フリーソフト R で学ぶ判別モデル [解説論文][*1]

Discriminant Models Using R

辻谷将明[*2]，田中祐輔[*3]，外山信夫[*4]

要約:フリーソフト R の飛躍的な発展とともに，判別モデルはマシンラーニングの枠組みでも益々脚光を浴びている。医療情報，画像処理，音声認識，テキストマイニングなど適用範囲も多岐に渡っている．本稿では従来の Fisher の判別分析や正準判別に留まらず，ロジスティック判別および非線形なカーネルロジスティック判別をも取り上げる．

1　2 群判別

2 つの群 (グループ)G_1，G_2 について，幾つかの変量に関する観測値がすでに与えられているとする．いま，新たに 1 つの観測値が得られたとき，このデータが G_1，G_2 のどちらの群に属するかを判別 (予測) したい．回帰分析で，目的変量が分類型になっているとみなせばよい [8, 15]．例えば，表 1.1 のような人工データを考える．胃潰瘍の患者グループ (第 1 群) と胃癌の患者グループ (第 2 群) に関する検査 A の結果である．

表 1.1　検査 A のデータ

	胃潰瘍 (第 1 群:◯ 印)	胃癌 (第 2 群:●印)
	0.1	0.8
	0.9	0.8
	0.2	0.7
	0.2	0.6
	0.6	0.7
	0.7	0.5
	0.3	0.8
	0.3	0.9
	0.3	0.9
	0.1	0.9
平均	0.37	0.76
分散	0.073	0.018

ここで，胃潰瘍か胃癌の疑いのある新患者について，検査 A に対する観測値が 0.6 とする．この患者が，胃潰瘍か胃癌かを判別したい．

[*1] 本解説論文は，[16] の第 4 章を大幅に加筆・修正したものである．なお，R の使い方については [17] に詳しい．

[*2] 大阪電気通信大学　情報通信工学部　情報工学科, 連絡先〒572-8530 寝屋川市初町 18-8;E-mail:ekaaf900@ricv.zaq.ne.jp

[*3] イーピエス株式会社 CRO 事業本部 臨床情報事業部 DS センター 統計解析 2 部，〒112-0004 大阪市淀川区宮原 3-4-30 ニッセイ新大阪ビル 11 階

[*4] 住宅金融支援機構 調査部，〒112-8570 東京都文京区後楽 1-4-10

2015 年 5 月 9 日受理

1.1　マハラノビスの汎距離

(1) 説明変数が 1 個の場合

新たに検査 A を受けた患者が，胃潰瘍か胃癌かを判別してみよう．第 1 群 (胃潰瘍) と第 2 群 (胃癌) に対する検査 A の観測値を 1 次元プロットすると図 1.1 のようになる．1 つの方法として，「第 1 群と第 2 群の平均

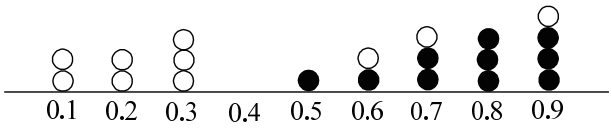

図 1.1　検査 A の 1 次元プロット

値に近いほうの群に属する」と判別する．第 1 群および第 2 群の検査 A の平均を区別するため，それぞれ右肩に (1)，(2) を付け

$$\bar{x}_A^{(1)} = 0.37, \qquad \bar{x}_A^{(2)} = 0.76 \tag{1.1}$$

と書く．これらの平均と新患者の検査値 $x_A = 0.6$ との距離は，それぞれ

$$\text{第 1 群}：|0.6 - 0.37| = 0.23,\ \text{第 2 群}：|0.6 - 0.76| = 0.16 \tag{1.2}$$

となり，第 2 群 (胃癌) に属すると判別できる．すなわち，2 点 A，B の距離を図 1.2 のように定義している．これは，われわれが日常使っている**ユークリッドの距離** (Euclidean distance) である．

図 1.2　線分 AB の距離

しかし，表 1.1 をみると第 1 群のデータは，第 2 群に比べてバラツキが大きい．分散は

$$\widehat{Var}\left(x_A^{(1)}\right) = 0.073,\ \widehat{Var}\left(x_A^{(2)}\right) = 0.018 \tag{1.3}$$

と推定される．明らかに，第 1 群のバラツキは第 2 群のそれの約 4 倍である．

第 1 群と第 2 群が正規分布に従うと仮定すると

$$x_A^{(1)} \sim N\left(0.37,\ 0.27^2\right), \quad x_A^{(2)} \sim N\left(0.76,\ 0.13^2\right) \tag{1.4}$$

と推定できる．バラツキを考慮するため，基準化

$$|\ \text{観測値} - \text{平均}\ |\,/\sqrt{\text{分散}} \tag{1.5}$$

を行うと，

$$\begin{cases} \text{第 1 群}：d_{(1)} = |\ \text{観測値} - \text{第 1 群の平均}\ |\,/\sqrt{\text{第 1 群の分散}} = |x - 0.37|\,/\sqrt{0.073} \\ \text{第 2 群}：d_{(2)} = |\ \text{観測値} - \text{第 2 群の平均}\ |\,/\sqrt{\text{第 2 群の分散}} = |x - 0.74|\,/\sqrt{0.018} \end{cases} \tag{1.6}$$

より，新患者について $d_{(1)} = 0.851 < 1.043 = d_{(2)}$ となる．よって，第 1 群 (胃潰瘍) に属すると判別する (ユークリッドの距離で測ると第 2 群に属した)．分散も考慮したこの距離を**マハラノビスの汎距離** (Mahalanobis's generalized distance) という．

(2) 説明変数が 2 個の場合

距離の考え方を説明変数が 2 個の場合へ拡張しよう．表 1.1 において，検査 A の他に検査 B の結果も得られたとする．それを表 1.2 に示す．

表 1.2　検査 A，B のデータ

	胃潰瘍 (第 1 群:◯ 印)		胃癌 (第 2 群:●印)	
	検査 A	検査 B	検査 A	検査 B
	0.1	0.4	0.8	0.3
	0.9	0.8	0.8	0.4
	0.2	0.8	0.7	0.6
	0.2	0.5	0.6	0.2
	0.6	0.8	0.7	0.4
	0.7	0.9	0.5	0.8
	0.3	0.8	0.8	0.6
	0.3	0.7	0.9	0.3
	0.3	0.5	0.9	0.4
	0.1	0.6	0.9	0.6
平均	0.37	0.68	0.76	0.46
分散	0.073	0.028	0.018	0.034

図 1.3 の横軸と縦軸は，それぞれ検査 A および B の値である．先ほどの新患者の検査 B の観測値は，$x_B = 0.5$ であった (図 1.3 の△印)．検査 B に対するマハラノビスの汎距離は

$$\begin{cases} \text{第 1 群}: d_{(1)} = |\,\text{観測値} - \text{第 1 群の平均}\,| \,/\sqrt{\text{第 1 群の分散}} = |x - 0.68| \,/\sqrt{0.028} \\ \text{第 2 群}: d_{(2)} = |\,\text{観測値} - \text{第 2 群の平均}\,| \,/\sqrt{\text{第 2 群の分散}} = |x - 0.46| \,/\sqrt{0.034} \end{cases} \tag{1.7}$$

となる．$d_{(1)} = 1.076 > 0.217 = d_{(2)}$ より，第 2 群 (胃癌) に属すると判別する．よって，検査 A，B それぞれについて，マハラノビスの汎距離を計算すると表 1.3 が得られる．この結果，検査 A では第 1 群 (胃潰瘍)，検査 B では第 2 群 (胃癌) と判別され，異なった結果となる．

1.2　線形判別関数

前節では，検査項目 A，B について個別に判別を行なうと，新患者が属する群が異なった．そこで，検査項目 A，B を同時に考慮した判別方式について考える．いま，第 1 群における検査 A と B との相関係数は，$\hat{\rho}^{(1)} = 0.666$，第 2 群のそれは $\hat{\rho}^{(2)} = 0.296$ と推定される．これらの相関係数は

$$\rho = \frac{Cov[x_A,\, x_B]}{\sqrt{Var[x_A]}\sqrt{Var[x_B]}} \tag{1.8}$$

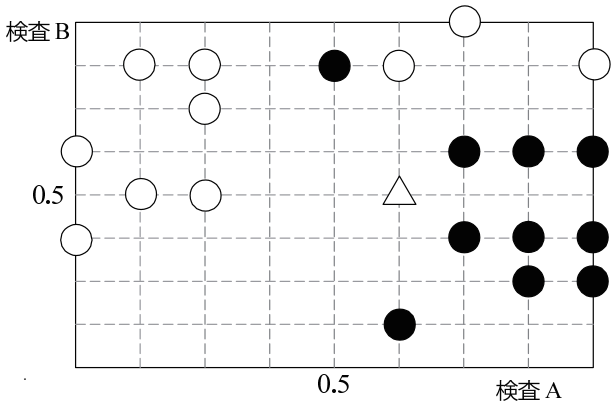

図 1.3　検査 A，B のプロット

表 1.3　検査 A，B それぞれのマハラノビスの汎距離

	第 1 群からの距離	第 2 群からの距離	判別
検査 A	$\frac{\lvert 0.6-0.37\rvert}{\sqrt{0.037}} = 0.851$	$\frac{\lvert 0.6-0.74\rvert}{\sqrt{0.018}} = 1.043$	第 1 群
検査 B	$\frac{\lvert 0.5-0.68\rvert}{\sqrt{0.028}} = 1.076$	$\frac{\lvert 0.5-0.46\rvert}{\sqrt{0.034}} = 0.217$	第 2 群

で与えられる．(1.8) 式の分子を x_A と x_B との**共分散** (covariance) と呼ぶ[*5]．この共分散を含めた分散共分散行列 $\mathbf{S}$ を

$$\mathbf{S} = \begin{bmatrix} \text{検査 } A \text{ の分散} & \text{検査 } A,\ B \text{ の共分散} \\ \text{検査 } A,\ B \text{ の共分散} & \text{検査 } B \text{ の分散} \end{bmatrix} \tag{1.9}$$

と定義する．

表 1.2 から，第 1 群における検査 A について

$$\hat{\mu}_A^{(1)} \equiv \bar{x}_A^{(1)} = 0.37, \hat{\sigma}_A^{(1)^2} \equiv \widehat{Var}[x_A] = 0.073 \tag{1.10}$$

となり，密度関数は

$$f\left(x_A^{(1)}\right) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_A^{(1)}} \exp\left\{ -\frac{\left(x_A^{(1)} - \hat{\mu}_A^{(1)}\right)^2}{2\hat{\sigma}_A^{(1)^2}} \right\} \tag{1.11}$$

と推定される．検査 B についても

$$\hat{\mu}_B^{(1)} \equiv \bar{x}_B^{(1)} = 0.68, \hat{\sigma}_B^{(1)^2} \equiv \widehat{Var}[x_B] = 0.018 \tag{1.12}$$

より，

$$f\left(x_B^{(1)}\right) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_B^{(1)}} \exp\left\{ -\frac{\left(x_B^{(1)} - \hat{\mu}_B^{(1)}\right)^2}{2\hat{\sigma}_B^{(1)^2}} \right\} \tag{1.13}$$

[*5] n 個のデータ $x_1, x_2, ..., x_n$ の平均と分散の推定量は $\bar{x} = \sum_{i=1}^{n} x_i/n$ および $\widehat{Var}(x) = \sum_{i=1}^{n} (x_i - \bar{x})^2 \Big/ (n-1)$ である．2 つの変数 x_1, x_2 のデータ $x_{11}, x_{21}, ..., x_{n1}$ および $x_{12}, x_{22}, ..., x_{n2}$ に対する共分散の推定量は $\widehat{Cov}(x_1, x_2) = \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)/(n-1)$ で与えられる．ここに，$\bar{x}_1 = \sum_{i=1}^{n} x_{i1}/n, \bar{x}_2 = \sum_{i=1}^{n} x_{i2}/n$ とする．

を得る．第 1 群の相関係数 $\hat{\rho}^{(1)} = 0.666$ より，$\left(x_A^{(1)},\, x_B^{(1)}\right)$ を同時に考慮した 2 変量正規分布は

$$
\begin{aligned}
f\left(x_A^{(1)},\, x_B^{(1)}\right) &= \frac{1}{2\pi\hat{\sigma}_A^{(1)}\hat{\sigma}_B^{(1)}} \times \\
&\exp\left[-\frac{1}{2\left(1-\hat{\rho}^{(1)2}\right)}\left\{\frac{\left(x_A-\hat{\mu}_A^{(1)}\right)^2}{\hat{\sigma}_A^{(1)2}} + \frac{\left(x_B-\hat{\mu}_B^{(1)}\right)^2}{\hat{\sigma}_B^{(1)2}} - \frac{2\hat{\rho}^{(1)}\left(x_A-\hat{\mu}_A^{(1)}\right)\left(x_B-\hat{\mu}_B^{(1)}\right)}{\hat{\sigma}_A^{(1)}\hat{\sigma}_B^{(1)}}\right\}\right] \\
&\equiv C_1 \exp\left[-d_{(1)}^2\right]
\end{aligned}
\tag{1.14}
$$

と推定される．ここに，

$$
C_1 = \frac{1}{2\pi\hat{\sigma}_A^{(1)}\hat{\sigma}_B^{(1)}} \tag{1.15}
$$

$$
d_{(1)}^2 = -\frac{1}{2\left(1-\hat{\rho}^{(1)2}\right)}\left\{\frac{\left(x_A-\hat{\mu}_A^{(1)}\right)^2}{\hat{\sigma}_A^{(1)2}} + \frac{\left(x_B-\hat{\mu}_B^{(1)}\right)^2}{\hat{\sigma}_B^{(1)2}} - \frac{2\hat{\rho}^{(1)}\left(x_A-\hat{\mu}_A^{(1)}\right)\left(x_B-\hat{\mu}_B^{(1)}\right)}{\hat{\sigma}_A^{(1)}\hat{\sigma}_B^{(1)}}\right\} \tag{1.16}
$$

とする．それを図示すると，図 1.4 が得られる．

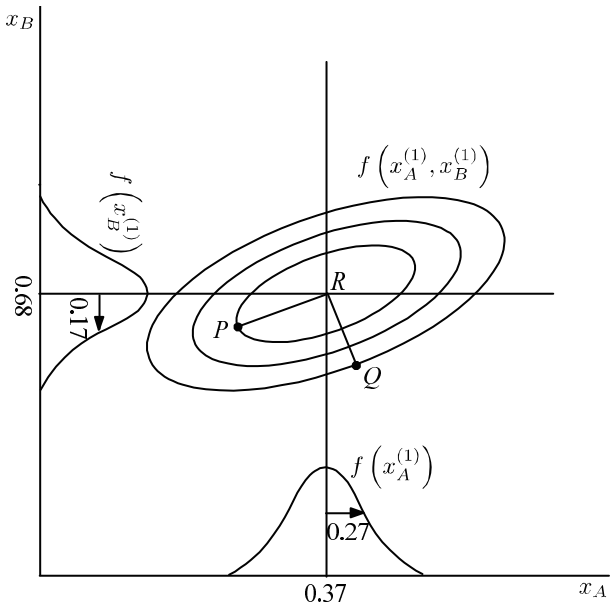

図 1.4　2 変量正規分布

検査 A, B を同時に考慮するためには，この 2 変量正規分布が必要になる．(1.16) 式の $d_{(1)}^2$ が一定 (すなわち，等確率) となる軌跡は楕円 (等確率楕円) で表される．この楕円は，平均 (重心)R= $\left(\bar{x}_A^{(1)}, \bar{x}_B^{(1)}\right) = (0.37,\ 0.68)$ に近いほど確率密度が高い等高線である．図 1.4 において，点 P，Q から重心 R までの直線距離 (ユークリッドの距離) は等しい．しかし，重心 R を山の頂上とみなし，楕円を段々畑とすれば，P は平面を歩けばよい．Q は段々畑の斜面を登るため，時間がかかり距離は遠くなる．すなわち，P はかなり高い確率楕円上にあるが，Q は低い確率楕円上にあり，それだけ重心 R から離れている．この勾配を考慮した距離が，マハラノビスの汎距離である (この 2 変量正規分布に基づく判別関数の基本原理については付録 A を参照されたい)．

新患者の観測値 $\mathbf{x}^t = (x_A,\, x_B)$ について，第 1 群の 2 変量の平均 (重心) までのマハラノビスの汎距離を計

算する．第 1 群の標本平均ベクトル $\bar{\mathbf{x}}^{(1)}$ は

$$\bar{\mathbf{x}}^{(1)} = \begin{bmatrix} \text{検査 } A \text{ の平均} \\ \text{検査 } B \text{ の平均} \end{bmatrix} = \begin{bmatrix} \bar{x}_A^{(1)} \\ \bar{x}_B^{(1)} \end{bmatrix} = \begin{bmatrix} 0.37 \\ 0.68 \end{bmatrix} \tag{1.17}$$

となる．次に，標本分散共分散行列 $\mathbf{S}^{(1)}$ を求めるための補助表 (表 1.4) を作成する．

表 1.4　標本分散共分散行列 $\mathbf{S}^{(1)}$ を求める補助表

No.	x_i(検査 A)	y_i(検査 B)	x_i^2	y_i^2	$x_i y_i$
1	0.1	0.4	0.01	0.16	0.04
2	0.9	0.8	0.81	0.64	0.72
.	.	.	.	.	.
.	.	.	.	.	.
10(=n)	0.1	0.6	0.01	0.36	0.06
計	A	B	C	D	E

このとき，$\mathbf{S}^{(1)}$ は

$$\mathbf{S}^{(1)} = \begin{bmatrix} \frac{1}{n-1}\left(C - \frac{A^2}{n}\right) & \frac{1}{n-1}\left(E - \frac{A\times B}{n}\right) \\ \frac{1}{n-1}\left(E - \frac{A\times B}{n}\right) & \frac{1}{n-1}\left(D - \frac{B^2}{n}\right) \end{bmatrix} = \begin{bmatrix} 0.0734 & 0.030 \\ 0.030 & 0.0284 \end{bmatrix}$$

と求められ，その逆行列は

$$\mathbf{S}^{(1)^{-1}} = \begin{bmatrix} 24.476 & -26.224 \\ -26.224 & 63.811 \end{bmatrix} \tag{1.18}$$

で与えられる. よって

$$\mathbf{x} = \begin{bmatrix} x_A \\ x_B \end{bmatrix}, \bar{\mathbf{x}}^{(1)} = \begin{bmatrix} \bar{x}_A^{(1)} \\ \bar{x}_B^{(1)} \end{bmatrix} = \begin{bmatrix} 0.37 \\ 0.68 \end{bmatrix}, \ \mathbf{S}^{(1)^{-1}} = \begin{bmatrix} 24.476 & -26.224 \\ -26.224 & 63.811 \end{bmatrix} \tag{1.19}$$

を用い，点 $\mathbf{x}$ と第 1 群の平均 $\bar{\mathbf{x}}^{(1)}$ とのマハラノビスの汎距離の 2 乗 (**マハラノビスの平方距離**と呼ぶ) を

$$\begin{aligned} D_{(1)}^2 &= \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right)^t \mathbf{S}^{(1)-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right) \\ &= \begin{bmatrix} x_A - 0.37,\ x_B - 0.68 \end{bmatrix} \begin{bmatrix} 24.0764 & -8.9464 \\ -8.9464 & 35.4684 \end{bmatrix} \begin{bmatrix} x_A - 0.37 \\ x_B - 0.68 \end{bmatrix} \end{aligned} \tag{1.20}$$

と定義する．$\mathbf{x}^t = (x_A,\ x_B) = (0.6,\ 0.5)$ を代入すると

$$D_{(1)}^2 = [0.23 \quad -0.18] \begin{bmatrix} 24.476 & -26.224 \\ -26.224 & 63.811 \end{bmatrix} \begin{bmatrix} 0.23 \\ -0.18 \end{bmatrix} = 5.534$$

となる[*6]．(1.14) 式と (1.20) 式との間には

$$f\left(x_A^{(1)},\ x_B^{(1)}\right) = C_1 \exp\left[-\frac{1}{2} D_{(1)}^2\right] \tag{1.21}$$

[*6] 1 変数 x_A のみの場合には，(1.5) 式より $D_{(1)}^2 = \left(\frac{\text{観測値-平均}}{\sqrt{\text{分数}}}\right)^2 = (x_A - 0.37)\, 0.037^{-1}\, (x_A - 0.37)$ と表現できる

という関係がある．同様に，点 $\mathbf{x}^t = (0.6,\ 0.5)$ と第 2 群の重心 $\bar{\mathbf{x}}^{(2)}$ とのマハラノビスの平方距離 $D^2_{(2)}$ を計算すると

$$D^2_{(2)} = [-0.16,\ 0.04] \begin{bmatrix} 60.391 & 12.433 \\ 12.433 & 31.972 \end{bmatrix} \begin{bmatrix} -0.16 \\ 0.04 \end{bmatrix} = 1.438 \tag{1.22}$$

