

研究简报

Fisher 识别的最佳投影平面

王景芳

(长岭炼油化工厂计算机应用研究所, 岳阳市 414012)

关键词: 模式识别, 信息提取, 投影平面, 判别向量.

一、引理

引理. $F(u) = \sum_{i=1}^m \frac{b_i^2}{\lambda_i - u}$, $b_i \neq 0, i = 1, 2, \dots, m$; 若 $\lambda_1, \lambda_2, \dots, \lambda_m$ 中有 $r (r > 1)$ 个互导, 即 $\lambda_{i_1} > \lambda_{i_2} > \dots > \lambda_{i_r}$, 那么 $F(u) = 0$ 在 $(\lambda_{i_{k+1}}, \lambda_{i_k})$, $k = 1, 2, \dots, r - 1$ 内有唯一解, 最大解 u_0 满足 $\lambda_{i_2} < u_0 < \lambda_{i_1}$.

证明. $\because F'(u) = \sum_{i=1}^m \left(\frac{b_i}{\lambda_i - u} \right)^2 > 0,$
 $\therefore u \rightarrow \lambda_{i_k} - 0, F(u) \rightarrow +\infty, u \rightarrow \lambda_{i_k} + 0,$
 $F(u) \rightarrow -\infty, (k = 1, 2, \dots, r),$
 $u \rightarrow \pm \infty, F(u) \rightarrow 0.$

$F(u)$ 的导函数大于 0, 表明它在每个连续区间内为增函数, 且每个区间 $(\lambda_{i_{k+1}}, \lambda_{i_k})$, $k = 1, 2, \dots, r - 1$ 内有一个零点且只有一个零点 ($F(u)$ 函数曲线见图 1). 引理得证.

