Discriminant Analysis

判别分析

肖磊，2026年5月21日

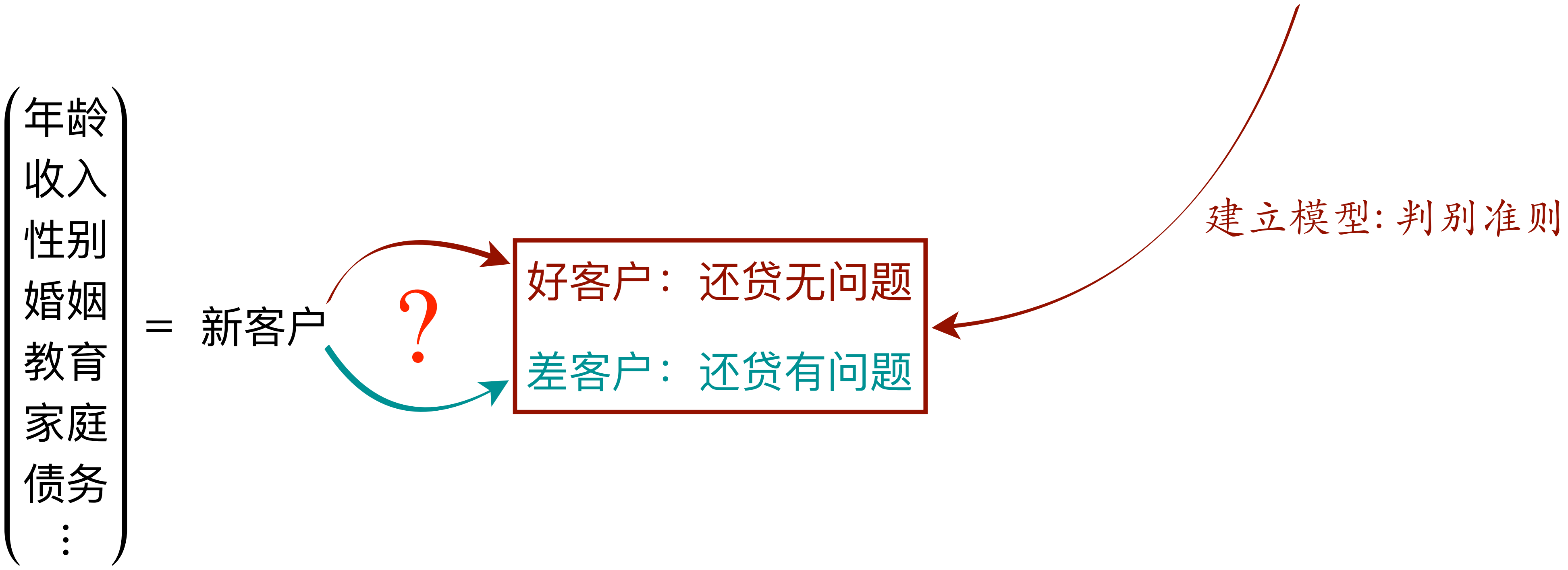
Outline

判别分析 (Discriminant Analysis)

- 分布已知的判别准则 (Allocation Rules for Known Distributions)
- 应用中的判别规则 (Discrimination Rules in Practice)
- R 中的判别分析

Preface 引言

- 判别分析用于分类结果事先已知的情形.
  - ▶ 判别分析的目的是将一个或若干个观察结果归到已知的类当中.
  - ▶ 例: 银行对客户的信用评价.
  - ▶ 银行数据：年龄，收入，性别，婚姻，教育程度，家庭成员，债务等.



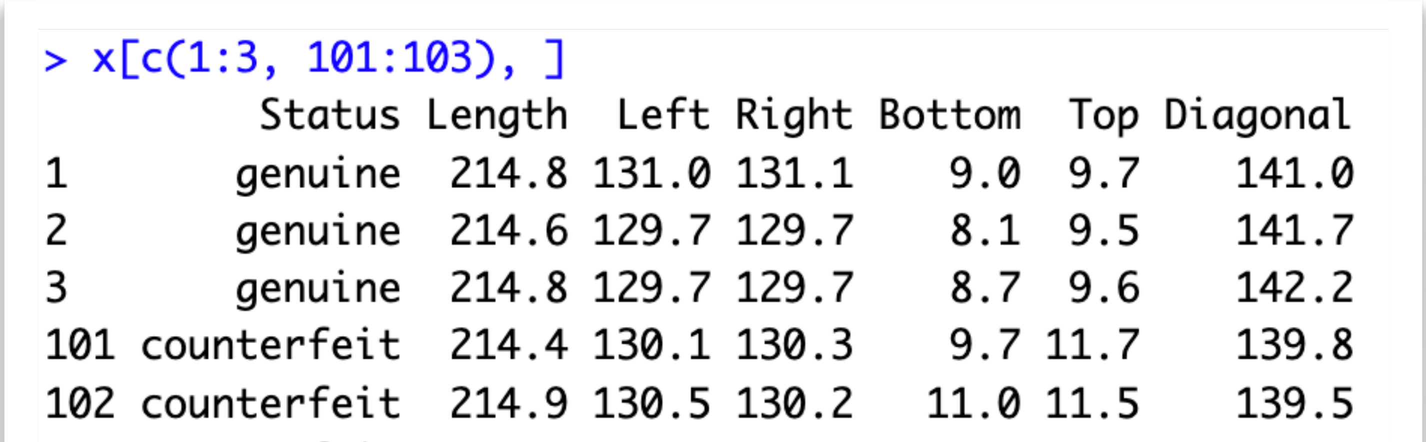
$$\begin{pmatrix} \text{年龄} \\ \text{收入} \\ \text{性别} \\ \text{婚姻} \\ \text{教育} \\ \text{家庭} \\ \text{债务} \\ \vdots \end{pmatrix} = \text{新客户}$$

  - ▶ 评价可能产生 “错判” 的风险.

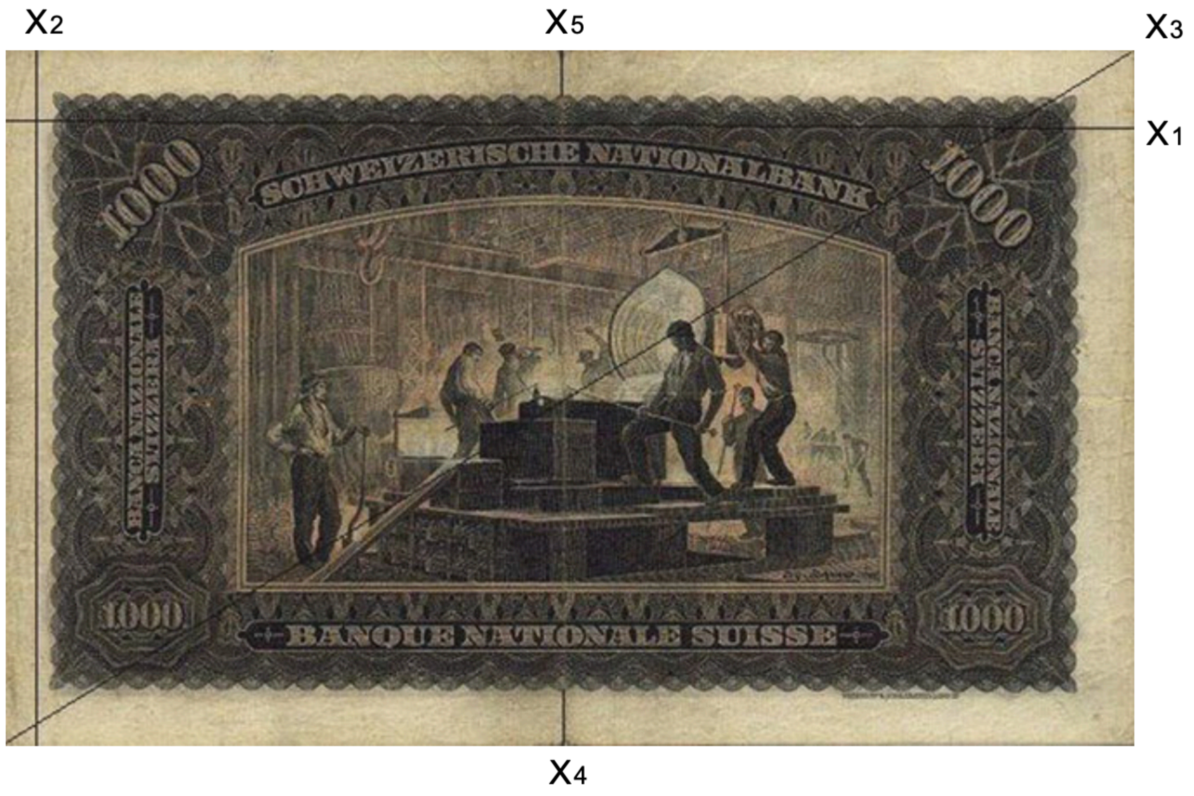
Allocation Rules for Known Distributions 分布已知的判别准则

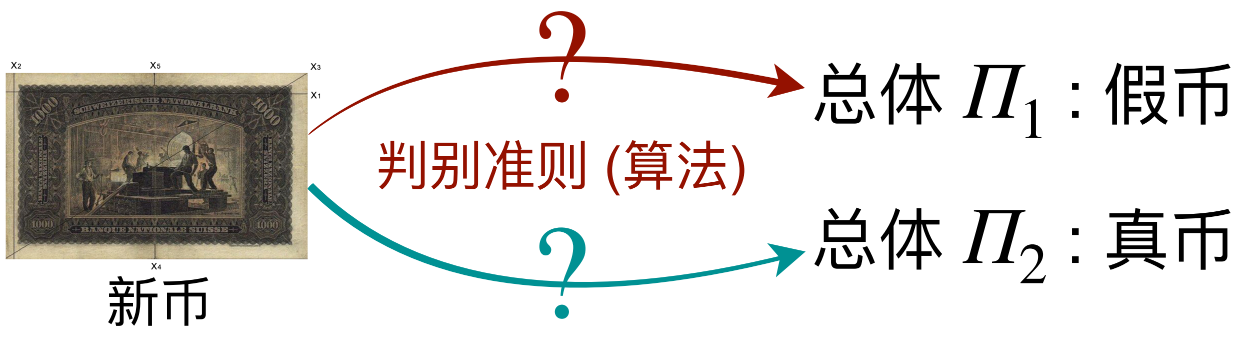
- 判别分析是用于区分一组总体 Π_j 并确定如何将新的观察结果归为哪一个总体的一系列方法.

```
library(mclust)
x = banknote
x[c(1:3, 101:103), ]
```



```
> x[c(1:3, 101:103), ]
         Status Length  Left Right Bottom  Top Diagonal
1       genuine  214.8 131.0 131.1    9.0  9.7    141.0
2       genuine  214.6 129.7 129.7    8.1  9.5    141.7
3       genuine  214.8 129.7 129.7    8.7  9.6    142.2
101 counterfeit  214.4 130.1 130.3    9.7 11.7    139.8
102 counterfeit  214.9 130.5 130.2   11.0 11.5    139.5
103 counterfeit  214.9 130.3 130.1    8.7 11.7    140.2
```





判别准则 (算法)

总体 Π_1 : 假币

总体 Π_2 : 真币

新币

- X_1 = 币的长度,
- X_2 = 币的宽度 (左),
- X_3 = 币的宽度 (右),
- X_4 = 币内框至下边界的距离,
- X_5 = 币内框至上边界的距离,
- X_6 = 币中心图片对角线的长度.

