Rによる判別分析

本稿のWebページ

古橋　武

2. 線形判別分析(2変量)

2.1 基礎理論

判別分析は判別線(面)を同定する.

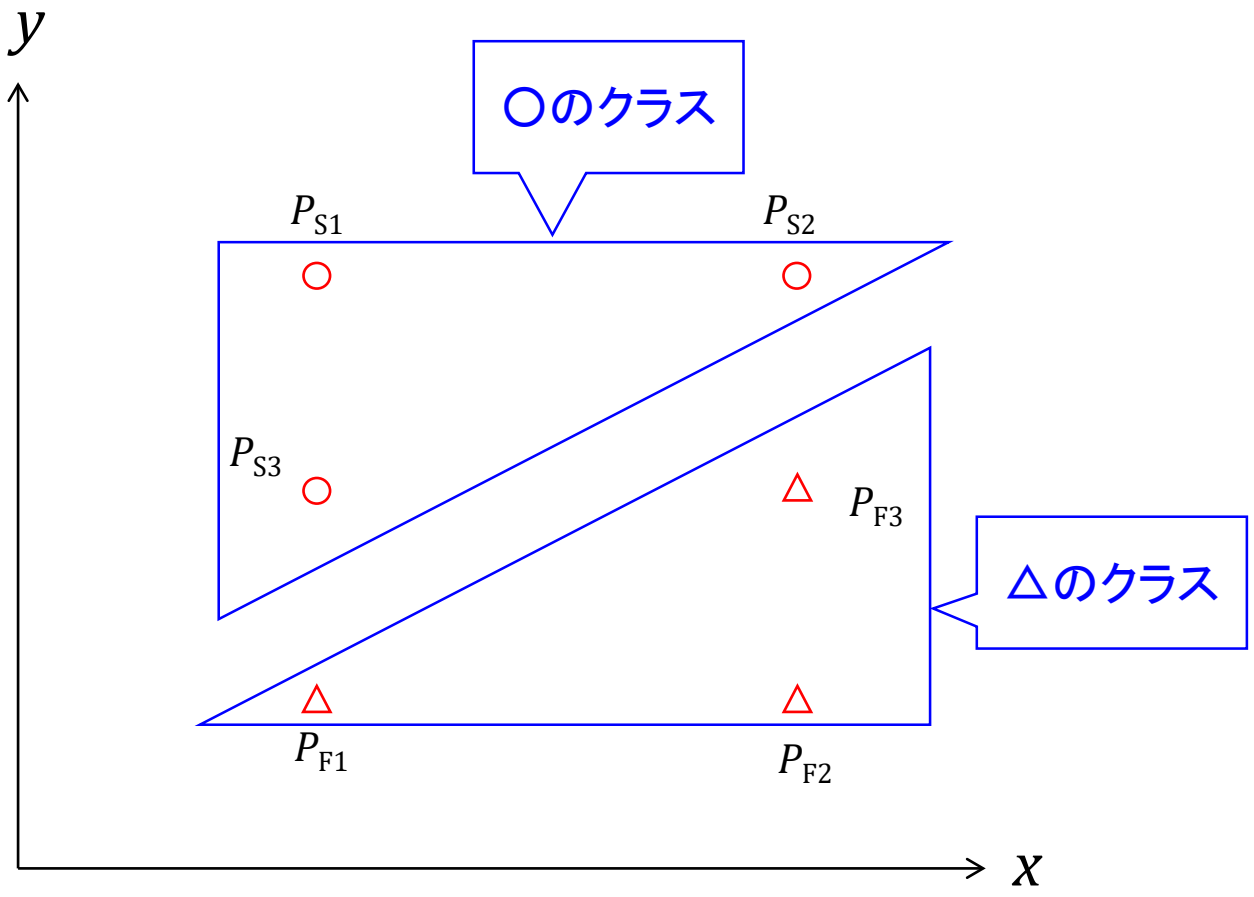


図2.1 判別分析の様子

二つのクラスを分ける線を同定する課題

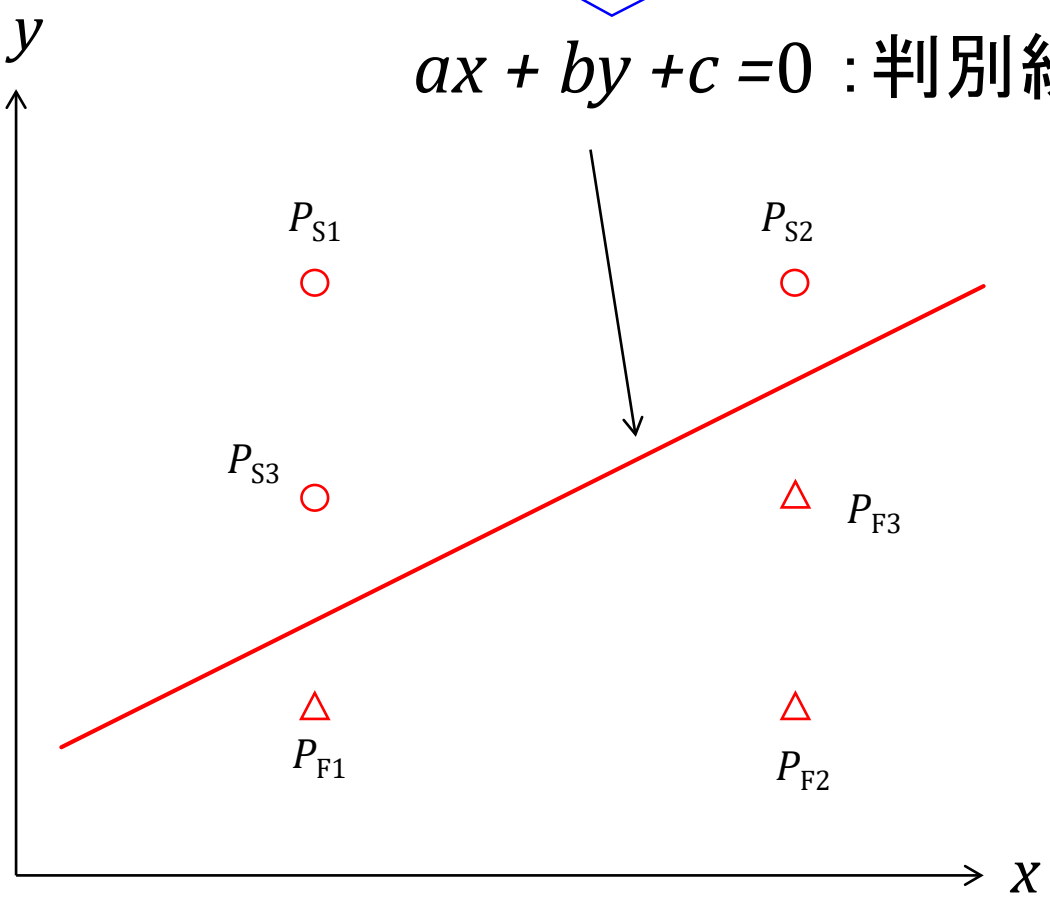
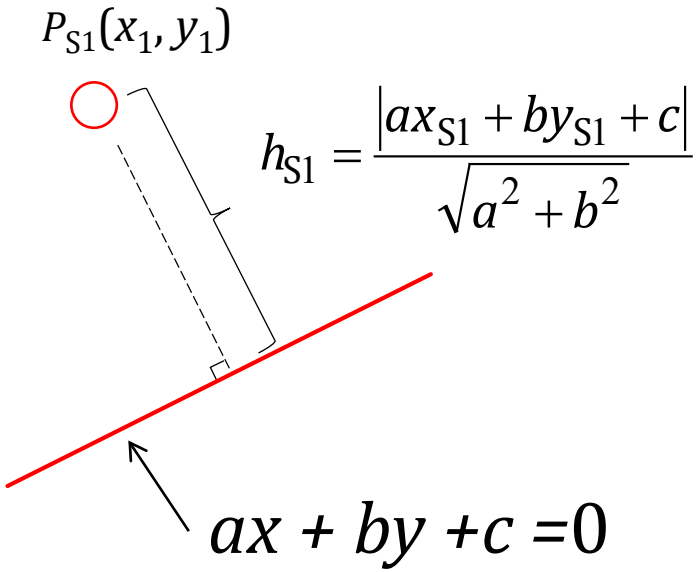


図2.1 判別分析の様子

では，判別線の基準は？

$$h_{S1} = \frac{|ax_{S1} + by_{S1} + c|}{\sqrt{a^2 + b^2}} \qquad (2.2)$$



$a^2 + b^2 = 1$

とすれば

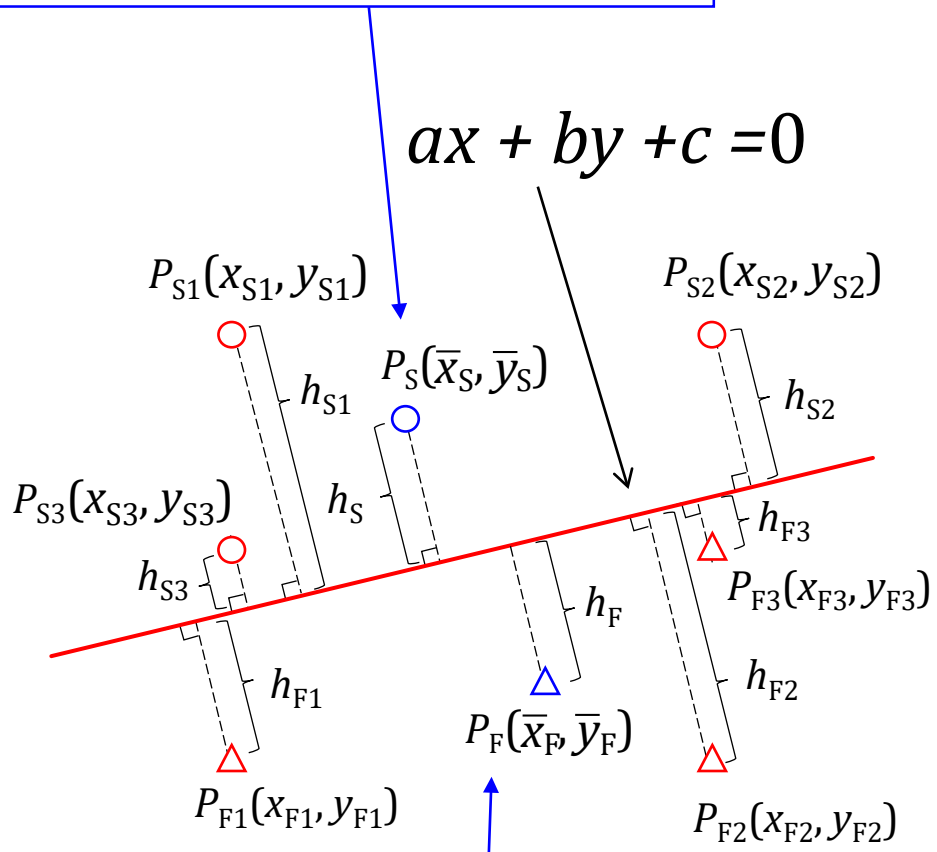
$$h_{S1} = |ax_{S1} + by_{S1} + c|$$

図2.2 データから判別線までの距離

$$\bar{x}_S = \frac{1}{n_S}\sum_{i=1}^{n_S} x_{Si},\quad \bar{y}_S = \frac{1}{n_S}\sum_{i=1}^{n_S} y_{Si}$$

クラス内散布 s_W^2

$$s_W^2 = (h_{S1} - h_S)^2 + (h_{S2} - h_S)^2 + (h_{S3} - h_S)^2 + (h_{F1} - h_F)^2 + (h_{F2} - h_F)^2 + (h_{F3} - h_F)^2$$



$$\bar{x}_F = \frac{1}{n_F}\sum_{i=1}^{n_F} x_{Fi},\quad \bar{y}_F = \frac{1}{n_F}\sum_{i=1}^{n_F} y_{Fi}$$