となる．よって，$D^2_{(1)} > D^2_{(2)}$ より，新たな観測値は，第 2 群に入ると判別する．

さて，マハラノビスの平方距離による判別では，どの説明変数が判別に寄与しているかが分からない．そこで，2 つの群の母集団の分散共分散行列 $\mathbf{\Sigma}^{(1)}$ と $\mathbf{\Sigma}^{(2)}$ が等しい (すなわち，$\mathbf{\Sigma}^{(1)} = \mathbf{\Sigma}^{(2)}$) と仮定する．そして，線形式

$$\begin{aligned} z &= a_0 + a_1 x_A + a_2 x_B \\ &= a_0 + a_1 \times (\text{検査 } A \text{ の値}) + a_2 \times (\text{検査 } B \text{ の値}) \end{aligned} \tag{1.23}$$

を導き

$$z = \begin{cases} \geq 0 & : \text{第 1 群に属する} \\ < 0 & : \text{第 2 群に属する} \end{cases} \tag{1.24}$$

と判別するため，データから係数 a_0，a_1，a_2 を推定しよう．

$$\text{第 1 群 :} \bar{\mathbf{x}}^{(1)} = \begin{bmatrix} \bar{x}^{(1)}_A \\ \bar{x}^{(1)}_B \end{bmatrix} = \begin{bmatrix} 0.37 \\ 0.68 \end{bmatrix},\ \mathbf{S}^{(1)^{-1}} = \begin{bmatrix} 24.476 & -26.224 \\ -26.224 & 63.811 \end{bmatrix} \tag{1.25}$$

$$\text{第 2 群 :} \bar{\mathbf{x}}^{(2)} = \begin{bmatrix} \bar{x}^{(2)}_A \\ \bar{x}^{(2)}_B \end{bmatrix} = \begin{bmatrix} 0.76 \\ 0.46 \end{bmatrix},\ \mathbf{S}^{(2)^{-1}} = \begin{bmatrix} 60.391 & 12.433 \\ 12.433 & 31.972 \end{bmatrix} \tag{1.26}$$

であった．ここで，$\mathbf{\Sigma}^{(1)} = \mathbf{\Sigma}^{(2)}$ と仮定して，$\mathbf{S}^{(1)}$，$\mathbf{S}^{(2)}$ から 2 つの群の分散共分散をプールした共通の標本分散共分散行列

$$\begin{aligned} \mathbf{S} &= \frac{1}{(n_1 - 1) + (n_2 - 1)} \left\{ (n_1 - 1)\,\mathbf{S}^{(1)} + (n_2 - 1)\,\mathbf{S}^{(2)} \right\} \\ &= \begin{bmatrix} 0.04583 & 0.01156 \\ 0.01156 & 0.03111 \end{bmatrix} \end{aligned} \tag{1.27}$$

を求める．$\mathbf{S}^{(1)}$, $\mathbf{S}^{(2)}$ の代りにこの $\mathbf{S}$ を用いたマハラノビスの平方距離

$$\begin{aligned} D^2_{(k)} &= \left(\mathbf{x} - \bar{\mathbf{x}}^{(k)}\right)^t \mathbf{S}^{-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(k)}\right) \\ &= -\frac{1}{2} \left(\mathbf{x}^t \mathbf{S}^{-1} \mathbf{x} - \mathbf{x}^t \mathbf{S}^{-1} \bar{\mathbf{x}}^{(k)} - \bar{\mathbf{x}}^{(k)^t} \mathbf{S}^{-1} \mathbf{x} + \bar{\mathbf{x}}^{(k)^t} \mathbf{S}^{-1} \bar{\mathbf{x}}^{(k)} \right), k = 1, 2 \end{aligned}$$

において，$\mathbf{x}^t \mathbf{S}^{-1} \mathbf{x}$ は k に依存しない。また，$\mathbf{x}^t \mathbf{S}^{-1} \bar{\mathbf{x}}^{(k)} = \bar{\mathbf{x}}^{(k)^t} \mathbf{S}^{-1} \mathbf{x}$ より

$$D^2_{(k)} = \bar{\mathbf{x}}^{(k)^t} \mathbf{S}^{-1} \mathbf{x} - \frac{1}{2} \bar{\mathbf{x}}^{(k)^t} \mathbf{S}^{-1} \bar{\mathbf{x}}^{(k)}, k = 1, 2 \tag{1.28}$$

を得る．(1.28) 式は $\mathbf{x}$ に関する線形関数で

$$\begin{cases} \text{傾き} = \bar{\mathbf{x}}^{(k)^t} \mathbf{S}^{-1} \\ \text{定数項} = -\frac{1}{2} \bar{\mathbf{x}}^{(k)^t} \mathbf{S}^{-1}\, \bar{\mathbf{x}}^{(k)} \end{cases}$$

となる．表 1.2 のデータについて

$$\begin{aligned} D^2_{(1)} &= \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right)^t \mathbf{S}^{-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right) \\ &= \begin{bmatrix} 0.23 & -0.18 \end{bmatrix} \begin{bmatrix} 24.0764 & -8.9464 \\ -8.9464 & 35.4684 \end{bmatrix} \begin{bmatrix} 0.23 \\ -0.18 \end{bmatrix} = 3.1636 \end{aligned} \tag{1.29}$$

$$D^2_{(2)} = \left(\mathbf{x} - \bar{\mathbf{x}}^{(2)}\right)^t \mathbf{S}^{-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(2)}\right) \tag{1.30}$$
$$= \begin{bmatrix} -0.16 & 0.04 \end{bmatrix} \begin{bmatrix} 24.0764 & -8.9464 \\ -8.9464 & 35.4684 \end{bmatrix} \begin{bmatrix} -0.16 \\ 0.04 \end{bmatrix} = 0.7804$$

を計算すると，$D^2_{(1)} > D^2_{(2)}$ であるから，第 2 群に属すると判別する．

新患者の検査値 $\mathbf{x}^t = (x_A,\, x_B)$ が得られたとき，これが G_1 に属すると判別されるのは，$D^2_{(1)} < D^2_{(2)}$(すなわち，$D^2_{(2)} - D^2_{(1)} > 0$) の場合である．ここで，

$$D^2_{(2)} - D^2_{(1)} = \left(\mathbf{x} - \bar{\mathbf{x}}^{(2)}\right)^t \mathbf{S}^{-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(2)}\right) - \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right)^t \mathbf{S}^{-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right)$$
$$= 2\left\{\left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right)^t \mathbf{S}^{-1}\mathbf{x} - \frac{1}{2}\left(\bar{\mathbf{x}}^{(1)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)}\right)\right\} \tag{1.31}$$

より[*7]，係数の 2 は正負の判別に無関係であるから

$$z = f\,(x_A,\, x_B) \equiv \frac{D^2_{(2)} - D^{(2)}_{(1)}}{2} = \left\{\mathbf{x} - \left(\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}\right)/2\right\}^t \mathbf{S}^{-1} \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right) \tag{1.32}$$
$$= -0.019 - 11.355x_A + 11.289x_B$$

となる[*8]．この (1.32) 式は **Fisher の線形判別関数**と呼ばれ，

$$f\,(x_A,\, x_B) = \begin{cases} \geq 0 & : \text{第 1 群に属する} \\ < 0 & : \text{第 2 群に属する} \end{cases} \tag{1.33}$$

と判別する．新患者の観測値 $(x_A,\, x_B) = (0.6,\, 0.5)$ を代入すると $f\,(0.6,\, 0.5) = -1.19$ より，第 2 群に属する．

一般に，表 1.5 のように I 個の説明変数 $(x_1,\, x_2,\, ...,\, x_I)$ をもつ 2 つの群から標本が得られているとする．平均 (重心) を

$$\bar{\mathbf{x}}^{(1)} = \left(\bar{x}^{(1)}_1, \bar{x}^{(1)}_2, \cdots, \bar{x}^{(1)}_I\right)^t, \bar{\mathbf{x}}^{(2)} = \left(\bar{x}^{(2)}_1, \bar{x}^{(2)}_2, \cdots, \bar{x}^{(2)}_I\right)^t \tag{1.34}$$

とし，2 つの群を併合した標本分散共分散行列を

$$\mathbf{S} = (s_{jj'}),\; s_{jj'} = \frac{1}{n_1 + n_2 - 2} \sum_{k=1}^{2} \sum_{i=1}^{n_k} \left(x^{(k)}_{ji} - \bar{x}^{(k)}_j\right)^2; j,\; j' = 1, \cdots,\; I \tag{1.35}$$

とする．

---

*7

$\left(\mathbf{x} - \bar{\mathbf{x}}^{(2)}\right)^t \mathbf{S}^{-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(2)}\right) - \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right)^t \mathbf{S}^{-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right)$
$= \left(\mathbf{x} - \bar{\mathbf{x}}^{(2)}\right)^t \left(\mathbf{S}^{-1}\mathbf{x} - \mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)}\right) - \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right)^t \left(\mathbf{S}^{-1}\mathbf{x} - \mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)}\right)$
$= \mathbf{x}^t\mathbf{S}^{-1}\mathbf{x} - \mathbf{x}^t\mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)} - \mathbf{x}^t\mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)} + \bar{\mathbf{x}}^{(2)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)} - \left(\mathbf{x}^t\mathbf{S}^{-1}\mathbf{x} - \mathbf{x}^t\mathbf{S}^{-1}\mathbf{x}^{(1)} - \mathbf{x}^t\mathbf{S}^{-1}\mathbf{x}^{(1)} + \bar{\mathbf{x}}^{(1)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)}\right)$
$= -\mathbf{x}^t\mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)} - \mathbf{x}^t\mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)} + \bar{\mathbf{x}}^{(2)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)} + \mathbf{x}^t\mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)} + \mathbf{x}^t\mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(1)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)}$
$= 2\left\{\left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right)^t \mathbf{S}^{-1}\mathbf{x} - \frac{1}{2}\left(\bar{\mathbf{x}}^{(1)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)}\right)\right\}$

となる．2 次形式 $\mathbf{x}^t\mathbf{S}^{-1}\mathbf{x}$ がキャンセルされていることに留意されたい．

*8

$\frac{D^2_{(2)} - D^{(2)}_1}{2} = \left(\mathbf{x} - \frac{\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}}{2}\right)^t \mathbf{S}^{-1} \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right)$
$= \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right)^t \mathbf{S}^{-1}\mathbf{x} - \frac{\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}}{2}\mathbf{S}^{-1} \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right)$
$= \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right)^t \mathbf{S}^{-1}\mathbf{x} - \left(\frac{\bar{\mathbf{x}}^{(1)^t}}{2}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)} - \frac{\bar{\mathbf{x}}^{(2)^t}}{2}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)} + \frac{\bar{\mathbf{x}}^{(2)^t}}{2}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)} - \frac{\bar{\mathbf{x}}^{(2)^t}}{2}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)}\right)$
$= \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right)^t \mathbf{S}^{-1}\mathbf{x} - \frac{1}{2}\left(\bar{\mathbf{x}}^{(1)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(2)}\right)$

表 1.5　2 群判別の一般型

	$x_1^{(1)}$	$x_2^{(1)}$	$\cdot$	$x_I^{(1)}$		$x_1^{(2)}$	$x_2^{(2)}$	$\cdot$	$x_I^{(2)}$
1	$x_{11}^{(1)}$	$x_{12}^{(1)}$	$\cdot$	$x_{1I}^{(1)}$	1	$x_{11}^{(2)}$	$x_{12}^{(2)}$	$\cdot$	$x_{1I}^{(2)}$
2	$x_{21}^{(1)}$	$x_{22}^{(1)}$	$\cdot$	$x_{2I}^{(1)}$	2	$x_{21}^{(2)}$	$x_{22}^{(2)}$	$\cdot$	$x_{2I}^{(2)}$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	
n_1	$x_{n_1 1}^{(1)}$	$x_{n_1 2}^{(1)}$	$\cdot$	$x_{n_1 I}^{(1)}$	n_2	$x_{n_2 1}^{(2)}$	$x_{n_2 2}^{(2)}$	$\cdot$	$x_{n_2 I}^{(2)}$
平均	$\bar{x}_1^{(1)}$	$\bar{x}_2^{(1)}$	$\cdot$	$\bar{x}_I^{(1)}$	平均	$\bar{x}_1^{(2)}$	$\bar{x}_2^{(2)}$	$\cdot$	$\bar{x}_I^{(2)}$

このとき，マハラノビスの平方距離

$$D_{(k)}^2 = \left(\mathbf{x} - \bar{\mathbf{x}}^{(k)}\right)^t \mathbf{S}^{-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(k)}\right),\ k = 1,\ 2 \tag{1.36}$$

を計算し，

$$D_{(2)}^2 - D_{(1)}^2 = \begin{cases} \geq 0 : \text{第 1 群} \\ < 0 : \text{第 2 群} \end{cases} \tag{1.37}$$

と判別する．あるいは，Fisher の線形判別関数

$$\begin{aligned} z &= \left\{\mathbf{x} - \left(\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}\right)/2\right\}^t \mathbf{S}^{-1} \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right) \\ &= a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_I x_I \end{aligned} \tag{1.38}$$

を計算し，

$$z = \begin{cases} \geq 0 & : \text{第 1 群に属する} \\ < 0 & : \text{第 2 群に属する} \end{cases} \tag{1.39}$$

としてもよい．なお，$\mathbf{S}$ の逆行列 $\mathbf{S}^{-1}$ が存在するためには，$(n_1 + n_2 - 2) > I$ でなければならない．境界線は

$$0 = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_I x_I \tag{1.40}$$

で与えられる．この $a_0,\ a_1,\ a_2, ...,\ a_I$ が定まると 2 つの群の観測値から，**判別スコア** (discriminant score)

$$z = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_I x_I,\ i = 1,\ 2, ...,\ n_k,\ k = 1,\ 2 \tag{1.41}$$

を計算する．

(1.38) 式で $\mathbf{S}^{-1}\left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right) = \mathbf{a} \equiv [a_1,\ a_2, ...,\ a_I]^t$ とおくと

$$z = \left(\mathbf{x} - \frac{\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}}{2}\right)^t \mathbf{a} \tag{1.42}$$

と書ける．よって，

$$z = a_1 \left(x_1 - \frac{\bar{x}_1^{(1)} + \bar{x}_1^{(2)}}{2}\right) + a_2 \left(x_2 - \frac{\bar{x}_2^{(1)} + \bar{x}_2^{(2)}}{2}\right) + \cdots + a_I \left(x_I - \frac{\bar{x}_I^{(1)} + \bar{x}_I^{(2)}}{2}\right) \tag{1.43}$$

を得る．(1.41) 式と (1.43) 式を比べると，a_0 は $(a_1,\ a_2, ...,\ a_I)$ を用い

$$\begin{aligned} a_0 &= -a_1 \frac{\bar{x}_1^{(1)} + \bar{x}_1^{(2)}}{2} - a_2 \frac{\bar{x}_2^{(1)} + \bar{x}_2^{(2)}}{2} - \cdots - a_I \frac{\bar{x}_I^{(1)} + \bar{x}_I^{(2)}}{2} \\ &= -\frac{\left(a_1 \bar{x}_1^{(1)} + a_2 \bar{x}_2^{(1)} + \cdots + a_I \bar{x}_I^{(1)}\right) + \left(a_1 \bar{x}_1^{(2)} + a_2 \bar{x}_2^{(2)} + \cdots + a_I \bar{x}_I^{(2)}\right)}{2} \end{aligned} \tag{1.44}$$

から計算される．a_0 は，両群の平均 $\left(\bar{x}_1^{(1)},\ \bar{x}_2^{(1)}\right)$，$\left(\bar{x}_1^{(2)},\ \bar{x}_2^{(2)}\right)$ に対する判別スコアの中点にマイナスを付けた量である．

表 1.2 のデータについて，Fisher の線形判別関数

$$z = -0.019 - 11.355x_A + 11.289x_B \tag{1.45}$$

が求まる．ただし，a_0 は (1.44) 式を用い

$$a_0 = -\frac{-11.355 \times 0.37 + 11.289 \times 0.68 - 11.355 \times 0.76 + 11.289 \times 0.46}{2} = -0.019$$

となる．図 1.3 にこの線形判別関数の境界線

$$0 = 0.019 - 11.355x_A + 11.289x_B \tag{1.46}$$

を記入すると図 1.5 のようになる．

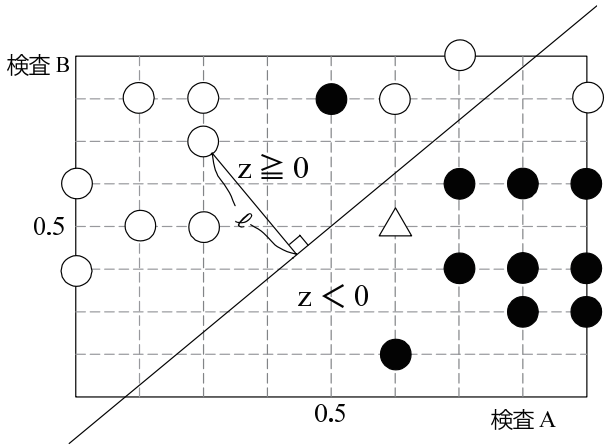

図 1.5　線形判別関数

手計算で (1.45) 式を算出してみよう．プログラムでは，データを

```
# ―――――――――――――― プログラム#(1.1) ――――――――――――――
A1<-c(0.1,0.9,0.2,0.2,0.6,0.7,0.3,0.3,0.3,0.1) #検査A(胃潰瘍)
B1<-c(0.4,0.8,0.8,0.5,0.8,0.9,0.8,0.7,0.5,0.6) #検査B(胃潰瘍)
A2<-c(0.8,0.8,0.7,0.6,0.7,0.5,0.8,0.9,0.9,0.9) #検査A(胃癌)
B2<-c(0.3,0.4,0.6,0.2,0.4,0.8,0.6,0.3,0.4,0.6) #検査B(胃癌)
```

と入力する．(1.25)，(1.26) 式の平均，および (1.27) 式の $\mathbf{S}^{(1)}$, $\mathbf{S}^{(2)}$, $\mathbf{S}$ を算出するには

```
# ―――――――――――――― プログラム#(1.2) ――――――――――――――
v1<-data.frame(A1,B1)
v2<-data.frame(A2,B2)
#データ数
n1<-nrow(v1)
n2<-nrow(v2)
#列平均 (ex:行平均->rowMeans)
vm1=colMeans(v1)                    # 第1群の2個の説明変数の平均
vm2=colMeans(v2)                    # 第2群の2個の説明変数の平均
s1=var(v1)                          # 行列S(1)
s2=var(v2)                          # 行列S(2)
s=((n1-1)*s1+(n2-1)*s2)/(n1+n2-2)   # 行列S
vm1
```

```
vm2
s1
s2
s
```

と続けると

```
> vm1
   A1    B1
0.37  0.68
> vm2
   A2    B2
0.76  0.46
> s1
             A1           B1
A1  0.07344444  0.03044444
B1  0.03044444  0.02844444
> s2
              A2             B2
A2   0.018222222  -0.007333333
B2  -0.007333333   0.033777778
> s
             A1           B1
A1  0.04583333  0.01155556
B1  0.01155556  0.03111111
```

と推定される. さらに

```
coe=solve(s) %*% (vm1-vm2)                # a1, a2の推定値
coe
```

から，a_1 および a_2 の推定値

```
         [,1]
A1  -11.35532
B1   11.28912
```

が求まり，a_0 の推定値は

```
vm=rbind(vm1,vm2)
z=-0.5 * sum(vm %*% coe)                  # a0の推定値
z
```

から

```
[1] -0.01904302
```

となる.