Allocation Rules for Known Distributions 分布已知的判别准则

- 判别分析是用于区分一组总体 $\varPi_j$ 并确定如何将新的观察结果归为哪一个总体的一系列方法.
  - 设有 J 个总体: $\varPi_1, \varPi_2, \ \ldots, \varPi_J$.
  - 拟将一个新的观测值 $\boldsymbol{x}$ 判定来自其中某一个总体.
  - 判别准则 (discriminant rule)：样本空间 $\mathbb{R}^p$ 的一个划分

$$R_1 \bigcup R_2 \bigcup \cdots \bigcup R_J = \mathbb{R}^p \ , \quad R_i \bigcap R_j = \varPhi \ (i \neq j)$$

$$\boldsymbol{x} \in R_j \qquad \Longrightarrow \qquad \boldsymbol{x} \in \varPi_j \ , \quad j = 1, 2, \ \ldots, J$$

  - 判别分析的主要任务：寻找一种 “好” 的划分，以使误判的概率尽可能小.

Allocation Rules for Known Distributions 分布已知的判别准则

- 极大似然判别准则 (Maximum Likelihood Discriminant Rule)
  - 记总体 Π_j 的密度函数为 $f_j(\boldsymbol{x}), j = 1, 2, \ldots, J.$
  - 极大似然判别准则 (ML 准则)：$L_j(\boldsymbol{x}) = f_j(\boldsymbol{x}) = \arg\max_i f_i(\boldsymbol{x}) \quad \Longrightarrow$ 判 $\boldsymbol{x}$ 属 Π_j .
  - 如果有若干个 $f_i(\boldsymbol{x})$ 同时使得目标函数达到最大，则任选其一判别即可.
  - 理论上，ML 准则确定的 $\mathbb{R}^p$ 的划分为

$$R_j = \left\{ \boldsymbol{x} \;\middle|\; L_j(\boldsymbol{x}) > L_i(\boldsymbol{x}) \;\;, i = 1, 2, \ldots, J \;, i \neq j \right\}$$

Allocation Rules for Known Distributions 分布已知的判别准则

- 极大似然判别准则 (Maximum Likelihood Discriminant Rule)
  - 在将一个观测值判别定给若干总体当中的一个时，可能会出现误判.
  - 当 $J = 2$ 时，如果 $\boldsymbol{x}$ 来自 Π_1 而我们判给了 Π_2，则误判的概率为

$$p_{21} = P\left(\boldsymbol{X} \in R_2 \middle| \Pi_1\right) = \int_{R_2} f_1(\boldsymbol{x})\, d\boldsymbol{x}$$

  - 同样，如果 $\boldsymbol{x}$ 来自 Π_2 而我们判给了 Π_1，则误判的概率为

$$p_{12} = P\left(\boldsymbol{X} \in R_1 \middle| \Pi_2\right) = \int_{R_1} f_2(\boldsymbol{x})\, d\boldsymbol{x}$$

Allocation Rules for Known Distributions 分布已知的判别准则

- 极大似然判别准则 (Maximum Likelihood Discriminant Rule)
  - 将来自总体 Π_j 的观测值 $\boldsymbol{x}$ 判给 R_i 时，我们用 $C\left(i \mid j\right)$ 表示误判的成本 (cost).

		判定总体	
		Π_1	Π_2
所属总体	Π_1	0	$C\left(2 \mid 1\right)$
	Π_2	$C\left(1 \mid 2\right)$	0

  - 误判期望成本 (expected cost of misclassification, ECM) 为

$$\mathrm{ECM} = C\left(2 \mid 1\right) \cdot p_{21} \cdot \underbrace{\boxed{\pi_1}}_{\text{来自总体 }\Pi_1\text{ 的先验概率}} + C\left(1 \mid 2\right) \cdot p_{12} \cdot \underbrace{\boxed{\pi_2}}_{\text{来自总体 }\Pi_2\text{ 的先验概率}}$$

  - 先验概率 π_j：随机选取的一个个体 $\boldsymbol{x}$ 属于总体 Π_j 的概率.

Allocation Rules for Known Distributions 分布已知的判别准则

- 极大似然判别准则 (Maximum Likelihood Discriminant Rule)
  - 我们关注 ECM 较小的判别准则，或若干个判别准则中 ECM 最小者.
  - 两个总体 ECM 最小的判别准则如下：

定理 14.1 对两个总体 Π_1 和 Π_2，使得误判期望成本 (ECM) 最小的判别准则为

$$R_1 = \left\{ \boldsymbol{x} \;\middle|\; \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} \geq \left(\frac{C(1|2)}{C(2|1)} \right) \left(\frac{\pi_2}{\pi_1} \right) \right\}$$

$$R_2 = \left\{ \boldsymbol{x} \;\middle|\; \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} < \left(\frac{C(1|2)}{C(2|1)} \right) \left(\frac{\pi_2}{\pi_1} \right) \right\}$$

  - 极大似然 (ML) 判别准则: ECM 等先验概率、等误判成本的一种特殊情形.

$$R_1 = \left\{ \boldsymbol{x} \;\middle|\; \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} \geq 1 \right\}, \quad R_2 = \left\{ \boldsymbol{x} \;\middle|\; \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} < 1 \right\}$$

Allocation Rules for Known Distributions 分布已知的判别准则

- Example: 假设

$$\begin{cases} \Pi_1 : \text{银行不良客户的总体} \\ \Pi_2 : \text{银行优质客户的总体} \end{cases} , \quad \begin{cases} C(2|1) : \text{不良客户判为优质客户产生的成本} \\ C(1|2) : \text{优质客户判为不良客户产生的成本} \end{cases}$$

- ▸ γ : 银行正确识别一个优质客户带来的收益.

$$G(R_2) = -C(2|1)\cdot\pi_1\cdot\int \mathrm{I}(\boldsymbol{x}\in R_2) f_1(\boldsymbol{x})\,d\boldsymbol{x} - C(1|2)\cdot\pi_2\cdot\int\left[1-\mathrm{I}(\boldsymbol{x}\in R_2)\right] f_2(\boldsymbol{x})\,d\boldsymbol{x} + \gamma\cdot\pi_2\cdot\int \mathrm{I}(\boldsymbol{x}\in R_2) f_2(\boldsymbol{x})\,d\boldsymbol{x}$$

$$= \boxed{-C(1|2)\cdot\pi_2} + \int \mathrm{I}(\boldsymbol{x}\in R_2)\left\{\boxed{-C(2|1)\,\pi_1 f_1(\boldsymbol{x}) + \left[C(1|2)+\gamma\right]\pi_2 f_2(\boldsymbol{x})}\right\} d\boldsymbol{x} \qquad \geq 0$$

常数

- ▸ 要收益最大，则

$$R_2 = \left\{ \boldsymbol{x} \;\middle|\; \frac{f_2(\boldsymbol{x})}{f_1(\boldsymbol{x})} \geq \frac{C(2|1)\cdot\pi_1}{\left[C(1|2)+\gamma\right]\cdot\pi_2} \right\} \xrightarrow{\gamma = 0} \text{定理 14.1}$$

Allocation Rules for Known Distributions 分布已知的判别准则

- Example: 假设 $x \in \{0\ ,\ 1\}$，并且

$$\begin{cases} \Pi_1 : P(X=0) = P(X=1) = \dfrac{1}{2} \\ \Pi_2 : P(X=0) = \dfrac{1}{4} = 1 - P(X=1) \end{cases}$$

ML 判别准则： $R_1 = \left\{ \boldsymbol{x} \;\middle|\; \dfrac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} \ge 1 \right\}$, $R_2 = \left\{ \boldsymbol{x} \;\middle|\; \dfrac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} < 1 \right\}$

$$\left. \begin{array}{l} x = 0 : \dfrac{f_1(0)}{f_2(0)} = \dfrac{\frac{1}{2}}{\frac{1}{4}} = 2 > 1 \\ x = 1 : \dfrac{f_1(1)}{f_2(1)} = \dfrac{\frac{1}{2}}{\frac{3}{4}} = \dfrac{2}{3} < 1 \end{array} \right\} \Longrightarrow \begin{array}{l} \{x=0\} \in \Pi_1 \\ \{x=1\} \in \Pi_2 \end{array} \Longrightarrow \begin{cases} R_1 = \{0\} \\ R_2 = \{1\} \end{cases}$$

Allocation Rules for Known Distributions 分布已知的判别准则

- Example: 假设有两个总体 $\begin{cases} \Pi_1 : N\left(\mu_1 , \sigma_1^2\right) \\ \Pi_2 : N\left(\mu_2 , \sigma_2^2\right) \end{cases}$

$$L_1(x) = f_1(x) = \frac{1}{\sqrt{2\pi}\ \sigma_1} \exp\left\{-\frac{\left(x-\mu_1\right)^2}{2\sigma_1^2}\right\} , \quad L_2(x) = f_2(x) = \frac{1}{\sqrt{2\pi}\ \sigma_2} \exp\left\{-\frac{\left(x-\mu_2\right)^2}{2\sigma_2^2}\right\}$$

ML 判别准则：$R_1 = \left\{ \boldsymbol{x} \;\middle|\; \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} \geq 1 \right\} , R_2 = \left\{ \boldsymbol{x} \;\middle|\; \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} < 1 \right\}$

$$x \in R_1 , \quad \frac{L_1(x)}{L_2(x)} = \frac{f_1(x)}{f_2(x)} = \frac{\sigma_2}{\sigma_1} \exp\left\{ -\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]\right\} \geq 1$$

$$\Longrightarrow \quad \ln\frac{\sigma_2}{\sigma_1} - \frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \left(\frac{x-\mu_2}{\sigma_2}\right)^2\right] \geq 0 \quad \Longrightarrow \quad \left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \left(\frac{x-\mu_2}{\sigma_2}\right)^2 \leq 2\ln\frac{\sigma_2}{\sigma_1}$$

$$\Longrightarrow \quad \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)x^2 - 2\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right)x + \left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2}\right) \leq 2\ln\frac{\sigma_2}{\sigma_1}$$