図2.3 各データと判別線までの距離

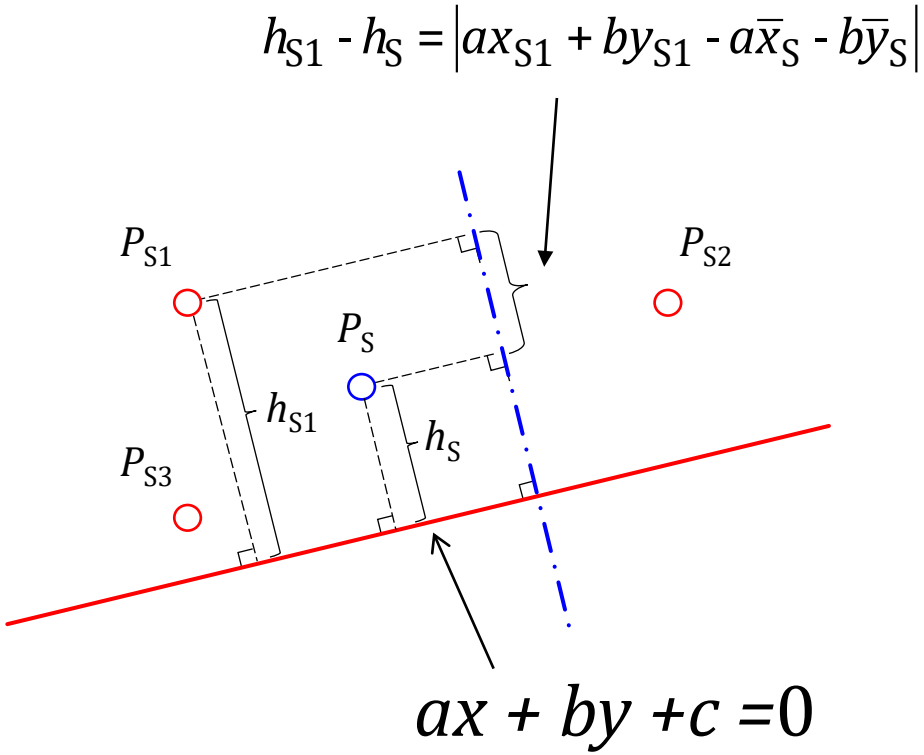
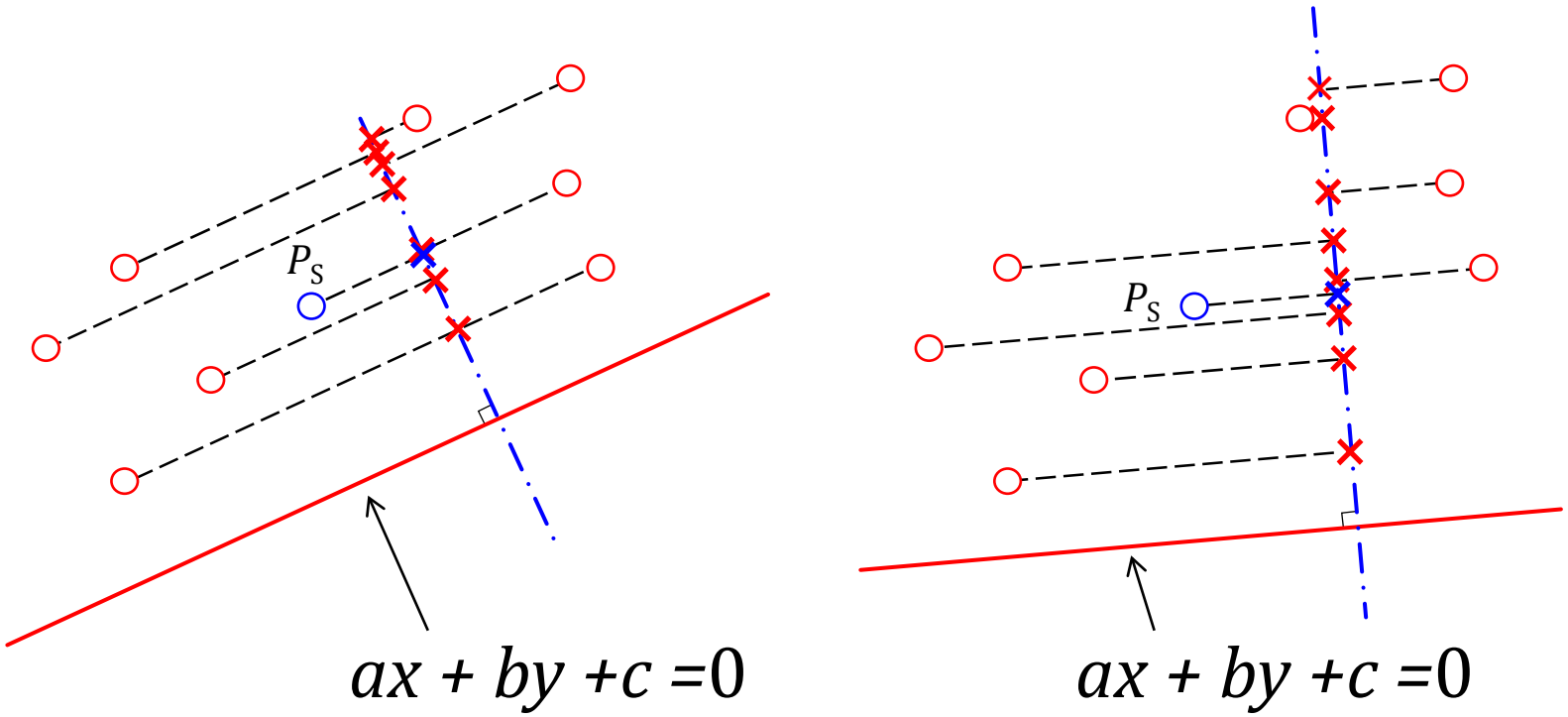


図2.4 〇のクラス内のデータの判別線との距離の差分

$$
\begin{aligned}
s_W^2 &= (h_{S1} - h_S)^2 \\
&\quad + (h_{S2} - h_S)^2 \\
&\quad + (h_{S3} - h_S)^2 \\
&\quad + (h_{F1} - h_F)^2 \\
&\quad + (h_{F2} - h_F)^2 \\
&\quad + (h_{F3} - h_F)^2
\end{aligned}
$$

$$
= \sum_{i=1}^{n_S} (ax_{Si} + by_{Si} - a\bar{x}_S - b\bar{y}_S)^2
$$

$$
+ \sum_{i=1}^{n_F} (ax_{Fi} + by_{Fi} - a\bar{x}_F - b\bar{y}_F)^2
$$



(a) クラス内散布小　　(b) クラス内散布大

図2.5 判別線の傾きとクラス内散布

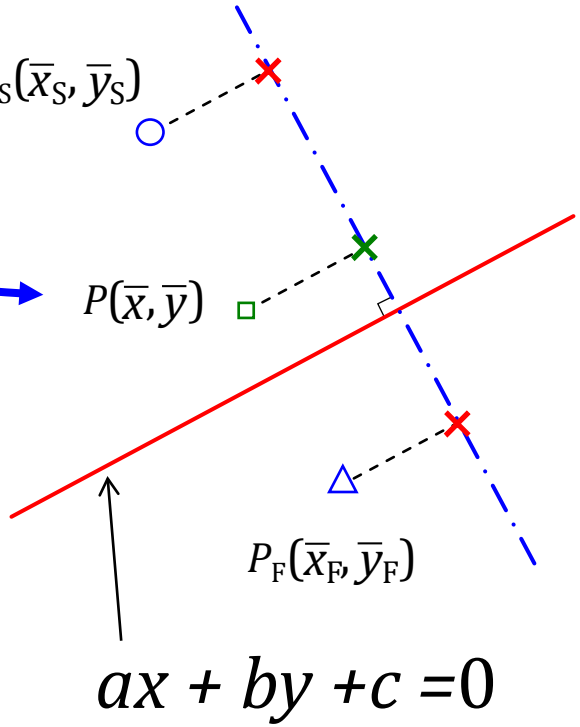
全体の平均

$$\bar{x} = \frac{1}{n}\left\{\sum_{i=1}^{n_S} x_{Si} + \sum_{i=1}^{n_F} x_{Fi}\right\}$$

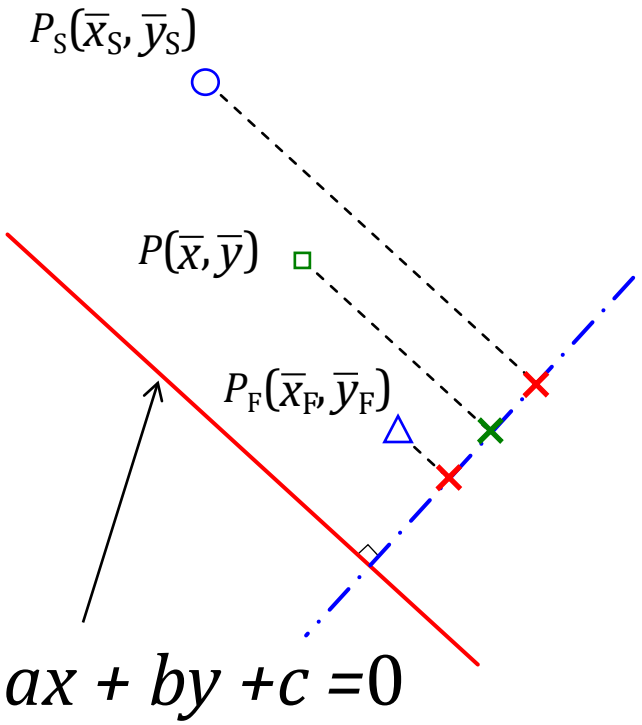
$$\bar{y} = \frac{1}{n}\left\{\sum_{i=1}^{n_S} y_{Si} + \sum_{i=1}^{n_F} y_{Fi}\right\}$$

クラス間散布 s_B^2

$$= n_S(a\bar{x}_S + b\bar{y}_S - a\bar{x} - b\bar{y})^2 + n_F(a\bar{x}_F + b\bar{y}_F - a\bar{x} - b\bar{y})^2$$

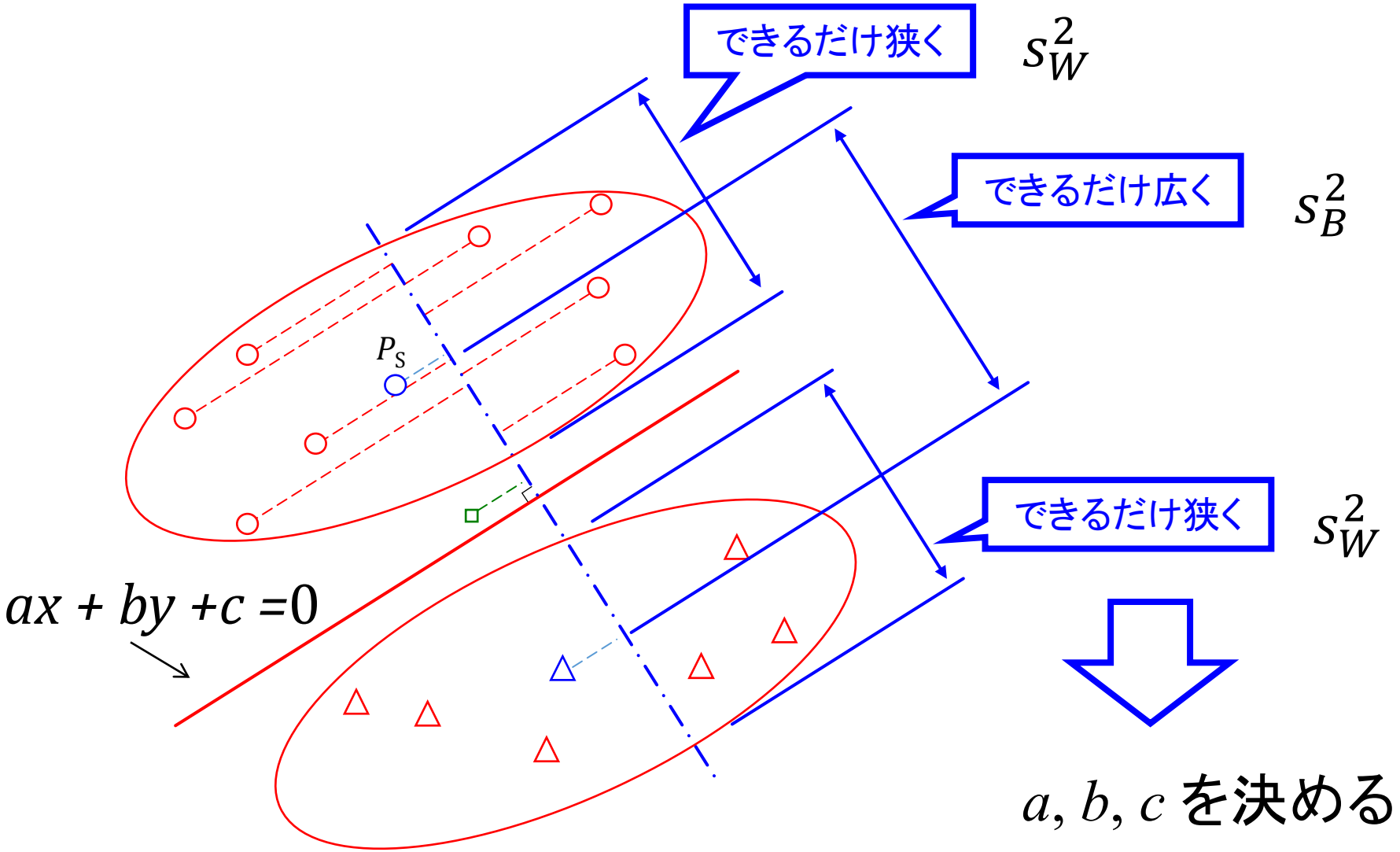


(a) クラス間散布大



(b) クラス間散布小

図2.6 判別線の傾きとクラス間散布


できるだけ狭く
s_W^2
できるだけ広く
s_B^2
P_S
できるだけ狭く
s_W^2
$ax + by + c = 0$
a, b, c を決める
指標：$F = \frac{s_W^2}{s_B^2}$
最小

