила $z = d^T B^{-1} x$ и $z_1 = d^T B_1^{-1} x$ линейно зависимы, то вектор $u = (d^T d)^{-1} d$ является нормированным собственным вектором $B$, $B_1$ и $D$.

Поскольку $z$ и $z_1$ линейно зависимы, то B, $B_1$ и D попарно коммутируют. Так как D — матрица ранга 1, то единственным ненулевым собственным числом D является $\lambda = d^T d$, и ему отвечает нормированный собственный вектор $u = (d^T d)^{-1} d$. Согласно теореме о приведении попарно коммутирующих матриц к диагональному виду, матрицы B, $B_1$ и D имеют общий набор собственных векторов, тогда среди них есть и вектор $u$.

Пример. Рассмотрим решающие правила $z_1 = e^T R_{\rho_1}^{-1} x$ и $z_2 = e^T R_{\rho_2}^{-1} x$, где $e^T = (1\ 1 \ldots 1)$ и $R_{\rho_i}$ — $(n \times n)$-корреляционная матрица вектора $x$ вида

$$R_{\rho_i} = \rho_i e e^T + (1 - \rho_i)\,I \qquad (i = 1,\ 2),$$

т. е. все коэффициенты корреляции в $R_{\rho_i}$ равны между собой. Согласно результатам приложения I,

$$R_{\rho_1}^{-1} = A\Lambda_{\rho_1}^{-1}A^T \ \text{ и } \ R_{\rho_2}^{-1} = A\Lambda_{\rho_2}^{-1}A^T .$$

Далее,

$$D = ee^T = A \begin{Vmatrix} 1 & 0 \ldots 0 \\ 0 & \\ \cdot & \\ \cdot & 0 \\ \cdot & \\ 0 & \end{Vmatrix} A^T .$$

Итак, $R_{\rho_1}$, $R_{\rho_2}$, D — попарно коммутативны, а $z_1$ и $z_2$ — линейно зависимы. Непосредственным вычислением или из (11) имеем

$$z_1 = \frac{1 + \rho_2 (n-1)}{1 + \rho_1 (n-1)}\, z_2.$$

2. Рассмотрим актуальную в практике дискриминантного анализа задачу отбора признаков для построения л. д. ф. Будем предполагать, что имеется два класса объектов, описываемых $n$-мерными нормальными распределениями с единичными ковариационными матрицами и вектором разности средних $d$, причем компоненты вектора $d$ положительны и упорядочены по убыванию, т. е. $d_1 \geqslant d_2 \geqslant \geqslant d_3 \geqslant \cdots \geqslant d_n$. Добавление новой переменной (или нового признака, участвующего в построении л. д. ф.) равносильно рассмотрению блочного $(n+1)$-мерного вектора $(x^T x_{n+1})$. Пусть $x_{n+1}$ — случайная нормальная величина с единичной дисперсией и разностью средних в обоих совокупностях $d_{n+1} \leqslant d_n$. Если новая переменная независима со всеми компонентами вектора $x$, то расстояние Махаланобиса для $(n+1)$-мерного вектора равно $d^T d + d_{n+1}^2$. Если $x_{n+1}$ коррелирует с вектором $x$, то, обозначив полученное расстояние через $\Delta_{n+1}$, будем говорить, что добавление нового признака полезно при

$$\delta = \Delta_{n+1} - (d^T d + d_{n+1}^2) > 0. \qquad (12)$$

Ковариационная матрица $(n+1)$-мерного вектора равна

$$R_{n+1} = \begin{Vmatrix} I & A_\rho \\ A_\rho^T & 1 \end{Vmatrix}, \tag{13}$$

где $A_\rho^T = (\rho_1\rho_2 \ldots \rho_n)$ и $\rho_i = \mathrm{cov}\,(x_i x_{n+1})$ $(i = \overline{1, n})$.

Воспользовавшись результатами приложения II, запишем (12) в виде

$$\delta = d^T A_\rho A_\rho^T d + d_{n+1}^2 A_\rho^T A_\rho - 2d_{n+1} A_\rho^T d > 0. \tag{14}$$

Для некоторых специальных случаев матрицы $A_\rho$ вид правила $\delta$ может быть упрощен.

а) $A_\rho = \rho e$. Решая (14), получаем значения $\rho$, при которых добавление нового признака полезно:

$$-1 < \rho < 0, \quad \frac{2d_{n+1}e^T d}{d^T ee^T d + nd_{n+1}^2} < \rho < 1. \tag{15}$$

б) Для случая добавления одной переменной к ранее рассмотренному одному признаку (14) равно

$$\rho^2 (d_1^2 + d_2^2) - 2\rho d_1 d_2 > 0. \tag{16}$$

Это означает, что необходимо добавлять отрицательно коррелированный признак с максимально возможной величиной $d_2$. Если (16) измерять относительно $d_1^2$, то имеем

$$(\rho - f)^2/(1 - \rho^2) > f^2,$$

где $f^2 = d_2^2/d_1^2 < 1$.