Allocation Rules for Known Distributions 分布已知的判别准则

- Example: 假设有两个总体 $\begin{cases} \Pi_1 : N\left(\mu_1 , \sigma_1^2\right) \\ \Pi_2 : N\left(\mu_2 , \sigma_2^2\right) \end{cases}$

  - 如果取：$\mu_1 = 0 , \sigma_1 = 1 , \mu_2 = 1 , \sigma_2 = \frac{1}{2}$

$$R_1 = \left\{ x \,\middle|\, x \le \frac{1}{3}\left(4 - \sqrt{4 + 6\ln 2}\right) \text{ 或 } x \ge \frac{1}{3}\left(4 + \sqrt{4 + 6\ln 2}\right) \right\}$$

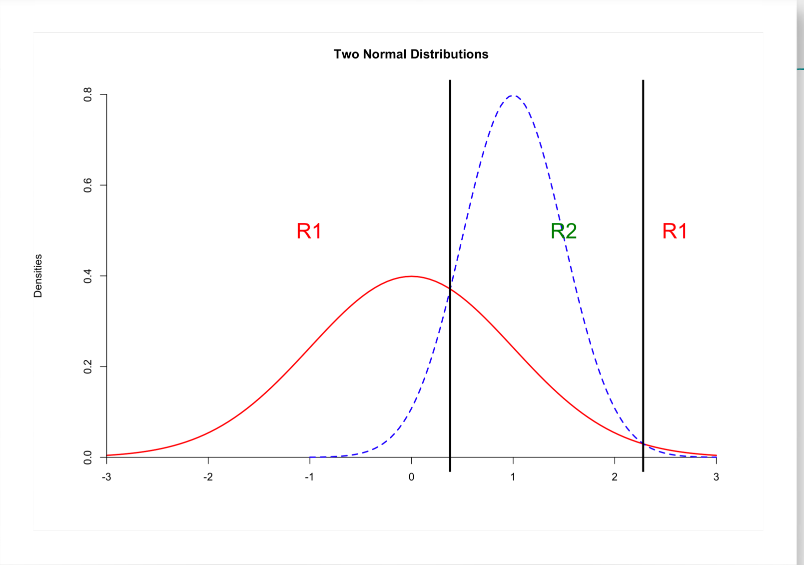
$$R_2 = \mathbb{R} \backslash R_1 = \left\{ x \,\middle|\, \frac{1}{3}\left(4 - \sqrt{4 + 6\ln 2}\right) < x < \frac{1}{3}\left(4 + \sqrt{4 + 6\ln 2}\right) \right\}$$

$$x \in R_1 \quad \Longrightarrow \quad \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)x^2 - 2\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right)x + \left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2}\right) \le 2\ln\frac{\sigma_2}{\sigma_1}$$

Allocation Rules for Known Distributions 分布已知的判别准则

- Example: 假设有两个总体 $\begin{cases} \Pi_1 : N\left(\mu_1 , \sigma_1^2\right) \\ \Pi_2 : N\left(\mu_2 , \sigma_2^2\right) \end{cases}$

```
## 方差不等
s1 = 1
mu1 = 0
s2 = 0.5
mu2    = 1
curve(dnorm(x, mu1, s1), -3, 3, col = "red", ylim = c(0, 0.8), axes = FALSE, xlab = "", ylab = "Densities", lwd = 2)
axis(1, at = -3:3, pos = 0)
axis(2, at = c(0, 0.2, 0.4, 0.6, 0.8), pos = -3)
curve(dnorm(x, mu2, s2), -1, 3, col = "blue", add = TRUE, lty = 2, lwd = 2)
title(main = "Two Normal Distributions")
abline(v = c(0.38, 2.28), lwd = 3)
text(-1, 0.5, "R1", col = "red", cex = 2)
text(2.6, 0.5, "R1", col = "red", cex = 2)
text(1.5, 0.5, "R2", col = "green4", cex = 2)
```



Allocation Rules for Known Distributions 分布已知的判别准则

- Example: 假设有两个总体 $\begin{cases} \Pi_1 : N\left(\mu_1 \text{ , } \sigma_1^2\right) \\ \Pi_2 : N\left(\mu_2 \text{ , } \sigma_2^2\right) \end{cases}$
  - 如果取：$\sigma_1 = \sigma_2$

$$R_1 = \left\{ x \,\middle|\, x \leq \frac{1}{2}\left(\mu_1 + \mu_2\right) \right\}$$

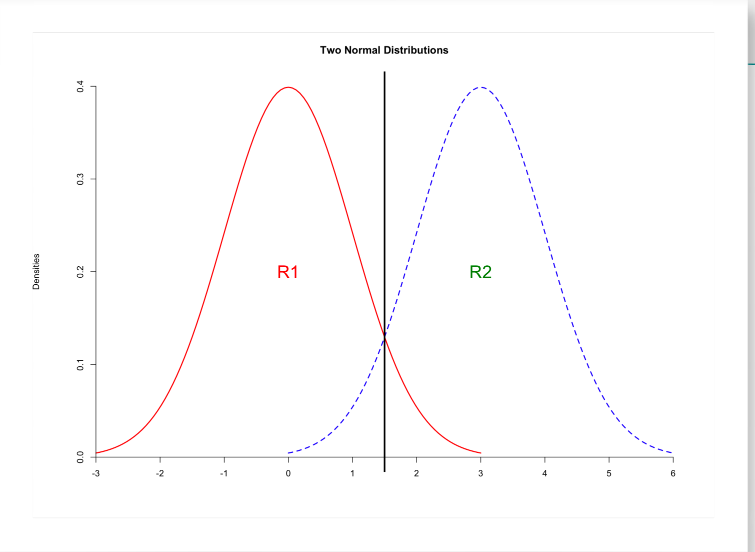
$$R_2 = \mathbb{R} \backslash R_1 = \left\{ x \,\middle|\, x > \frac{1}{2}\left(\mu_1 + \mu_2\right) \right\}$$

$$x \in R_1 \quad \Longrightarrow \quad \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) x^2 - 2\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right) x + \left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2}\right) \leq 2\ln\frac{\sigma_2}{\sigma_1}$$

Allocation Rules for Known Distributions 分布已知的判别准则

- **Example**: 假设有两个总体 $\begin{cases} \Pi_1 : N\left(\mu_1 \, , \, \sigma_1^2\right) \\ \Pi_2 : N\left(\mu_2 \, , \, \sigma_2^2\right) \end{cases}$

```
## 方差相等
s1 = 1
mu1 = 0
s2 = 1
mu2    = 3
curve(dnorm(x, mu1, s1), -3, 3, col = "red", xlim = c(-3, 6), ylim = c(0, 0.4), axes = FALSE,
     xlab = "", ylab = "Densities", lwd = 2)
axis(1, at = -3:6, pos = 0)
axis(2, at = c(0, 0.1, 0.2, 0.3, 0.4), pos = -3)
curve(dnorm(x, mu2, s2), 0, 6, col = "blue", add = TRUE, lty = 2, lwd = 2)
title(main = "Two Normal Distributions")
abline(v = 1.5, lwd = 3)
text(0, 0.2, "R1", col = "red", cex = 2)
text(3, 0.2, "R2", col = "green4", cex = 2)
```



Allocation Rules for Known Distributions 分布已知的判别准则

定理 14.2 假设 $\Pi_i \sim N_p(\boldsymbol{\mu}_i, \Sigma)$.

(a) 极大似然准则：如果 $\boldsymbol{x}$ 与 $\boldsymbol{\mu}_i$ 的 Mahalanobis 距离 $\delta(\boldsymbol{x}, \boldsymbol{\mu}_i)$ 的平方达到最小，则判 $\boldsymbol{x}$ 属 Π_i , $i = 1, 2, \ldots, J$, 其中

$$\delta^2(\boldsymbol{x}, \boldsymbol{\mu}_i) = (\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i), \quad i = 1, 2, \ldots, J$$

(b) 当 $J = 2$ 时，

$$\boldsymbol{x} \in R_1 \iff \boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu}) \geq 0$$

其中

$$\boldsymbol{\alpha} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \quad \boldsymbol{\mu} = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$$

Allocation Rules for Known Distributions 分布已知的判别准则

- 贝叶斯判别准则 (Bayes Discriminant Rule)
  - 记总体 Π_j 的密度函数为 $f_j(\boldsymbol{x}), j = 1, 2, \ \dots, J.$
  - 用 π_j 表示任意 $\boldsymbol{x} \in \mathbb{R}^p$ 来自总体 Π_j 的先验概率 (prior probability), $j = 1, 2, \ \dots, J.$

$$\sum_{j=1}^{J} \pi_j = 1$$

  - 贝叶斯判别准则： $\pi_j \cdot f_j(\boldsymbol{x}) = \max_{1 \le i \le J} \{\pi_i \cdot f_i(\boldsymbol{x})\} \quad \Longrightarrow \quad$ 判 $\boldsymbol{x}$ 属 Π_j
  - 贝叶斯判别准则确定的 $\mathbb{R}^p$ 的划分为

$$R_j = \left\{ \boldsymbol{x} \,\middle|\, \pi_j \cdot f_j(\boldsymbol{x}) \ge \pi_i \cdot f_i(\boldsymbol{x}) \ , i = 1, 2, \ \dots, J \right\}$$

$$\pi_1 = \pi_2 = \cdots = \pi_J = \frac{1}{J}$$

  - 极大似然判别准则 (ML 准则)： $L_j(\boldsymbol{x}) = f_j(\boldsymbol{x}) = \arg\max_i f_i(\boldsymbol{x}) \quad \Longrightarrow \quad$ 判 $\boldsymbol{x}$ 属 Π_j .

