回归模型和判别问题的关系

张 尧 庭

1. 引言。早在1936年，Fisher[1]就指出了可以用回归方程来得出两组判别的线性判别函数，[2]已将此结果写在书上。到1955年，Tatsuoka[4]等得到了用典型相关的方法导出多组判别的线性判别函数。由于见不到原文，仅在1968年 Glahn[4]中知道这一结果，也没有给出证明；[4]第一次讨论了典型相关、回归、判别之间的关系。[4]没有使用矩阵的广义逆，因此，证明的结果并不一般。杨自强在[5]中证明了[4]的结果，并给出了用回归的算法来解判别问题的计算方法。本文使用广义逆的工具，很方便证得[3]的结果，导出[1]这样的特殊情况，并且在一般的情况下，讨论回归模型和判别分析的关系，最后提出了使用双重筛选的多元逐步回归方法来解决判别问题的新方法。以下用大写的拉丁字母表示矩阵，小写的字母表示数或向量，A^- 表示矩阵 A 的广义逆，有关性质参看[6]。

2. 回归和判别的关系。

设 k 个总体相应的 $n_1, \cdots, n_k$ 个样品为：

$$x_{(1)}^{(1)},\ x_{(2)}^{(1)},\ \cdots,\ x_{(n_1)}^{(1)};\ \cdots;\ x_{(1)}^{(k)},\ x_{(2)}^{(k)},\ \cdots,\ x_{(n_k)}^{(k)}.$$

每一个样品是一个 p 维的向量，用矩阵的形式写出，得

$$\underset{n_i\times p}{X_i} = \begin{pmatrix} x_{(1)}^{(i)\prime} \\ \vdots \\ x_{(n_i)}^{(i)\prime} \end{pmatrix} \quad i = 1, 2, \cdots, k,$$

记 $\underset{n\times p}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$，其中 $n = \sum_{i=1}^{k} n_i$，

令 $\bar{x}^{(i)} = \frac{1}{n_i} X_i' 1_{n_i}$，$i = 1, 2, \cdots, k$，其中 $1_{n_i} = \left.\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}\right\} n_i$ 个 1

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{k} n_i \bar{x}^{(i)},$$

$$L_{jj}=X_j'\left(I-\frac{1}{n_j}J\right)X_j,\ \text{其中}\ J=11',$$

于是，组内差矩阵 $W=\sum_{i=1}^{k}L_{ii}$，组间差矩阵

$$B=X'\left(I-\frac{1}{n}J\right)\begin{pmatrix}n_1^{-1}J & & 0\\ & \ddots & \\ 0 & & n_k^{-1}J\end{pmatrix}\left(I-\frac{1}{n}J\right)X.$$

由判别分析的 Fisher 准则知道相应的判别函数是

$$(B-\lambda W)a=0 \qquad (1)$$

的解，即 B 对于 W 而言的相对特征根 λ_i 所相应的特征向量 a_i。现在，我们把判别问题看作是一个多元的 0－1 回归问题。令

$$\underset{n_i\times k}{Y_i}=(\underbrace{0\ 0\cdots 0}_{i-1}\ 1\ \underbrace{0\cdots 0}_{n-i})\quad i=1,2,\cdots,k$$

$$\underset{n\times n}{Y}=\begin{pmatrix}Y_1\\ \vdots\\ Y_k\end{pmatrix},$$

于是，可以认为

$$Y=(1X)\beta+\varepsilon,$$

这就是回归问题。记 $L_{xy}=X'\left(I-\frac{1}{n}J\right)Y$，类似地 L_{xx}，L_{yy} 的符号就不重复了。由熟知的公式，得 β 的最小二乘估计 $\hat{\beta}$，