2.2 行列・ベクトルによる表現

$$F = \frac{s_W^2}{s_B^2}$$

$$= \frac{\sum_{i=1}^{n_S}(ax_{Si} + by_{Si} - a\bar{x}_S - b\bar{y}_S)^2 + \sum_{i=1}^{n_F}(ax_{Fi} + by_{Fi} - a\bar{x}_F - b\bar{y}_F)^2}{n_S(a\bar{x}_S + b\bar{y}_S - a\bar{x} - b\bar{y})^2 + n_F(a\bar{x}_F + b\bar{y}_F - a\bar{x} - b\bar{y})^2}$$

⇩ 行列，ベクトルにより表現する.

$$\boldsymbol{W} = \begin{pmatrix} x_{S1} - \bar{x}_S & y_{S1} - \bar{y}_S \\ x_{S2} - \bar{x}_S & y_{S2} - \bar{y}_S \\ \vdots & \vdots \\ x_{Sn_S} - \bar{x}_S & y_{Sn_S} - \bar{y}_S \\ x_{F1} - \bar{x}_F & y_{F1} - \bar{y}_F \\ x_{F2} - \bar{x}_F & y_{F2} - \bar{y}_F \\ \vdots & \vdots \\ x_{Fn_F} - \bar{x}_F & y_{Fn_F} - \bar{y}_F \end{pmatrix}$$

$$s_W^2 = \boldsymbol{Z}^t\boldsymbol{Z} = \boldsymbol{v}^t\boldsymbol{W}^t\boldsymbol{W}\boldsymbol{v} = \boldsymbol{v}^t\boldsymbol{U}_W\boldsymbol{v} \qquad (2.34)$$

$$\boldsymbol{U}_W = \boldsymbol{W}^t\boldsymbol{W} \qquad (2.35)$$

$$\boldsymbol{v} = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{v} \qquad (2.33)$$

$$\boldsymbol{B} = \begin{pmatrix} \bar{x}_{\mathrm{S}} - \bar{x} & \bar{y}_{\mathrm{S}} - \bar{y} \\ \bar{x}_{\mathrm{S}} - \bar{x} & \bar{y}_{\mathrm{S}} - \bar{y} \\ \vdots & \vdots \\ \bar{x}_{\mathrm{S}} - \bar{x} & \bar{y}_{\mathrm{S}} - \bar{y} \\ \bar{x}_{\mathrm{F}} - \bar{x} & \bar{y}_{\mathrm{F}} - \bar{y} \\ \bar{x}_{\mathrm{F}} - \bar{x} & \bar{y}_{\mathrm{F}} - \bar{y} \\ \vdots & \vdots \\ \bar{x}_{\mathrm{F}} - \bar{x} & \bar{y}_{\mathrm{F}} - \bar{y} \end{pmatrix} \quad (2.36)$$

$$\begin{aligned} s_B^2 &= Y^t \boldsymbol{Y} \\ &= \boldsymbol{v}^t \boldsymbol{B}^t \boldsymbol{B} \boldsymbol{v} \\ &= \boldsymbol{v}^t \boldsymbol{U_B} \boldsymbol{v} \end{aligned} \quad (2.37)$$

$$\boldsymbol{U_B} = \boldsymbol{B}^t \boldsymbol{B} \quad (2.38)$$

$$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{v}$$

$$F = \frac{s_{\mathrm{W}}^2}{s_{\mathrm{B}}^2} = \frac{\boldsymbol{v}^t \boldsymbol{U_W} \boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{U_B} \boldsymbol{v}} \quad (2.39)$$

⇩ F を最小とする v

$$\boldsymbol{U_W}^{-1} \boldsymbol{U_B} \boldsymbol{v} = \lambda \boldsymbol{v} \quad (2.47)$$

⇨ $\boldsymbol{v} = \begin{pmatrix} a \\ b \end{pmatrix}$ a, b の決定

2.3 Rによる行列計算

判別分析_行列計算_入社試験の成績.R

x <- read.csv("C:/Users/Furuhashi/Documents/日本テクノセンター/判別分析/test_for_employment.csv")

表2.1 入社試験の成績と入社後の評価

学科	面接	10年後の評価
6	6	1
3	6	1
3	3	1
6	3	2
6	0	2
3	0	2

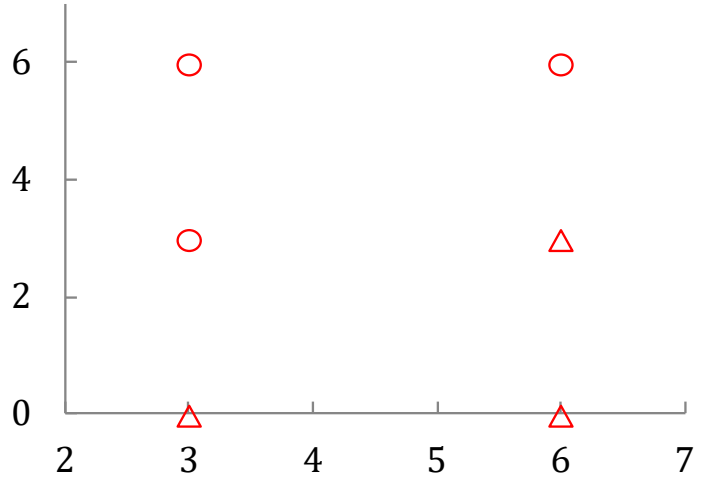


図2.7入社試験の成績と入社後の評価

x_S <- subset(x[,-3], x$入社後の評価 < 1.5)

	学科	面接
1	6	6
2	3	6
3	3	3

$$\mathbf{x\_S} = \begin{pmatrix} x_{S1} & y_{S1} \\ x_{S2} & y_{S2} \\ x_{S3} & y_{S3} \end{pmatrix}$$

x_F <- subset(x[,-3], x$入社後の評価 > 1.5)

	学科	面接
4	6	3
5	6	0
6	3	0

$$\mathbf{x\_S} = \begin{pmatrix} x_{S1} & y_{S1} \\ x_{S2} & y_{S2} \\ x_{S3} & y_{S3} \end{pmatrix}$$

WS <- scale(x_S,center = TRUE, scale = FALSE)

	学科	面接
1	2	1
2	-1	1
3	-1	-2

$$\boldsymbol{W\_S} = \begin{pmatrix} x_{S1} - \bar{x}_S & y_{S1} - \bar{y}_S \\ x_{S2} - \bar{x}_S & y_{S2} - \bar{y}_S \\ x_{S3} - \bar{x}_S & y_{S3} - \bar{y}_S \end{pmatrix}$$

WF <- scale(x_F,center = TRUE, scale = FALSE)

	学科	面接
4	1	2
5	1	-1
6	-2	-1

$$\boldsymbol{W\_S} = \begin{pmatrix} x_{S1} - \bar{x}_S & y_{S1} - \bar{y}_S \\ x_{S2} - \bar{x}_S & y_{S2} - \bar{y}_S \\ x_{S3} - \bar{x}_S & y_{S3} - \bar{y}_S \end{pmatrix}$$

UW <- t(WS) %*% WS + t(WF) %*% WF

$$\boldsymbol{U_W = W^t W}$$

UB <- num_S*((mean_S-mean_全体) %*% t(mean_S-mean_全体)) +
num_F*((mean_F-mean_全体) %*% t(mean_F-mean_全体))

$$\boldsymbol{U_B = B^t B}$$

inv_UW <- solve(UW)

$$U_W^{-1}$$

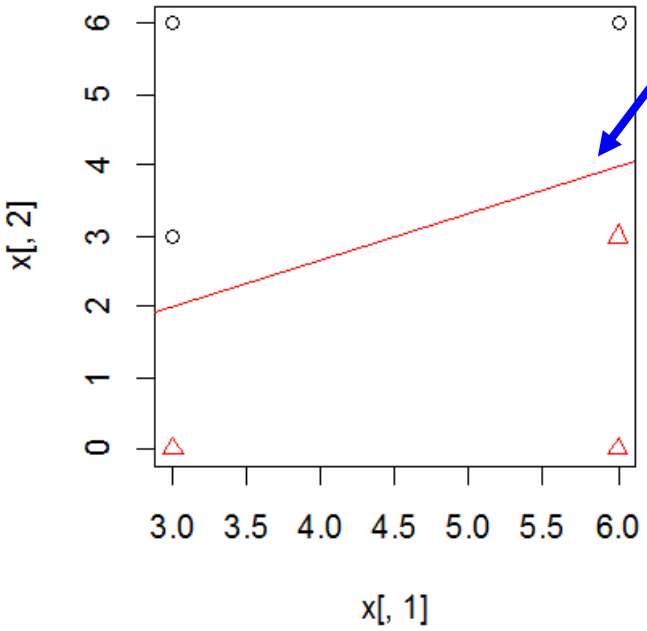
Eigen_value <- eigen(inv_UW %*% UB)　　*a, b の決定*

⇨ $U_W^{-1}U_B$ の固有値，固有ベクトル計算　$v = \begin{pmatrix} a \\ b \end{pmatrix}$

abline(0,-Re(Eigen_value$vectors[1,1])/Re(Eigen_value$vectors[2,1]), col = "red")