Allocation Rules for Known Distributions 分布已知的判别准则

- 贝叶斯判别准则 (Bayes Discriminant Rule)
  - 确定性判别准则 (deterministic discriminant rule)：

$$\phi_j(\boldsymbol{x}) = \begin{cases} 1, & \pi_j \cdot f_j(\boldsymbol{x}) = \max\limits_{1\le i\le J} \left\{ \pi_i \cdot f_i(\boldsymbol{x}) \right\} \\ 0, & \text{其它} \end{cases}, \quad j = 1, 2, \ldots, J$$

  - 随机化判别准则 (randomized discriminant rule)：

$$\phi_j(\boldsymbol{x}) = P\left(\boldsymbol{x} \in R_j\right), \quad j = 1, 2, \ldots, J, \quad \text{where} \quad \sum_{j=1}^{J} \phi_j(\boldsymbol{x}) = 1.$$

$$\phi_j(\boldsymbol{x}) = \max_{1\le i\le J} \left\{ \phi_i(\boldsymbol{x}) \right\} \quad \Longrightarrow \quad \text{判 } \boldsymbol{x} \text{ 属 } \mathit{\Pi}_j$$

Allocation Rules for Known Distributions 分布已知的判别准则

- 贝叶斯判别准则 (Bayes Discriminant Rule)
  - 哪一个判别准则最佳?
  - 用 p_{ij} 表示来自总体 Π_j 的个体 $\boldsymbol{x}$ 但判给总体 Π_i 的概率：

$$p_{ij} = \int \phi_i(\boldsymbol{x}) \cdot f_j(\boldsymbol{x}) d\boldsymbol{x} \ , \quad i,\ j = 1,\ 2,\ \ldots,\ J\ .$$

  - 假设有两个判别准则：(1) 概率 $\left\{p_{ij} \mid i,\ j = 1,\ 2,\ \ldots,\ J\right\}$;

    (2) 概率 $\left\{p'_{ij} \mid i,\ j = 1,\ 2,\ \ldots,\ J\right\}$.

  - 准则 (1) 优于准则 (2)：对所有的 $i = 1,\ 2,\ \ldots,\ J$ 都有 $p_{ii} \geq p'_{ii}$ ，且至少对某一个 i 有 $p_{ii} > p'_{ii}$ 严格成立.

Allocation Rules for Known Distributions 分布已知的判别准则

- 贝叶斯判别准则 (Bayes Discriminant Rule)
  - 对于某一个判别准则而言，如果不存在优于它的其它判别准则，则称该判别准则是容许的 (admissible).

定理 14.3 所有贝叶斯判别准则（包括极大似然 (ML) 准则）都是容许的.

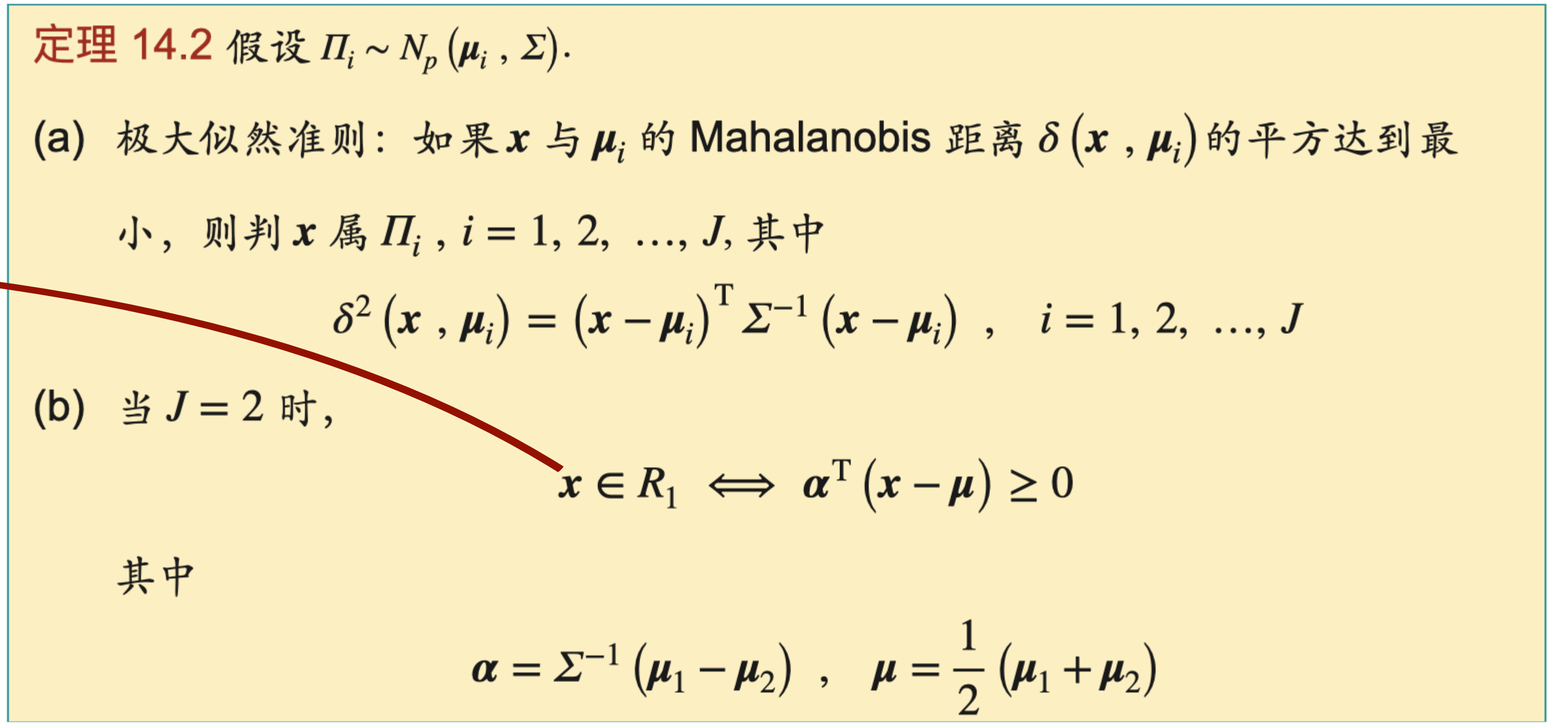
Allocation Rules for Known Distributions 分布已知的判别准则

- ML 准则 $(J = 2)$ 的误判概率

$$p_{12} = P\left(\boldsymbol{x} \in R_1 \,\middle|\, \Pi_2\right)$$

$$= P\left(\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu}) > 0 \,\middle|\, \Pi_2\right)$$

$$= P\left(\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu}) > 0\right)$$



定理 14.2 假设 $\Pi_i \sim N_p(\boldsymbol{\mu}_i, \Sigma)$.

(a) 极大似然准则：如果 $\boldsymbol{x}$ 与 $\boldsymbol{\mu}_i$ 的 Mahalanobis 距离 $\delta(\boldsymbol{x}, \boldsymbol{\mu}_i)$ 的平方达到最小，则判 $\boldsymbol{x}$ 属 Π_i，$i = 1, 2, \ldots, J$, 其中

$$\delta^2(\boldsymbol{x}, \boldsymbol{\mu}_i) = (\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i), \quad i = 1, 2, \ldots, J$$

(b) 当 $J = 2$ 时，

$$\boldsymbol{x} \in R_1 \iff \boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu}) \geq 0$$

其中

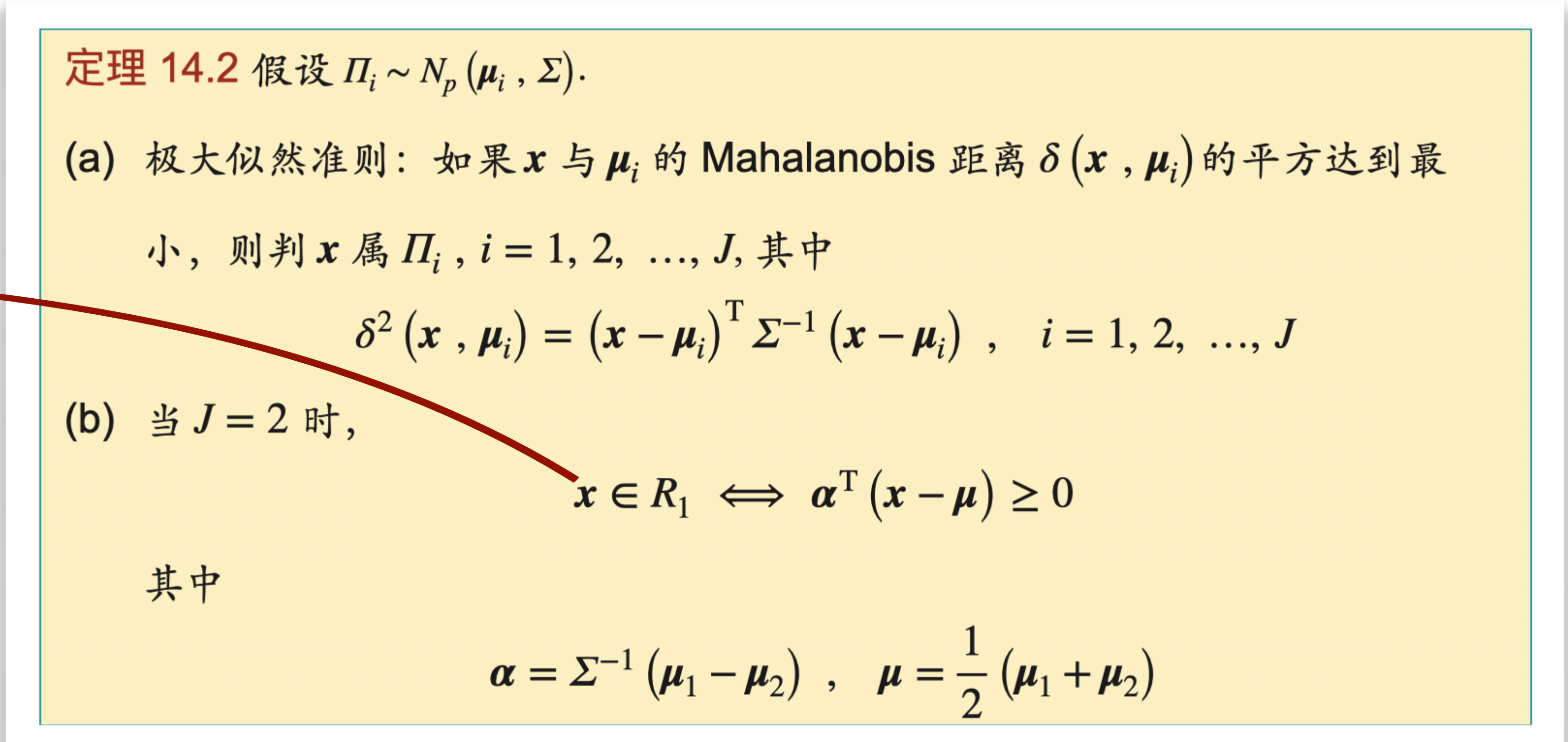
$$\boldsymbol{\alpha} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \quad \boldsymbol{\mu} = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$$

如果 $\boldsymbol{x}$ 来自 Π_2 $\Longrightarrow$ $\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu}) \sim N_1\left(-\frac{1}{2}\delta^2, \delta^2\right)$

$\boldsymbol{\mu}_1$ 与 $\boldsymbol{\mu}_2$ 的 Mahalnobis 距离：$\delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$

Allocation Rules for Known Distributions 分布已知的判别准则

- ML 准则 ($J = 2$) 的误判概率



定理 14.2 假设 $\Pi_i \sim N_p(\boldsymbol{\mu}_i\,, \Sigma)$.