次に，R の関数 {lda} を用いて Fisher の線形判別関数を導く．プログラムでは、まず入力データ

```
# ---------- プログラム#(1.3) ----------
検査A <- c(0.1,0.9,0.2,0.2,0.6,0.7,0.3,0.3,0.3,0.1,0.8,0.8,0.7,0.6,0.7,0.5,0.8,0.9,0.9,0.9)
検査B <- c(0.4,0.8,0.8,0.5,0.8,0.9,0.8,0.7,0.5,0.6,0.3,0.4,0.6,0.2,0.4,0.8,0.6,0.3,0.4,0.6)
疾病 <- factor(c(rep("胃潰瘍", 10), rep("胃癌", 10)),levels=c("胃潰瘍","胃癌"))
train <- data.frame(検査A, 検査B, 疾病)
```

を作成する．プログラム#(1.3) に続けて

```
# ---------- プログラム#(1.4) ----------
library(MASS)
kfit <- lda(疾病~検査A+検査B,data=train)
```

```
print(kfit)
apply(-kfit$means%*%kfit$scaling,2,mean)
# 見掛け上の誤判別表
table(train$疾病,predict(kfit)$class)
# 1例消去CV
kfit <- lda(疾病~検査A+検査B,data=train,CV=TRUE)
table(train$疾病,kfit$class)
```

と入力すると

```
Prior probabilities of groups:
胃潰瘍   胃癌
  0.5    0.5
Group means:
       検査A 検査B
胃潰瘍  0.37  0.68
胃癌    0.76  0.46
Coefficients of linear discriminants: # 判別関数の傾き
           LD1
検査A  4.319085
検査B -4.293906
> apply(-kfit$means%*%kfit$scaling,2,mean)
       LD1
0.007243166                          # 判別関数の定数項
```

を得る．このアウトプットから，Fisher の判別関数の境界線は

$$0 = 0.0072 + 4.3191x_A - 4.2939x_B \tag{1.47}$$

となる．(1.47) 式は，マハラノビスの平方距離に基づいて手計算から得られた (1.45) 式とは異なる．しかし、両式とも

$$x_B = 1.0058638x_A + 0.00168$$

となる．これは，プログラム #(1.4) の判別関数は固有方程式から求められており，固有ベクトルが一意に決まらないことに起因する ([18] の p.74 を参照)．なお，相関比に基づく線形判別関数は付録 B に詳しい．

(1.36) 式のマハラノビスの平方距離に基づく Fisher の線形判別関数のプログラムは，R の関数 LdaClassic を採用する．#(1.3) に続けて

```
#------------------ プログラム#(1.5) -------------------
library(rrcov)
#当てはめ
kfit <- LdaClassic(疾病 ~ 検査A + 検査B, data=train)
summary(kfit)
```

と入力すると

```
Prior Probabilities of Groups:
胃潰瘍   胃癌
  0.5    0.5
Group means:
       検査A 検査B
胃潰瘍  0.37  0.68
胃癌    0.76  0.46
Within-groups Covariance Matrix:
           検査A       検査B
検査A 0.04583333 0.01155556
検査B 0.01155556 0.03111111
Linear Coeficients:
```

```
              検査A       検査B
胃潰瘍   2.826793  20.807191
胃癌    14.182110   9.518073
Constants:
    胃潰瘍        胃癌
-8.290549  -8.271506
```

となる．このアウトプットの Constatnts と Linear Coefficients の値から

$$\begin{aligned} z &= -8.291 - (-8.272) + (2.827 - 14.182)x_A + (20.807 - 9.518)x_B \\ &= -0.019 - 11.355x_A + 11.289x_B \end{aligned}$$

を得，(1.45) 式の Fisher の判別関数が求まる．また，プログラム#(1.5) に続けて

```
# ―――――――――― プログラム#(1.6) ――――――――――
#判別境界
x1.new <- seq(0, 1, length=200)              #x1
coef.diff <- fit.lda@ldf[1, ] - fit.lda@ldf[2, ]        #係数の差
const.diff <- fit.lda@ldfconst[1] - fit.lda@ldfconst[2] #切片の差
x2.new <- -(coef.diff[1] * x1.new + const.diff)/(coef.diff[2]) #x2
#プロット
op <- par(cex.lab=1.3)
plot(train$検査A[疾病=="胃潰瘍"], train$検査B[疾病=="胃潰瘍"],
        xlab="検査A", ylab="検査B", xlim=c(0, 1), ylim=c(0, 1),
        pch=1, col="black", cex=2)
points(train$検査A[疾病=="胃癌"], train$検査B[疾病=="胃癌"],
        pch=16, col="black", cex=2)
lines(x1.new, x2.new, lty=2, lwd=3, col="black")            #判別境界
legend("bottomright", c("胃潰瘍", "胃癌"), col=c("black", "black"),
        pch=c(1, 16), cex=2)
grid()
par(op)
```

と入力すれば図 1.6 が描かれる．

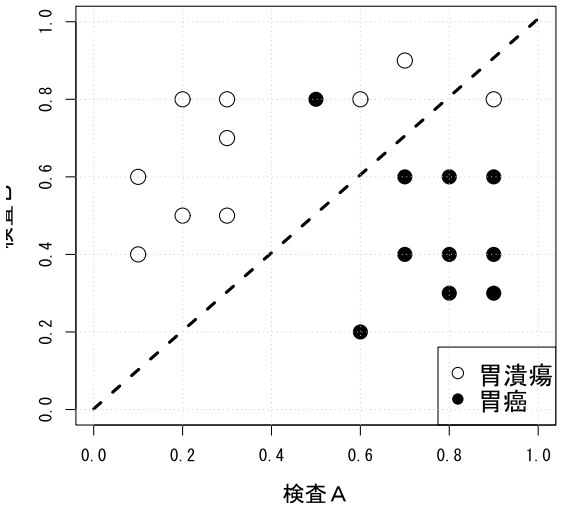

図 1.6　Fisher の線形判別関数

表 1.2 のデータについて，判別スコアを算出すると表 1.6 が得られる．

表 1.6　線形判別スコア

	第 1 群 (白丸)			第 2 群 (黒丸)		
i	x_1	x_2	$z_i^{(1)}$	x_1	x_2	$z_i^{(2)}$
1	0.1	0.4	3.361	0.8	0.3	-5.717
2	0.9	0.8	-1.208 ※	0.8	0.4	-4.588
3	0.2	0.8	6.741	0.7	0.6	-1.194
4	0.2	0.5	3.354	0.6	0.2	-4.574
5	0.6	0.8	2.199	0.7	0.4	-3.452
6	0.7	0.9	2.192	0.5	0.8	3.335 ※
7	0.3	0.8	5.606	0.8	0.6	-2.33
8	0.3	0.7	4.477	0.9	0.3	-6.852
9	0.3	0.5	2.219	0.9	0.4	-5.723
10	0.1	0.6	5.619	0.9	0.6	-3.465

例えば，第 1 群の観測値 (0.1, 0.4) について，判別スコアは (1.45) 式から

$$z = -0.019 - 11.355 \times 0.1 + 11.289 \times 0.4 = 3.361 \tag{1.48}$$

となる．各スコアに関して

$$z_i^{(k)} \begin{cases} \geq 0 : \text{第 1 群に属する} \\ < 0 : \text{第 2 群に属する} \end{cases}$$

と判別すると，白丸 1 個，黒丸 1 個が誤判別 (表 1.5 の※印) される．この判別スコア $z_i^{(k)}$ と境界線 (1.46) 式との関係を調べてみよう．任意の点 (x_A, x_B) と境界線 (1.46) 式との距離は，ヘッセの公式より

$$\ell = \frac{|-0.019 - 11.355x_A + 11.289x_B|}{\sqrt{(-11.355)^2 + 11.289^2}} \tag{1.49}$$

で与えられる．よって，判別スコアの絶対値は，この距離の $\sqrt{(-11.355)^2 + 11.289^2}$ 倍になっている．例えば，第 1 群の 8 番目の観測値 (0.3, 0.7) と境界線 (1.46) 式との距離 (図 1.5 の ℓ) は

$$\frac{|-0.019 - 11.355 \times 0.3 + 11.289 \times 0.7|}{\sqrt{(-11.355)^2 + 11.289^2}} = 5.309 \tag{1.50}$$

となる．この判別スコアの変動に基づいて線形判別関数を導出することもできる (詳細は，付録 B を参照されたい)．

2 つの群に属する**事前確率**が等しいとき，それが第 k 群に属するベイズの**事後確率**は

$$\Pr(k|\mathbf{x}) = \frac{\exp\left(-D_{(k)}^2/2\right)}{\exp\left(-D_{(1)}^2/2\right) + \exp\left(-D_{(2)}^2/2\right)}, k = 1, 2 \tag{1.51}$$

で与えられる (詳細は，付録 A を参照されたい)．よって，$D_{(k)}^2$ が最小になる k は，ベイズの事後確率が最大になる k と同一である．例えば，第 1 群の観測値 (0.1,0.4) について，マハラノビスの平方距離

$$D_{(1)}^2 = \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right) \mathbf{S}^{-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(1)}\right) \tag{1.52}$$

を算出する．

$$\mathbf{x}=\begin{bmatrix}0.1\\0.4\end{bmatrix},\quad \bar{\mathbf{x}}^{(1)}=\begin{bmatrix}0.37\\0.68\end{bmatrix},\bar{\mathbf{S}}=\begin{bmatrix}0.046 & 0.012\\0.012 & 0.031\end{bmatrix} \tag{1.53}$$

とおけば

$$D^2_{(1)}=3.183,\ \exp\left(-D^2_{(1)}\Big/2\right)=0.204 \tag{1.54}$$

となる．同様に

$$D^2_{(2)}=9.905,\ \exp\left(-D^2_{(2)}\Big/2\right)=0.007 \tag{1.55}$$

を得る．よって，ベイズの事後確率は，(1.51) 式から

$$\begin{cases}\Pr(1\,|\mathbf{x})=0.204/(0.204+0.007)=0.967\\ \Pr(2\,|\mathbf{x})=0.007/(0.204+0.007)=0.033\end{cases} \tag{1.56}$$

と求まり，$P(1\,|\mathbf{x})>P(2\,|\mathbf{x})$ より第 1 群に属する．

プログラム #(1.4) に続けて

```
# ―――――――――― プログラム#(1.7) ――――――――――
#  ベイズの事後確率
kfit <- lda(疾病~検査A+検査B,data=train)
print(predict(kfit,train)$posterior)
```

と入力すれば， ベイズの事後確率

```
            1           2
1  0.966465550 0.033534450
2  0.230137654 0.769862346
3  0.998820150 0.001179850
4  0.966250335 0.033749665
5  0.900165155 0.099834845
6  0.899568659 0.100431341
7  0.996336463 0.003663537
8  0.988757460 0.011242540
9  0.901935780 0.098064220
10 0.996384474 0.003615526
11 0.003280213 0.996719787
12 0.010074228 0.989925772
13 0.232491791 0.767508209
14 0.010207125 0.989792875
15 0.030705765 0.969294235
16 0.965596664 0.034403336
17 0.088682728 0.911317272
18 0.001056124 0.998943876
19 0.003258641 0.996741359
20 0.030314147 0.969685853
```

が得られる．R プログラムを採用した線形判別分析は [18] の 4 章に詳しい.

線形判別分析では，与えられたデータを直線で分けるため，間違って判別されることもある．すなわち，G_1 に属しているのに G_2 と誤って判別したり，逆に G_2 に属しているのに G_1 と誤って判別することがある．この誤判別率は

i) 見掛け上の誤判別

ii)1 例消去クロスバリデーション法

などにより推定される．

i) 見掛け上の誤判別

表 1.7　誤判別表

	第 1 群 (G_1)	第 2 群 (G_2)
群のサンプル数	n_1	n_2
誤判別個数	m_1	m_2
誤判別率	$\hat{P}_1 = m_1/n_1$	$\hat{P}_2 = m_2/n_2$

誤判別個数および誤判別率を表 1.7 のように定義する．表 1.2 のデータについて，判別を行うと表 1.8 のようになる．よって $\hat{P}_1 = 1/10 = 0.1$，$\hat{P}_2 = 1/10 = 0.1$ を得る．$\hat{P}_1$ と $\hat{P}_2$ は見掛け上の誤判別率 (apparent error rate) と呼ばれている．これは，同一のデータから判別式と誤判別率の両方を算出しているため，誤判別率を過小評価する傾向がある．判別分析の本来の目的は，手元にあるデータから，新たに得られた観測値を正しく判別 (予測) することにある．判別関数を求めたデータを再利用して判別を行い，誤判別率を推定する点に問題がある．

表 1.8　表 1.2 のデータに対する見掛け上の誤判別表 (線形判別)

もとの群＼予測群	第 1 群	第 2 群
第 1 群	9	1
第 2 群	1	9

プログラム #(1.7) に

```
# ―――――――――― プログラム#(1.8) ――――――――――
table(train$疾病 ,predict(kfit)$class)
```

と続ければ，表 1.8 の誤判別表

```
   1 2
1 9 1
2 1 9
```

が得られる．

***ii)* 1 例消去クロスバリデーション法**

見掛け上の誤判別の欠点を回避するため，計算量は膨大になるが，有効な方法として 1 例消去 (leaving-one-out) クロスバリデーション (CV:Cross-Validation) 法がある．具体的には，第 1 群の n_1 個から 1 個の観測値を除去した残りの $n_1 - 1$ 個，および第 2 群の n_2 個の計 $n_1 - 1 + n_2$ 個に基づいて判別関数を推定し，除去した観測値の判別を行う．これを，n_1 個のすべてのデータについて判別を行い，誤判別個数 m_1^* を数える．同様に，第 2 群からの n_2 個のデータについても誤判別個数 m_2^* を計算する．そして，誤判別率を

$$\hat{P}_1^* = m_1^*/n_1, \hat{P}_2^* = m_2^*/n_1 \tag{1.57}$$

から推定する．この推定値の偏りはほとんどないことが検証されている．表 1.2 のデータについて，

$$\hat{P}_1^* = 1/10 = 0.1, \hat{P}_2^* = 1/10 = 0.1 \tag{1.58}$$

を得る．

プログラム #(1.8) に

```
# ———————— プログラム#(1.9) ————————
kfit <- lda(疾病~検査A+検査B,data=train,CV=TRUE)
table(train$疾病,kfit$class)
```

と続ければ，1 例消去クロスバリデーションによる誤判別表

```
    1 2
  1 9 1
  2 1 9
```

が得られる．表 1.2 の場合，見かけ上および 1 例消去クロスバリデーションによる誤判別表は同一になる．

1.3 検定

(1) 分散共分散行列 $\boldsymbol{\Sigma}$ の同等性に関する Bartlett の検定

2 つの群の母集団の分散共分散行列 $\boldsymbol{\Sigma}^{(1)}$ と $\boldsymbol{\Sigma}^{(2)}$ が等しいと仮定し，(1.27) 式の共通の標本分散共分散行列 $\mathbf{S}$ を計算した．実際のデータに適用する際，それらが等しいか否かを検定しなければならない．そのため，

$$\text{帰無仮説 } H_0 : \boldsymbol{\Sigma}^{(1)} = \boldsymbol{\Sigma}^{(2)},\ \text{対立仮説 } H_1 : \boldsymbol{\Sigma}^{(1)} \neq \boldsymbol{\Sigma}^{(2)}$$

に関して，検定統計量

$$\chi_0^2 = \left[1 - \left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1} - \frac{1}{n_1 + n_2 - 2}\right) \frac{\left(2I^2 + 3I - 1\right)}{6\left(I + 1\right)}\right] \times \ln\left\{\frac{|\mathbf{S}|^{n_1+n_2-2}}{\left|\mathbf{S}^{(1)}\right|^{n_1-1} \left|\mathbf{S}^{(2)}\right|^{n_2-1}}\right\} \tag{1.59}$$

を計算する．H_0 のもとで，χ_0^2 は自由度 $I\left(I+1\right)/2$ のカイ二乗分布に従うから，

$$\chi_0^2 \geq \chi^2\left(I\left(I+1\right)/2, \alpha\right) \tag{1.60}$$

なら帰無仮説 H_0 を棄却する．

表 1.2 のデータについて，(1.60) 式の値を計算すると

$$\chi_0^2 = 7.42 < \chi^2\left(3, 0.05\right) = 7.81$$

を得，分散共分散行列は等しいとみなせる．

プログラムは，#(1.3) に続けて

```
# ———————— プログラム#(1.10) ————————
cov1 <- cov(data[train$疾病=="胃潰瘍",1:2])
cov2 <- cov(data[train$疾病=="胃癌",1:2])
n1 <- length(train$疾病[train$疾病=="胃潰瘍"])
n2 <- length(train$疾病[train$疾病=="胃癌"])
I <- nrow(cov1)
cov.pool <- ((n1-1)*cov1+(n2-1)*cov2)/(n1+n2-2)
fact1 <- 1/(n1-1) + 1/(n2-1) - 1/(n1+n2-1)
fact2 <- (2*I^2+3*I-1)/(6*(I+1))
fact3 <- 1 - fact1*fact2
fact4 <- log(det(cov.pool)^(n1+n2-2)) - log(det(cov1)^(n1-1)) - log(det(cov2)^(n2-1))
chi <- fact3 * fact4
chi  # カイ 2 乗値
```

と入力すると

```
[1] 7.42
```

となり，(1.60) 式の χ_0^2 値が得られる．

(2) 線形判別関数の係数に関する検定

I 個の説明変数から求めた (1.38) 式の線形判別関数において，各係数 a_1，a_2,...，a_I がゼロ (すなわち，各変量 x_1，x_2,...，x_I が判別に寄与していない) か否かを検定するため

$$\text{帰無仮説 } H_0 : a_i = 0,\ \text{対立仮説 } H_1 : a_i \neq 0 \tag{1.61}$$

に関して，検定統計量

$$F_0 = \frac{(n_1+n_2-I-1)\, n_1 n_2 \left(\Delta^2 - \Delta_{[i]}^2\right)}{(n_1+n_2)(n_1+n_2-2) + n_1 n_2 d^2 (x_1,\ x_2,...,\ x_{i-1},\ x_{i+1},...,\ x_I)} \tag{1.62}$$

を計算する ([5] の 3.1 節)．ただし，I 個の説明変数すべてを用いたときの 2 つの群の平均間のマハラノビスの平方距離を**判別効率** (discriminant efficiency)

$$\Delta^2 = \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right)^t \mathbf{S}^{-1} \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}\right) \tag{1.63}$$

と定義する．

$$\Delta_{[i]}^2 = \left(\bar{\mathbf{x}}_{[i]}^{(1)} - \bar{\mathbf{x}}_{[i]}^{(2)}\right)^t \mathbf{S}^{-1} \left(\bar{\mathbf{x}}_{[i]}^{(1)} - \bar{\mathbf{x}}_{[i]}^{(2)}\right) \tag{1.64}$$

は，第 i 番目の変数を除いた判別効率である．ここに，$\bar{\mathbf{x}}_{[i]}^{(1)}$，$\bar{\mathbf{x}}_{[i]}^{(2)}$ は，第 i 番目の変数を除去したベクトルを表す．

帰無仮説のもとで，F_0 は自由度 $(1, n_1+n_2-I-1)$ の F 分布に従うから

$$F_0 \geq F\,(1,\ n_1+n_2-I-1;\alpha) \tag{1.65}$$

なら，H_0 を棄却する．

表 1.2 のデータについて

$$\text{帰無仮説 } H_0 : a_1 = 0,\ \text{対立仮説 } H_1 : a_1 \neq 0$$

は，$F_0 = 17.662^{**} > F\,(1,\ 17; 0.01) = 8.40$ より，帰無仮説を棄却する．同様に，

$$\text{帰無仮説 } H_0 : a_2 = 0,\ \text{対立仮説 } H_1 : a_2 \neq 0$$

は，$F_0 = 8.830^{**} > F\,(1,\ 17; 0.01) = 8.40$ より，この帰無仮説も棄却する．

プログラムは#(1.10) に続けて

```
# ―――――――― プログラム#(1.11) ――――――――
#準備
S <- ((n1-1)*cov1+(n2-1)*cov2)/(n1+n2-2)
mu1 <- apply(data[train$疾病=="胃潰瘍",1:2], 2, mean)
mu2 <- apply(data[train$疾病=="胃癌",1:2], 2, mean)
#マハラノビスの平方距離
d.square.all <- mahalanobis(mu1, mu2, S)
#変数 1 を除く
d.square.1 <- mahalanobis(mu1[2], mu2[2], S[-1, -1])
num <- (n1+n2-2-1)*n1*n2*(d.square.all-d.square.1)
denom <- (n1+n2)*(n1+n2-2)+n1*n2*d.square.1
(f.value <- num/denom)
#変数 2 を除く
d.square.2 <- mahalanobis(mu1[1], mu2[1], S[-2, -2])
num <- (n1+n2-2-1)*n1*n2*(d.square.all-d.square.2)
denom <- (n1+n2)*(n1+n2-2)+n1*n2*d.square.2
(f.value <- num/denom)
```

と入力すると，検査 A および B の有意性検定に関する F_0 値

```
> "検査Aの有意性検定"
> 17.66194
> "検査Bの有意性検定"
> 8.830148
```

が得られる．

1.4　2 次判別関数

線形判別では，2 つの群の分散共分散が等しい (すなわち，$\mathbf{\Sigma}^{(1)} = \mathbf{\Sigma}^{(2)}$) と仮定し，$\mathbf{S}^{(1)}$ と $\mathbf{S}^{(2)}$ を合併した (1.27) 式の $\mathbf{S}$ を用い，(1.38) 式の Fisher の線形判別関数を導いた．しかし $\mathbf{\Sigma}^{(1)} \neq \mathbf{\Sigma}^{(2)}$ の場合，**2 次判別関数** (quadratic discriminant function)

$$
\begin{aligned}
&2Q\left(x_A,\, x_B\right) \equiv D^2_{(2)} - D^2_{(1)} \\
&= \mathbf{x}^t \left(\mathbf{S}^{(2)^{-1}} - \mathbf{S}^{(1)^{-1}}\right)\mathbf{x} - 2\mathbf{x}^t \left(\mathbf{S}^{(2)^{-1}}\bar{\mathbf{x}}^{(2)} - \mathbf{S}^{(1)^{-1}}\bar{\mathbf{x}}^{(1)}\right) + \left(\bar{\mathbf{x}}^{(2)^t}\mathbf{S}^{(2)^{-1}}\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)^t}\mathbf{S}^{(1)^{-1}}\bar{\mathbf{x}}^{(1)}\right)
\end{aligned}
\tag{1.66}
$$

となり 2 次の項が残る．これは，両群のデータのバラツキ具合が異なることを考慮し，1 次式ではなく 2 次式で判別することに狙いがある．

表 1.2 のデータについて 1.3 節の結果から，2 つの群の母集団の分散共分散行列 $\mathbf{\Sigma}^{(1)}$ と $\mathbf{\Sigma}^{(2)}$ は等しいとみなせた．ちなみに，このデータについて，2 次判別関数を推定すると

$$
D^2_{(2)} - D^2_{(1)} = 15.667 - 60.463x_A + 8.435x_B + 17.829x_A^2 - 15.377x_B^2 + 39.250x_Ax_B \tag{1.67}
$$