-c/b

-a/b



$ax + by + c = 0$

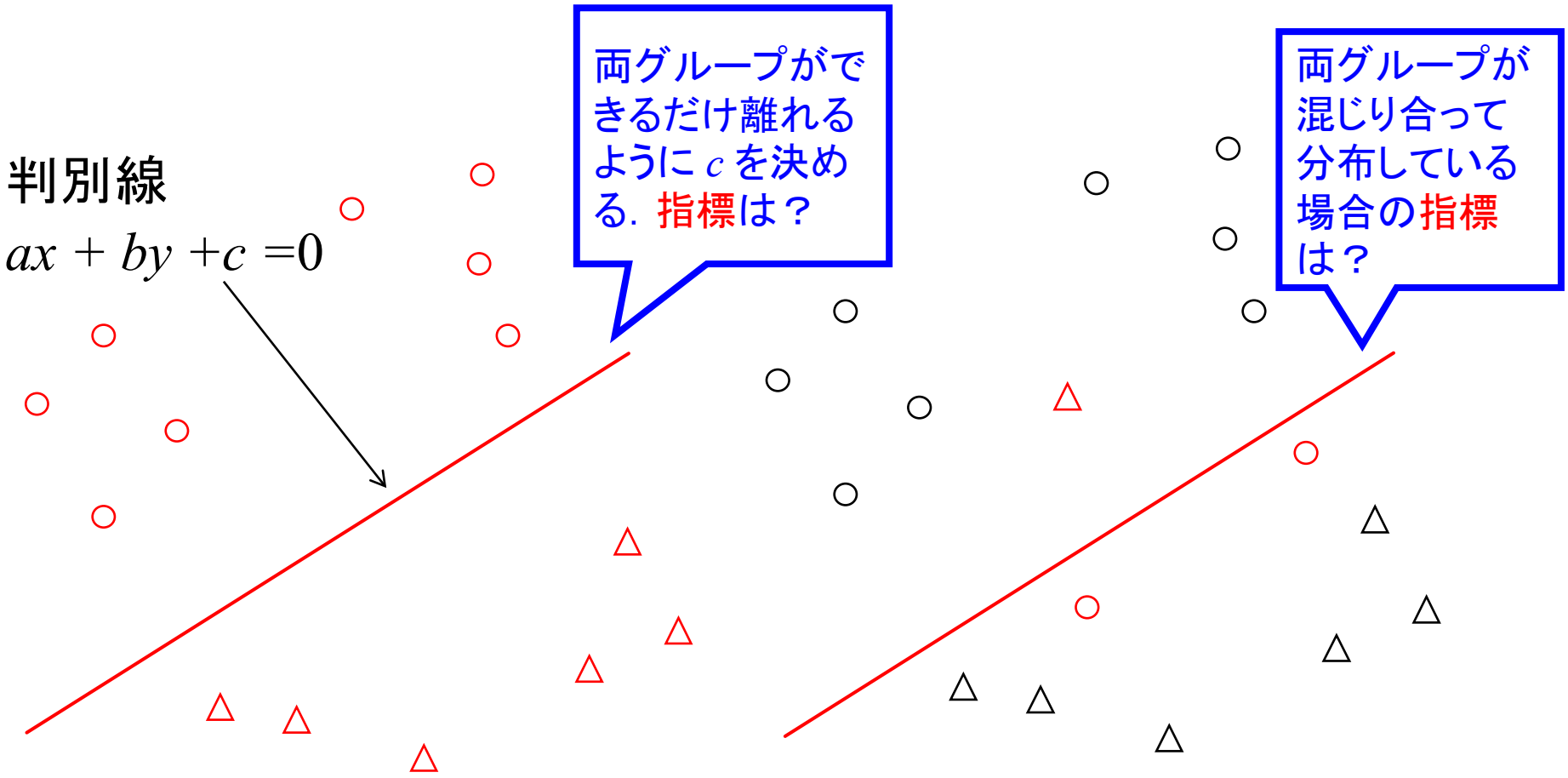
$$y = -\frac{a}{b}x - \frac{c}{b}$$

図2.11 abline()の実行結果

３．定数項cの決め方

3.1 マハラノビス距離

判別線の定数項cはどう決めればよいか？



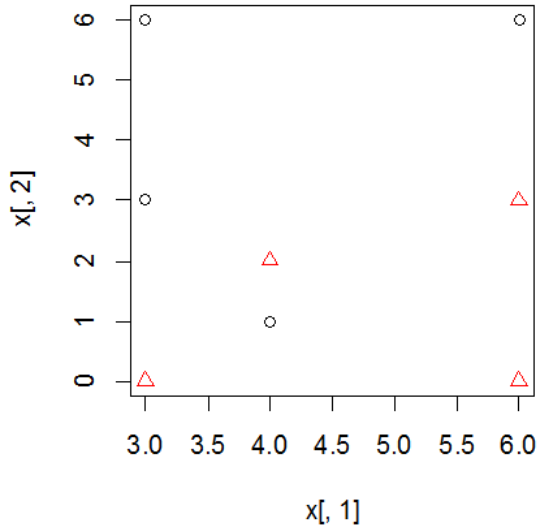


図3.8 表3.1のデータ分布の例

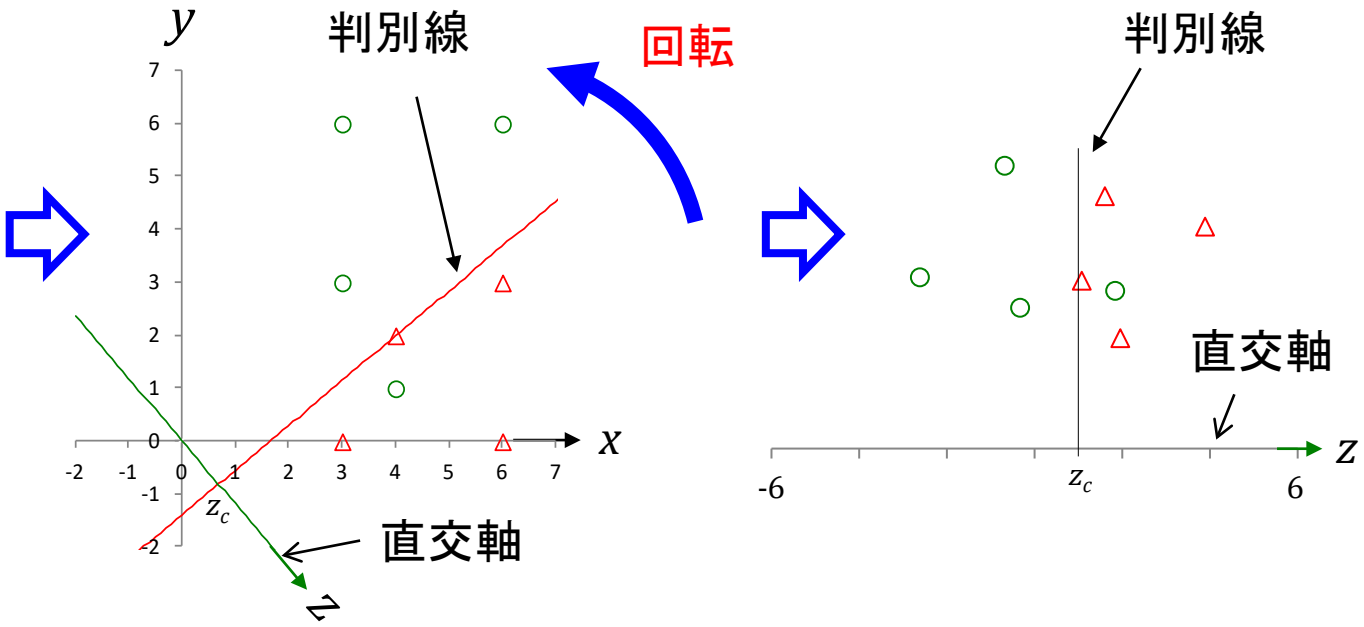
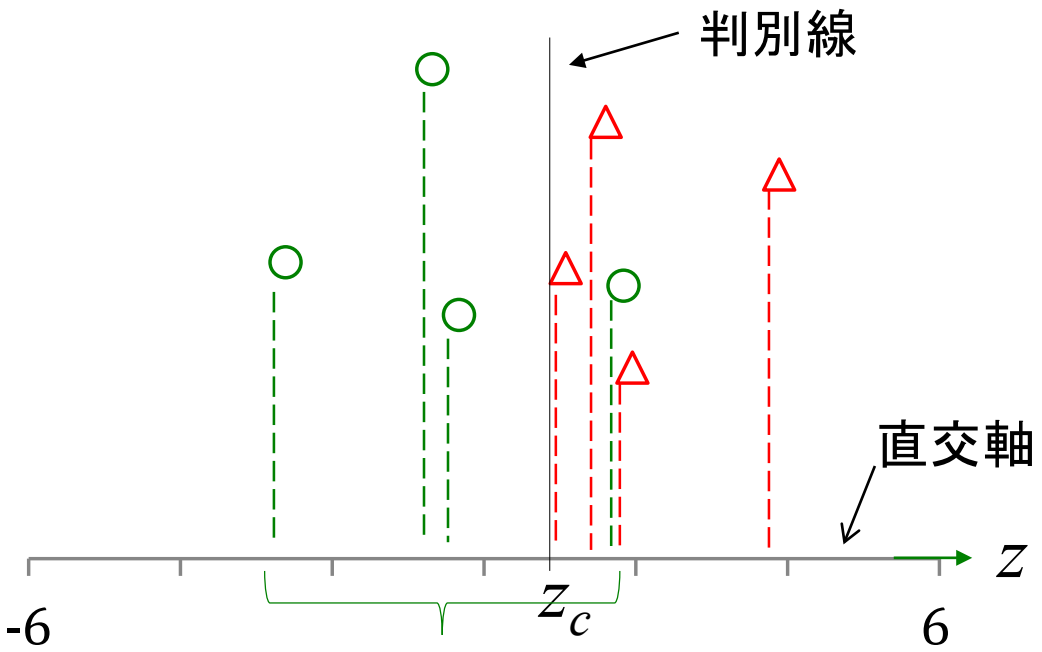


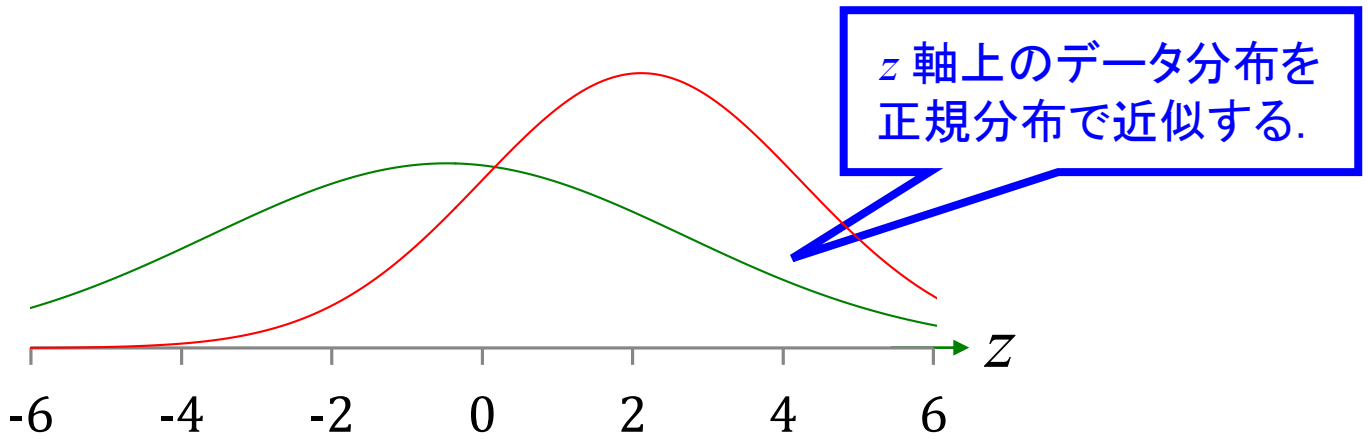
図3.10 判別線と直交軸 z



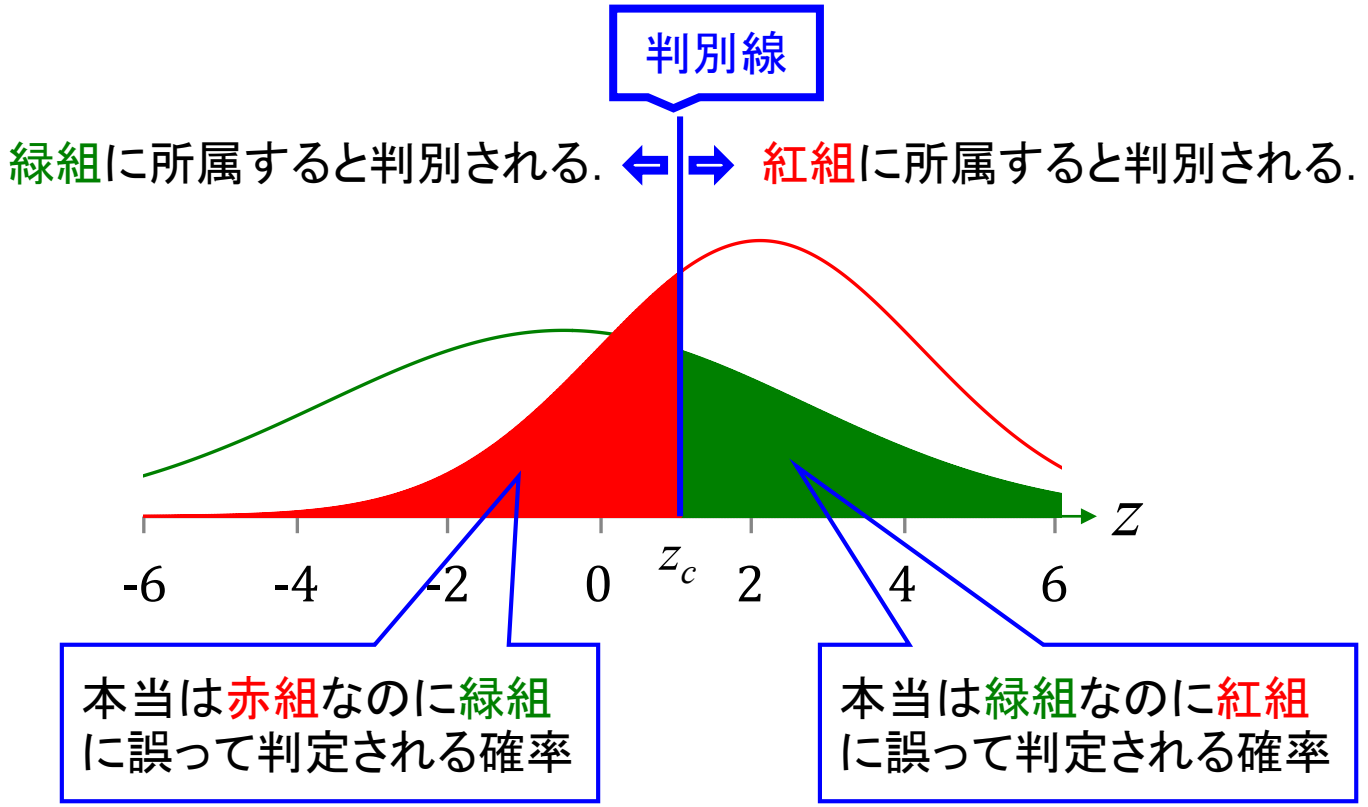
平均　$\bar{z}_S = a\bar{x}_S + b\bar{y}_S + c_0$

不偏分散　$v_{eS}^2 = \dfrac{1}{n_S - 1}\sum_{i=1}^{n_S}(z_{Si} - \bar{z}_S)^2$

$$f_S(z) = \frac{1}{\sqrt{2\pi v_{eS}^2}} e^{\left\{-\frac{(z-\bar{z}_S)^2}{2v_{eS}^2}\right\}}$$



新しいデータが来たときにそのデータの所属先を決めるには？



z_c はそれぞれの間違い確率を同じにする.