(a) 极大似然准则: 如果 $\boldsymbol{x}$ 与 $\boldsymbol{\mu}_i$ 的 Mahalanobis 距离 $\delta(\boldsymbol{x}\,, \boldsymbol{\mu}_i)$ 的平方达到最小, 则判 $\boldsymbol{x}$ 属 Π_i, $i = 1, 2, \dots, J$, 其中

$$\delta^2(\boldsymbol{x}\,, \boldsymbol{\mu}_i) = (\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_i) \;, \quad i = 1, 2, \dots, J$$

(b) 当 $J = 2$ 时,

$$\boldsymbol{x} \in R_1 \iff \boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) \geq 0$$

其中

$$\boldsymbol{\alpha} = \Sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \;, \quad \boldsymbol{\mu} = \frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)$$

$$p_{12} = P\left(\boldsymbol{x} \in R_1 \,\middle|\, \Pi_2\right)$$

$$= P\left(\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) > 0 \,\middle|\, \Pi_2\right)$$

$$= P\left(\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) > 0\right)$$

$$= P\left(\frac{\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) - \left(-\frac{1}{2}\delta^2\right)}{\delta} > \frac{0 - \left(-\frac{1}{2}\delta^2\right)}{\delta}\right)$$

$$= P\left(\frac{\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) + \frac{1}{2}\delta^2}{\delta} > \frac{1}{2}\delta\right) = 1 - \Phi\left(\frac{1}{2}\delta\right) = \Phi\left(-\frac{1}{2}\delta\right)$$

$$\Longrightarrow \quad \boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) \sim N_1\left(-\frac{1}{2}\delta^2 \,, \delta^2\right)$$

$$\Longrightarrow \quad \frac{\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) + \frac{1}{2}\delta^2}{\delta} \sim N_1(0 \,, 1)$$

▸ 同理可得: $p_{21} = \Phi\left(-\frac{1}{2}\delta\right)$.

Allocation Rules for Known Distributions 分布已知的判别准则

- 协方差矩阵不等时的判别
  - 此时 ECM (expected cost of misclassification, 误判期望成本) 依赖于密度之比 $\dfrac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}$.
  - 或等价地，依赖于 $\ln\left\{f_1(\boldsymbol{x})\right\} - \ln\left\{f_2(\boldsymbol{x})\right\}$.
  - 当两个总体的协方差矩阵不等时，判别准则变得更为复杂：

$$R_1 = \left\{ \boldsymbol{x} \,\middle|\, \underbrace{-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\left(\Sigma_1^{-1} - \Sigma_2^{-1}\right)\boldsymbol{x} + \left(\boldsymbol{\mu}_1^{\mathrm{T}}\Sigma_1^{-1} - \boldsymbol{\mu}_2^{\mathrm{T}}\Sigma_2^{-1}\right)\boldsymbol{x} - k}_{\text{二次函数}} \geq \ln\left[\frac{c\left(1\,|\,2\right)}{c\left(2\,|\,1\right)}\cdot\frac{\pi_2}{\pi_1}\right] \right\}$$

$$R_2 = \left\{ \boldsymbol{x} \,\middle|\, -\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\left(\Sigma_1^{-1} - \Sigma_2^{-1}\right)\boldsymbol{x} + \left(\boldsymbol{\mu}_1^{\mathrm{T}}\Sigma_1^{-1} - \boldsymbol{\mu}_2^{\mathrm{T}}\Sigma_2^{-1}\right)\boldsymbol{x} - k < \ln\left[\frac{c\left(1\,|\,2\right)}{c\left(2\,|\,1\right)}\cdot\frac{\pi_2}{\pi_1}\right] \right\}$$

$$k = \frac{1}{2}\ln\left(\frac{\left|\Sigma_1\right|}{\left|\Sigma_2\right|}\right) + \frac{1}{2}\ln\left(\boldsymbol{\mu}_1^{\mathrm{T}}\Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^{\mathrm{T}}\Sigma_2^{-1}\boldsymbol{\mu}_2\right)$$

Discrimination Rules in Practice 应用当中的判别准则

- 如果总体的分布与参数均已知，则使用 ML 准则.
  - 假设数据来自多元正态总体 $N_p\left(\boldsymbol{\mu}_j\,,\,\Sigma\right)$, $j = 1, 2, \ldots, J$.

$$\begin{cases} \boldsymbol{x}_{11}, \boldsymbol{x}_{12}, \ \ldots, \boldsymbol{x}_{1n_1} \overset{\text{i.i.d.}}{\sim} N_p\left(\boldsymbol{\mu}_1\,,\,\Sigma\right) \\ \boldsymbol{x}_{21}, \boldsymbol{x}_{22}, \ \ldots, \boldsymbol{x}_{2n_2} \overset{\text{i.i.d.}}{\sim} N_p\left(\boldsymbol{\mu}_2\,,\,\Sigma\right) \\ \qquad\vdots \qquad\qquad\qquad\qquad \vdots \\ \boldsymbol{x}_{J1}, \boldsymbol{x}_{J2}, \ \ldots, \boldsymbol{x}_{Jn_J} \overset{\text{i.i.d.}}{\sim} N_p\left(\boldsymbol{\mu}_J\,,\,\Sigma\right) \end{cases}$$

$$\Longrightarrow \quad \widehat{\boldsymbol{\mu}}_j = \overline{\boldsymbol{x}}_j = \frac{1}{n_j}\sum_{k=1}^{n_j} \boldsymbol{x}_{kj}\,, \quad j = 1, 2, \ \ldots, J$$

$$\mathscr{S}_j = \frac{1}{n_j}\sum_{k=1}^{n_j}\left(\boldsymbol{x}_{jk} - \overline{\boldsymbol{x}}_j\right)\left(\boldsymbol{x}_{jk} - \overline{\boldsymbol{x}}_j\right)^{\mathrm{T}}$$

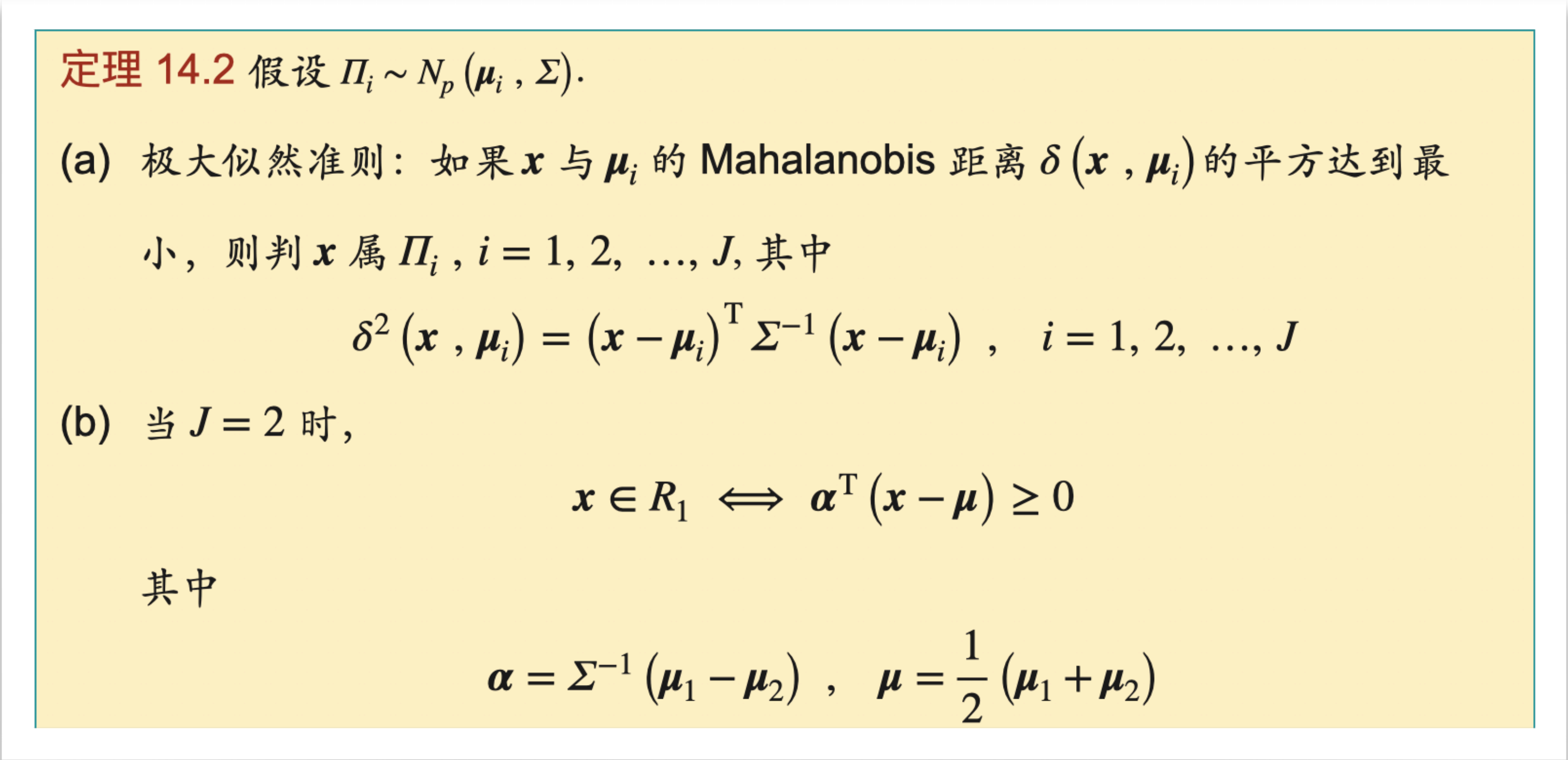
$$\widehat{\Sigma} = \mathscr{S}_u = \sum_{j=1}^{J} n_j\left(\frac{\mathscr{S}_j}{n-J}\right)$$

$$n = \sum_{j=1}^{J} n_j$$

$$= \frac{1}{n-J}\sum_{j=1}^{J}\sum_{k=1}^{n_j}\left(\boldsymbol{x}_{jk} - \overline{\boldsymbol{x}}_j\right)\left(\boldsymbol{x}_{jk} - \overline{\boldsymbol{x}}_j\right)^{\mathrm{T}}$$

Discrimination Rules in Practice 应用当中的判别准则

- 如果总体的分布与参数均已知，则使用 ML 准则.
    - 假设数据来自多元正态总体 $N_p\left(\boldsymbol{\mu}_j\ , \Sigma\right)\ , j = 1, 2,\ \ldots, J.$



定理 14.2 假设 $\Pi_i \sim N_p\left(\boldsymbol{\mu}_i\ , \Sigma\right)$.