$$\hat{\beta}=L_{xx}^{-}L_{xy}.\ （或\ L_{xx}\hat{\beta}=L_{xy}）$$

今

$$\begin{aligned}L_{xx}&=X'\left(I-\frac{1}{n}J\right)X\\ &=X'\left(I-\frac{1}{n}J\right)\left(I-\frac{1}{n}J\right)X\\ &=\sum_{i=1}^{k}(X_i-1\bar{x}')'(X_i-1\bar{x})\\ &=\sum_{i=1}^{k}(X_i-1\bar{x}^{(i)\prime}+1(\bar{x}^{(i)}-\bar{x})')'(X_i-1\bar{x}^{(i)\prime}+1(\bar{x}^{(i)}-\bar{x})')\\ &=\sum_{i=1}^{k}(X_i-1\bar{x}^{(i)\prime})'(X_i-1\bar{x}^{(i)\prime})+\sum_{i=1}^{k}n_i(\bar{x}^{(i)}-\bar{x})(\bar{x}^{(i)}-\bar{x})'\\ &=W+B,\end{aligned}$$

注意到

$$(X_i - 1\bar{x}')'Y_i = (X_i' - \bar{x}1')(\underbrace{0\cdots0}_{i-1}\ 1\ \underbrace{0\cdots0}_{n-i})$$

$$= (\underbrace{0\cdots0}_{i-1}\ n_i(\bar{x}^{(i)} - \bar{x})\ \underbrace{0\cdots0}_{n-i}),$$

$$i = 1, 2, \cdots, k,$$

因此

$$L_{xy} = \sum_{i=1}^{k}(X_i - 1\bar{x}')'Y_i$$

$$= (n_1(\bar{x}^{(1)} - \bar{x}),\ n_2(\bar{x}^{(2)} - \bar{x}),\ \cdots,\ n_k(\bar{x}^{(k)} - \bar{x}))$$

由正规方程

$$(W + B)\hat{\beta} = L_{xy}$$

得

$$W\hat{\beta} = L_{xy} - B\hat{\beta}$$

$$= L_{xy} - \sum_{i=1}^{k} n_i(\bar{x}^{(i)} - \bar{x})(\bar{x}^{(i)} - \bar{x})'(\hat{\beta}_1\cdots\hat{\beta}_k),$$

比较上式两端的各列，就有：

$$W\hat{\beta}_i = n_i(\bar{x}^{(i)} - \bar{x}) - \sum_{j=1}^{k} n_j(\bar{x}^{(j)} - \bar{x})(\bar{x}^{(j)} - \bar{x})'\hat{\beta}_i$$

$$i = 1, 2, \cdots, k.$$

记

$$h_{ij} = \delta_{ij} - \bar{x}^{(j)\prime}\hat{\beta}_i,$$

就有

$$W\hat{\beta}_i = \sum_{j=1}^{k} h_{ij}n_j(\bar{x}^{(j)} - \bar{x}), \quad i = 1, 2, \cdots, k. \tag{2}$$

与(1)相比较，由(1)导出

$$Wa_i = \lambda_i Ba_i = \lambda_i \sum_{j=1}^{k} n_j(\bar{x}^{(j)} - \bar{x})(\bar{x}^{(j)} - \bar{x})'a_i,$$

记 $g_{ij} = \lambda_i\bar{x}^{(j)\prime}a_i$ 后，就有

$$Wa_i = \sum_{j=1}^{k} g_{ij}n_j(\bar{x}^{(j)} - \bar{x}), \quad i = 1, 2, \cdots, k \tag{3}$$

一般说来，(2)与(3)是不同的，当 $k=2$ 时，两者只差一个常数乘子，因而可以由 $\hat{\beta}$ 导出 a 来。我们将在 4 中详细讨论(2)与(3)的差异。

3. 典型相关和判别的关系。

我们知道，典型相关变量的系数 b 满足方程：

$$\mu^2 L_{xx}b - L_{xy}L_{\bar{y}\bar{y}}L_{yx}b = 0.$$

今已知 $L_{xx} = W + B$，现在来求 $L_{xy}L_{\bar{y}\bar{y}}L_{yx}$ 和 W、B 的关系式。今

$$L_{xy}L_{\bar{y}\bar{y}}L_{yx}=X'\left(I-\frac{1}{n}J\right)Y\left(Y'\left(I-\frac{1}{n}J\right)Y\right)^{-}Y'\left(I-\frac{1}{n}J\right)X,$$