を得、判別スコアは表 1.9 のようになる．

表 1.9　2 次判別スコア

	第 1 群 (白丸)			第 2 群 (黒丸)		
i	x_1	x_2	$z_i^{(1)}$	x_1	x_2	$z_i^{(2)}$
1	0.1	0.4	12.283	0.8	0.3	-10.726
2	0.9	0.8	0.859	0.8	0.4	-7.819
3	0.2	0.8	7.475	0.7	0.6	-1.910
4	0.2	0.5	8.586	0.6	0.2	-8.410
5	0.6	0.8	1.555	0.7	0.4	-6.017
6	0.7	0.9	1.944	0.5	0.8	2.500 ※
7	0.3	0.8	5.460	0.8	0.6	-2.927
8	0.3	0.7	5.746	0.9	0.3	-12.564
9	0.3	0.5	5.394	0.9	0.4	-9.264
10	0.1	0.6	11.680	0.9	0.6	-3.587

見掛け上の誤判別個数は，表 1.10 の通りである．線形判別に比べて，誤判別個数は 1 個減っている．表 1.8 の線形判別では，第 1 群の 2 番目の観測値が誤判別されたが，表 1.10 の 2 次判別では正しく判別されている．

プログラムは，#(1.3) に続けて

表 1.10　誤判別表 (2 次判別)

もとの群＼予測群	第 1 群	第 2 群
第 1 群	10	0
第 2 群	1	9

```
# ―――――――― プログラム#(1.12) ――――――――
library(MASS)
kfit <- qda(疾病~検査A+検査B,data=train)
print(kfit)
print(predict(kfit,train)$posterior)
```

と入力すると

```
Prior probabilities of groups:
胃潰瘍   胃癌
   0.5    0.5
Group means:
       検査A 検査B
胃潰瘍  0.37  0.68
胃癌    0.76  0.46
> print(predict(kfit,train)$posterior)
         胃潰瘍          胃癌
1  9.999933e-01 6.658282e-06
2  6.214701e-01 3.785299e-01
3  9.991850e-01 8.150338e-04
4  9.997316e-01 2.683769e-04
5  7.670307e-01 2.329693e-01
6  8.292140e-01 1.707860e-01
7  9.939199e-01 6.080078e-03
8  9.954231e-01 4.576944e-03
9  9.935070e-01 6.492951e-03
10 9.999878e-01 1.217241e-05
11 1.527612e-05 9.999847e-01
12 2.795469e-04 9.997205e-01
13 9.334683e-02 9.066532e-01
14 1.547685e-04 9.998452e-01
15 1.692053e-03 9.983079e-01
16 8.944209e-01 1.055791e-01
17 3.590269e-02 9.640973e-01
18 2.431332e-06 9.999976e-01
19 6.589215e-05 9.999341e-01
20 1.887715e-02 9.811228e-01
```

となり，事後確率が表示される．更に，プログラムの#(1.12) に続けて

```
#(1.13) 見掛け上の誤判別表
table(train$疾病,predict(kfit)$class)
```

と入力すれば，見かけ上の誤判別個数

```
    0  1
0  10  0
1   1  9
```

が得られる．誤判別は 1 個である．1 例消去 CV 法による誤判別の判別は，プログラムの#(1.13) に続けて

```
#―――――――――― #(1.14) ――――――――――――――――――
kfit <- qda(疾病~検査A+検査B,data=train,CV=TRUE)
table(train$疾病,kfit$class)
```

と入力すれば，1 例消去 CV 法による誤判別表

```
   0 1
0  9 1
1  1 9
```

が得られる．誤判別個数は 1 個増える．

2　多群判別

2.1　線形判別

(1)Wilks の Λ

線形判別では，新しく得られた観測値と各群の平均とのマハラノビスの平方距離を計算し，小さいほうの群に属すると判別した．それを $K\,(\geq 3)$ 個の群から成る多群判別へ拡張しよう．(1.28) 式のマハラノビスの平方距離を考える．各群の分散共分散行列が等しい K 個の群から，大きさ $n_1, n_2, ..., n_K$ の標本が表 2.1 のように得られたとする．

表 2.1　K 群判別の一般型

	$x_1^{(1)}$	$x_2^{(1)}$	$\cdot$	$x_I^{(1)}$	$\cdot$		$x_1^{(K)}$	$x_2^{(K)}$	$\cdot$	$x_I^{(K)}$
1	$x_{11}^{(1)}$	$x_{12}^{(1)}$	$\cdot$	$x_{1I}^{(1)}$	$\cdot$	1	$x_{11}^{(K)}$	$x_{12}^{(K)}$	$\cdot$	$x_{1I}^{(K)}$
2	$x_{21}^{(1)}$	$x_{22}^{(1)}$	$\cdot$	$x_{2I}^{(1)}$	$\cdot$	2	$x_{21}^{(K)}$	$x_{22}^{(K)}$	$\cdot$	$x_{2I}^{(K)}$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
n_1	$x_{n_1 1}^{(1)}$	$x_{n_1 2}^{(1)}$	$\cdot$	$x_{n_1 I}^{(1)}$	$\cdot$	n_K	$x_{n_K 1}^{(K)}$	$x_{n_K 2}^{(K)}$	$\cdot$	$x_{n_K I}^{(K)}$
平均	$\bar{x}_1^{(1)}$	$\bar{x}_2^{(1)}$	$\cdot$	$\bar{x}_I^{(1)}$	$\cdot$	平均	$\bar{x}_1^{(K)}$	$\bar{x}_2^{(K)}$	$\cdot$	$\bar{x}_I^{(K)}$

新しく観測値 $\mathbf{x}$ が得られたとき，第 k 群 $(k = 1, ..., K)$ の重心 (平均) ベクトル $\bar{\mathbf{x}}^{(k)}$ とのマハラノビスの平方距離

$$D_{(k)}^2 = \left(\mathbf{x} - \bar{\mathbf{x}}^{(k)}\right)^t \mathbf{S}^{-1} \left(\mathbf{x} - \bar{\mathbf{x}}^{(k)}\right), k = 1, 2, ..., K \tag{2.1}$$

が最小になる群に属すると判別する．ここに，第 k 群の i 番目の観測ベクトル $\mathbf{x}_i^{(k)}$ について，

$$\bar{\mathbf{x}}^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_i^{(k)} \tag{2.2}$$

$$\mathbf{S} = \frac{1}{n-K} \sum_{k=1}^{K} (n_k - 1)\, \mathbf{S}_{(k)} \tag{2.3}$$

とする．(2.1) 式の 2 次の項は k に依存しないから，(2.1) 式が最小になる k を求めることは，線形関数

$$f_k(\mathbf{x}) = \mathbf{x}^t \mathbf{S}^{-1} \bar{\mathbf{x}}^{(k)} - \frac{1}{2} \bar{\mathbf{x}}^{(k)t} \mathbf{S}^{-1} \bar{\mathbf{x}}^{(k)} \tag{2.4}$$

が最大になる k を見つければよい[*9].

さて，I 個の説明変数について，群内平方和積和行列 $\boldsymbol{W}$ および総平方和積和行列 $\boldsymbol{T}$

$$\mathbf{W}=\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(\mathbf{x}_i^{(k)}-\bar{\mathbf{x}}^{(k)}\right)\left(\mathbf{x}_i^{(k)}-\bar{\mathbf{x}}^{(k)}\right)^t,\quad \mathbf{T}=\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(\mathbf{x}_i^{(k)}-\bar{\mathbf{x}}\right)\left(\mathbf{x}_i^{(k)}-\bar{\mathbf{x}}\right)^t, \tag{2.5}$$

の行列式の値を求め，その比 (**Wilks の** Λ と呼ぶ)

$$\Lambda=\frac{|\mathbf{W}|}{|\mathbf{T}|} \tag{2.6}$$

を計算する．ここに，$\bar{\mathbf{x}}=\frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathbf{x}_i^{(k)}, n=\sum_{i=1}^{K}n_i$ とする．Λ は，

$$0<\Lambda<1$$

で，0 に近いほど，うまく判別されている．Λ と相関比 (付録 B) との関係については [10] の 2.2.1 節に詳しい．

(2) 変数選択

判別分析でも重回帰分析と同様の変数選択を行うことができる．ここでは，(2.6) 式の Wilks の Λ を用いた線形判別の変数選択を取り上げる．ある特定の変数 x_{q+1} が，判別に寄与しているか否かを検定したい．説明変数 $x_1, x_2, ..., x_q$ を用いたときの (2.6) 式の Λ を $\Lambda(x_1, x_2, ..., x_q)$，新たに変数 x_{q+1} を加えたときのそれを $\Lambda(x_1, x_2, ..., x_q, x_{q+1})$ と書く．この x_{q+1} を加えることによる判別力の増加は，**偏** (partial)Λ

$$\Lambda^*_{[in]}=\frac{\Lambda(x_1, x_2, ..., x_q, x_{q+1})}{\Lambda(x_1, x_2, ..., x_q)} \tag{2.7}$$

で測ることができる．このとき，偏 F 値

$$F=\frac{(n-K-q)}{K-1}\frac{(1-\Lambda^*_{[in]})}{\Lambda^*_{[in]}} \tag{2.8}$$

は，新たに加えた x_{q+1} が判別に寄与しないという帰無仮説のもとで，自由度 $(K-1, n-q-K)$ の F 分布に従う．

(2.8) 式を用いて変数選択を行うことができる．いま，$x_1, x_2, ..., x_q$ を用いて判別式を求めたとき，ある新しい変数 x_{q+1} を加えるか否かは，(2.8) 式の偏 F 値を計算し，

$$F\geq F_{in} \tag{2.9}$$

なら，x_{q+1} を加える．また，$x_1, x_2, ..., x_q$ から，ある変数 x_i を除去すべきか否かは

$$\Lambda^*_{[out]}=\frac{\Lambda(x_1, x_2, ..., x_q)}{\Lambda(x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_q)} \tag{2.10}$$

を計算する．(2.10) 式の分子は，$x_1, x_2, ..., x_q$ すべてを用いたときの Λ，分母は $x_1, x_2, ..., x_q$ から x_i を除去したときの Λ である．この $\Lambda^*_{[out]}$ を用い，

$$F=\frac{(n-K-q+1)}{K-1}\frac{(1-\Lambda^*_{[out]})}{\Lambda^*_{[out]}} \tag{2.11}$$

[*9] $D^2_{(k)}=\left(\mathbf{x}-\bar{\mathbf{x}}^{(k)}\right)^t\mathbf{S}^{-1}\left(\mathbf{x}-\bar{\mathbf{x}}^{(k)}\right)=\left(\mathbf{x}-\bar{\mathbf{x}}^{(k)}\right)^t\left(\mathbf{S}^{-1}\mathbf{x}-\mathbf{S}^{-1}\bar{\mathbf{x}}^{(k)}\right)=\mathbf{x}^t\mathbf{S}^{-1}\mathbf{x}-\mathbf{x}^t\mathbf{S}^{-1}\bar{\mathbf{x}}^{(k)}-\mathbf{x}^t\mathbf{S}^{-1}\bar{\mathbf{x}}^{(k)}+\bar{\mathbf{x}}^{(k)t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(k)}$
$\equiv -2\mathbf{x}^t\mathbf{S}^{-1}\bar{\mathbf{x}}^{(k)}+\bar{\mathbf{x}}^{(k)^t}\mathbf{S}^{-1}\bar{\mathbf{x}}^{(k)}$ より，$f_k(\mathbf{x})=-\frac{1}{2}D^2_{(k)}$ となる．

を計算し，

$$F < F_{out} \tag{2.12}$$

なら x_i を除去する．(2.9)，(2.11) 式の F_{in}，F_{out} はあらかじめ指定された限界値で，F 分布の上側 15% 点，あるいは

$$F_{in} = F_{out} = 2.0 \tag{2.13}$$

とする．

(2.6) 式の Wilks の Λ を用い

帰無仮説 H_0 : 群間に差がない (I 個の説明変数は，K 個の群の判別に寄与しない)

を検定することもできる．Λ を用いた **Rao の近似**

$$F = \frac{f_2}{f_1}\frac{(1-\Lambda^{1/c})}{\Lambda^{1/c}} \tag{2.14}$$

を採用する [9, 11]．ただし，

$$\begin{aligned}
&f_1 = aI \\
&f_2 = bc + 1 - aI/2 \\
&a = K - 1 \\
&n = \sum_{k=1}^{K} n_k \\
&b = n - 1 - (I + K)/2 \\
&c = \begin{cases} \sqrt{\frac{I^2a^2-4}{I^2+a^2-5}} \ : \ I^2 + a^2 \neq 5 \\ 1 \ : \ I^2 + a^2 = 5 \end{cases}
\end{aligned}$$

とする．帰無仮説のもとで，(2.14) 式は自由度 (f_1，f_2) の F 分布に従う．

(3) 分散共分散行列の同等性に関する Bartlett の検定

1.3 節における 2 群の分散共分散行列性の同等性の検定を多群の場合へ拡張し，

帰無仮説 H_0 : $\mathbf{\Sigma}^{(1)} = \mathbf{\Sigma}^{(2)} = \cdots = \mathbf{\Sigma}^{(K)}$

を検定する [3, 10]．第 k 群の標本分散共分散行列を $\mathbf{S}^{(k)}$ $(k = 1, 2, ..., K)$，全データからの標本分散共分散行列を

$$\mathbf{S} = \frac{1}{(n-K)}\sum_{k=1}^{K}(n_k - 1)\mathbf{S}^{(k)}$$

とすると

$$\chi_0^2 = \left[1 - \frac{2I^2 + 3I - 1}{6\,(I+1)\,(K-1)}\left\{\sum_{k=1}^{K}\frac{1}{(n_k-1)} - \frac{1}{(n-K)}\right\}\right] \times \ln\left\{\frac{|\mathbf{S}|^{n-K}}{\prod_{k=1}^{K}\left|\mathbf{S}^{(k)}\right|}\right\} \tag{2.15}$$

は，漸近的に自由度 $(K-1)\,I\,(I+1/2)$ のカイ二乗分布に従う．よって

$$\chi_0^2 \geq \chi^2\left((K-1)\,I\,(I+1)/2, \alpha\right) \tag{2.16}$$

なら，帰無仮説を棄却する．

2.2　適用例

(1) 糖尿病データ

3 群判別として表 2.2 の糖尿病データ [1] を取り上げる．145 名の対象者に施した 5 種類の検査

表 2.2　糖尿病データ

相対体重 (X1)	空腹時血糖値 (X2)	ブドウ糖値 (X3)	インシュリン値 (X4)	SSPG(X5)	判別群 (GROUP)
0.92	300	1468	28	455	1
0.86	303	1487	23	327	1
0.85	125	714	232	279	1
.	.	.	.	.	
0.9	213	1025	29	209	1
1.11	328	1246	124	442	1
0.74	346	1568	15	253	1
0.99	98	478	151	122	2
1.02	88	439	208	244	2
1.19	100	429	201	194	2
.	.	.	.	.	
0.94	88	423	212	156	2
0.91	114	643	155	100	2
0.83	103	533	120	135	2
0.81	80	356	124	55	3
0.95	97	289	117	76	3
0.94	105	319	143	105	3
.	.	.	.	.	
0.89	99	398	76	108	3
1.11	93	393	490	259	3
1.18	89	318	73	220	3

$$
\begin{cases}
x_1：\text{相対体重} \\
x_2：\text{空腹時血糖値} \\
x_3：\text{ブドウ糖を空腹時に負荷し，3 時間を 30 分間隔で血漿ブドウ糖値を記録したとき，} \\
\quad\ \text{検査値曲線の下方面積の値} \\
x_4：\text{ブドウ糖を空腹時に負荷し，3 時間を 30 分間隔で血漿インシュリン値を記録したとき，} \\
\quad\ \text{検査値曲線の下方面積の値} \\
x_5：\text{一定量のインシュリンとブドウ糖を静脈内に注入後，平衡状態に達したときの血漿ブドウ糖値}
\end{cases}
$$

に基づいて，正常，化学的糖尿病，および臨床的糖尿病の 3 つの群に判別する．(2.4) 式の線形関数は

$$
\begin{cases}
f_1(\mathbf{x}) = -79.962 + 96.655x_1 - 0.173x_2 + 0.113x_3 + 0.033x_4 - 0.065x_5 \\
f_2(\mathbf{x}) = -65.963 + 99.538x_1 - 0.133x_2 + 0.085x_3 + 0.045x_4 - 0.082x_5 \\
f_3(\mathbf{x}) = -49.330 + 89.967x_1 + 0.0022x_2 + 0.046x_3 + 0.034x_4 - 0.088x_5
\end{cases}
\tag{2.17}
$$

となる．よって，新たに得られた観測値 $\mathbf{x}^t = (x_1, x_2, x_3, x_4, x_5)$ について，$f_k(\mathbf{x})$ を計算し，最大値を与える群に判別する．表 2.3 から見掛け上の誤判別個数は 13 となる．また，表 2.4 から 1 例消去 CV 法による誤判別個数も 13 となる．ちなみに，2 次判別関数による見掛け上の誤判別個数は表 2.5 から 7 で，かなり少なくなる．また，判別力を示す (2.6) 式の値は $\Lambda = 0.105$ で 0 に近くうまく判別されており，(2.14) 式より $F = 57.489 > F\,(10, 276; 0.05)$ を得，判別式は有意である．

表 2.3　見掛け上の誤判別個数 (線形判別)

予測群 / もとの群	第 1 群	第 2 群	第 3 群
1	27	5	1
2	0	34	2
3	0	5	71

表 2.4　1 例消去 CV 法による誤判別個数 (線形判別)

予測群 / もとの群	第 1 群	第 2 群	第 3 群
1	27	5	1
2	0	34	2
3	0	5	71

表 2.5　見掛け上の誤判別個数 (2 次判別)

予測群 / もとの群	第 1 群	第 2 群	第 3 群
1	73	3	0
2	1	35	0
3	0	3	30

ここでは，事前確率が一様であると仮定した．一方，各群のデータ数が (76,36,33) であるから，事前確率として各群の割合 (76/145,36/145,33/145)=(0.2276,0.2483,0.5241) を採用すると，見かけ上の誤判別個数は 14 となる．また，1 例消去 CV 法による誤判別個数は 16 となり，誤判別個数は 2 個増える．ちなみに，2 次判別関数による見かけ上の誤判別個数は 7 で，かなり少なくなる．事前確率については付録 D を参照されたい．

固有値に基づく Fisher の線形判別関数のプログラムは

```
# ———————— プログラム#(2.1) ————————————
library(MASS)
library(rrcov)
train <- read.csv("E:\\糖尿病.csv",header=TRUE)
GR <-factor(train$GROUP)
kfit<-lda(GR~X1+X2+X3+X4+X5, data=train,prior=c(1,1,1)/3)
kfit
```

と書ける．その結果，

```
Prior probabilities of groups:
        1         2         3
0.3333333 0.3333333 0.3333333

Group means:
         X1        X2        X3       X4       X5
1 0.9839394 217.66667 1043.7576 106.0000 318.8788
2 1.0558333  99.30556  493.9444 288.0000 208.9722
3 0.9372368  91.18421  349.9737 172.6447 114.0000

Coefficients of linear discriminants:
             LD1           LD2
X1 -0.9835559784 -3.8998182874
X2  0.0298817074  0.0397658702
X3 -0.0118177013 -0.0082948376
X4  0.0007090488 -0.0061334218
X5 -0.0043341699  0.0007111392

Proportion of trace:
   LD1    LD2
0.8713 0.1287
```

を得る．見掛け上の誤判別とベイズの事後確率は

```
# ---------- プログラム#(2.2) ----------
#見かけ上の誤判別
table(GR, predict(kfit)$class)
#ベイズの事後確率
kfit<-lda(GR~X1+X2+X3+X4+X5, data=train, prior=c(1,1,1)/3)
print(predict(kfit, train)$posterior)
```

と続ければ

```
> #見かけ上の誤判別
> table(GR, predict(kfit)$class)
GR   1  2  3
  1 27  5  1
  2  0 34  2
  3  0  5 71
> #ベイズの事後確率
> kfit<-lda(GR~X1+X2+X3+X4+X5, data=train, prior=c(1,1,1)/3)
> print(predict(kfit, train)$posterior)
                1            2            3
1    1.000000e+00 4.058162e-09 2.364691e-14
2    1.000000e+00 1.716562e-08 3.231608e-13
3    5.276168e-01 4.718298e-01 5.534645e-04
4    1.000000e+00 6.041723e-09 6.405458e-15
5    9.999846e-01 1.543307e-05 1.913083e-09
 .           .            .            .
140  2.468071e-06 3.873044e-02 9.612671e-01
141  9.183590e-07 2.070697e-02 9.792921e-01
142  1.108037e-07 1.222205e-02 9.877778e-01
143  2.013256e-06 1.113799e-02 9.888600e-01
144  7.093879e-06 9.764483e-01 2.354461e-02
145  4.952880e-06 6.111271e-02 9.388823e-01
```