(a) 极大似然准则：如果 $\boldsymbol{x}$ 与 $\boldsymbol{\mu}_i$ 的 Mahalanobis 距离 $\delta\left(\boldsymbol{x}\ , \boldsymbol{\mu}_i\right)$ 的平方达到最小，则判 $\boldsymbol{x}$ 属 $\Pi_i\ , i = 1, 2,\ \ldots, J,$ 其中

$$\delta^2\left(\boldsymbol{x}\ , \boldsymbol{\mu}_i\right) = \left(\boldsymbol{x} - \boldsymbol{\mu}_i\right)^{\mathrm{T}} \Sigma^{-1} \left(\boldsymbol{x} - \boldsymbol{\mu}_i\right)\ , \quad i = 1, 2,\ \ldots, J$$

(b) 当 $J = 2$ 时，

$$\boldsymbol{x} \in R_1 \iff \boldsymbol{\alpha}^{\mathrm{T}}\left(\boldsymbol{x} - \boldsymbol{\mu}\right) \geq 0$$

其中

$$\boldsymbol{\alpha} = \Sigma^{-1}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)\ , \quad \boldsymbol{\mu} = \frac{1}{2}\left(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2\right)$$

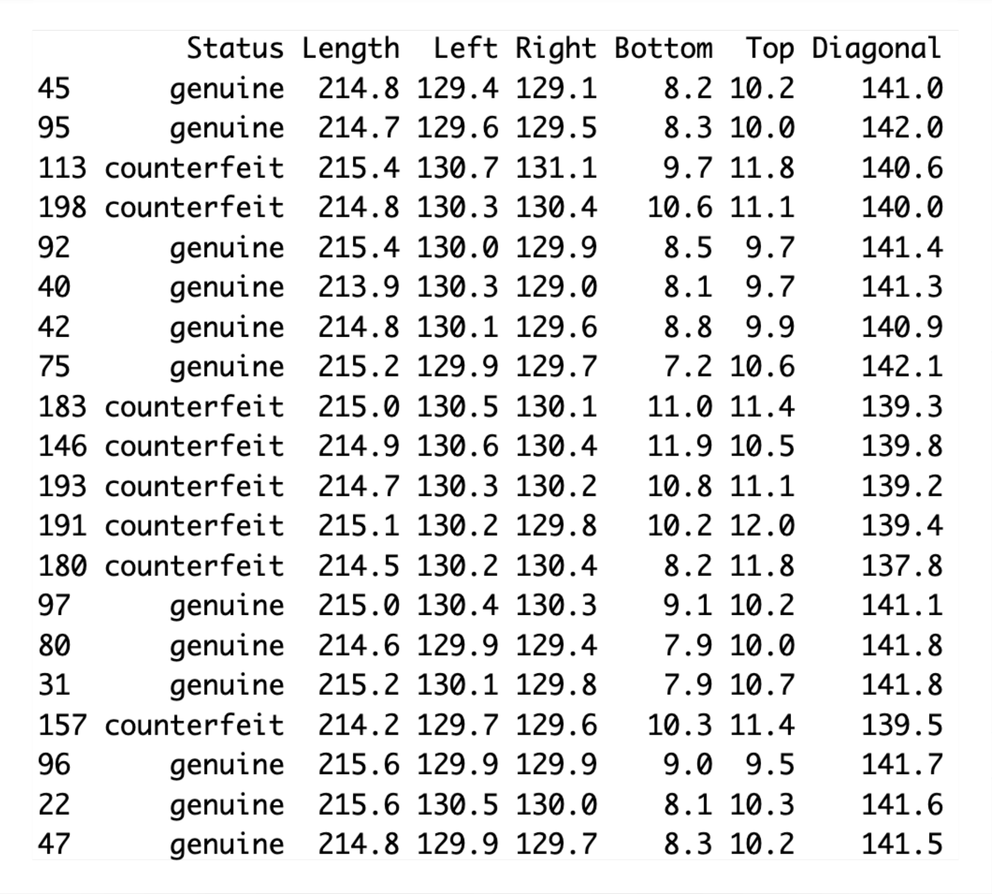
$$\left(\boldsymbol{x} - \bar{\boldsymbol{x}}_k\right) \mathcal{S}_u^{-1} \left(\boldsymbol{x} - \bar{\boldsymbol{x}}_k\right)^{\mathrm{T}} = \min_{1 \leq i \leq J} \left\{ \left(\boldsymbol{x} - \bar{\boldsymbol{x}}_i\right) \mathcal{S}_u^{-1} \left(\boldsymbol{x} - \bar{\boldsymbol{x}}_i\right)^{\mathrm{T}} \right\} \quad \Longrightarrow \quad \boldsymbol{x} \in R_k$$

即，判 $\boldsymbol{x}$ 属 Π_k

Discrimination Rules in Practice 应用当中的判别准则

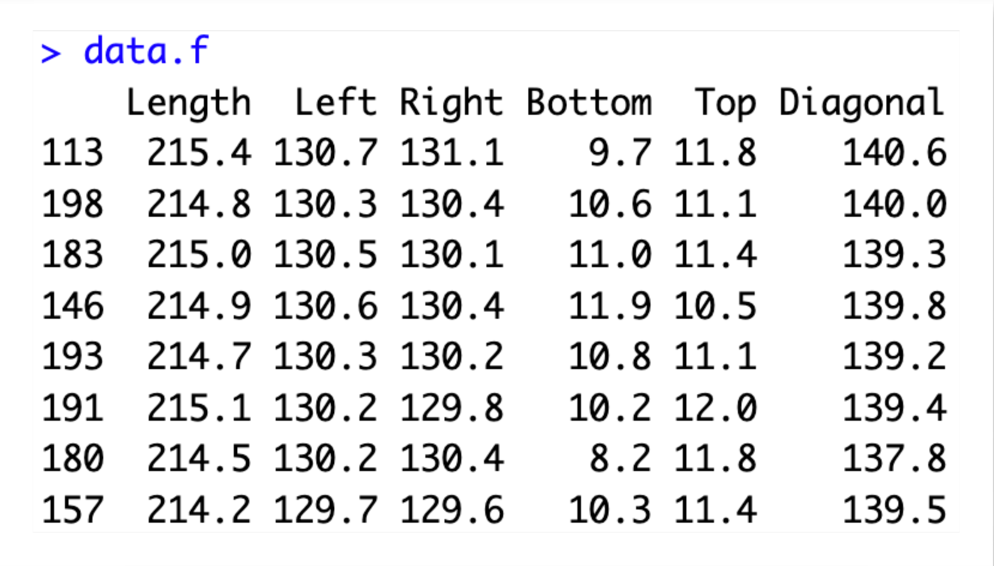
- Example: 对真伪钞票数据集应用 ML 判别准则.
  - 在真伪钞票数据集当中随机选择 20 个数据：

```
library(mclust)
data(banknote)
set.seed(1993)
n0 = sample(1:200, size = 20, replace = FALSE)
x = banknote[n0, ]
x
```

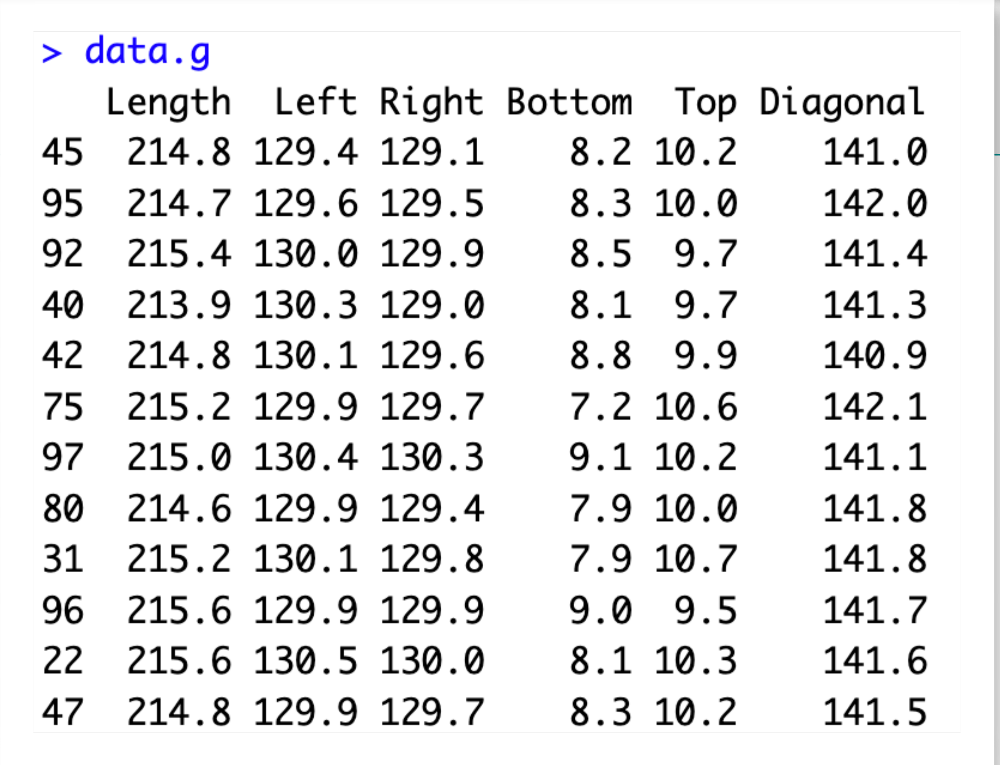


```
         Status Length  Left Right Bottom  Top Diagonal
45      genuine  214.8 129.4 129.1    8.2 10.2    141.0
95      genuine  214.7 129.6 129.5    8.3 10.0    142.0
113 counterfeit  215.4 130.7 131.1    9.7 11.8    140.6
198 counterfeit  214.8 130.3 130.4   10.6 11.1    140.0
92      genuine  215.4 130.0 129.9    8.5  9.7    141.4
40      genuine  213.9 130.3 129.0    8.1  9.7    141.3
42      genuine  214.8 130.1 129.6    8.8  9.9    140.9
75      genuine  215.2 129.9 129.7    7.2 10.6    142.1
183 counterfeit  215.0 130.5 130.1   11.0 11.4    139.3
146 counterfeit  214.9 130.6 130.4   11.9 10.5    139.8
193 counterfeit  214.7 130.3 130.2   10.8 11.1    139.2
191 counterfeit  215.1 130.2 129.8   10.2 12.0    139.4
180 counterfeit  214.5 130.2 130.4    8.2 11.8    137.8
97      genuine  215.0 130.4 130.3    9.1 10.2    141.1
80      genuine  214.6 129.9 129.4    7.9 10.0    141.8
31      genuine  215.2 130.1 129.8    7.9 10.7    141.8
157 counterfeit  214.2 129.7 129.6   10.3 11.4    139.5
96      genuine  215.6 129.9 129.9    9.0  9.5    141.7
22      genuine  215.6 130.5 130.0    8.1 10.3    141.6
47      genuine  214.8 129.9 129.7    8.3 10.2    141.5
```

```
data.g = subset(x, Status == "genuine", select = 2:7) # 真钞子集
data.g
data.f = subset(x, Status == "counterfeit", select = 2:7) # 伪钞子集
data.f
```



```
> data.f
    Length  Left Right Bottom  Top Diagonal
113  215.4 130.7 131.1    9.7 11.8    140.6
198  214.8 130.3 130.4   10.6 11.1    140.0
183  215.0 130.5 130.1   11.0 11.4    139.3
146  214.9 130.6 130.4   11.9 10.5    139.8
193  214.7 130.3 130.2   10.8 11.1    139.2
191  215.1 130.2 129.8   10.2 12.0    139.4
180  214.5 130.2 130.4    8.2 11.8    137.8
157  214.2 129.7 129.6   10.3 11.4    139.5
```