Решение задачи отбора признаков в таком виде дает Кочрен [3].

В приложении III указывается ортогональное преобразование, приводящее (13) к диагональному виду. Таким образом, последовательное применение неравенства (14) и приведение ковариационной матрицы (13) к диагональному виду может рассматриваться как алгоритм отбора признаков для построения л. д. ф.

## ПРИЛОЖЕНИЕ

I. Матрица

$$R_\rho = \rho ee^T + (1 - \rho) I$$

имеет определитель $(1-\rho)^{n-1}[1 + (n-1)\rho]$ и обратную матрицу

$$R_\rho^{-1} = (1-\rho)^{-1}\left[I - \rho\,[1 + \rho\,(n-1)]^{-1} ee^T\right].$$

Собственные значения $R_\rho$ равны $\lambda_1 = 1 + \rho\,(n-1)$ и $\lambda_2 = \lambda_3 = \cdots = \lambda_n = 1 - \rho$. Им отвечает система собственных векторов

$$
\begin{aligned}
\alpha_1^T &= (1 \quad 1 \quad 1 \quad 0 \ldots \quad 1 \quad 1)/\sqrt{n}\,,\\
\alpha_2^T &= (1 \quad -1 \quad 0 \quad 0 \ldots \quad 0 \quad 0)/\sqrt{2}\,,\\
\alpha_3^T &= (1 \quad 1 \quad -2 \quad 0 \ldots \quad 0 \quad 0)/\sqrt{6}\,,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\alpha_n^T &= (1 \quad 1 \quad 1 \quad 1 \ldots -(n-1) \quad 0)/\sqrt{n\,(n-1)}\,,
\end{aligned}
$$

определяющая ортогональное преобразование Хельмерта [4]. Если ввести диагональную матрицу $\Lambda_\rho = \{1 + \rho\,(n-1)\ 1-\rho \ldots 1-\rho\}$ и $A = (\alpha_1\alpha_2 \ldots \alpha_n)$, то $R_\rho = A\Lambda_\rho A^T$.

II. Определитель матрицы (13) равен

$$\| R_{n+1} \| = 1 - A_\rho' A_\rho\,. \tag{17}$$

Поскольку $\| R_{n+1} \| > 0$, то (17) накладывает ограничения на $A_\rho$;

$$R_{n+1}^{-1} = \begin{Vmatrix} A_\rho A_\rho^T & -A_\rho \\ -A_\rho & A_\rho^T A_\rho \end{Vmatrix} + (1 - A_\rho^T A_\rho)\,I.$$

III. Решая задачу об определении собственных значений матрицы (13), имеем

$$\lambda_1 = 1 + \{A_\rho^T A_\rho\}^{\frac{1}{2}},$$

$$\lambda_2 = 1 - \{A_\rho^T A_\rho\}^{\frac{1}{2}},$$

$$\lambda_3 = \lambda_4 = \cdots\ \lambda_{n+1} = 1.$$

Собственным значениям отвечают векторы

$$
\begin{aligned}
B_1^T &= \left(\rho_1 \quad \rho_2 \quad \rho_3 \quad \ldots \quad \rho_{n-1} \quad \rho_n \quad \sqrt{\sum_{i=1}^{n} \rho_i^2}\right),\\
B_2^T &= \left(\rho_1 \quad \rho_2 \quad \rho_3 \quad \ldots \quad \rho_{n-1} \quad \rho_n \quad -\sqrt{\sum_{i=1}^{n} \rho_i^2}\right),\\
B_3^T &= (\rho_1\rho_2 \quad -\rho_1^2 \quad 0 \quad \ldots \quad 0 \quad 0 \quad 0),\\
B_4^T &= (\rho_1\rho_3 \quad \rho_2\rho_3 \quad -(\rho_1^2 + \rho_2^2) \ldots \quad 0 \quad 0 \quad 0),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
B_n^T &= \left(\rho_1\rho_n \quad \rho_2\rho_n \quad \rho_3\rho_n \quad \ldots -\sum_{i=1}^{n-1} \rho_i^2 \quad 0 \quad 0\right).
\end{aligned}
$$

Ортогональность $B_i$ легко проверяется; нормируя множество векторов $B_i$, получаем ортогональное преобразование, приводящее (13) к диагональному виду.

## ЛИТЕРАТУРА

1. Fisher R. A. The use of multiple measurements in taxonomic problems.— Ann. Eugen., N 7, 1936, p. 179—88.
2. Гантмахер Ф. Р. Теория матриц. М., «Наука», 1967.
3. Cochran W. G. On the performance on the linear discriminant function.— Technometrics, 6, N 2, 1964.
4. Кендалл Дж., Стьюарт А. Теория распределений. М., «Наука», 1966.

B. Z. Doktorow

## TWO PROBLEMS FROM THE THEORY OF DISCRIMINANT ANALYSIS

Summary

The paper contains the analysis of the situation, received from application of the linear discriminant function (l. d. f.) different from the optimal one and the problem of the selection of the variables for l. d. f.

Поступила в редколлегию 8.IV. 1969.