となる．

マハラノビスの平方距離に基づく判別関数は，プログラム

```
# ———————— プログラム#(2.3) ————————————
library(rrcov)
train <- read.csv("E:\\糖尿病.csv",header=TRUE)
GR <-factor(train$GROUP)
kfit<-LdaClassic(GR~X1+X2+X3+X4+X5, data=train,prior=c(1,1,1)/3)
kfit
```

を採用すると

```
Prior Probabilities of Groups:
        1         2         3
0.3333333 0.3333333 0.3333333
Group means:
         X1        X2        X3       X4       X5
1 0.9839394 217.66667 1043.7576 106.0000 318.8788
2 1.0558333  99.30556  493.9444 288.0000 208.9722
3 0.9372368  91.18421  349.9737 172.6447 114.0000
Within-groups Covariance Matrix:
           X1           X2           X3           X4          X5
X1  0.0145041   -0.4303672    -2.938608     1.328559    3.134166
X2 -0.4303672 1378.9464315  5222.495161  -961.429763  908.704030
X3 -2.9386081 5222.4951610 23046.745753 -3901.371201 3656.126583
X4  1.3285591 -961.4297628 -3901.371201 10610.897239 1464.429577
X5  3.1341661  908.7040297  3656.126583  1464.429577 4392.299207

Linear Coeficients:
        X1          X2         X3         X4          X5
1 96.65486 -0.17317749 0.11273845 0.03265619 -0.06527225
2 99.53831 -0.13318740 0.08487645 0.04507648 -0.08157455
3 89.96717  0.00223123 0.04598939 0.03432152 -0.08842841
Constants:
       1         2         3
-79.96201 -65.96315 -49.33044
```

となり，アウトプットの Linear Coeficients と Constants の値から (2.17) 式が得られる．

変数増加法による変数選択を行う。プログラム

```
# ———————— プログラム#(2.4) ————————————
library(MASS)
library(klaR)
library(rrcov)
train <- read.csv("E:\\train.csv",header=TRUE)
kfit <- greedy.wilks(factor(G) ~ X1+X2+X3+X4+X5, data=train,niveau=0.15)
kfit
```

を採用すると

```
Formula containing included variables:
factor ~ X3 + X2 + X4 + X1 + X5
Values calculated in each step of the selection procedure:
  vars Wilks.lambda F.statistics.overall p.value.overall F.statistics.diff
1   X3    0.2262307            242.83889    1.489919e-46        242.838894
2   X2    0.1535618            109.40676    3.725810e-56         33.362184
3   X4    0.1262404             84.67653    5.087132e-60         15.149650
4   X1    0.1094834             70.27208    2.428376e-62         10.637319
5   X5    0.1052131             57.48913    1.143238e-61          2.800505
  p.value.diff
```

```
1  1.489919e-46
2  1.326717e-12
3  1.096929e-06
4  5.004230e-05
5  6.421277e-02
```

を得る。アウトプットの factor~ $X3+X2+X4+X1+X5$ は，X3，X2，X4，X1，X5 の順に説明変数が取り込まれることを意味している．また，F.statistics.diff と p.value.diff から変数選択後の説明変数の有意性検定は表 2.6 のようになる．

表 2.6　説明変数の有意性検定

ステップ	追加数	入力済	**F 値**	**Pr > F**
1	1	X3	242.839	<<0.0001
2	2	X2	33.362	<<0.0001
3	3	X4	15.150	<<0.0001
4	4	X1	10.637	<<0.0001
5	5	X5	2.801	0.064

(2) 低体重データ

低体重児 (出生体重 2500g 未満) の発生要因として妊娠中のリスクが関連していると考えられる [13]．群を

$$\begin{cases} 2500g \text{ 未満} = 1 \\ 2500g \text{ 以上} = 0 \end{cases}$$

とし、説明変数

$$\begin{cases} X1 : \text{年齢 } (age) \\ X2 : \text{最終月経時の体重 } (lwt) \\ X3 : \text{人種 } (race) : \text{白人} = 1, \text{黒人} = 2, \text{その他} = 3 \\ X4 : \text{妊娠時の喫煙} : \text{喫煙} = 1, \text{非喫煙} = 0 \\ X5 : \text{妊娠 28 週以前の労働回数 (連続変数)} \\ X6 : \text{高血圧歴} : \text{有り} = 1, \text{なし} = 0 \\ X7 : \text{子宮の痛み} : \text{有り} = 1, \text{なし} = 0 \end{cases}$$

を取り上げる．

変数増加法による変数選択のプログラム

```
# ———————— プログラム#(2.5) ————————————
library(MASS)
library(klaR)
library(rrcov)
X <- read.csv("G:\\train.csv",header=TRUE)
test.lda <- greedy.wilks(G ~ X1+X2+X3+X4+X5+X6+X7, data=X,niveau=0.15)
test.lda
```

を採用すると、

```
G ~ X5 + X6 + X2 + X7 + X4 + X3
Values calculated in each step of the selection procedure:
```

```
  vars Wilks.lambda F.statistics.overall p.value.overall F.statistics.diff
1   X5    0.9615498             7.477710    0.0068474412          7.477710
2   X6    0.9373980             6.210791    0.0024486994          4.792229
3   X2    0.9029447             6.628395    0.0002816056          7.058969
4   X7    0.8855900             5.942775    0.0001610925          3.605822
5   X4    0.8727574             5.336053    0.0001328557          2.690737
6   X3    0.8510840             5.307490    0.0000455421          4.634750
  p.value.diff
1  0.006847441
2  0.029826841
3  0.008572585
4  0.059133984
5  0.102641522
6  0.032639877
```

が得られる．取り込まれる説明変数の順序は X5，X6，X2，X7，X4，X3 である．変数選択後の説明変数の有意性検定の結果を表 2.7 に示す．

表 2.7　説明変数の有意性検定

ステップ	入力済	F 値	Pr > F
1	X5	7.48	0.0068
2	X6	4.79	0.0298
3	X2	7.06	0.0086
4	X7	3.61	0.0591
5	X4	2.69	0.1027
6	X3	4.63	0.0326

3　正準判別分析

説明変数の個数 I が 2 あるいは 3 なら，それらの値を 2 次元平面あるいは 3 次元空間にプロットし，データの特徴を視覚的に捉えることができる．しかし，I が 4 以上になるとグラフ表現はできない．1.2 節の線形判別関数は，マハラノビスの平方距離を用いて導出された．しかし，付録 B のような相関比から導くこともできる．本章では多群判別について，相関比に基づく正準判別分析を取り上げる．それは，説明変数の 1 次結合からなる新しい変量 (**正準変量**) を構成し，群間の相違を 2 次元あるいは 3 次元に縮約する方法である．

3.1　定式化

2 群判別の場合と同様に，線形結合 z により K 個の群をうまく判別したい．そのため，相関比

$$\eta^2 = S_B/S_T \tag{3.1}$$

が最大になる正準判別係数 $a_1, ..., a_I$ を求める．S_B, S_T は群間平方和および総平方和である．これは，**群間平方和** S_B と**群内平方和** S_W との比

$$\lambda = S_B/S_W = \mathbf{a}^t\mathbf{B}\mathbf{a}/\mathbf{a}^t\mathbf{W}\mathbf{a}\,,\ \mathbf{a} = (a_1, a_2, ..., a_I)^t \tag{3.2}$$

を最大化することと同等である (付録 B を参照されたい)．ただし，$\mathbf{B}$ は**群間平方和積和行列**，$\mathbf{W}$ は**群内平方和積和行列** ((2.5) 式を参照) と呼ばれ，

$$\mathbf{B} = (b_{jj'}) \ , \ b_{jj'} = \sum_{k=1}^{K} n_k \left(\bar{x}_j^{(k)} - \bar{x}_j \right) \left(\bar{x}_{j'}^{(k)} - \bar{x}_{j'} \right) \tag{3.3}$$

$$\mathbf{W} = (w_{jj'}) \ , \ w_{jj'} = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left(x_{ji}^{(k)} - \bar{x}_j^{(k)} \right) \left(x_{j'i}^{(k)} - \bar{x}_{j'}^{(k)} \right) \tag{3.4}$$

とする．ただし，$\bar{x}_j = \sum_{k=1}^{K} \bar{x}_j^{(k)} / K$ である．

付録 B の固有方程式 (B.8) 式を解くと $r = \min(K-1, I)$ 個の非負の固有値が得られる．ゆえに，最大の固有値 λ_1 に対応する固有ベクトル $\mathbf{a}_1 = (a_{11}, ..., a_{I1})^t$ が求める正準判別係数になる．この正準判別分析と Fisher の線形判別分析との関係については [10] の 2.2.1 節に詳しい．

この固有ベクトルを用い，線形結合

$$z_1 = a_{11}x_1 + \cdots + a_{I1}x_I - (a_{11}\bar{x}_1 + \cdots + a_{I1}\bar{x}_I) \tag{3.5}$$

を**第 1 正準変量**と呼ぶ．なお，正準変量の平均は 0 になるようにした．

1 つの正準変量のみで K 個の群を十分に判別できなければ，2 番目の正準変量を求めればよい．2 番目に大きい固有値 λ_2 に対応する固有ベクトル $\mathbf{a}_2 = (a_{12}, \ ..., \ a_{I2})^t$ を用いて線形結合

$$z_2 = a_{12}x_1 + \cdots + a_{I2}x_I - (a_{12}\bar{x}_1 + \cdots + a_{I2}\bar{x}_I) \tag{3.6}$$

を決める．z_2 を第 2 正準変量と呼ぶ．以下，同様にして r 個の正準変量を導くことができる．このとき，$\lambda_1 \Big/ \sum_{i=1}^{r} \lambda_i$, $(\lambda_1 + \lambda_2) \Big/ \sum_{i=1}^{r} \lambda_i$,... をそれぞれ，第 1 正準変量，第 2 正準変量，... の**累積寄与率**という．

3.2 適用例

表 2.2 の糖尿病データについて，正準判別分析を適用しよう．群間および群内平方和積和行列 $\mathbf{B}$, $\mathbf{W}$ は，それぞれ

$$\mathbf{B} = \begin{bmatrix} 0.345 & & & & \\ 50.627 & 392732.0 & & & \\ 558.752 & 2073040.0 & 11193300.0 & & \\ 311.510 & -304616.0 & -1306220.0 & 599298.0 & \\ 313.289 & 569350.0 & 3211650.0 & -193336.0 & 995189.0 \end{bmatrix} \tag{3.7}$$

$$\mathbf{W} = \begin{bmatrix} 2.060 & & & & \\ -61.112 & 195810.0 & & & \\ -417.283 & 741594.0 & 3272640.0 & & \\ 188.655 & -136523.0 & -553995.0 & 1506750.0 & \\ 445.052 & 129036.0 & 519170.0 & 207949.0 & 623706.0 \end{bmatrix} \tag{3.8}$$

となる．付録 B の (B.15) 式へ $\mathbf{B}$, $\mathbf{W}$ を代入して一般固有値問題を解くと $r = \min(2, 5) = 2$ であるから固有値 $(\lambda_1, \lambda_2) = (4.780, 0.644)$ が求まる．$\lambda_1/(\lambda_1 + \lambda_2) = 0.88$ より，第 1 正準変量の寄与率は 88% である．

プログラムは

```
# ———————— プログラム#(3.1) ————————————
library(MASS)
train<-read.csv("E:\\train.csv",header=FALSE)
# 事前確率が一様なら、
```

```
    # prior=c(1,1)/2  ;2群の場合
    # prior=c(1,1,1)/3   ;3群の場合
    # 指定しなければ、proportion
kfit<-lda(V6~V1+V2+V3+V4+V5,data=train,prior=c(1,1,1)/3)
print(kfit)
apply(-kfit$means%*%kfit$scaling,2,mean)
```

と書ける．その結果，

```
Prior probabilities of groups:
        1         2         3
0.3333333 0.3333333 0.3333333
Group means:
         V1        V2        V3       V4       V5
1 0.9839394 217.66667 1043.7576 106.0000 318.8788
2 1.0558333  99.30556  493.9444 288.0000 208.9722
3 0.9372368  91.18421  349.9737 172.6447 114.0000
Coefficients of linear discriminants:
             LD1           LD2
V1 -0.9835559784 -3.8998182874
V2  0.0298817074  0.0397658702
V3 -0.0118177013 -0.0082948376
V4  0.0007090488 -0.0061334218
V5 -0.0043341699  0.0007111392
Proportion of trace:
   LD1    LD2
0.8713 0.1287
> apply(-kfit$means%*%kfit$scaling,2,mean)
     LD1       LD2
5.139915 4.685363
```

を得る．よって，2 つの正準変量は，アウトプットの Coefficients of linear discriminants と最終行の値から

$$\begin{cases} \text{第 1 正準変量}: z_1 = 5.1399 - 0.9836x_1 + 0.02988x_2 - 0.01182x_3 + 0.00071x_4 - 0.00433x_5 \\ \text{第 2 正準変量}: z_2 = 4.6853 - 3.8998x_1 + 0.03977x_2 - 0.00833x_3 - 0.00613x_4 + 0.00071x_5 \end{cases} \tag{3.9}$$

となる．

プログラム #(3.1) に

```
# ———————— プログラム#(3.2) ————————
predict(kfit)$x
```

と続ければ，正準変量の値

```
            LD1         LD2
1   -6.10102382  1.00230239
2   -5.62567316  1.13762848
3   -1.44346708 -0.80580857
4   -6.29894366  0.32999683
 .        .           .
 .        .           .
```

が求まる．例えば，最初の観測値の x 座標の値 $(= -6.10101)$ は

$$\text{第 1 正準変量}: z_1 = 5.1399 - 0.9836x_1 + 0.02988x_2 - 0.01182x_3 + 0.00071x_4 - 0.00433x_5$$

の $(x_1, x_2, x_3, x_4, x_5)$ に表 2.2 の最初の観測値 $(0.92, 300, 1468, 28, 455)$ を代入した

$$-6.10101 = 5.1399 - 0.9836 \times 0.92 + 0.02988 \times 300 - 0.01182 \times 1468 + 0.00071 \times 28 - 0.00433 \times 455$$

から求まる．そして

```
# ---------- プログラム#(3.3) ----------
plot(kfit,dimen=2)
```

と続ければ，5 変数で構成された糖尿病データに対する正準変量の 2 次元プロットが得られる (図 3.1)．第 1 群と第 2 群の重なり部分が多く，お互いの誤判別が多いことが視覚的にも明白である．

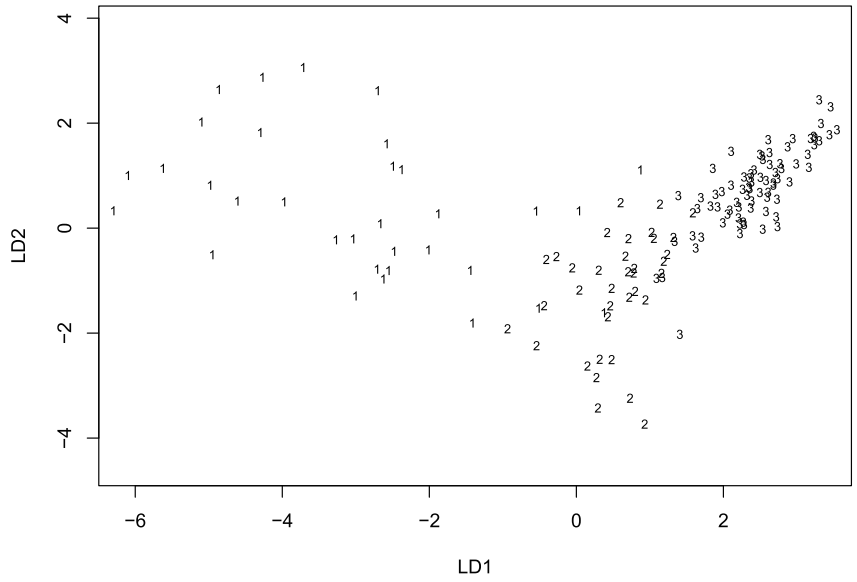

図 3.1　正準変量の 2 次元プロット

プログラム#(3.3) に

```
# ---------- プログラム#(3.4) ----------
# 見掛け上の誤判別表
table(train$V6,predict(kfit)$class)
```

と続ければ，見かけ上の誤判別表

```
    1  2  3
1  27  5  1
2   0 34  2
3   0  5 71
```

が得られる．

1 例消去 CV 法による誤判別表は，プログラム#(3.4) に

```
# ---------- プログラム#(3.5) ----------
# 1例消去CV
kfit <- lda(V6~V1+V2+V3+V4+V5,train,CV=TRUE)
table(train$V6,kfit$class)
```

と続ければよい．その結果，

```
    1  2  3
1  26  6  1
2   0 30  6
3   0  3 73
```

となり，誤判別個数が 3 個増える．

4　ロジスティック判別

4.1　定式化

ロジスティック判別は，Fisher の線形判別とともに広範に活用されてきた．Fisher の線形判別と異なり，

i) 説明変数に多変量正規分布を仮定する必要がない

ii) 説明変数に離散型データが含まれていてもよい

などの利点を有する．Fisher の線形判別では

i) 説明変数に多変量正規分布を仮定する

ii) 2 つの群間の分散共分散行列が同等である

iii) 平均と分散共分散行列を標本から推定しなければならない

などの条件を満たさなければならない．この意味からもロジスティック判別は，Fisher の線形判別より適用範囲が広いといえる．

一般に，I 個の説明変数 $(x_1, x_2, ..., x_I)$ が観測されているとする．第 1 群と第 2 群に対応するクラス (母集団)C_1，C_2 から，観測ベクトル $\mathbf{x}^t = (x_1, x_2, ..., x_I)$ が抽出される事前確率を π_1, π_2 とする．クラス C_k において，観測ベクトル $\mathbf{x}^t$ が確率密度 $f_k(\mathbf{x})$ の分布に従うなら，$k = 1, 2$ について

$$\begin{cases} \Pr(C_k) = \pi_k \\ \Pr(\mathbf{x}\,|C_k) = f_k(\mathbf{x}) \end{cases} \tag{4.1}$$

と書ける．よって，$\mathbf{x}$ がクラス C_k から得られる事後確率 $\Pr(C_k\,|x)$ は，ベイズの定理を用いると

$$\Pr(C_k\,|\mathbf{x}) = \frac{\Pr(\mathbf{x}\,|C_k)\Pr(C_k)}{\Pr(\mathbf{x}\,|C_1)\Pr(C_1) + \Pr(\mathbf{x}\,|C_2)\Pr(C_2)} \tag{4.2}$$

で与えられる．(4.2) 式へ (4.1) 式を代入すると

$$\begin{aligned} \Pr(C_k\,|\mathbf{x}) &= \frac{\pi_k f_k(\mathbf{x})}{\pi_1 f_1(\mathbf{x}) + \pi_2 f_2(\mathbf{x})} \\ &= \frac{\left(\frac{\pi_k}{\pi_1}\right)\left\{\frac{f_k(\mathbf{x})}{f_1(\mathbf{x})}\right\}}{1 + \left(\frac{\pi_2}{\pi_1}\right)\left\{\frac{f_2(\mathbf{x})}{f_1(\mathbf{x})}\right\}}\ , k = 1, 2 \end{aligned} \tag{4.3}$$

を得る．

分布 $f_1(\mathbf{x})$ と $f_2(\mathbf{x})$ との比 $f_1(\mathbf{x})/f_2(\mathbf{x})$ の対数が，観測ベクトル $\mathbf{x}^t = (x_1, x_2, ..., x_I)$ の線形式

$$\ln\left\{\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\right\} = \boldsymbol{\beta}^t\mathbf{x} \tag{4.4}$$

で表されると仮定する．ここに，$\boldsymbol{\beta}^t = (\beta_1, \beta_2, ..., \beta_I)$ は未知パラメータの係数ベクトルである．$\beta_0 \equiv \ln(\pi_1/\pi_2)$ とおいて，(4.3) 式へ (4.4) 式を代入すると

$$\Pr(C_1\,|\mathbf{x}) = \frac{\exp\left(\beta_0 + \boldsymbol{\beta}^t\mathbf{x}\right)}{1 + \exp\left(\beta_0 + \boldsymbol{\beta}^t\mathbf{x}\right)} \tag{4.5}$$

すなわち

$$\ln\left\{\frac{\Pr(C_1\,|\mathbf{x})}{1 - \Pr(C_1\,|\mathbf{x})}\right\} = \beta_0 + \boldsymbol{\beta}^t\mathbf{x} \tag{4.6}$$

となる．Fisher の線形判別と同様に，判別ルール「観測値ベクトル $\mathbf{x}$ が得られたとき，事後確率 $\Pr(C_k|\mathbf{x})$ が最大になる群に，その個体が属する」を採用する．ゆえに

$$\Pr(C_1|\mathbf{x}) > \Pr(C_2|\mathbf{x}) \tag{4.7}$$

なら，$\mathbf{x}$ は第 1 群に属する．そうでなければ，第 2 群に属する．$\Pr(C_1|\mathbf{x}) + \Pr(C_2|\mathbf{x}) = 1.0$ であるから，$\pi_1 = \pi_2$ なら (4.7) 式は

$$\Pr(C_1|\mathbf{x}) > 0.5 \tag{4.8}$$

となる．

特に，$f_1(\mathbf{x}), f_2(\mathbf{x})$ が多変量正規分布

$$f_k(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}^{(k)}\right)^t \mathbf{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}^{(k)}\right)\right], k = 1, 2 \tag{4.9}$$