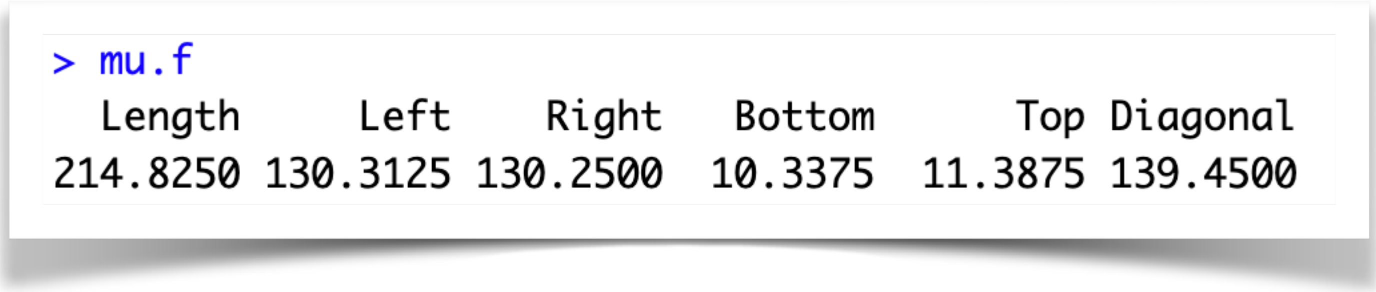
```
> data.g
   Length  Left Right Bottom  Top Diagonal
45  214.8 129.4 129.1    8.2 10.2    141.0
95  214.7 129.6 129.5    8.3 10.0    142.0
92  215.4 130.0 129.9    8.5  9.7    141.4
40  213.9 130.3 129.0    8.1  9.7    141.3
42  214.8 130.1 129.6    8.8  9.9    140.9
75  215.2 129.9 129.7    7.2 10.6    142.1
97  215.0 130.4 130.3    9.1 10.2    141.1
80  214.6 129.9 129.4    7.9 10.0    141.8
31  215.2 130.1 129.8    7.9 10.7    141.8
96  215.6 129.9 129.9    9.0  9.5    141.7
22  215.6 130.5 130.0    8.1 10.3    141.6
47  214.8 129.9 129.7    8.3 10.2    141.5
```

Discrimination Rules in Practice 应用当中的判别准则

- Example: 对真伪钞票数据集应用 ML 判别准则.
    - 计算两个数据集的样本均值向量:

```
mu.g = apply(data.g, 2, mean) # 真钞的均值向量
mu.g
mu.f = apply(data.f, 2, mean) # 伪钞的均值向量
mu.f
```

```
> mu.g
    Length       Left      Right     Bottom        Top   Diagonal
214.966667 130.000000 129.658333   8.283333  10.083333 141.516667
```

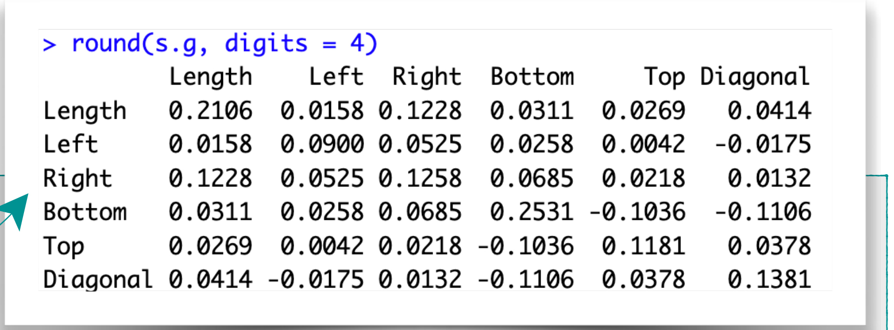


```
> mu.f
  Length     Left    Right   Bottom      Top Diagonal
214.8250 130.3125 130.2500  10.3375  11.3875 139.4500
```

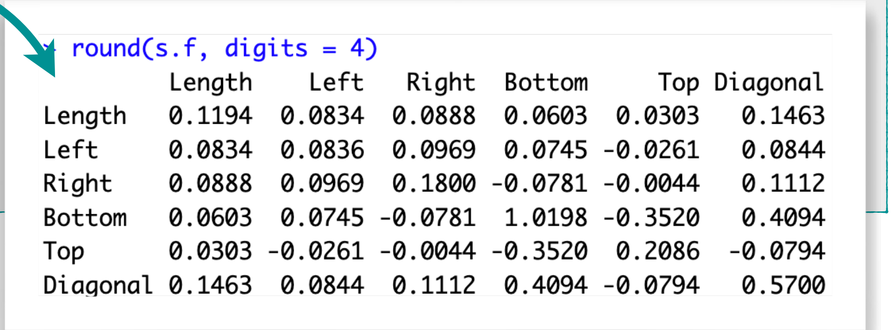
Discrimination Rules in Practice 应用当中的判别准则

- Example: 对真伪钞票数据集应用 ML 判别准则.
  - 计算合并的样本协方差矩阵： $\widehat{\Sigma} = \mathcal{S}_u = \sum_{j=1}^{J} n_j \left(\frac{\mathcal{S}_j}{n - J} \right)$

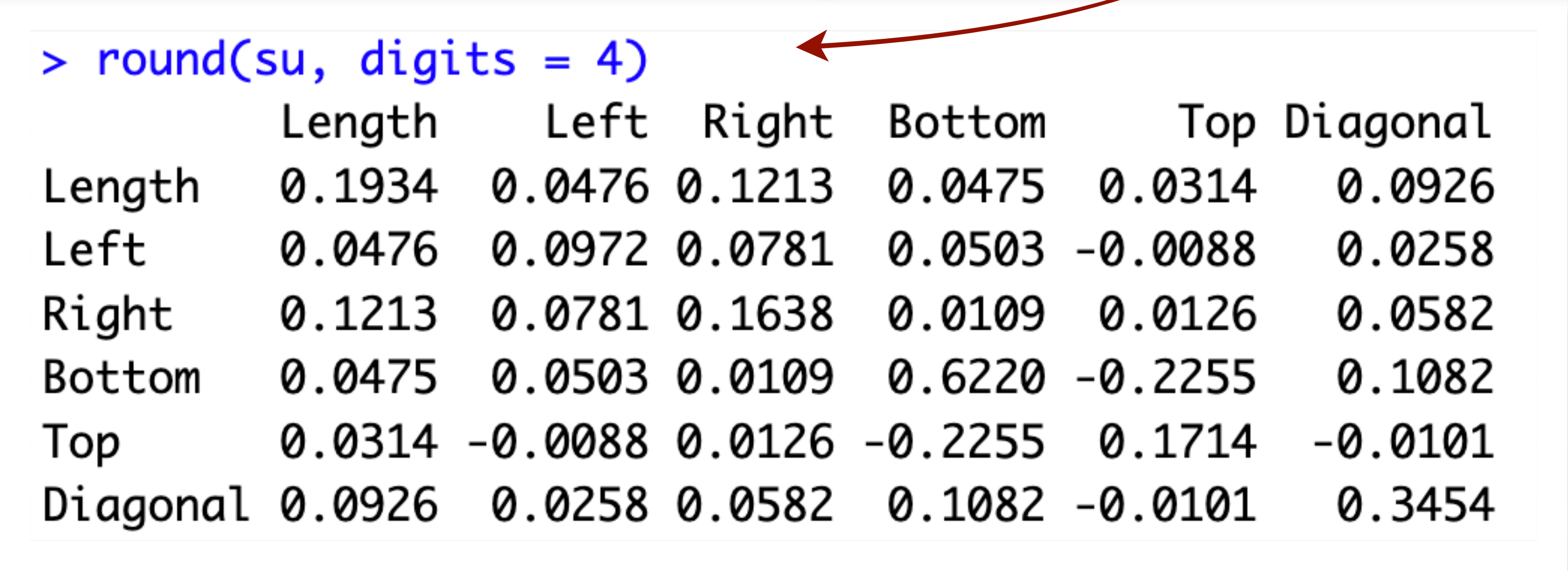
```
s.g = cov(data.g) * (dim(data.g)[1] - 1) / (dim(data.g)[1]) # 真钞的样本协方差矩阵
round(s.g, digits = 4)
s.f = cov(data.f) * (dim(data.f)[1] - 1) / (dim(data.f)[1]) # 伪钞的样本协方差矩阵
round(s.f, digits = 4)
su = ((dim(data.g)[1]) * s.g + (dim(data.f)[1]) * s.f) / (20 - 2) # 合并计算的样本协方差矩阵
round(su, digits = 4)
```



```
> round(s.g, digits = 4)
         Length    Left  Right  Bottom     Top Diagonal
Length   0.2106  0.0158 0.1228  0.0311  0.0269   0.0414
Left     0.0158  0.0900 0.0525  0.0258  0.0042  -0.0175
Right    0.1228  0.0525 0.1258  0.0685  0.0218   0.0132
Bottom   0.0311  0.0258 0.0685  0.2531 -0.1036  -0.1106
Top      0.0269  0.0042 0.0218 -0.1036  0.1181   0.0378
Diagonal 0.0414 -0.0175 0.0132 -0.1106  0.0378   0.1381
```



```
> round(s.f, digits = 4)
         Length    Left   Right  Bottom     Top Diagonal
Length   0.1194  0.0834  0.0888  0.0603  0.0303   0.1463
Left     0.0834  0.0836  0.0969  0.0745 -0.0261   0.0844
Right    0.0888  0.0969  0.1800 -0.0781 -0.0044   0.1112
Bottom   0.0603  0.0745 -0.0781  1.0198 -0.3520   0.4094
Top      0.0303 -0.0261 -0.0044 -0.3520  0.2086  -0.0794
Diagonal 0.1463  0.0844  0.1112  0.4094 -0.0794   0.5700
```