又

$$Y'Y=\begin{pmatrix}n_1 & & 0\\ & \ddots & \\ 0 & & n_k\end{pmatrix},\quad Y'1=\begin{pmatrix}n_1\\ \vdots\\ n_k\end{pmatrix},\quad Y1=1,$$

于是有

$$Y'\left(I-\frac{1}{n}J\right)Y(Y'Y)^{-1}Y'\left(I-\frac{1}{n}J\right)Y$$

$$=\left(Y'Y-Y'\left(\frac{1}{n}J\right)Y\right)(Y'Y)^{-1}Y'\left(I-\frac{1}{n}J\right)Y$$

$$=\left(I-\frac{1}{n}Y'1\,1'Y\right)(Y'Y)^{-1}Y'\left(I-\frac{1}{n}J\right)Y$$

而

$$\frac{1}{n}Y'1\,1'Y(Y'Y)^{-1}Y'\left(I-\frac{1}{n}J\right)Y$$

$$=\frac{1}{n}\begin{pmatrix}n_1\\ \vdots\\ n_k\end{pmatrix}(n_1\cdots n_k)\begin{pmatrix}n_1^{-1} & & 0\\ & \ddots & \\ 0 & & n_k^{-1}\end{pmatrix}Y'\left(I-\frac{1}{n}J\right)Y$$

$$=\frac{1}{n}\begin{pmatrix}n_1\\ \vdots\\ n_k\end{pmatrix}1'Y'\left(I-\frac{1}{n}J\right)Y=0$$

因此

$$Y'\left(I-\frac{1}{n}J\right)Y(Y'Y)^{-1}Y'\left(I-\frac{1}{n}J\right)Y=Y'\left(I-\frac{1}{n}J\right)Y,$$

也即$\left(Y'\left(I-\frac{1}{n}J\right)Y\right)^{-}$之一就是$(Y'Y)^{-1}$，而$L_{xy}L_{\bar{y}\bar{y}}L_{yx}$与$L_{\bar{y}\bar{y}}$的选法无关，因此

$$L_{xy}L_{\bar{y}\bar{y}}L_{yx}=X'\left(I-\frac{1}{n}J\right)Y(Y'Y)^{-1}Y'\left(I-\frac{1}{n}J\right)X$$

$$=X'\left(I-\frac{1}{n}J\right)\begin{pmatrix}Y_1\\ \vdots\\ Y_k\end{pmatrix}\begin{pmatrix}n_1^{-1} & & 0\\ & \ddots & \\ 0 & & n_k^{-1}\end{pmatrix}(Y_1'\cdots Y_k')\left(I-\frac{1}{n}\right)X$$

$$=X'\left(I-\frac{1}{n}J\right)\begin{pmatrix}n_1^{-1}J & & 0\\ & \ddots & \\ 0 & & n_k^{-1}J\end{pmatrix}\left(I-\frac{1}{n}J\right)X$$

$$=\sum_{i=1}^{k}\frac{1}{n_i}(X_i-1\bar{x}')'11'(X_i-1\bar{x}')$$

$$=\sum_{i=1}^{k}n_i(\bar{x}^{(i)}-\bar{x})(\bar{x}^{(i)}-\bar{x})'$$

$$=B.$$

因此，b 满足

$$\mu^2(W+B)b-Bb=0,$$

即

$$Wb=\frac{1-\mu^2}{\mu^2}Bb. \tag{4}$$

将(1)、(4)比较一下，就知道 a 与 b 只是形式上的差异，λ_i 与 $\frac{1-\mu_i^2}{\mu_i^2}$ 相当，这样就证明了[5]的结果。