に従うなら，(4.3) 式から (4.6) 式の左辺は

$$\ln\left\{\frac{\Pr(C_1|\mathbf{x})}{1-\Pr(C_1|\mathbf{x})}\right\} = \ln\left(\frac{\pi_1}{\pi_2}\right) - \frac{1}{2}\left(\boldsymbol{\mu}^{(1)^t}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)^t}\mathbf{\Sigma}^{-1}\mu^{(2)}\right) + \mathbf{x}^t\mathbf{\Sigma}^{-1}\left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}\right) \tag{4.10}$$

となる．

さて，(4.6) 式の未知パラメータ $\beta_0, \boldsymbol{\beta}$ を推定しよう．n 組の観測ベクトル $\mathbf{x}_i^t = (x_{i1}, x_{i2}, ..., x_{iI}), i = 1, 2, ..., n$ が与えられたとき，それが第 1 群に属する確率を

$$\Pr\{t_i = 0|\mathbf{x}_i\} = \Pr(C_1|\mathbf{x}_i) = p_i \tag{4.11}$$

で示すと

$$\Pr\{t_i = 1|\mathbf{x}_i\} = \Pr(C_2|\mathbf{x}_i) = 1 - p_i \tag{4.12}$$

となる．よって，ベルヌーイ分布

$$f(t_i|\mathbf{x}_i) = p_i^{t_i}(1-p_i)^{1-t_i} \tag{4.13}$$

を得，尤度関数は

$$L(\boldsymbol{\beta}; \mathbf{x}, \mathbf{t}) = \prod_{i=1}^{n} p_i^{t_i}(1-p_i)^{1-t_i} \tag{4.14}$$

と書ける．(4.14) 式へ (4.5) 式を代入すると

$$\begin{aligned}\ln L(\boldsymbol{\beta}; \mathbf{x}, \mathbf{t}) &= \sum_{i=1}^{n}\{t_i \ln p_i + (1-t_i)\ln(1-p_i)\} \\ &= \sum_{i=1}^{n}\left[t_i \ln\left\{\frac{\exp(\beta_0 + \boldsymbol{\beta}\mathbf{x}^t)}{1+\exp(\beta_0+\boldsymbol{\beta}\mathbf{x}^t)}\right\} + (1-t_i)\ln\left\{\frac{1}{1+\exp(\beta_0+\boldsymbol{\beta}\mathbf{x}^t)}\right\}\right]\end{aligned} \tag{4.15}$$

となる．この (4.15) 式を最大にする $\beta_0, \boldsymbol{\beta}$ が最尤推定量である．

4.2 適用例

表 1.2 のデータについて，(4.6) 式の $\beta_0, \beta_1, \beta_2$ を求めると

$$\ln\left\{\frac{\Pr(C_1|\mathbf{x})}{1-\Pr(C_1|\mathbf{x})}\right\} = 1.097 + 7.931x_1 - 9.420x_2 \tag{4.16}$$

を得，判別結果は表 1.7 と同じになった．プログラム

```
# ---------- プログラム#(4.1) ----------
train<-read.csv("E:\\train.csv",header=FALSE)
kfit<-glm(V3~V1+V2,data=train,family=binomial)
kfit$coefficients
```

から，係数 $(\beta_0, \beta_1, \beta_2)$

```
(Intercept)          V1          V2
   1.096710    7.930602   -9.419720
```

が得られる．#(4.1) に

```
# ---------- プログラム#(4.2) ----------
kpred<-predict(kfit, type="response")
(tab<-table(train$V3,kpred>0.5))
#誤判別率
(error<-(tab[1,2]+tab[2,1])/sum(tab))
```

と続ければ，誤判別表と誤判別率

```
    FALSE TRUE
  0     9    1
  1     1    9
> #誤判別率
> (error<-(tab[1,2]+tab[2,1])/sum(tab))
[1] 0.1
```

が得られる．

5　カーネルロジスティック判別

5.1　定式化

(4.13) 式のベルヌーイ分布に対する (負の) 対数尤度

$$-\ln\left\{p_i^{t_i}\left(1-p_i\right)^{1-t_i}\right\} = -t_i \ln p_i - (1-t_i)\ln\left(1-p_i\right) \tag{5.1}$$

について，$t_i = (y_i+1)/2, y_i = \pm 1$ とおくと

$$-\ln\left\{p_i^{t_i}\left(1-p_i\right)^{1-t_i}\right\} = \begin{cases} -\ln p_i : y_i = +1 (i.e., t_i = 1) \\ -\ln\left(1-p_i\right) : y_i = -1 (i.e., t_i = 0) \end{cases} \tag{5.2}$$

となる．ロジット変換

$$f_i = f\left(\mathbf{x}_i\right) = \ln\left(\frac{p_i}{1-p_i}\right) \tag{5.3}$$

より，

$$p_i = \frac{e^{f_i}}{1+e^{f_i}} = \frac{1}{1+e^{-f_i}} \tag{5.4}$$

となり，

$$\begin{cases} -\ln p_i = -\ln\left(\frac{e^{f_i}}{1+e^{f_i}}\right) = \ln\left(1+e^{-f_i}\right) \\ -\ln\left(1-p_i\right) = -\ln\left(1-\frac{e^{f_i}}{1+e^{f_i}}\right) = \ln\left(1+e^{f_i}\right) \end{cases} \tag{5.5}$$

を得る．損失関数は

$$\ln\left(1+e^{-y_i f_i}\right) = \begin{cases} \ln\left(1+e^{-f_i}\right) = -\ln p_i : y_i = +1 \\ \ln\left(1+e^{f_i}\right) = -\ln\left(1-p_i\right) : y_i = -1 \end{cases} \tag{5.6}$$

となる．(5.2) 式と (5.6) 式は等価であるからベルヌーイ分布の損失関数は，y_i と f_i を用い

$$-\ln\left\{p_i^{t_i}\left(1-p_i\right)^{1-t_i}\right\}=\ln\left(1+e^{-y_i f_i}\right) \tag{5.7}$$

と書け，$t_i\in\{0,1\}$ を $y_i\in\{+1,-1\}$ に変換できる．

よって，**再生核ヒルベルト空間** (Reproducing Kernel Hilbert Space:RKHS) における最適化問題

$$\underset{\alpha}{Min}\left[\sum_{i=1}^{n} ln\left\{1+e^{-y_i f(x_i)}\right\}-\frac{n}{2}\lambda\|f\|_{\mathcal{H}_K}^2\right] \tag{5.8}$$

を解けば，ペナルティ付きロジスティック判別モデルが得られる．ロジスティック判別モデルにペナルティ項を付加することによって，**過学習** (overfitting) を回避できる．

(5.8) 式の最適解は，representer 定理より $f\left(\mathbf{x}\right)=\sum_{i=1}^{n}\alpha_i\phi\left(\mathbf{x}_i\right)$ で与えられる．$\|f\|_{\mathcal{H}_K}^2=\alpha^t K\alpha$ より

$$\underset{\alpha}{Min}\left[\sum_{i=1}^{n} ln\left\{1+e^{-y_i f(x_i)}\right\}-\frac{n}{2}\lambda\alpha^t K\alpha\right] \tag{5.9}$$

と書ける．ここに，$\mathbf{K}$ は (i,j) 要素が $K\left(\mathbf{x}_i,\mathbf{x}_j\right)$ の $n\times n$ 行列 $\mathbf{K}=\left[K\left(\mathbf{x}_i,\mathbf{x}_j\right)\right]_{i,j=1}^{n}$ である．カーネル関数 $K\left(\mathbf{x}_i,\mathbf{x}_j\right)$ は，入力空間における 2 点 $\mathbf{x},\mathbf{x}'$ について定義され，

$$\begin{cases}\text{ユークリッド内積:}K\left(\mathbf{x},\mathbf{x}'\right)=\left\langle\mathbf{x},\mathbf{x}'\right\rangle=\mathbf{x}^t\mathbf{x}'\\ \text{d 次多項式:}K\left(\mathbf{x},\mathbf{x}'\right)=\left(1+\left\langle\mathbf{x},\mathbf{x}'\right\rangle\right)^d=\left(1+\mathbf{x}^t\mathbf{x}'\right)^d\\ \text{動径基底:}K\left(\mathbf{x},\mathbf{x}'\right)=\exp\left(-\gamma\left\|\mathbf{x}-\mathbf{x}'\right\|^2\right),\left\|\mathbf{x}-\mathbf{x}'\right\|^2=\left\langle\mathbf{x}-\mathbf{x}',\mathbf{x}-\mathbf{x}'\right\rangle\\ \text{ニューラルネット:}K\left(\mathbf{x},\mathbf{x}'\right)=\tanh\left(\omega_1\left\langle\mathbf{x},\mathbf{x}'\right\rangle+\omega_2\right)=\tanh\left(\omega_1\mathbf{x}^t\mathbf{x}'+\omega_2\right)\end{cases} \tag{5.10}$$

などがある．カーネル関数は 2 点 $\mathbf{x},\mathbf{x}'$ の**類似度** (similarity measure) を表している．よって，カーネルロジスティック判別モデル

$$f\left(\mathbf{x}\right)=\ln\left\{\frac{p\left(y=+1\left|\mathbf{x}\right.\right)}{p\left(y=-1\left|\mathbf{x}\right.\right)}\right\}=\sum_{i=1}^{n}\alpha_i K\left(\mathbf{x},x_i\right)+b \tag{5.11}$$

を得る．$y=+1$ のクラスに属する確率は

$$\hat{p}\left(y=+1\left|\mathbf{x}\right.\right)=\frac{1}{1+\exp\left[-\left\{\sum_{i=1}^{n}\hat{\alpha}_i K\left(\mathbf{x},x_i\right)+\hat{b}\right\}\right]} \tag{5.12}$$

で推定される．

(5.11) 式のカーネル関数 $K\left(\mathbf{x},x_i\right)$ として，動径基底関数

$$K\left(\mathbf{x},\mathbf{x}'\right)=\exp\left(-\gamma\left\|\mathbf{x}-\mathbf{x}'\right\|^2\right),\left\|\mathbf{x}-\mathbf{x}'\right\|^2=\left\langle\mathbf{x}-\mathbf{x}',\mathbf{x}-\mathbf{x}'\right\rangle \tag{5.13}$$

を採用する [10]．(5.9) 式のチューニングパラメータ λ と (5.13) 式の γ を決定するため，n **重 CV 法** (n-fold cross-validation) を採用する．n 組のデータを $\mathbf{X}=\left(\mathbf{X}_1,\mathbf{X}_2,...,\mathbf{X}_n\right)$ とする．ここに $\mathbf{X}_i=\left(x_{1i},x_{2i},...,x_{Ii},y_i\right)$，$i=1,2,\ldots,n$ とする．

ステップ 1：全データ $\mathbf{X}$ から i 番目の $\mathbf{X}_i$ を除いた $\mathbf{X}_{[i]}=\left(\mathbf{X}_1,\mathbf{X}_2,...,\mathbf{X}_{i-1},\mathbf{X}_{i+1},...,\mathbf{X}_n\right)$ を用い，(5.11) 式のカーネルロジスティック判別モデルを構築する．そして除いた i 番目のデータ $\mathbf{X}_i$ に対する (5.12) 式の予測値 $\hat{p}_{[i]}$ を算出する．

ステップ 2：すべての $i = 1, 2, ..., n$ について繰り返し，CV スコア

$$CV = \sum_{i=1}^{n} \left\{ t_i \ln \hat{p}_{[i]} + (1 - t_i) \ln \left(1 - \hat{p}_{[i]} \right) \right\} \tag{5.14}$$

を求める．

ステップ 3：CV 値が最小になる (λ, γ) の組合せをグリッド検索する．

5.2 適用例

適用例として脊柱後湾症データ (表 5.1) を取り上げる ([7] の p.103)．同表は，椎弓切除術を施した 83 例について、脊柱後湾症が認められた (群=1) か否 (群=0) かを調べたデータである．説明変数として，x_1：手術時の年齢 (月齢)，x_2：何番目の脊椎から先を手術したか，および x_3：手術した脊椎の個数を考える．

表 5.1 脊柱後湾症データ

患者番号	年齢	何番目の脊椎	脊椎の個数	群
1	71	5	3	0
2	158	14	3	0
⋮	⋮	⋮	⋮	⋮
82	42	6	7	1
83	36	13	4	0

この脊柱後湾症データについて，λ と γ を決定する n 重 CV 法のプログラムは

```
# ―――――――― プログラム#(5.1) ――――――――
DS <- read.csv("E:\\train.csv",header=F)  #データ
DS$Subj <- c(1:dim(DS)[1])
head(DS)
size <- 83.0 #データ数
gamma <- 2^(-20:11)
lambda <- (size/2)*2^(-10:-20) #チューニングパラメータ
#結果を格納するデータフレーム
ANS <- data.frame(
  rep(gamma,each=length(lambda )),
  rep(lambda,times=length(gamma)),
  numeric(length(gamma)*length(lambda)),
  numeric(length(gamma)*length(lambda)),
  numeric(length(gamma)*length(lambda))
)
names(ANS) <- c("gamma","lambda","CV","Error","ErrorRate")
#
X <- model.matrix(~V1+V2+V3,data=DS)
#応答変数
y <- DS$V4
n<-nrow(X)  #データ数
logiti <- function(x){ 1/(1+exp(-1*x))}  #逆リンク関数
Id <- unique(DS$Subj)
num <- length(Id)
LL <-numeric(num)
PP <-numeric(num)
k=1
```

```
for(l in 1:length(gamma)){
mykernel <- function(x1,x2){ exp(-gamma[l]*t(x1-x2)%*%(x1-x2)) }
K <- matrix(numeric(n*n),nrow=n)
for(i in 1:n){
  for(j in 1:n){
    K[i,j]<-mykernel(X[i,],X[j,])
  }
}
G <- cbind(K,matrix(rep(1,n),nrow=n))
R <- rbind(cbind(K,matrix(rep(0,n),nrow=n)),matrix(rep(0,n+1),nrow=1))
for(m in 1:length(lambda)){
 print(paste(l,m))
for(d in 1:num){
del<- as.numeric(rownames(DS[DS$Subj==Id[d],]))
Gc<-G[(-1*del),(-1*del)]
Rc<-R[(-1*del),(-1*del)]
Gd<-G[     del ,(-1*del)]
yc<-y[(-1*del)]
yd<-y[     del ]
theta<-numeric(n+1-length(del))
iter <- 0#反復回数
end  <- 0#終了スイッチ
while(end==0){
eta <- Gc%*%theta
pi  <- logiti(eta)
w <- pi*(1-pi)
W <- diag(c(w))
z <- eta+(yc-pi)/w
theta_new <- solve(t(Gc)%*%W%*%Gc+size*lambda[m]*Rc)%*%t(Gc)%*%W%*%z
convp <- max(abs(theta_new-theta))  #収束判別

convd<- abs(-2*sum( yc*eta-log(1+exp(eta)) -
(yc*Gc%*%theta_new-log(1+exp(Gc%*%theta_new))) ) )
if(convp<1e-6||convd<1e-4){
  end=1
}
if(iter>100000000){
  end=1
  print("Maximum number of iterations exceeded.")
 }
theta <- theta_new
iter  <- iter+1
}
eta <- Gd%*%theta
LL[d] <- -2*sum(yd*eta-log(1+exp(eta)))
PP[d]  <- logiti(eta)
}
ANS[k,]$CV <- sum(LL)
tab <- table(y, as.integer(PP>=0.5))
ANS[k,]$Error     <- sum(tab) - sum(diag(tab))
ANS[k,]$ErrorRate <- (sum(tab) - sum(diag(tab))) / sum(tab)
    k=k+1
}
}
tab
ANS
write.csv(ANS,file="E:\\KLR_CV.csv")
```

と書ける．その結果，$\left(\hat{\lambda},\hat{\gamma}\right)=\left(2^{-17},2^{-11}\right)$ となり，CV 値 $=72.603$ を得る．$\left(\hat{\lambda},\hat{\gamma}\right)=\left(2^{-17},2^{-11}\right)$ とし，(5.11) 式のカーネルロジスティック判別モデルで誤判別率を計算するプログラムは

```
# ―――――――― プログラム#(5.2) ――――――――――

DS <- read.csv("J:\\train.csv",header=F)
#チューニングパラメータ
size <- 83.0
gamma <- 2^(-11)                    # 最適値
lambda <- (83.0/2)*2^(-17)     # 最適値
X <- model.matrix(~V1+V2+V3,data=DS)
y <- DS$V4
n<-nrow(X)
mykernel <- function(x1,x2){ exp(-gamma*t(x1-x2)%*%(x1-x2)) }
logiti <- function(x){ 1/(1+exp(-1*x)) }
K <- matrix(numeric(n*n),nrow=n)
for(i in 1:n){
  for(j in 1:n){
    K[i,j]<-mykernel(X[i,],X[j,])
  }
}
G <- cbind(K,matrix(rep(1,n),nrow=n))
R <- rbind(cbind(K,matrix(rep(0,n),nrow=n)),matrix(rep(0,n+1),nrow=1))
theta<-numeric(n+1)
iter <- 0#反復回数
end  <- 0#終了スイッチ
while(end==0){
eta <- G%*%theta
pi  <- logiti(eta)
w <- pi*(1-pi)
W <- diag(c(w))
z <- eta+(y-pi)/w
theta_new <- solve(t(G)%*%W%*%G+size*lambda*R)%*%t(G)%*%W%*%z
convp <- max(abs(theta_new-theta))
convd<- abs(-2*sum( y*eta-log(1+exp(eta)) - (y*G%*%theta_new-log(1+exp(G%*%theta_new)))))
if(convp<1e-6||convd<1e-4){
  end=1
}
if(iter>100000000){
  end=1
  print("Maximum number of iterations exceeded.")
 }
theta <- theta_new
iter  <- iter+1
}
eta <- G%*%theta
pi  <- logiti(eta)
w <- pi*(1-pi)
W <- diag(c(w))
H <- G%*%solve(t(G)%*%W%*%G+size*lambda*R)%*%t(G)%*%W
df = sum(diag(H))
Dev <- -2*sum(y*eta-log(1+exp(eta)))
Dev
df
DS$pi <- pi
table(DS$V4, as.integer(DS$pi>=0.5))
```

と書ける．その結果

```
      0    1
0    62    3
1     8   10
```

を得る．よって，誤判別率は 0.205 である．線形判別の誤判別率 $(= 0.217)$ より小さくなる．

付録 A　判別関数の基本原理

分散共分散行列が，ともに $\mathbf{\Sigma}$ である多変量正規母集団 $G_1 : N\left(\mu^{(1)}, \mathbf{\Sigma}\right)$ および $G_2 : N\left(\mu^{(2)}, \mathbf{\Sigma}\right)$ から，胃潰瘍と胃癌の表 1.2 のような標本が得られているとする．このとき，新しく得られた観測値 $\mathbf{X} = \mathbf{x}$ が，G_1, G_2 のどちらに入るかを判別する．2 変数の場合なら，G_1, G_2 を境界線 z により，領域 R_1, R_2 に分割すればよい．

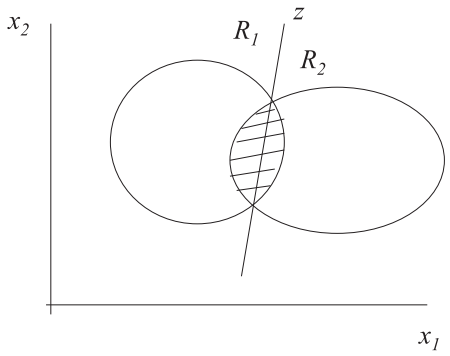

図 A.1　2 群判別

しかし，図 A.1 の斜線部では

i)$\mathbf{x}$ が G_1 からのサンプルにもかかわらず $\mathbf{x} \in R_2$ と判別する

ii)$\mathbf{x}$ が G_2 からのサンプルにもかかわらず $\mathbf{x} \in R_1$ と判別する

の 2 種類の誤り (誤判別) が起こる．*i*) の場合の損失を $c\,(2\,|1\,)$，*ii*) の場合のそれを $c\,(1\,|2\,)$ とする．表 1.2 のデータなら

i) $\mathbf{x}$ が胃潰瘍にもかかわらず胃癌と判別する

ii) $\mathbf{x}$ が胃癌にもかかわらず胃潰瘍と判別する

となる．一般には，$c\,(2\,|1\,) \neq c\,(1\,|2\,)$ である．表 1.2 のデータでは，胃癌であることを発見できなければ致命的になるが，誤って胃癌と診断しても大きな影響を及ばさない．胃癌にもかかわらず胃潰瘍と判別したときの損失 $c\,(1\,|2\,)$ は，胃潰瘍を胃癌と判別したときの損失 $c\,(2\,|1\,)$ より大きい．