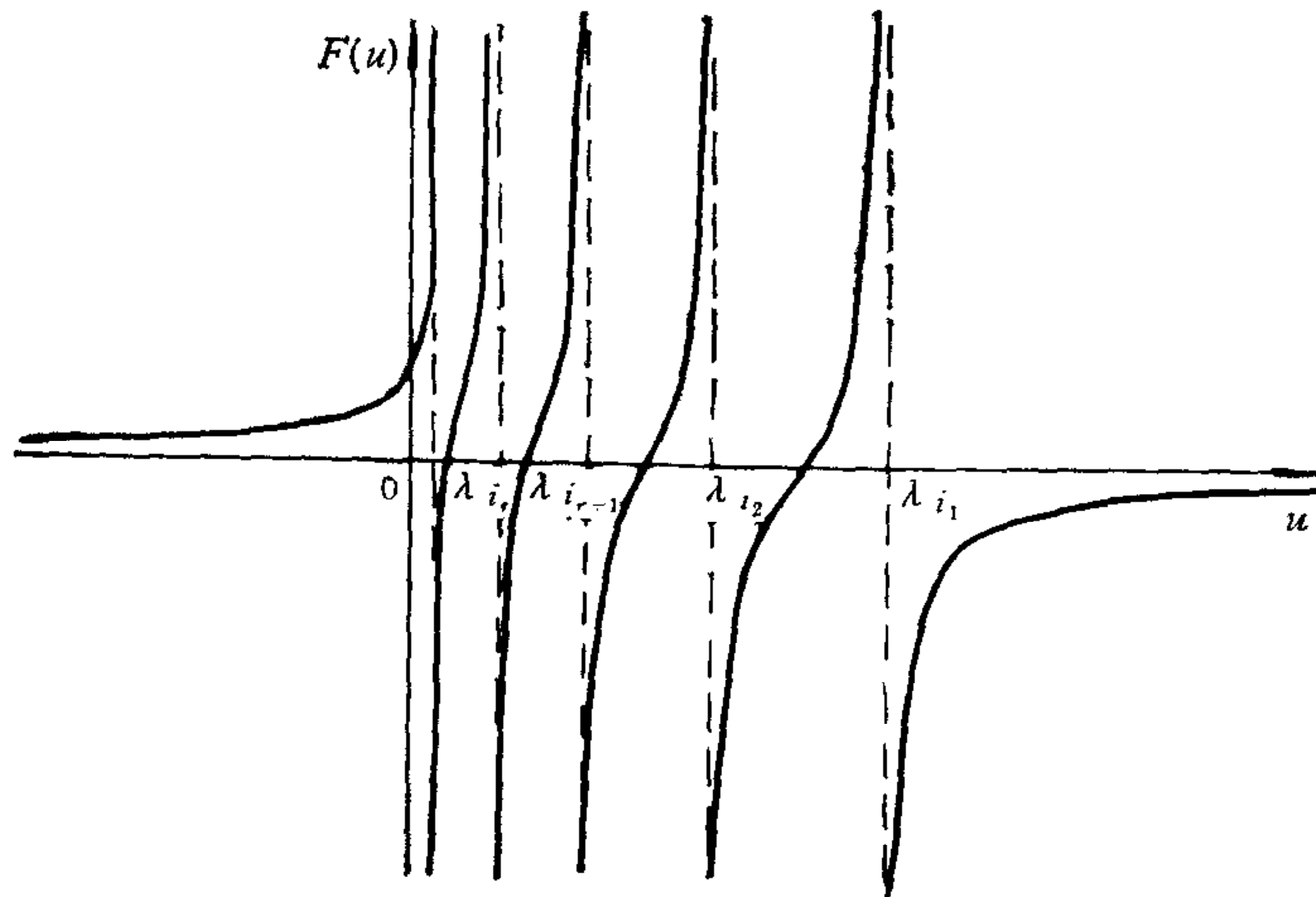


图 1 函数 $F(u)$ 曲线图

二、定理与证明

若 k 类样本集 X 的类间差阵为 B , 类内差阵为 E , 则 X 在向量 η 上的投影的类间差为 $\eta' B \eta$, 类内差为 $\eta' E \eta$. 若这 k 类有显著性差异, 即组间分开, 则 $\eta' B \eta$ 应充分大; 若组内团聚, 则 $\eta' E \eta$ 应尽可能小. 综合即 Fisher 信息量 $\Delta(\eta) = \eta' B \eta / \eta' E \eta$ 应充分大, 使得 $\Delta(\cdot)$ 极大的值为 $|\lambda E - B| = 0$ 的最大特征根. $\Delta(\eta)$ 的大小用来衡量识别函数 $\mu(x) = \eta' x$ 的判别效果.

定理. m 阶方阵 $E > 0, B \geq 0$, 特征方程 $|\lambda E - B| = 0$ 的特征根 $\lambda_1, \lambda_2, \dots, \lambda_m$ 对应的特征向量分别为 $\eta_1, \eta_2, \dots, \eta_m$; 不妨记 $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_m$, 可行集

$$S = \left\{ \eta \mid \eta = \sum_{i=1}^m z_i \eta_i, \eta_i' \eta = 0, \|z\| = 1, z \in R^m, z = (z_1, z_2, \dots, z_m)' \right\},$$

则存在唯一的 $\eta_0 \in S$, 使 $\Delta(\eta_0)$ 最大, 且 $\lambda_2 \leq \Delta(\eta_0) < \lambda_1$, 等号只有当 $\eta_i' \eta = 0$ 时才可能取到.

注. 因为在求广义特征值与特征向量时可使

$$\begin{aligned} (\eta_1, \eta_2, \dots, \eta_m)' B (\eta_1, \eta_2, \dots, \eta_m) &= \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\}, \\ (\eta_1, \eta_2, \dots, \eta_m)' E (\eta_1, \eta_2, \dots, \eta_m) &= I \text{ (单位阵)}, \end{aligned}$$

所以

$$\Delta(\eta) \triangleq \eta' B \eta / \eta' E \eta = z' \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\} z.$$

证明. 记 $b = (b_1, b_2, \dots, b_m)', b_i = \eta_i' \eta_i, i = 1, 2, \dots, m$,

$$\max z' \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\} z,$$

$$\begin{cases} b' z = 0, \\ \|z\| = 1, \end{cases} \quad (1)$$

记 $G(z) = z' \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\} z + u(1 - z' z) + v b' z$, 运用 Lagrange 方法求极值

$$\frac{\partial G}{\partial z} = 2 \times \text{diag}\{\lambda_1 - u, \lambda_2 - u, \dots, \lambda_m - u\} z + v b = 0, \quad (2)$$

$$\frac{\partial G}{\partial u} = 1 - z' z = 0, \quad (3)$$

$$\frac{\partial G}{\partial v} = b' z = 0, \quad (4)$$

由(2)式得

$$z = 0.5 \times v \times \text{diag}\{(\lambda_1 - u)^{-1}, \dots, (\lambda_m - u)^{-1}\} b. \quad (5)$$

将(5)式代入(4)式得

$$F(u) = \sum_{i=1}^m \frac{b_i^2}{\lambda_i - u} = 0. \quad (6)$$

1) $b_2 \neq 0$ 时, 由引理得满足(6)式的唯一最大解 $u_0, \lambda_2 < u_0 < \lambda_1$.

$$\begin{aligned}
 \text{令} \quad d &= \left(\sum_{i=1}^m \left(\frac{b_i}{\lambda_i - u_0} \right)^2 \right)^{-0.5}, \\
 z_i &= \frac{b_i \cdot d}{\lambda_i - u_0}, \quad i = 1, 2, \dots, m. \\
 \eta_0 &= \sum_{i=1}^m z_i \eta_i \implies \eta_0 \in S, \\
 \Delta(\eta_0) &= z' \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\} z \\
 &= z' \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\} z + u_0(1 - z'z) \\
 &= z' \text{diag}\{\lambda_1 - u_0, \dots, \lambda_m - u_0\} z + u_0 \\
 &= d^2 F(u_0) + u_0 \\
 &= u_0
 \end{aligned} \tag{7}$$