4. 典型相关与回归的关系

[4]中提出了用典型相关的方法来作多指标和多个自变量之间的回归。此时可以不考虑样本的资料，只讨论随机向量之间的关系，在模型上更清楚一些。

所谓回归就是考虑随机向量 $\underset{p\times 1}{y}$ 在随机向量 $\underset{k\times 1}{x}$ 所张成的线性空间 $L(x)$ 上的投影，从投影的公式我们都知道，用 $\hat{P}(y|x)$ 表示 y 在 $L(x)$ 上的投影时，就有

$$\hat{P}(y|x)=E(y)+\operatorname{cov}(y,x)V^{-}(x)(x-E(x)). \tag{5}$$

(5)式就是通常所说的回归方程。

典型相关讨论 $L(y)$ 与 $L(x)$ 中有代表性的变量之间关系，即存在一组典型变量

$$(a_i'x,\ b_i'y),\quad i=1,2,\cdots,r,\quad r=rk\,V^{+}(x)\operatorname{cov}(x,y)V^{+}(y)\operatorname{cov}(y,x)$$

它们彼此是不相关的，而 $a_i'x$ 与 $b_i'y$ 的相关系数 ρ_i 的平方是 $V^{+}(x)\operatorname{cov}(x,y)V^{+}(y)\operatorname{cov}(y,x)$ 的非 0 特征根。于是就有回归方程

$$b_i'(y-E(y))=\rho_i a_i'(x-E(x)),\qquad i=1,2,\cdots,r$$

写成矩阵的形式，记 $B=(b_1\cdots b_r)$，$A=(a_1\cdots a_r)$，

$$\rho=\begin{pmatrix}\rho_1 & & 0\\ & \ddots & \\ 0 & & \rho_r\end{pmatrix},$$

就得

$$\begin{cases}\underset{r\times p}{B'}(y-E(y))=\underset{r\times k}{\rho}\,A'(x-E(x)),\\ B'V(y)B=I_r,\quad A'V(x)A=I_r,\\ B'\operatorname{cov}(y,x)A=\begin{pmatrix}\rho_1^2 & & 0\\ & \ddots & \\ 0 & & \rho_r^2\end{pmatrix}.\end{cases} \tag{6}$$

考虑新的“回归方程”，形式上从(6)的第一式解出 y，就用 $\hat{y}$ 表示它，得

$$\hat{y}=B'^{+}\rho A'(x-E(x))+E(y) \tag{7}$$

当 $r=p=k$ 时，从(6)式得

$$V(y)=B'^{-1}B^{-1}=(BB')^{-1},\quad V(x)=(AA')^{-1},$$

$$B'\operatorname{cov}(y,x)A=\rho,$$

于是

$$B'^{+}\rho A' = B'^{-1}B'\,\mathrm{cov}(y, x)AA'$$

$$= \mathrm{cov}(y, x)V^{-1}(x)$$

因此，（7）式和（5）式给出完全相同的方程。当 $r<p$ 或 $r<k$，而且 $rkV(y)>r$ 或 $rkV(x)>r$ 时，（7）式给出的方程自然没有（5）式好（在最小二乘的意义下，即在均方误差的意义下），此时（7）式与（5）式并不能证明它们是完全一样的。由此可知，实际上我们毋需考虑从典型相关的角度求出典型变量再进行反解，只要直接考虑回归就可以了。

因此，从均方误差最小的意义上来看，把多组判别看成是一个回归问题是合适的。[7]中提出了双重筛选的多指标逐步回归方法，很自然，用到多组判别问题上去就得到逐步分组的逐步筛选因子的判别方法。

参 考 文 献

[1] Fisher, R. A. (1936), "The use of multiple measurments in taxonomic problems", Ann, Eugen., 9, 179—188.

[2] Anderson, T. W. (1958), An Introduction to Multivariate Statistical Analysis.

[3] Tatsuoka, M. M. (1955), "The relationship between canonical correlation and discriminant analysis; and a proposal for uitlizing qualitative deta in discriminant analysis", Cambridge, Educational Research Corporation, 47p. p..

[4] Glahn, H. R. (1968), "Canonical correlation and its relationship to discriminant analysis and multiple regtession", J. Atomospheric Sci., 25(1), 23—31.

[5] 杨自强，"判别分析和回归分析"（待发表）。

[6] Rao, C. R. (1973), Linear Statistical Inference and its Applications, Second Edition.

[7] 张尧庭，赵凌，"双重筛选的逐步回归"（将发表）。