緑と赤の塗りつぶした箇所の面積を一緒にする z_c は

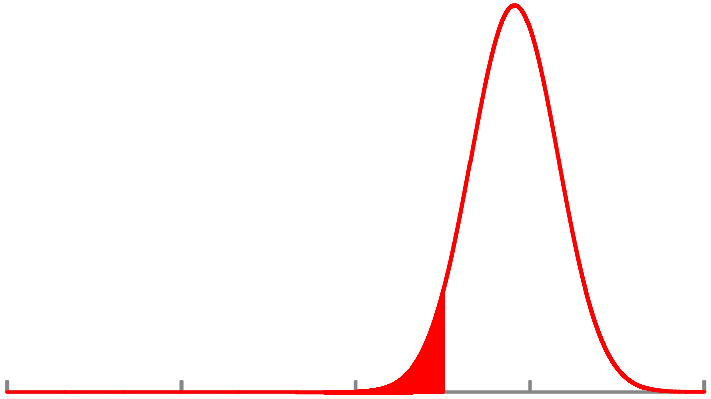
マハラノビス距離：

$$\frac{z_c - \bar{z}_{\mathrm{S}}}{\sqrt{v_{e\mathrm{S}}^2}} = \frac{-z_c + \bar{z}_{\mathrm{F}}}{\sqrt{v_{e\mathrm{F}}^2}}$$

⇩ z_c の決定

$$z_c = \frac{\sqrt{v_{e\mathrm{S}}^2}\bar{z}_{\mathrm{F}} + \sqrt{v_{e\mathrm{F}}^2}\bar{z}_{\mathrm{S}}}{\sqrt{v_{e\mathrm{S}}^2} + \sqrt{v_{e\mathrm{F}}^2}}$$

マハラノビス距離より大きな値をとる確率が同じである理由



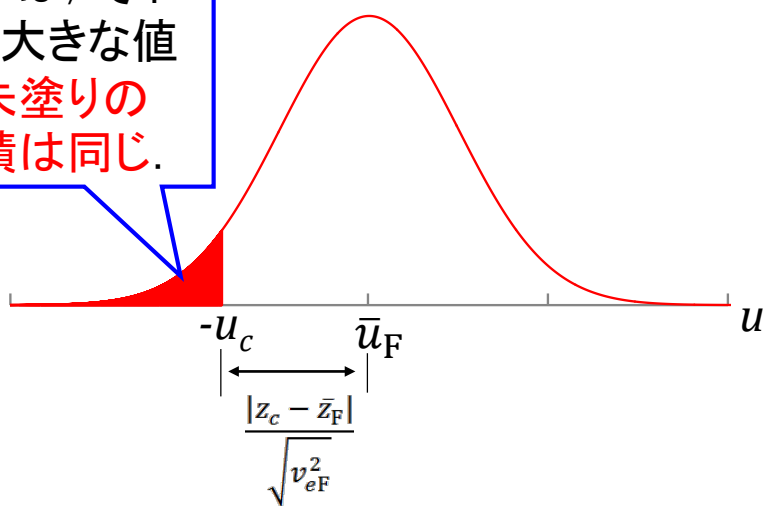
$$f_{\mathrm{F}}(z) = \frac{1}{\sqrt{2\pi v_{e\mathrm{F}}^2}} e^{\left\{-\frac{(z-\bar{z}_{\mathrm{F}})^2}{2v_{e\mathrm{F}}^2}\right\}}$$

マハラノビス距離が同じであれば，それより大きな値の朱塗りの面積は同じ．

$$u = \frac{z - \bar{z}_{\mathrm{F}}}{v_{e\mathrm{F}}}$$

$$dz = v_{e\mathrm{F}} du$$

マハラノビス距離の変数変換により



平均0，分散1の標準分布に変換される．

$$f_{\mathrm{F}}(z) = \frac{1}{\sqrt{2\pi}} e^{\left\{-\frac{u^2}{2}\right\}}$$

判別分析_定数項の決定_入社試験の成績.R

表3.1 入社試験の成績と入社後の評価(II)

学科	面接	入社後の評価
6	6	1
3	6	1
3	3	1
6	3	2
6	0	2
3	0	2
4	1	1
4	2	2

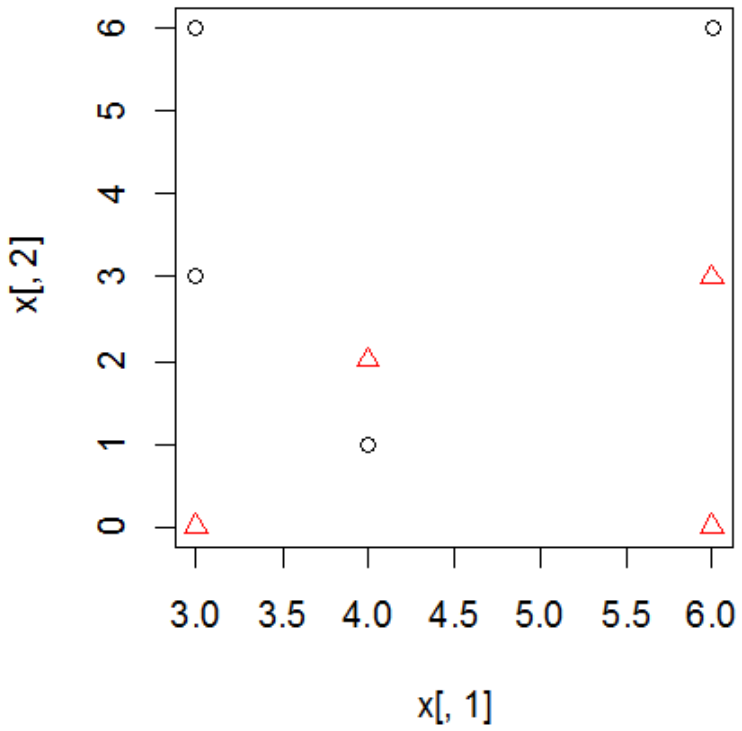
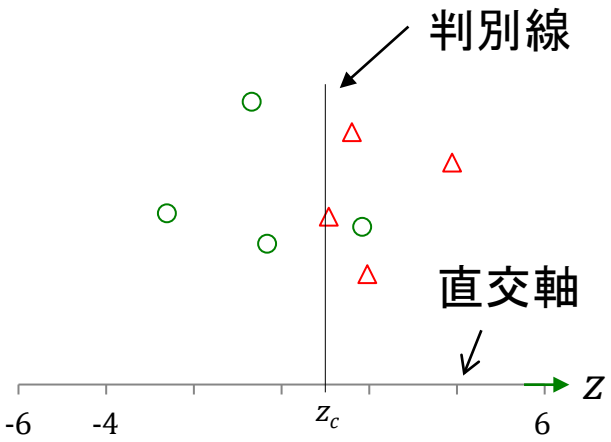


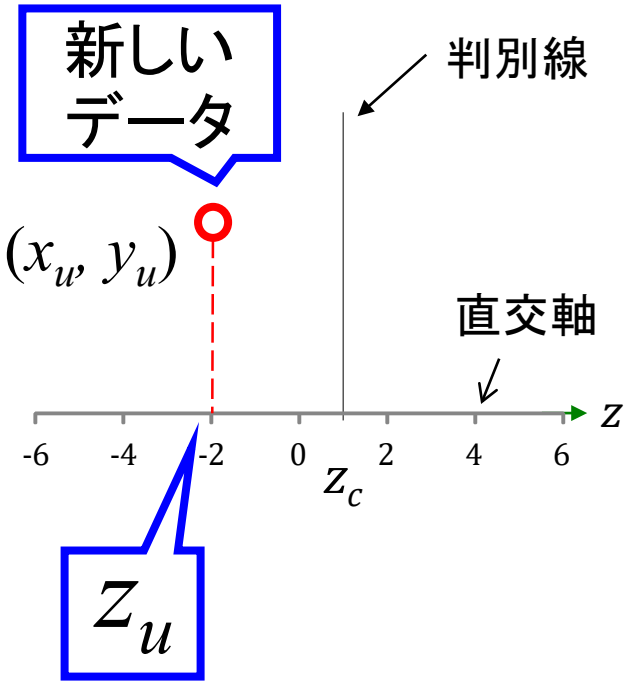
図3.8 表3.1のデータ分布

```
ve_S <- (t(v) %*% t(WS) %*% WS %*% v)/(num_S-1)          #veS^2の計算

ve_F <- (t(v) %*% t(WF) %*% WF %*% v)/(num_F-1)          #veF^2の計算

mean_zS <- t(v) %*% mean_S          #zSの平均値の計算(c = 0の仮定の下で)

mean_zF <- t(v) %*% mean_F          #zFの平均値の計算(c = 0の仮定の下で)
```




```
zc <- (sqrt(ve_F)*mean_zS + sqrt(ve_S)*mean_zF)/(sqrt(ve_F) + sqrt(ve_S))

                                    #zcの計算
```

3.3 未知データの判別



$$z_u = ax_u + by_u$$

$z_u > z_c$: 低評価

$z_u < z_c$: 高評価

4. 線形判別分析（多変量）

4.1 行列による表現

二つのクラスを分ける面を同定する課題

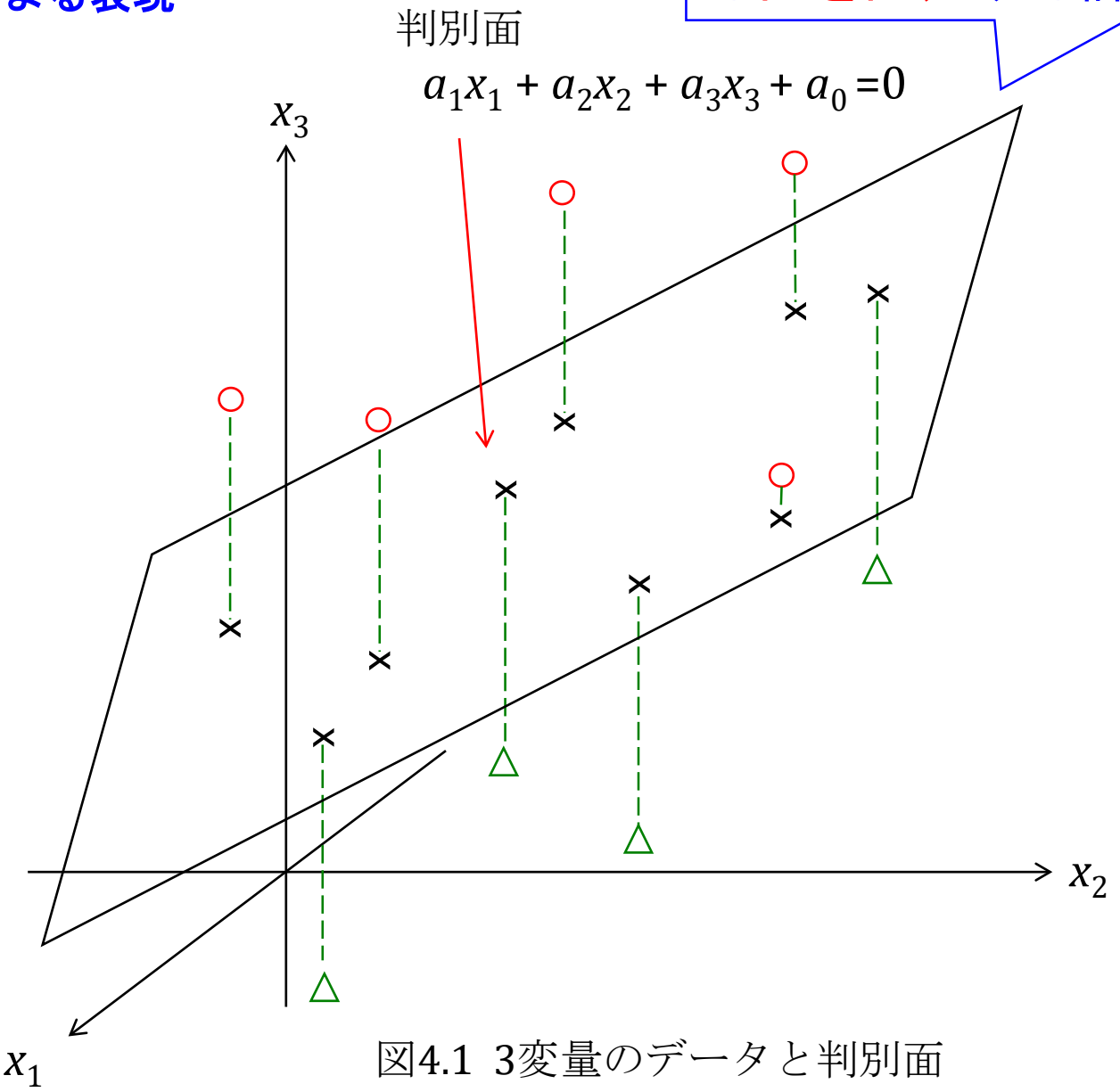


図4.1 3変量のデータと判別面

$$\boldsymbol{W}=\begin{pmatrix} x_{\mathrm{S}11}-\bar{x}_{\mathrm{S}1} & x_{\mathrm{S}21}-\bar{x}_{\mathrm{S}2} & \cdots & x_{\mathrm{S}k1}-\bar{x}_{\mathrm{S}k} \\ x_{\mathrm{S}12}-\bar{x}_{\mathrm{S}1} & x_{\mathrm{S}22}-\bar{x}_{\mathrm{S}2} & \cdots & x_{\mathrm{S}k2}-\bar{x}_{\mathrm{S}k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\mathrm{S}1n_{\mathrm{S}}}-\bar{x}_{\mathrm{S}1} & x_{\mathrm{S}2n_{\mathrm{S}}}-\bar{x}_{\mathrm{S}2} & \cdots & x_{\mathrm{S}kn_{\mathrm{S}}}-\bar{x}_{\mathrm{S}k} \\ x_{\mathrm{F}11}-\bar{x}_{\mathrm{F}1} & x_{\mathrm{F}21}-\bar{x}_{\mathrm{F}2} & \cdots & x_{\mathrm{F}k1}-\bar{x}_{\mathrm{F}k} \\ x_{\mathrm{F}12}-\bar{x}_{\mathrm{F}1} & x_{\mathrm{F}22}-\bar{x}_{\mathrm{F}2} & \cdots & x_{\mathrm{F}k2}-\bar{x}_{\mathrm{F}k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\mathrm{F}1n_{\mathrm{F}}}-\bar{x}_{\mathrm{F}1} & x_{\mathrm{F}2n_{\mathrm{F}}}-\bar{x}_{\mathrm{F}2} & \cdots & x_{\mathrm{F}kn_{\mathrm{F}}}-\bar{x}_{Fk} \end{pmatrix} \quad (4.2)$$