さて，G_1, G_2 の母集団が多変量正規分布に従うなら，G_1, G_2 の密度関数 $f_1\,(\mathbf{x})\,, f_2\,(\mathbf{x})$ は

$$f_k\,(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\,|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}\left(\mathbf{x} - \boldsymbol{\mu}^{(k)}\right)^t \mathbf{\Sigma}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}^{(k)}\right)\right], k = 1, 2 \tag{A.1}$$

と書ける．ある観測値 $\mathbf{x}$ が G_1 から得られたとき，正しく R_1 と判別される条件付き確率は

$$\Pr\,(1\,|1\,) = \Pr\,(\mathbf{x} \in R_1\,|G_1\,) = \int_{R_1} f_1\,(\mathbf{x}) d\mathbf{x} \tag{A.2}$$

である．(A.2) 式の積分は，領域 R_1 で密度関数 $f_1\,(\mathbf{x})$ によって構成される体積を求めている．誤って R_2 と

判別される条件付き確率は

$$\Pr(2|1) = \Pr(\mathbf{x} \in R_2 | G_1) = \int_{R_2} f_1(\mathbf{x}) d\mathbf{x} \tag{A.3}$$

である．また，$\mathbf{x}$ が G_2 から得られたとき，正しく R_2 と判別される条件付き確率は

$$\Pr(2|2) = \Pr(\mathbf{x} \in R_2 | G_2) = \int_{R_2} f_2(\mathbf{x}) d\mathbf{x} \tag{A.4}$$

であり，誤って R_1 と判別される条件付き確率は

$$\Pr(1|2) = \Pr(\mathbf{x} \in R_1 | G_2) = \int_{R_1} f_2(\mathbf{x}) d\mathbf{x} \tag{A.5}$$

である．更に，胃潰瘍と胃癌の疾病率には大きな差があり，同等に扱うのが理に合わない (すなわち，胃潰瘍の母集団のほうが胃癌のそれより大きいため，胃潰瘍の可能性が高い) 場合，それらを**事前確率** (**事前**の生起**確率**:a priori probability)π_1, π_2 $(\pi_1 + \pi_2 = 1)$ として与える．すなわち，新しく得られた観測値 $\mathbf{X} = \mathbf{x}$ が，観測する以前にもっている $\mathbf{x} \in G_k$ となる確率

$$\pi_1 = \Pr\{\mathbf{x} \in G_1\}, \pi_2 = \Pr\{\mathbf{x} \in G_2\} \tag{A.6}$$

である．

よって，観測値 $\mathbf{X} = \mathbf{x}$ が第 1 母集団 G_1 から取られたとき，正しく判別される確率は，事前確率と条件付き確率との積

$$\begin{aligned} \Pr(G_1\text{として正しく判別される}) &= \Pr(\text{観測値が } G_1\text{から取られ，かつ } G_1\text{として正しく判別される}) \\ &= \Pr(\mathbf{x} \in G_1) \Pr(\mathbf{x} \in R_1 | G_1) = \pi_1 \Pr(1|1) \end{aligned} \tag{A.7}$$

で与えられる．誤判別される確率は $\pi_1 \Pr(2|1)$ となる．また，観測値 $\mathbf{X} = \mathbf{x}$ が第 2 母集団 G_2 から取られたとき，正しく判別される確率は $\pi_2 \Pr(2|2)$ で，誤判別される確率は $\pi_2 \Pr(1|2)$ となる．よって，誤判別したときの期待損失 (損失の期待値) は

$$c(2|1) \pi_1 \Pr(2|1) + c(1|2) \pi_2 \Pr(1|2) \tag{A.8}$$

となる．ゆえに，(A.3)，(A.5) 式を (A.8) 式へ代入すると，

$$c(2|1) \pi_1 \int_{R_2} f_1(\mathbf{x}) dx + c(1|2) \pi_2 \int_{R_1} f_2(\mathbf{x}) d\mathbf{x} = \int_{R_2} \{\pi_1 c(2|1) f_1(\mathbf{x}) - \pi_2 c(1|2) f_2(\mathbf{x})\} d\mathbf{x} + \pi_2 c(1|2) \tag{A.9}$$

を得る．右辺の第 2 項は，定数であるから第 1 項が最小になるのは R_2 が

$$\pi_1 c(2|1) f_1(\mathbf{x}) - \pi_2 c(1|2) f_2(\mathbf{x}) < 0 \tag{A.10}$$

の点 $\mathbf{x}$ のみを含む場合である．すなわち，(A.10) 式が最小になる領域 R_2 として

$$R_2 = \left\{ x \,\middle|\, \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{c(1|2) \pi_2}{c(2|1) \pi_1} \right\} \tag{A.11}$$

を得る．同様に

$$R_1 = \left\{ \mathbf{x} \,\middle|\, \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{c(1|2) \pi_2}{c(2|1) \pi_1} \right\} \tag{A.12}$$

となる．R_1 と R_2 を決める際，事前確率や損失の値そのものではなく，それぞれの比が分かればよい (ただし，事前確率の比は，密度関数 $f_k(\mathbf{x})$ の比の順序と逆である)．よって，期待損失 (A.10) 式を最小にする領域 R_1 と R_2 は

$$\begin{cases} R_1 : \ln\left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\right) \geq \ln\kappa \\ R_2 : \ln\left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\right) < \ln\kappa \end{cases} \tag{A.13}$$

で与えられる．ここに

$$\kappa \equiv \frac{c\,(1\,|2\,)\,\pi_2}{c\,(2\,|1\,)\,\pi_1}$$

を判別の**分岐点** (cut-off point) という．この κ は

$$\kappa = \begin{cases} c(1 \mid 2)/c(2 \mid 1) : \text{事前確率が等しい場合} \\ \pi_1/\pi_2 : \text{損失が等しい場合} \\ 1 : \text{事前確率が等しく，損失も等しい場合} \end{cases} \tag{A.14}$$

となる．

また，次のようにして (A.13) 式を導くこともできる．新しく得られた観測値 $\mathbf{X} = \mathbf{x}$ は「最大の事後確率をもつ母集団に属する」とみなす．**ベイズの定理**より観測値 $\mathbf{X} = \mathbf{x}$ を得たとき，これが母集団 G_1 からの観測値である条件付き確率 (事後確率)$\Pr(G_1\,|\mathbf{x}\,)$ は

$$\begin{aligned} \Pr(G_1\,|\mathbf{x}\,) &= \frac{\Pr(G_1\text{が起こり，かつ }\mathbf{x}\text{ を観測})}{\mathbf{Pr}(\mathbf{x}\text{ を観測})} \\ &= \frac{\Pr\left(\mathbf{x}\text{ を観測}\middle|\mathbf{G_1}\right)\Pr(G_1)}{\mathbf{Pr}(\mathbf{x}\text{ を観測}\ |\mathbf{G_1}\,)\Pr(G_1)+\Pr(\mathbf{x}\text{ を観測}\ |G_2\,)\Pr(G_2)} \\ &= \frac{\pi_1 f_1(\mathbf{x})}{\pi_1 f_1(\mathbf{x})+\pi_2 f_2(\mathbf{x})} \end{aligned} \tag{A.15}$$

となる．同様に，

$$\Pr(G_2\,|\mathbf{x}\,) = \frac{\pi_2 f_2\,(\mathbf{x})}{\pi_1 f_1\,(x) + \pi_2 f_2\,(\mathbf{x})} \tag{A.16}$$

を得る．ゆえに，事後確率の大きい群に属すると考えると

$$\Pr(G_1\,|\mathbf{x}\,) \geq \Pr(G_2\,|\mathbf{x}\,)$$

すなわち

$$\pi_1 f_1\,(\mathbf{x}) \geq \pi_2 f_2\,(\mathbf{x})$$

のとき，$\mathbf{x}$ は G_1 に属するとみなす．同様に，

$$\pi_1 f_1\,(\mathbf{x}) < \pi_2 f_2\,(\mathbf{x})$$

のとき，$\mathbf{x}$ は G_2 に属するとみなす．このような決め方を**ベイズの判別ルール** (Bayes discriminant rule) と呼ぶ．これは，R_1, R_2 において，損失が等しい (すなわち，$c\,(1\,|2\,) = c\,(2\,|1\,)$ のとき) 場合と同等になる．更に $\kappa = 1$ の場合，

$$\begin{cases} f_1\,(\mathbf{x}) < f_2\,(\mathbf{x}) \text{ なら } G_1\text{に属する} \\ f_1\,(\mathbf{x}) \geq f_2\,(\mathbf{x}) \text{ なら } G_2\text{に属する} \end{cases} \tag{A.17}$$

と判別する．

(A.13) 式に (A.1) 式を代入すると

$$\begin{aligned} \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &= \frac{\exp\left[-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}^{(1)}\right)^t\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}^{(1)}\right)\right]}{\exp\left[-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}^{(2)}\right)^t\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}^{(2)}\right)\right]} \\ &= \exp\left[-\frac{1}{2}\left\{\left(\mathbf{x}-\boldsymbol{\mu}^{(1)}\right)^t\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}^{(1)}\right) - \left(\mathbf{x}-\boldsymbol{\mu}^{(2)}\right)^t\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}^{(2)}\right)\right\}\right] \\ &= \exp\left[\mathbf{x}^t\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}\right) - \frac{1}{2}\left(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)}\right)^t\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}\right)\right] \end{aligned} \tag{A.18}$$

を得る．よって，(A.13) 式と (A.14) 式は

$$\begin{cases} R_1 = \left\{ \mathbf{x} \left| \mathbf{x}^t \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) - \frac{1}{2} \left(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)} \right)^t \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) \geq \ln k \right. \right\} \\ R_2 = \left\{ \mathbf{x} \left| \mathbf{x}^t \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) - \frac{1}{2} \left(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)} \right)^t \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) < \ln k \right. \right\} \end{cases} \tag{A.19}$$

となる．ここで，

$$z^* = \mathbf{x}^t \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) - \frac{1}{2} \left(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)} \right)^t \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) \tag{A.20}$$

を**真の線形判別関数** (true linear discriminant function) と呼ぶ．

次に，$f(\mathbf{x})$ に多変量正規分布を仮定できるなら，データに依存しない誤判別確率 $P(i|j)$ を推定できる ([10] の 2.1.3 節)．$\mathbf{X}$ は多変量正規分布に従うため，その 1 次結合である (A.11) 式も多変量正規分布になる．まず，新しく得られた観測値 $\mathbf{X}$=$\mathbf{x}$ が母集団 G_1 に属する場合，

$$\begin{cases} E_1[z^*] = \boldsymbol{\mu}^{(1)^t} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) - \frac{1}{2} \left(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)} \right)^t \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) \\ \quad = \frac{1}{2} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right)^t \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) \\ Var[z^*] = \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right)^t \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) \end{cases} \tag{A.21}$$

となる．ここで，

$$\Delta^{*2} \equiv \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right)^t \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} \right) \tag{A.22}$$

とおけば

$$z^* \sim N \left(\frac{1}{2} \Delta^{*2}, \Delta^{*2} \right) \tag{A.23}$$

を得る．(A.22) 式の推定量は (1.63) 式で与えられる．同様に，$\boldsymbol{X}$=$\boldsymbol{x}$ が母集団 G_2 に属する場合，

$$z^* \sim N \left(-\frac{1}{2} \Delta^{*2}, \Delta^{*2} \right) \tag{A.24}$$

となる．(A.22) 式の Δ^{*2} は $\boldsymbol{\mu}^{(1)}$ と $\boldsymbol{\mu}^{(2)}$ との間のマハラノビスの平方距離で，判別の良さ (判別力) を示す 1 つの指標となる．Δ^{*2} が大きいほど，2 つの群の重心間のマハラノビスの汎距離が離れており，判別の効率が良くなる．

ゆえに，

$$\Pr(2|1) = \int_{R_2} f_1(\mathbf{x}) d\mathbf{x} = \frac{1}{\sqrt{2\pi}\Delta^*} \int_{-\infty}^{\kappa} \exp \left\{ -\frac{1}{2} \left(\frac{u - \frac{1}{2}\Delta^{*2}}{\Delta^*} \right) \right\} du = \Phi \left(\frac{\kappa^* - \frac{1}{2}\Delta^{*2}}{\Delta^*} \right) \tag{A.25}$$

$$\Pr(1|2) = \int_{R_1} f_2(\mathbf{x}) d\mathbf{x} = \frac{1}{\sqrt{2\pi}\Delta^*} \int_{\kappa}^{\infty} \exp \left\{ -\frac{1}{2} \left(\frac{u + \frac{1}{2}\Delta^{*2}}{\Delta^*} \right) \right\} du = 1 - \Phi \left(\frac{\kappa^* + \frac{1}{2}\Delta^{*2}}{\Delta^*} \right) \tag{A.26}$$

となる．ただし，$\kappa^* = \ln \kappa$ とし，$\Phi(z^*)$ は標準正規分布で

$$\Phi(z^*) = \int_{-\infty}^{z^*} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} u^2 \right) du \tag{A.27}$$

である．よって，期待される誤判別率は

$$P = \pi_1 \Phi \left(\frac{\kappa^* - \frac{1}{2}\Delta^{*2}}{\Delta^*} \right) + \pi_2 \left\{ 1 - \Phi \left(\frac{\kappa^* + \frac{1}{2}\Delta^{*2}}{\Delta^*} \right) \right\} \tag{A.28}$$

となる．特に，$\pi_1=\pi_2$，$c\,(2\,|1\,)=c\,(1\,|2\,)$ なら，(A.27) 式の κ は 1 となるから

$$P=\frac{1}{2}\pi_1\Phi\left(-\frac{\Delta^*}{2}\right)+\frac{1}{2}\pi_2\left\{1-\Phi\left(\frac{\Delta^*}{2}\right)\right\}=1-\Phi\left(\frac{\Delta^*}{2}\right) \tag{A.29}$$

を得る．

母平均 $\boldsymbol{\mu}^{(1)}=\left(\mu_1^{(1)},\boldsymbol{\mu}_2^{(1)},...,\mu_I^{(1)}\right)$，$\boldsymbol{\mu}^{(2)}=\left(\mu_1^{(2)},\mu_2^{(2)},...,\mu_I^{(2)}\right)$ および $\mathbf{S}$ の不偏推定量は

$$\hat{\mu}_i^{(1)}=\frac{1}{n_1}\sum_{j=1}^{n_1}x_{ij}^{(1)},\ \hat{\mu}_i^{(2)}=\frac{1}{n_2}\sum_{j=1}^{n_2}x_{ij}^{(2)},\ \hat{\mathbf{S}}=\mathbf{S}=\left(s_{jj'}\right) \tag{A.30}$$

となる．よって，(A.29) 式を (A.19) 式に代入 (plug-in) すると (1.32) 式が得られる．これを，**標本線形判別関数** (sample linear discriminant function) という．

表 1.2 のデータについて，$c\,(1\,|2\,)=c\,(2\,|1\,),\Pr\,(G_1)=\Pr\,(G_2)=1/2$ の場合，(A.28) 式の誤判別率を求めよう．(A.22) 式の推定値は (1.63) 式から $\Delta^2=4.140$ となり，$\kappa^*=\ln\kappa=0$ であるから

$$\frac{\kappa^*-\frac{1}{2}\Delta^2}{\Delta}=-1.02 \tag{A.31}$$

を得る．ゆえに，(A.25)，(A.26) 式より

$$\begin{cases}\Pr\,(2\,|1\,)=\Phi\,(-1.02)=0.154\\ \Pr\,(1\,|2\,)=0.154\end{cases} \tag{A.32}$$

となり，**期待される誤判別率**は

$$P=\frac{1}{2}\times 0.154+\frac{1}{2}\times 0.154=0.154 \tag{A.33}$$

と推定される．

付録 B　相関比に基づく線形判別式の導出

Fisher の線形判別関数では，I 個の説明変数について，1 次式

$$z=a_0+a_1x_1+a_2x_2+\cdots+a_Ix_I \tag{B.1}$$

の係数 $a_0,a_1,a_2,...,a_I$ を観測されたデータから求めた．この $a_0,a_1,a_2,...,a_I$ が定まると 2 つの群の観測値から，判別スコア

$$z_i^{(k)}=a_0+a_1x_{1i}^{(k)}+a_2x_{2i}^{(k)}+\cdots+a_Ix_{Ii}^{(k)}\ \ i=1,2,...,n_k;k=1,2 \tag{B.2}$$

を計算する．第 k 群における標本平均ベクトルと標本分散共分散行列は

$$\begin{cases}\text{標本平均ベクトル:}\bar{\mathbf{x}}_k=\frac{1}{n_k}\sum\limits_{i=1}^{n_k}\mathbf{x}_i^{(k)}\\ \text{標本分散共分散行列:}\mathbf{S}_k=\frac{1}{n_k-1}\sum\limits_{i=1}^{n_k}\left(\mathbf{x}_i^{(k)}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_i^{(k)}-\bar{\mathbf{x}}_i\right)^t\end{cases} \tag{B.3}$$

と書ける．よって，(B.2) 式の左辺の標本平均と標本分散は

$$\begin{cases}\text{標本平均}:\bar{z}^{(k)}=\frac{1}{n_k}\sum\limits_{i=1}^{n_k}z_i^{(k)}=\frac{1}{n_k}\sum\limits_{i=1}^{n_k}\mathbf{a}^t\mathbf{x}_i^{(k)}=\mathbf{a}^t\bar{\mathbf{x}}_k\\ \text{標本分散:}\frac{1}{n_k-1}\sum\limits_{i=1}^{n_k}\left(z_i^{(k)}-\bar{z}^{(k)}\right)^2=\mathbf{a}^t\left\{\frac{1}{n_k-1}\sum\limits_{i=1}^{n_k}\left(\mathbf{x}_i^{(k)}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_i^{(k)}-\bar{\mathbf{x}}_i\right)^t\right\}\mathbf{a}=\mathbf{a}^t\mathbf{S}_j\mathbf{a}\end{cases} \tag{B.4}$$

となる．また，全データの標本平均は

$$\bar{z}=\frac{1}{n}\sum_{i=1}^{n_k} z_i^{(k)} =\frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n_k} z_i^{(k)} = \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n_k} \mathbf{a}^t\mathbf{x}_i^{(k)} =\mathbf{a}^t\bar{\mathbf{x}} \tag{B.5}$$

となる．

(B.2) 式の判別スコア $z_i^{(k)}$ の変動を調べる．そのため，分散分析でよく用いられる平方和の分解

$$\underbrace{\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(z_i^{(k)}-\bar{z}\right)^2}_{\text{総平方和 } S_T} = \underbrace{\sum_{k=1}^{K} n_k\left(\bar{z}^{(k)}-\bar{z}\right)^2}_{\text{群間平方和 } S_B} + \underbrace{\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(z_i^{(k)}-\bar{z}^{(k)}\right)^2}_{\text{群内平方和 } S_W} \tag{B.6}$$

を行う．ここで，総平均 $\bar{z}$ は $z_1^{(1)},...,z_I^{(1)},...,z_1^{(K)},...,z_I^{(K)}$ の平均，$\bar{z}^{(k)}$ は，第 k 群の平均である．(B.6) 式の左辺の S_T は，各個体の判別スコア $z_i^{(k)}$ と判別スコアの総平均 $\bar{z}$ との差の平方和 (総平方和) で，全変動を表す．(B.6) 式の右辺第 1 項の S_B は，第 k 群の判別スコアの $\bar{z}^{(k)}$ と全スコアの総平均 $\bar{z}$ との差の平方和 (群間平方和) で，群間の変動を表す．S_W は各個体の判別スコア $z_i^{(k)}$ と第 k 群の判別スコアの $\bar{z}^{(k)}$ との差の平方和 (群内平方和) で，群内の変動を表す．全変動を

$$\text{全変動 } (S_T) = \text{群間変動 } (S_B) + \text{群内変動 } (S_W) \tag{B.7}$$

に分解する．よって，2 つの群をうまく分離させるためには，全体での変動の中で，群間の変動をできるだけ大きくすればよい．群間平方和 S_B の総平方和 S_T に対する相対的な大きさ，すなわち相関比

$$\eta^2 = S_B/S_T \tag{B.8}$$

が最大になる係数 $a_1,a_2,...,a_I$ を求める．それは，群間平方和 S_B と群内平方和 S_W の比

$$\lambda = S_B/S_W \tag{B.9}$$

を最大化することと同等である．すなわち，群間平方和をできるだけ大きく，群内平方和をできるだけ小さくなるように $a_1,a_2,...,a_I$ を求める．なお，(B.8) 式と (B.9) 式との間には

$$\eta^2 = \frac{S_B}{S_T} = \frac{S_B}{S_B+S_W} = \frac{\lambda}{1+\lambda} \tag{B.10}$$

が成り立つ．

群間平方和は

$$S_B = \sum_{k=1}^{K} n_k\left(\bar{z}^{(k)}-\bar{z}\right)^2 = \mathbf{a}^t\left\{\sum_{k=1}^{K} n_k\left(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}\right)\left(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}\right)^t\right\}\mathbf{a} = \mathbf{a}^t\mathbf{B}\mathbf{a} \tag{B.11}$$

となる．ただし $\mathbf{B} = \sum_{k=1}^{K} n_k\left(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}\right)\left(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}\right)^t$ は，群間平方和積和行列である．さらに群内平方和は

$$\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(z_i^{(k)}-\bar{z}^{(k)}\right)^2 = \mathbf{a}^t\left\{\sum_{k=1}^{K}\left(n_k-1\right)S_k\right\}\mathbf{a} = \mathbf{a}^t\mathbf{W}\mathbf{a} \tag{B.12}$$