由 u_0 的最大解及唯一性使得定理得证。

2) 若有某 $i_0 (2 < i_0 \leq m)$ 使得 $b_{i_0} \neq 0$, 而 $b_i = 0, i = 2, 3, \dots, i_0 - 1$. 仿 1) 证, 有满足(6)式的唯一最大解 $u_0, \lambda_{i_0} < u_0 < \lambda_1$. 由(7)式及 $\eta_2 \in S$ 得 $\Delta(\eta_0) \geq \Delta(\eta_2)$, 从而

$$u_0 \geq \lambda_2.$$

3) $b_i = 0$, 即 $\eta_i' \eta_i = 0$, 且有 $\eta_i \in S, i = 2, 3, \dots, m$.

因任一 $\eta \in S$, 均有 $\eta_i' \eta = 0 \implies \eta_i' \eta_1 \times z_1 = 0$, 从而 $z_1 = 0$.

又 $\Delta(\eta_2) = \lambda_2, \Delta(\eta) = \sum_{i=1}^m \lambda_i z_i^2 \leq \lambda_2 + (\lambda_1 - \lambda_2) \times z_1^2 = \lambda_2$, 取 $\eta_0 = \eta_2$, 定理得

证(这时 $\lambda_2 > \lambda_3$, 唯一性才成立).

若 $\det(E) = 0$, 那么在上述定理中用总离差阵 $E + B$ 代替 E , 其正确性不变, 且一对正交投影向量亦不变.

三、识别过程与准则

在用测试样本识别前, 将训练样本集的样本中心标准化预处理, 记均值向量

$$U = (u_1, u_2, \dots, u_m)',$$

标准差 $V = (v_1, v_2, \dots, v_m)'$; 再依上述定理求得一对最佳投影向量 η_1, η_0 及它们归一化的 Fisher 信息量 $\Delta(\eta_1), \Delta(\eta_0)$. 若待识样本 $X = (x_1, x_2, \dots, x_m)'$, 先进行

$$x'_i = \frac{(x_i - u_i)}{v_i}, \quad i = 1, 2, m,$$

它仍用 X 记; 再计算投影

$$z_1 = \eta_1' X, \quad z_2 = \eta_0' X. \tag{8}$$

若训练样本有 k 类, k 类均值经(8)式投影为 $(z_1^{(i)}, z_2^{(i)}), i = 1, 2, \dots, k$, 那么识别准则为

$$D_i = \Delta(\eta_1) \cdot |z_1 - z_1^{(i)}| + \Delta(\eta_0) \cdot |z_2 - z_2^{(i)}| \quad i = 1, 2, \dots, k. \tag{9}$$

若 $D_{i_0} = \min\{D_1, D_2, \dots, D_k\}$, 那么 X 层于 i_0 类. (9)式充分利用了 Fisher 信息

量.

为了验证定理与识别准则的有效性, 将它应用于 IRIS 数据的模式问题. 实验选自 R. A. Fisher 论文^[3]中所列的 IRIS 数据(见表 1).

表 1 Fisher 数据集

1 类样本					2 类样本					3 类样本				
No.	X1	X2	X3	X4	No.	X1	X2	X3	X4	No.	X1	X2	X3	X4
1	5.1	3.5	1.4	0.2	51	7.0	3.2	4.7	1.4	101	6.3	3.3	6.0	2.5
2	4.9	3.0	1.4	0.2	52	6.4	3.2	4.5	1.5	102	5.8	2.7	5.1	1.9
...
50	5.0	3.3	1.4	0.2	100	5.7	2.8	4.1	1.3	150	5.9	3.0	5.1	1.8

试验与文献[5]对照. 试验(1)从样本集中选取45个(每类各前15个)为训练样本集, 测试余下的105个样本, 识别率为 $101/105=96.19\%$ 与^[5]类同. 试验(2)是小样本集, 从样本集中选6个(每类各前2个)为训练样本集, 测试余下的144个, 识别率为 $138/144=95.83\%$; 这时文献[5]的识别率为86%.

参 考 文 献

- [1] 边肇祺, 模式识别, 清华大学出版社, 1988.
- [2] 方开泰, 实用多元统计分析, 华东师范大学出版社, 1989.
- [3] Fisher, R. A., The Use of Multiple Measurements in Taxonomic Problems, *Ann. Eugenics*, 7(1936), 178—188.
- [4] Sammon, J. W., Jr, An Optimal Discriminant Plane, *IEEE Trans. Comput.* 19(1970), 826—829.
- [5] 程永清等, 一种改进的 Fisher 判别准则, 计算机研究与发展, 28(1991), (6), 43—48.

OPTIMAL PROJECTIVE PLANE OF FISHER-RECOGNITION

WANG JINGFANG

(Computer Application Institute, Changling Refining & Chemical Works.
Yueyang. Hunan 414012, P. R. China)

Key words: Pattern recognition; information extraction; projective plane; discriminant vector.