$$\boldsymbol{v}=\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} \quad (4.3)$$

$$\boldsymbol{U_W}=\boldsymbol{W^tW} \quad (4.4)$$

$$\boldsymbol{B}=\begin{pmatrix} \bar{x}_{\mathrm{S}1}-\bar{x} & \bar{x}_{\mathrm{S}2}-\bar{x} & \cdots & \bar{x}_{\mathrm{S}k}-\bar{x} \\ \bar{x}_{\mathrm{S}1}-\bar{x} & \bar{x}_{\mathrm{S}2}-\bar{x} & \cdots & \bar{x}_{\mathrm{S}k}-\bar{x} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_{\mathrm{S}1}-\bar{x} & \bar{x}_{\mathrm{S}2}-\bar{x} & \cdots & \bar{x}_{\mathrm{S}k}-\bar{x} \\ \bar{x}_{\mathrm{F}1}-\bar{x} & \bar{x}_{\mathrm{F}2}-\bar{x} & \cdots & \bar{x}_{\mathrm{F}k}-\bar{x} \\ \bar{x}_{\mathrm{F}1}-\bar{x} & \bar{x}_{\mathrm{F}2}-\bar{x} & \cdots & \bar{x}_{\mathrm{F}k}-\bar{x} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{x}_{\mathrm{F}1}-\bar{x} & \bar{x}_{\mathrm{F}2}-\bar{x} & \cdots & \bar{x}_{\mathrm{F}k}-\bar{x} \end{pmatrix} \quad (4.5)$$

$$\boldsymbol{U_B}=\boldsymbol{B^tB} \quad (4.6)$$

次数が増えても表式は同じ

$$\boldsymbol{U_W}^{-1}\boldsymbol{U_B}\boldsymbol{v}=\lambda\boldsymbol{v} \quad (4.7)$$

4.2 定数項a_0の決定

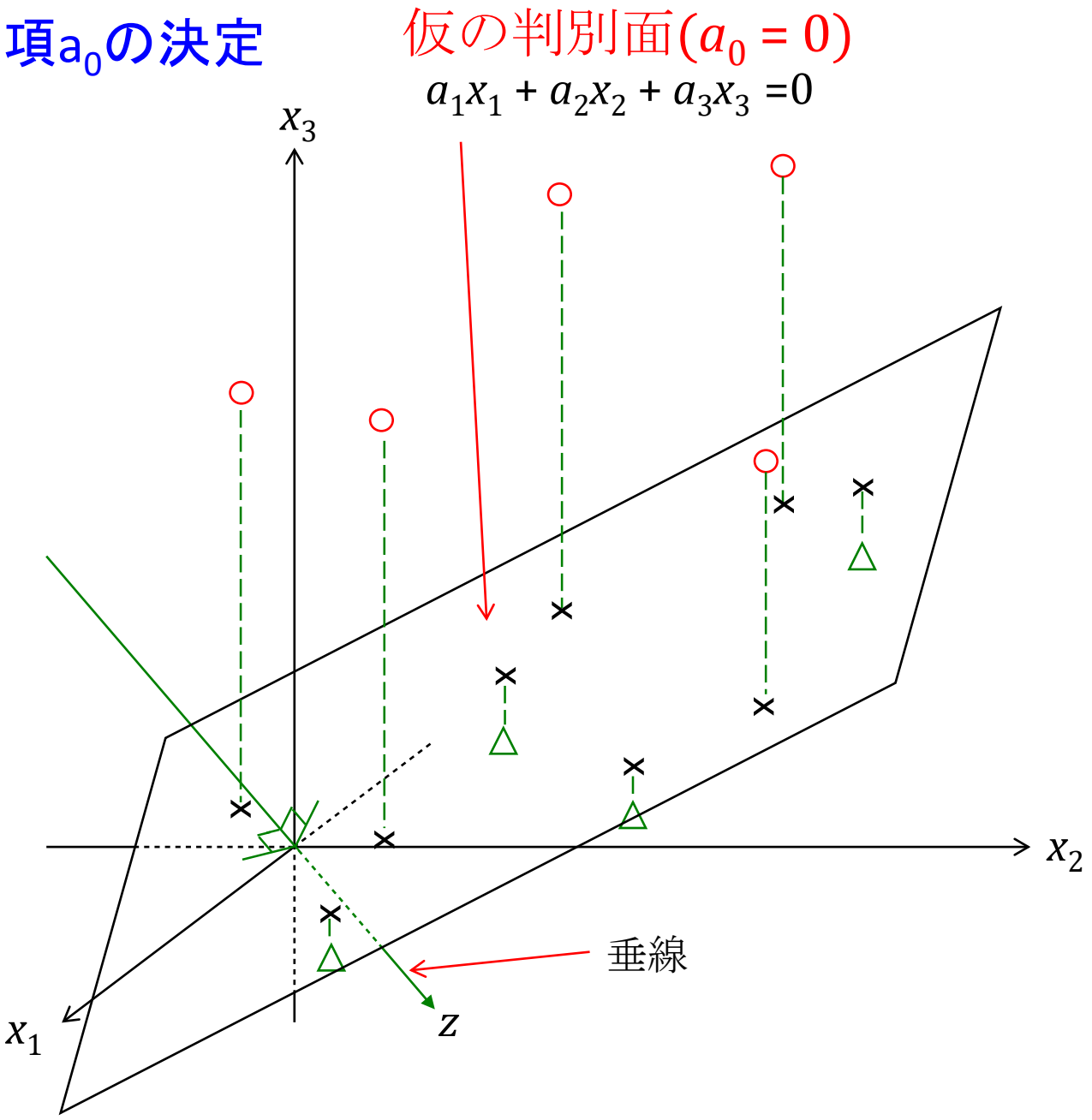


図4.2 仮の判別面($a_0 = 0$)と垂線

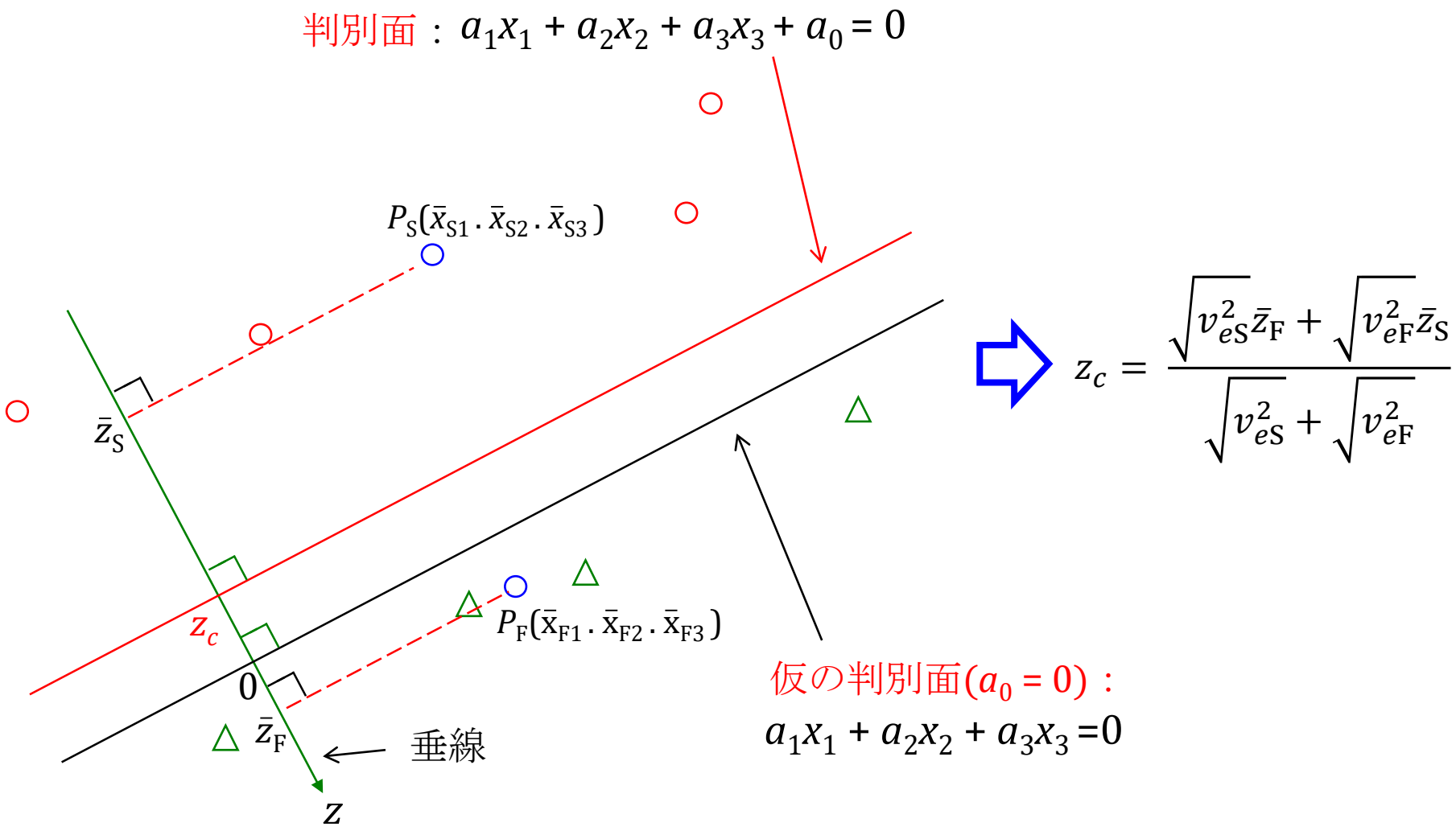
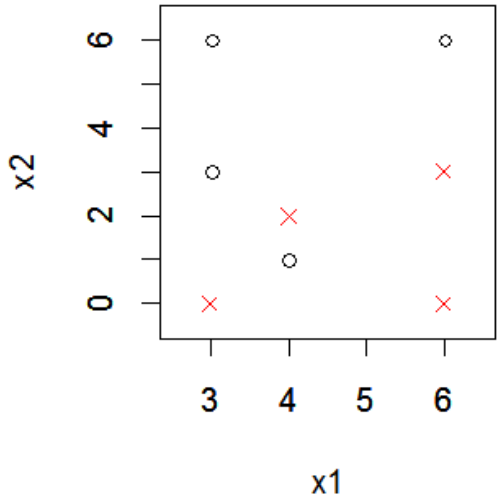