となる．ただし，$\mathbf{W}$ は (2.5) 式の群内平方和積和行列である．

このとき，総平均からの差をとっているから定数項 a_0 は消える．ゆえに，群間平方和 S_B の群内平方和 S_W に対する比 (Rayleigh 係数，分散比)

$$\lambda = \frac{S_B}{S_W} = \frac{a^t\mathbf{B}a}{a^t\mathbf{W}a} \tag{B.13}$$

が最大となる $a_1, a_2, ..., a_I$ を求める．これは 行列 $\mathbf{W}^{-1}\mathbf{B}$ の固有値問題に帰着される．

(B.13) 式を最大にすることは，制約条件 $S_W = 1$ のもとで $S_B = \mathbf{a}^t\mathbf{B}\mathbf{a}$ を最大にすることと同等である ([10] の 1.2 節)．よって，ラグランジュ未定乗数 λ_0 を用い

$$f = S_B - \lambda_0 (S_W - 1) = \mathbf{a}^t\mathbf{B}\mathbf{a} - \lambda_0 \left(\mathbf{a}^t\mathbf{W}\mathbf{a} - 1\right) \tag{B.14}$$

を $\mathbf{a}$ について偏微分した

$$\frac{\partial f}{\partial \mathbf{a}} = 2 (\mathbf{B} - \lambda_0\mathbf{W})\, \mathbf{a} = \mathbf{0} \tag{B.15}$$

すなわち

$$(\mathbf{B} - \lambda_0\mathbf{W})\, \mathbf{a} = \mathbf{0} \tag{B.16}$$

を解けばよい．$\mathbf{W}$ は正則行列より，(B.18) 式の両辺に $\mathbf{W}^{-1}$ を掛けると

$$\left(\mathbf{W}^{-1}\mathbf{B} - \lambda_0\mathbf{I}\right) \mathbf{a} = \mathbf{0} \tag{B.17}$$

となり，$\mathbf{a} \neq \mathbf{0}$ より固有方程式

$$\left|\mathbf{W}^{-1}\mathbf{B} - \lambda_0\mathbf{I}\right| = \mathbf{0} \tag{B.18}$$

を得る．(B.18) 式は $r = \min(K-1, I)$ の非負の固有値 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$ をもつ．(B.16) 式の両辺に $\mathbf{a}^t$ を掛けると

$$\mathbf{a}^t\mathbf{B}\mathbf{a} - \lambda_0\mathbf{a}^t\mathbf{W}\mathbf{a} = \mathbf{0} \tag{B.19}$$

となり

$$\lambda_0 = \frac{\mathbf{a}^t\mathbf{B}\mathbf{a}}{\mathbf{a}^t\mathbf{W}\mathbf{a}} = \frac{S_B}{S_W} = \lambda \tag{B.20}$$

と書け，ラグランジュ未定乗数 λ_0 が Rayleigh 係数 λ である．よって，最大固有値 λ_1 に対する固有ベクトルが線形判別関数の係数 $\mathbf{a}$ と一致する．

次に，相関比に基づく線形判別関数の導出について述べる ([11] の 4.2 節，[8] の pp.284-287)．表 1.2 のデータの判別スコアは，表 B.1 のようにまとめられ，

$$\bar{z} = a_0 + 0.565a_1 + 0.57a_2 \tag{B.21}$$

となる．

(B.6) 式の S_T，S_B，S_W を書き下すと

$$S_T = \left(z_1^{(1)} - \bar{z}\right)^2 + \cdots + \left(z_I^{(1)} - \bar{z}\right)^2 + \left(z_1^{(2)} - \bar{z}\right)^2 + \cdots + \left(z_I^{(2)} - \bar{z}\right)^2 \tag{B.22}$$

$$S_B = 10 \times \left(\bar{z}^{(1)} - \bar{z}\right)^2 + 10 \times \left(\bar{z}^{(2)} - \bar{z}\right)^2 \tag{B.23}$$

$$\begin{aligned} S_W = &\left(z_1^{(1)} - \bar{z}^{(1)}\right)^2 + \left(z_2^{(1)} - \bar{z}^{(1)}\right)^2 + \cdots + \left(z_{10}^{(1)} - \bar{z}^{(1)}\right)^2 \\ &+ \left(z_1^{(2)} - \bar{z}^{(2)}\right)^2 + \left(z_2^{(2)} - \bar{z}^{(2)}\right)^2 + \cdots + \left(z_{10}^{(2)} - \bar{z}^{(2)}\right)^2 \end{aligned} \tag{B.24}$$

となる．(B.8) 式の相関比 η^2

$$\eta^2 = S_B / S_T \tag{B.25}$$

が最大になる a_1，a_2 を求める．そこで，η^2 を a_1, a_2 について偏微分して 0 とおいた連立方程式

$$\begin{cases} s_{11}a_1 + s_{12}a_2 = \bar{x}_1^{(1)} - \bar{x}_1^{(2)} \\ s_{12}a_1 + s_{22}a_2 = \bar{x}_2^{(1)} - \bar{x}_2^{(2)} \end{cases} \tag{B.26}$$

表 B.1　判別スコア

第 1 群の判別スコア	第 2 群の判別スコア
$z_1^{(1)} = a_0 + 0.1a_1 + 0.4a_2$	$z_1^{(2)} = a_0 + 0.8a_1 + 0.3a_2$
$z_2^{(1)} = a_0 + 0.9a_1 + 0.8a_2$	$z_2^{(2)} = a_0 + 0.8a_1 + 0.4a_2$
$z_3^{(1)} = a_0 + 0.2a_1 + 0.8a_2$	$z_3^{(1)} = a_0 + 0.7a_1 + 0.6a_2$
$z_4^{(1)} = a_0 + 0.2a_1 + 0.5a_2$	$z_4^{(2)} = a_0 + 0.6a_1 + 0.2a_2$
$z_5^{(1)} = a_0 + 0.6a_1 + 0.8a_2$	$z_5^{(2)} = a_0 + 0.7a_1 + 0.4a_2$
$z_6^{(1)} = a_0 + 0.7a_1 + 0.9a_2$	$z_6^{(2)} = a_0 + 0.5a_1 + 0.8a_2$
$z_7^{(1)} = a_0 + 0.3a_1 + 0.8a_2$	$z_7^{(2)} = a_0 + 0.8a_1 + 0.6a_2$
$z_8^{(1)} = a_0 + 0.3a_1 + 0.7a_2$	$z_8^{(1)} = a_0 + 0.9a_1 + 0.3a_2$
$z_9^{(1)} = a_0 + 0.3a_1 + 0.5a_2$	$z_9^{(1)} = a_0 + 0.9a_1 + 0.4a_2$
$z_{10}^{(1)} = a_0 + 0.1a_1 + 0.6a_2$	$z_{10}^{(1)} = a_0 + 0.9a_1 + 0.6a_2$
$\bar{z}^{(1)} = a_0 + 0.37a_1 + 0.68a_2$	$\bar{z}^{(2)} = a_0 + 0.76a_1 + 0.46a_2$

を解けばよい．ただし，

$$s_{ij} = \frac{1}{n_1 + n_2 - 2} \left\{ \sum_{k=1}^{n_1} \left(x_{ik}^{(1)} - \bar{x}_i^{(1)} \right) \left(x_{jk}^{(1)} - \bar{x}_j^{(1)} \right) + \sum_{k=1}^{n_2} \left(x_{ik}^{(2)} - \bar{x}_i^{(2)} \right) \left(x_{jk}^{(2)} - \bar{x}_j^{(2)} \right) \right\} \tag{B.27}$$

とし，定数項 a_0 は (1.44) 式から求める．これは，

$$a_0 + a_1 x_1 + a_2 x_2 \begin{cases} \geq 0 : \text{第 1 群に属する} \\ < 0 : \text{第 2 群に属する} \end{cases} \tag{B.28}$$

とすれば，(1.23) 式に一致する．相関比に基づく判別関数の導出は [8] に詳しい．

付録 C　判別効率に基づく変数選択

q 個の説明変数 $(x_1, x_2, ..., x_q)$ を用いた判別効率を，(1.63) 式から

$$\Delta_q^2 = \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)} \right)^t \mathbf{S}^{-1} \left(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)} \right) \tag{C.1}$$

と書く．(C.1) 式は，2 つの群の平均 $\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}$ 間のマハラノビス平方距離になる．ここに $\mathbf{S}$ は，(1.27) 式の 2 つの群をプールしたときの分散共分散行列である．いま，新たに r 個の変量 $(x_{q+1}, x_{q+2}, ..., x_{q+r})$ を追加したときの判別効率を Δ_{q+r}^2 とすると

$$\Delta_{q+r}^2 - \Delta_q^2 \geq 0$$

で，r 個の説明変数を追加することにより，常に判別効率は増加する．しかし，この増分は単に誤差変動によるのか，追加した変量が真に判別に寄与しているのかを調べなければならない．そこで，

$$\text{帰無仮説 } H_0 : \Delta_{q+r}^2 = \Delta_q^2 \,(\text{すなわち，} x_{q+1}, x_{q+2}, ..., x_{q+r} \text{は判別に寄与していない})$$

のもとで，

$$F = \frac{f - q - r + 1}{r} \times \frac{\Delta_{q+r}^2 - \Delta_q^2}{\frac{f(f+2)}{n_1 + n_2} + \Delta_q^2} \tag{C.2}$$

は，自由度 $(r, f-q-r+1)$ の F 分布に従う．ただし，$f = n_1 + n_2 - 2$ である．

変数選択では，1 個づつ変数を増加あるいは減少させるので，$r = 1$ である．よって，

$$F = \frac{(n-I)\left(\Delta_{q+1}^2 - \Delta_q^2\right)}{\frac{n(n+2)}{n_1+n_2} + \Delta_q^2} \tag{C.3}$$

が，自由度 $(1, f-q)$ の F 分布に従うことを利用する．重回帰分析と同様に，棄却点は F 分布の上側 15% 点，あるいは f が大きければ $F = 2.0$ (F 分布の上側 16% あるいは 17% 点に対応) にとる ([8] の 2.3 節)．

適用例 (1)

がくの長さ，がくの巾，花弁の長さ，花弁の巾の測定値からアヤメ科の 2 種類の植物，イリスヴェルシコール，イリスヴィルジニカを判別する．データは“iris”と R で入力すれば得られる．ここでは，変数増加法で変数選択を行う．

ステップ 1 : 表 C.1 は，4 つの説明変数それぞれについての (C.3) 式の F 値である．例えば，x_4 に対する F 値は，(C.3) 式において $n_1 = n_2 = 50$, $I = 0, \Delta_1^2 = 8.556, \Delta_0^2 = 0$ より

$$F = \frac{(50+50-2)\times(8.556-0)}{\frac{100\times 98}{50+50} + 0} = 213.901$$

となる．ステップ 1 では説明変数が 0 個から 1 個増やすので $q = 0, r = 1$ とする．

表 C.1　F 値

変数	F 値
x_1	31.688
x_2	10.277
x_3	158.855
x_4	213.901

ステップ 2 : 表 C.1 で F 値が最大 (213.901 で 2.0 より大きい) となる x_4 を選択する．このとき，Fisher の線形判別関数は

$$z = 20.486 - 12.223x_4$$

となる．x_4 を取入れた後の F 値は，表 C.2 で与えられる．

表 C.2　x_4 を取入れた後の F 値

変数	F 値
x_1	0.003
x_2	13.370
x_3	11.754

ステップ 3 : 表 C.2 より，x_2 を取入れる (F 値が 13.370 で 2.0 より大きい)．このとき線形判別関数は

$$z = 12.508 + 5.074x_2 - 16.158x_4$$

で与えられる．x_2 を取入れた後の F 値は表 C.3 となる．

表 C.3　x_4, x_2 を取入れた後の F 値

変数	F 値
x_1	1.798
x_3	17.838

ステップ 4：表 C.3 より，x_3 を取入れる (F 値が 17.838 で 2.0 より大きい)．このとき線形判別関数は

$$z = 21.664 + 6.757x_2 - 3.780x_3 - 13.439x_4$$

で与えられる．x_4, x_2, x_3 を取入れた後の F 値は，表 C.4 となる．

表 C.4　x_4, x_2, x_3 を取入れた後の F 値

変数	F 値
x_1	7.368

ステップ 5：表 C.4 より，残りの x_1 に対する F 値は 7.368 で 2.0 より大きいので x_1 も取入れる．このとき線形判別関数は

$$z = 16.663 + 3.556x_1 + 5.579x_2 - 6.970x_3 - 12.386x_4$$

で与えられる．最終的な F 値を表 C.5 に与えておく．参考までに，見かけ上の誤判別の個数を表 C.6 に示す．

表 C.5　x_4, x_2x_3, x_1 を取入れた後の F 値

変数	F 値
x_1	7.368
x_2	10.587
x_3	24.157
x_4	37.092

表 C.6　誤判別表

もとの群＼予測群	第 1 群	第 2 群
第 1 群	48	2
第 2 群	1	49

適用例 (2)

表 2.2 の糖尿病データについて，変数増加法を適用してみよう (変数減少法は，この逆の手順をふめばよい)．まず，説明変数間の相関係数は，表 C.7 のようになる．(x_2, x_3), (x_2, x_5), (x_3, x_5) の間に高い相関がみられる．

ステップ 1：各変数に対する (C.3) 式の F 値を計算すると表 C.8 のようになる．

表 C.7　相関係数表

	x_1	x_2	x_3	x_4	x_5
x_1	1.0	-0.009	0.024	0.222	0.384
x_2	-	1.0	0.965	-0.396	0.715
x_3	-	-	1.0	-0.337	0.771
x_4	-	-	-	1.0	0.008

表 C.8　各変数に対する F 値

変数	F 値
x_1	11.909
x_2	142.403
x_3	242.839
x_4	28.240
x_5	113.288

よって，変数 x_3 が取り込まれる．Wilks の $\Lambda = 0.226$ より，(2.8) 式は $F = 242.839 > F\,(0.01; 2, 142)$ となり，判別式は高度に有意である．

ステップ 2 : 変数 x_3 の次に取り込むべき変数を選択するため，(C.3) 式の F 値を計算すると表 C.9 のようになる．よって，変数 x_2 が取り込まれる．$\Lambda = 0.153$ より，(2.8) 式は $F = 109.407 > F\,(0.01; 4, 282)$ となり，判別式は高度に有意である．

表 C.9　変数 x_3 の次に取り込むべき変数を選択するための F 値

変数	F 値
x_1	13.960
x_2	33.362
x_4	22.646
x_5	15.023

ステップ 3 : 変数 x_2, x_3 の次に取り込むべき変数を選択するため，(C.3) 式の F 値を計算すると表 C.10 のようになる．よって，変数 x_4 が取り込まれる．$\Lambda = 0.126$ より，(2.8) 式は $F = 84.676 > F\,(0.01; 6, 280)$ と

表 C.10　変数 x_2, x_3 の次に取り込むべき変数を選択するための F 値

変数	F 値
x_1	14.322
x_4	15.150
x_5	13.241

なり，判別式は高度に有意である．

ステップ 4 : 変数 x_2, x_3, x_4 の次に取り込むべき変数を選択するため，(C.3) 式の F 値を計算すると表 C.11

のようになる．よって，変数 x_1 が取り込まれる．$\Lambda = 0.109$ より，(2.8) 式は $F = 70.272 > F\,(0.01; 8, 278)$ となり，判別式は高度に有意である．

表 C.11　変数 x_2, x_3, x_4 の次に取り込むべき変数を選択するための偏 F 値

変数	F 値
x_1	10.637
x_5	7.648

ステップ 5：変数 x_2, x_3, x_4, x_1 の次に取り込むべき変数を選択するため，(C.3) 式の F 値を計算すると表 C.12 のようになる．

表 C.12　変数 x_2, x_3, x_4, x_1 の次に取り込むべき変数を選択するための F 値

変数	F 値
x_5	2.801

よって，x_5 が取り込まれる．$\Lambda = 0.105$ より，(C.3) 式は $F = 57.489 > F\,(0.01; 10, 276)$ となり，判別式は高度に有意である．

付録 D　2 つの群の大きさが異なる場合のベイズ判別

2 つの群の大きさが異なる場合の線形判別について述べる．例えば，胃潰瘍患者群は胃癌患者群より母集団の人数は多い. この疾病率の差を判別関数に取り込むため事前確率を用いる．I 個の説明変数が観測されているとする．第 1 群と第 2 群に対応するクラス (母集団)C_1，C_2 から，観測ベクトル $\mathbf{x}^t = (x_1, x_2, ..., x_I)$ が抽出される事前確率 $Pr(C_1)$ を π_1, π_2 とする．クラス C_k において，観測ベクトル $\mathbf{x}^t$ が確率密度 $f_k(\mathbf{x})$ の分布に従うなら，$k = 1, 2$ について (4.1) 式となる．$\mathbf{x}$ がクラス C_k から得られる事後確率は，ベイズの定理を用いると

$$\Pr\left(C_k \,|\mathbf{x}\right) = \frac{\pi_k f_k(\mathbf{x})}{\pi_1 f_1\left(\mathbf{x}\right) + \pi_2 f_2\left(\mathbf{x}\right)}, k = 1, 2 \tag{D.1}$$

となる．

(D.1) 式の右辺の分母は，群 k に依存しないから，分子の $\pi_k f_k(\mathbf{x})$ を最大にすればよい．$f_k(\mathbf{x})$ が (4.9) 式の多変量正規分布に従うとき，$f_k(\mathbf{x})$ の右辺の中の $\frac{1}{(2\pi)^{n/2}\left|\mathbf{\Sigma}\right|^{1/2}}$ も k に依存しない．(4.9) 式の $\mu^{(k)}, \mathbf{\Sigma}$ の推定量はそれぞれ $\hat{\mu}^{(k)} = \bar{\mathbf{x}}^{(k)}, \hat{\mathbf{\Sigma}} = \mathbf{S}$ で与えられるから，$\ln\{\pi_k f_k(\mathbf{x})\}$ すなわち

$$\ln \pi_k - \frac{1}{2} D^2_{(k)} = \ln \pi_i - \frac{1}{2}\left(\mathbf{x} - \bar{\mathbf{x}}^{(k)}\right)^t \mathbf{S}^{-1}\left(\mathbf{x} - \bar{\mathbf{x}}^{(k)}\right), k = 1, 2 \tag{D.2}$$

が最大となる群 k に属すると判別する．π_i が未知なら，各群のデータ数の割合 (proportion)，2 群の場合

$$\begin{cases} \hat{\pi}_1 = n_1/(n_1 + n_2) \\ \hat{\pi}_2 = n_2/(n_1 + n_2) \end{cases} \tag{D.3}$$

とする．

参考文献

[1] Andrews,D.F. and Herzberg, A.M.:Data-A Collection of Problems from Many Fields for the Student and resarch Worker, Springer,1985.

[2] Bartlett,M.S.(1947):Multivariate Analysis,J.Roy.Statist.Soc.,Supplement,**9**,176-197.

[3] Box,G.E.P.:A general distribution theory for a class of likelihood criteria,Biometrika,**36**,317-346.

[4] Fisher,R.A.:The use of multiple measurements in taxonomic problems, Ann. Eugen., 7,179-188,1936.

[5] Hawkins,D.M.,Topics in Applied Multivariate Analysis,Cambridge University Press,1982.
医学統計研究会訳. 多変量解析の理論と実際. マール社,1988.

[6] 小西貞則: 多変量解析入門-線形から非線形へ- , 岩波書店，2010.

[7] 小西貞則，北川源四郎: 情報量規準, 朝倉書店，2004.

[8] 奥野忠一，久米均，芳賀敏朗，吉澤正：多変量解析法，日科技連出版，1981

[9] Rao,C.R.:Linear Statistical Inference and It's Application(2nd ed.), John Wiley & Sons,1973.
(奥野忠一他訳:統計的推測とその応用, 東京図書)

[10] 佐藤義治:多変量データの分類-判別分析・クラスター分析-線形から非線形へ-，朝倉書店 ,2009.

[11] 田中豊, 垂水共之:Windows 版統計解析ハンドブック多変量解析，共立出版,1995.

[12] 田中豊, 垂水共之, 脇本和昌:パソコン統計解析ハンドブック,V, 多変量分散分析線形モデル偏，共立出版,1990.

[13] 丹後俊郎, 山岡和枝, 高木晴良:新版ロジスティック回帰分析，2013.

[14] 竹澤邦夫，みんなのためのノンパラメトリック回帰 (下) 第 3 版，吉岡書店，2007

[15] 辻谷将明:関西多変量解析法セミナー入門コーステキスト (第 5 章　判別と予測)，日本科学技術連盟，2004.

[16] 辻谷将明，竹澤邦夫:R で学ぶデータサイエンス マシンラーニング (第 2 版)，共立出版，2015.

[17] 外山信夫，辻谷将明:実践 R 統計分析, オーム社,2015.

[18] 中村永友:R で学ぶデータサイエンス　多次元データ解析法，共立出版,2009.

[19] 宮原英夫, 丹後俊郎偏:医学統計学ハンドブック，朝倉書店，1995.