図4.3 仮の判別面の垂線上への射影

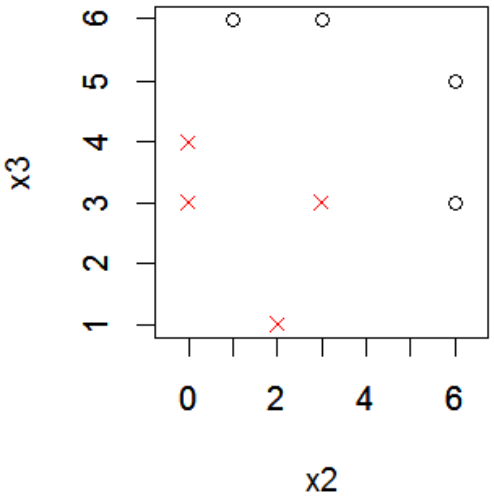
判別分析_3変量_入社試験の成績.R

表4.1 入社試験の成績と入社後の評価
（3変量）

学科	面接	実技	入社後の評価
6	6	5	1
3	6	3	1
3	3	6	1
6	3	3	2
6	0	4	2
3	0	3	2
4	1	6	1
4	2	1	2



(a) $x_1 - x_2$



(b) $x_2 - x_3$

図4.4 表4.1のデータ分布

```
UB <- num_S*((mean_S-mean_全体) %*% t(mean_S-mean_全体)) +
                num_F*((mean_F-mean_全体) %*% t(mean_F-mean_全体))
                                        # UB行列を求める
UW <- t(WS) %*% WS + t(WF) %*% WF       # UW行列を求める
inv_UW <- solve(UW)                     # UW行列の逆行列を求める

Eigen_value <- eigen(inv_UW %*% UB)     # UW^(-1) UB の固有値計算

v <- Re(Eigen_value$vectors[,1])        #固有ベクトルをvに代入
```

判別分析

```
ve_S <- (t(v) %*% t(WS) %*% WS %*% v)/(num_S-1)        #veS^2の計算
ve_F <- (t(v) %*% t(WF) %*% WF %*% v)/(num_F-1)        #veF^2の計算
mean_zS <- t(v) %*% mean_S                  #zSの平均値の計算(c = 0)
mean_zF <- t(v) %*% mean_F                  #zFの平均値の計算(c = 0)

zc <- (sqrt(ve_F)*mean_zS + sqrt(ve_S)*mean_zF)/(sqrt(ve_F) + sqrt(ve_S))
                                                       #zcの計算
```

定数項の計算

```
x1_cd <- seq(min_x1, max_x1, length = 50)
                # x1軸の[min, max]区間を50分割座標の生成
x2_cd <- seq(min_x2, max_x2, length = 50)
                # x2軸の[min, max]区間を50分割座標の生成
x1 <- matrix(0, 2500)
x2 <- matrix(0, 2500)
y <- matrix(0, 2500)
for(i in 1:50) for(j in 1:50){x1[(i-1)*50+j] <- x1_cd[i];
x2[(i-1)*50+j] <- x2_cd[j];  y[(i-1)*50 + j] <-  x3_intercept - v[1]/v[3]*x1_cd[i] -
v[2]/v[3]*x2_cd[j]}
                # x1-x2平面の50*50 = 2500の格子点上のyの値の計算
plot3d(x1, x2, y, col="blue", pch=2, add = 1)
```

判別面の描画

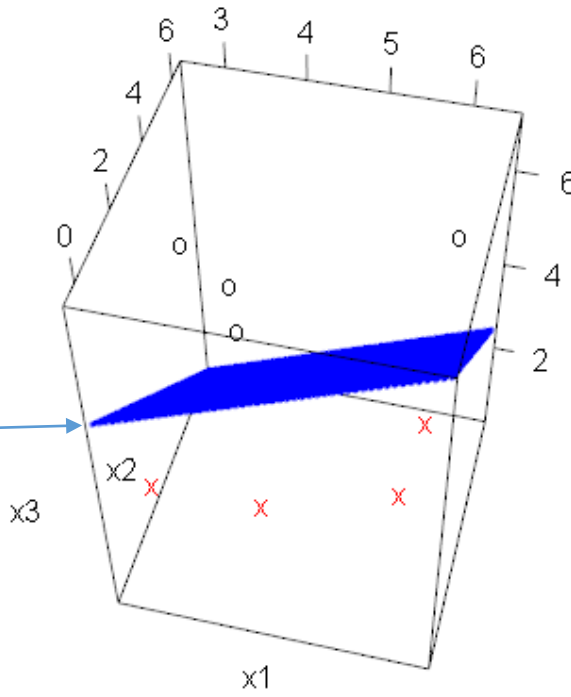


図4.5　判別分析結果

5. 非線形判別分析

5.2 マハラノビス距離

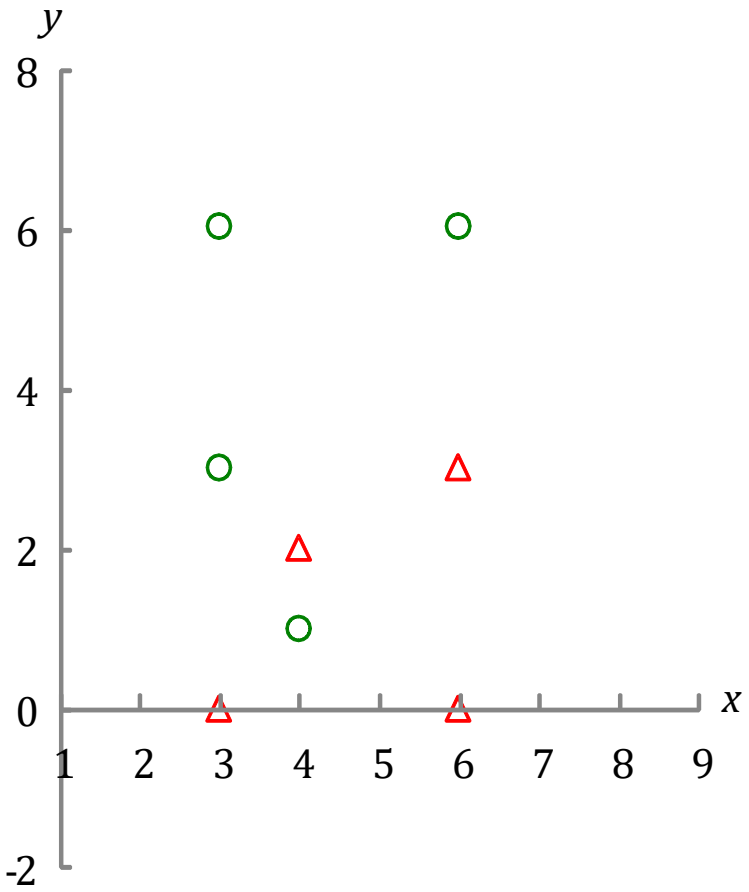


図5.12 入社試験の成績

判別面をどのように引けばそれぞれに間違う確率を同じにできるか？

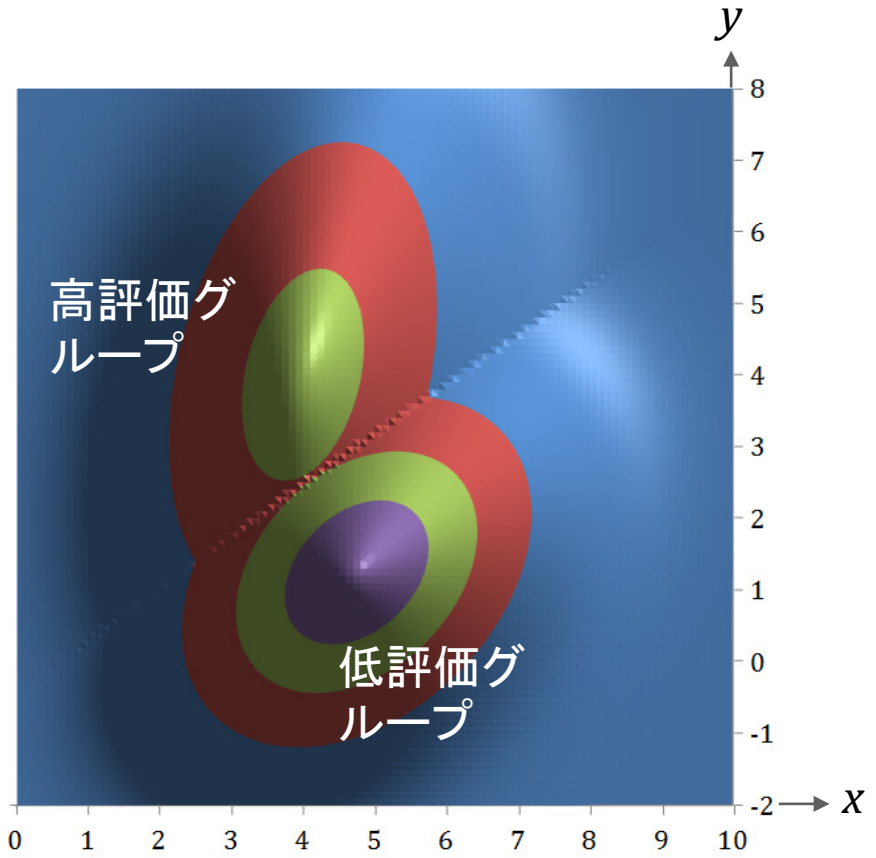


図5.9 2変量正規分布でデータ分布を近似

$x_S = 4, \quad y_S = 4, \quad v_{xS}{}^2 = 2, \quad v_{yS}{}^2 = 6, \quad \rho = 0.289$

$x_F = 4.75, y_F = 1.25, v_{xF}{}^2 = 2.25, v_{yF}{}^2 = 1.25, \rho = 0.333$

2変量を x, y とし，クラスSに属するデータの平均 $\bar{x}_{\mathrm{S}}, \bar{y}_{\mathrm{S}}$ と不偏分散 $v_{x\mathrm{S}}^2, v_{y\mathrm{S}}^2$ を

$$\bar{x}_{\mathrm{S}} = \frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}} x_{\mathrm{S}i}, \quad \bar{y}_{\mathrm{S}} = \frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}} y_{\mathrm{S}i} \qquad (5.1)$$

$$v_{x\mathrm{S}}^2 = \frac{1}{n_{\mathrm{S}}-1}\sum_{i=1}^{n_{\mathrm{S}}}(x_{\mathrm{S}i} - \bar{x}_{\mathrm{S}})^2 \qquad (5.2)$$

$$v_{y\mathrm{S}}^2 = \frac{1}{n_{\mathrm{S}}-1}\sum_{i=1}^{n_{\mathrm{S}}}(y_{\mathrm{S}i} - \bar{y}_{\mathrm{S}})^2$$

とすると

$$f_S(x,y) = \frac{1}{2\pi v_{x\mathrm{S}} v_{y\mathrm{S}}\sqrt{(1-\rho^2)}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\bar{x}_{\mathrm{S}})^2}{v_{x\mathrm{S}}^2} - 2\rho\frac{(x-\bar{x}_{\mathrm{S}})(y-\bar{y}_{\mathrm{S}})}{v_{x\mathrm{S}}v_{y\mathrm{S}}} + \frac{(y-\bar{y}_{\mathrm{S}})^2}{v_{y\mathrm{S}}^2}\right\}\right\} \qquad (5.3)$$

ρ は変量 x, y 間の相関係数

$$\rho = \frac{v_{xy\mathrm{S}}}{v_{x\mathrm{S}}v_{y\mathrm{S}}} \qquad (5.4)$$

ただし，$v_{xy\mathrm{S}} = \frac{1}{n_{\mathrm{S}}-1}\sum_{i=1}^{n_{\mathrm{S}}}(x_{\mathrm{S}i} - \bar{x}_{\mathrm{S}})(y_{\mathrm{S}i} - \bar{y}_{\mathrm{S}})$

5.4　2変量正規分布の行列表現

分散共分散行列 Σ を

$$\Sigma_S = \begin{pmatrix} v_{xS}^2 & v_{xyS} \\ v_{xyS} & v_{yS}^2 \end{pmatrix} \qquad (5.18)$$

$$\boldsymbol{X}_S = \begin{pmatrix} x - \bar{x}_S \\ y - \bar{y}_S \end{pmatrix}$$

と定義すると，(5.3)式の2次元正規分布 $f_S(x, y)$ は

マハラノビス距離の2乗値

$$f_S(x, y) = \frac{1}{2\pi\sqrt{|\Sigma_S|}} \exp\left\{-\frac{1}{2}\boldsymbol{X}_S^t \Sigma_S^{-1} \boldsymbol{X}_S\right\} \qquad (5.19)$$

ただし，$|\Sigma|$ は行列 Σ の行列式，Σ^{-1} は行列 Σ の逆行列

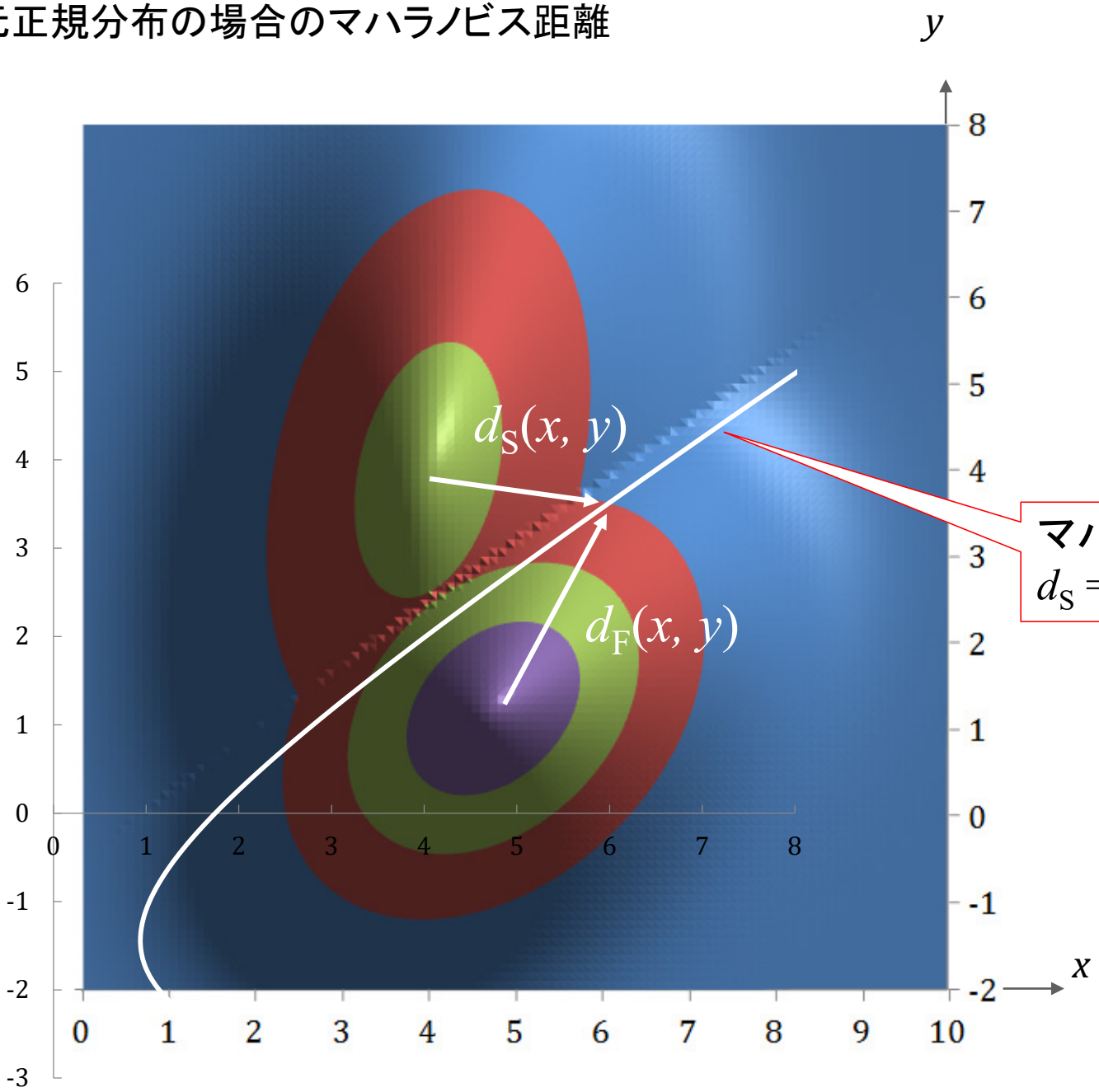
$$|\Sigma_S| = v_{xS}^2 v_{yS}^2 - v_{xyS}^2 \qquad (5.20)$$

$$\Sigma_S^{-1} = \frac{1}{v_{xS}^2 v_{yS}^2 - v_{xyS}^2} \begin{pmatrix} v_{yS}^2 & -v_{xyS} \\ -v_{xyS} & v_{xS}^2 \end{pmatrix} \qquad (5.21)$$

マハラノビス距離 $d_S(x, y)$

$$d_S(x, y)^2 = \boldsymbol{X}_S^t \Sigma_S^{-1} \boldsymbol{X}_S \qquad (5.22)$$

2次元正規分布の場合のマハラノビス距離



判別分析_マハラノビス距離による判別_行列表現_入社試験の成績.R

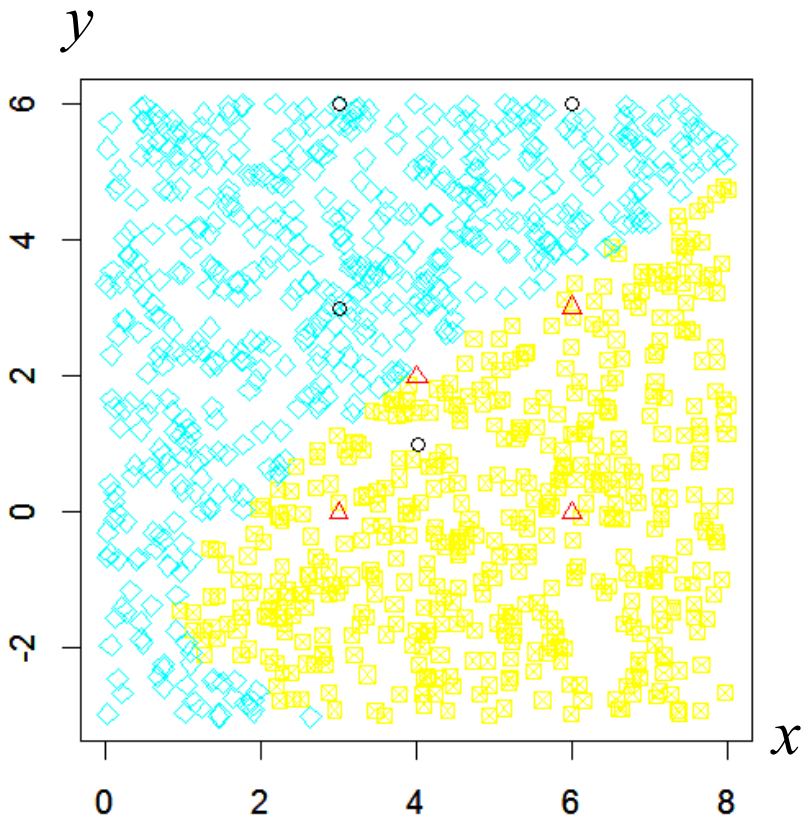


図5.15 入社試験の成績とマハラノビス距離による判別結果

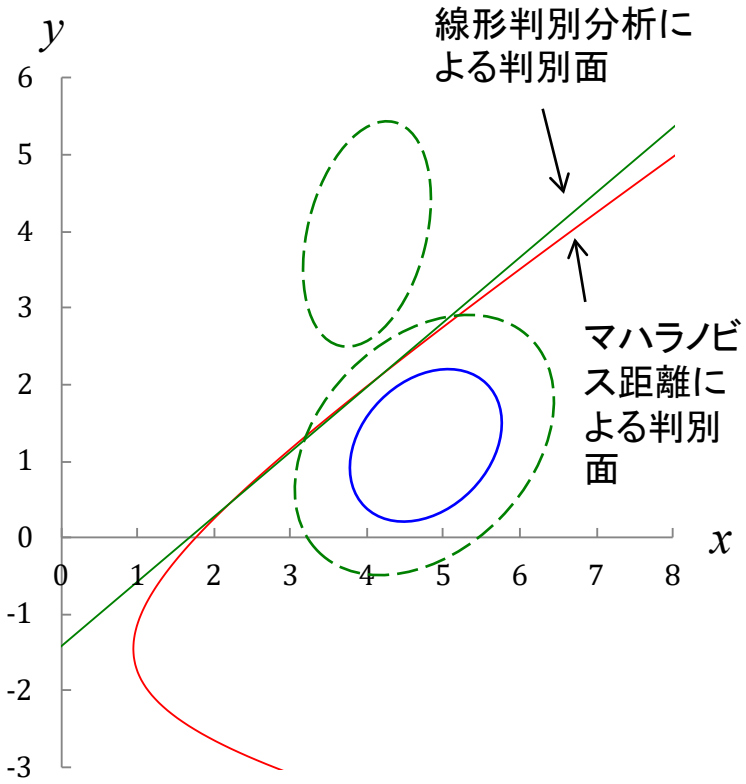


図5.14 マハラノビス距離による判別面と線形判別分析による判別面の比較

5.7 Rによるアヤメのデータの判別

判別分析_マハラノビス距離による判別_行列表現_アイリス.R

xx <- data.frame(iris[,-5], as.integer(iris[,5]))

Rにはirisという名前でアヤメのデータが組み込まれている.

データの5列目はアヤメの種類

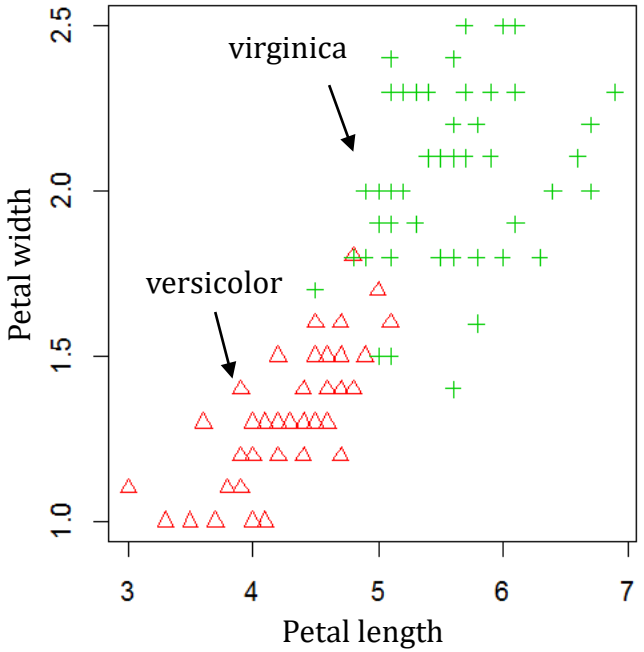


図5.16 アヤメのデータ
(versicolorとvirginica)

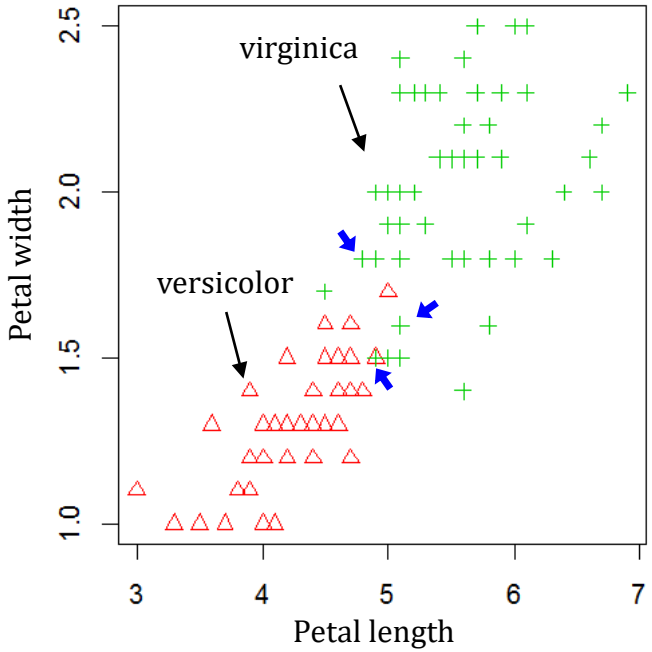


図5.17 アヤメのデータのマハラノビス距離による再判別結果

判別分析_lda_アイリス.R

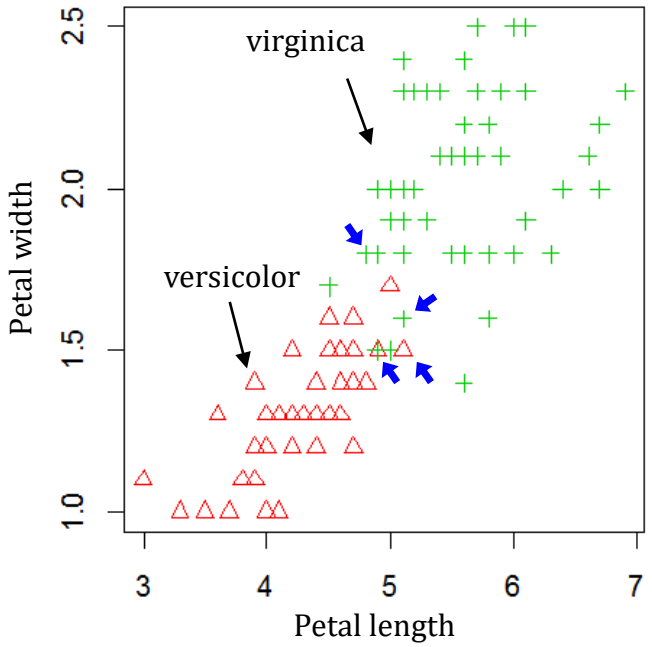


図5.18 アヤメのデータの線形判別分析とマハラノビス距離に基づくz_c決定法による再判別結果

本稿の詳細な解説は

多変量解析の基礎 Ⅲ（判別分析）－理論とRによる演習－

と題してアマゾンよりkindle版として出版しています.