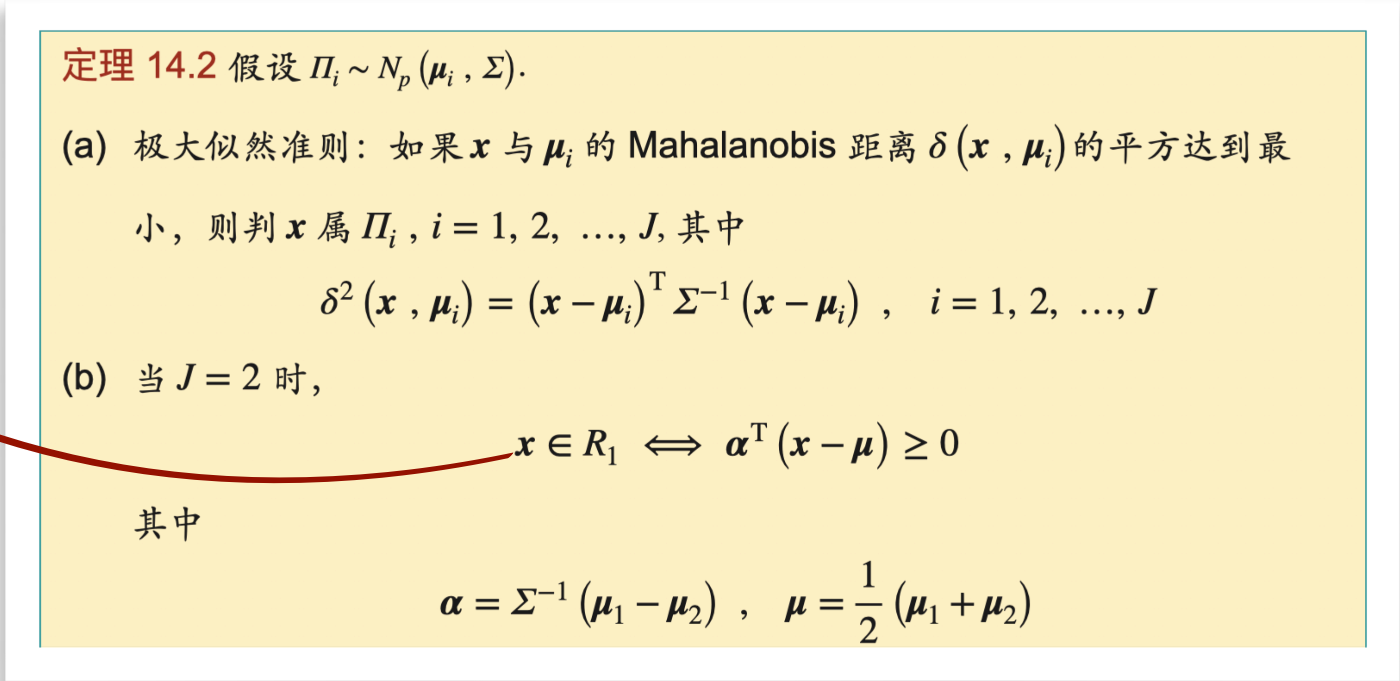
```
> round(su, digits = 4)
         Length    Left  Right  Bottom     Top Diagonal
Length   0.1934  0.0476 0.1213  0.0475  0.0314   0.0926
Left     0.0476  0.0972 0.0781  0.0503 -0.0088   0.0258
Right    0.1213  0.0781 0.1638  0.0109  0.0126   0.0582
Bottom   0.0475  0.0503 0.0109  0.6220 -0.2255   0.1082
Top      0.0314 -0.0088 0.0126 -0.2255  0.1714  -0.0101
Diagonal 0.0926  0.0258 0.0582  0.1082 -0.0101   0.3454
```

Discrimination Rules in Practice 应用当中的判别准则

- Example: 对真伪钞票数据集应用 ML 判别准则.
  - 划分两个总体的超平面：

$$\hat{\alpha}^{\mathrm{T}}\left(\boldsymbol{x}-\hat{\boldsymbol{\mu}}\right)=0$$

```
mu.g = as.matrix(mu.g)
mu.f = as.matrix(mu.f)
su = as.matrix(su)
alpha = solve(su) %*% (mu.g - mu.f)
round(alpha, digits = 4)
mu = (mu.g + mu.f) / 2
round(mu, digits = 4)
```



定理 14.2 假设 $\Pi_i \sim N_p(\boldsymbol{\mu}_i, \Sigma)$.

(a) 极大似然准则：如果 $\boldsymbol{x}$ 与 $\boldsymbol{\mu}_i$ 的 Mahalanobis 距离 $\delta(\boldsymbol{x}, \boldsymbol{\mu}_i)$ 的平方达到最小，则判 $\boldsymbol{x}$ 属 Π_i, $i = 1, 2, \ldots, J$, 其中

$$\delta^2(\boldsymbol{x}, \boldsymbol{\mu}_i) = (\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i), \quad i=1,2,\ldots,J$$

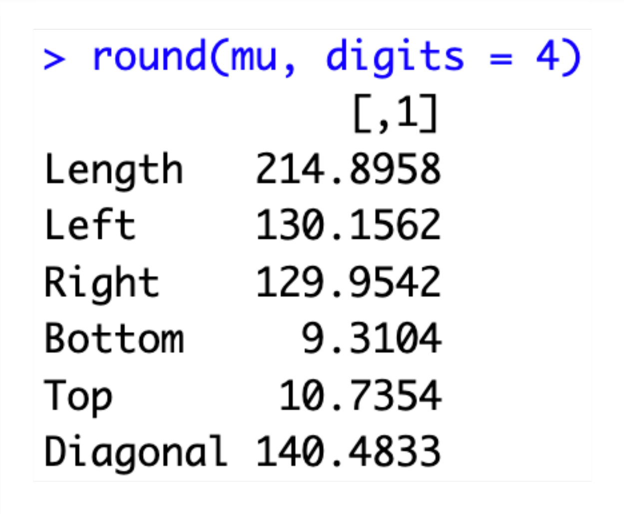
(b) 当 $J=2$ 时，

$$\boldsymbol{x}\in R_1 \iff \boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) \geq 0$$

其中

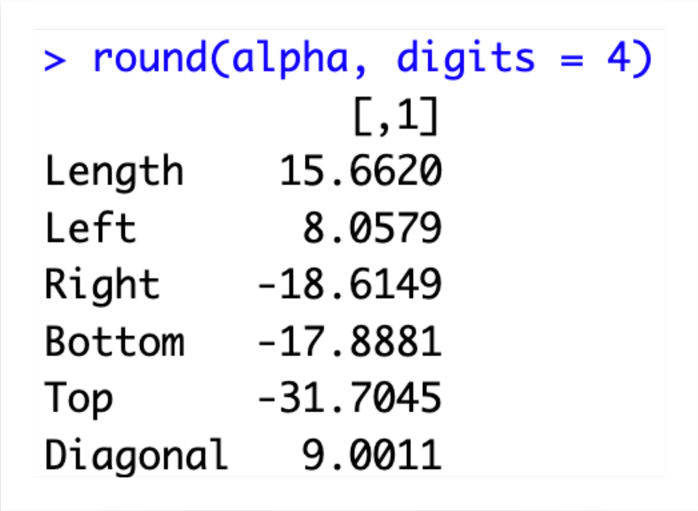
$$\boldsymbol{\alpha}=\Sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2), \quad \boldsymbol{\mu}=\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)$$

$$\hat{\boldsymbol{\mu}}=\frac{1}{2}\left(\hat{\boldsymbol{\mu}}_1+\hat{\boldsymbol{\mu}}_2\right)=$$



```
> round(mu, digits = 4)
             [,1]
Length   214.8958
Left     130.1562
Right    129.9542
Bottom     9.3104
Top       10.7354
Diagonal 140.4833
```

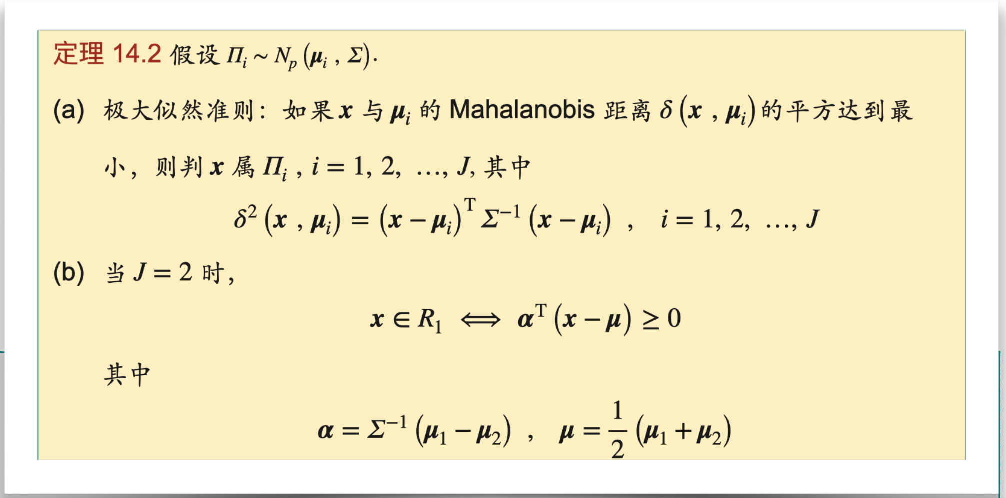
$$\hat{\boldsymbol{\alpha}}=\hat{\Sigma}^{-1}\left(\hat{\boldsymbol{\mu}}_1-\hat{\boldsymbol{\mu}}_2\right)=$$



```
> round(alpha, digits = 4)
            [,1]
Length   15.6620
Left      8.0579
Right   -18.6149
Bottom  -17.8881
Top     -31.7045
Diagonal  9.0011
```

Discrimination Rules in Practice 应用当中的判别准则

- Example: 对真伪钞票数据集应用 ML 判别准则.
    - 对真伪钞票数据集所有个体的回判：



定理 14.2 假设 $\Pi_i \sim N_p(\boldsymbol{\mu}_i, \Sigma)$.

(a) 极大似然准则：如果 $\boldsymbol{x}$ 与 $\boldsymbol{\mu}_i$ 的 Mahalanobis 距离 $\delta(\boldsymbol{x}, \boldsymbol{\mu}_i)$ 的平方达到最小，则判 $\boldsymbol{x}$ 属 Π_i, $i = 1, 2, \dots, J$, 其中

$$\delta^2(\boldsymbol{x}, \boldsymbol{\mu}_i) = (\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i), \quad i = 1, 2, \dots, J$$

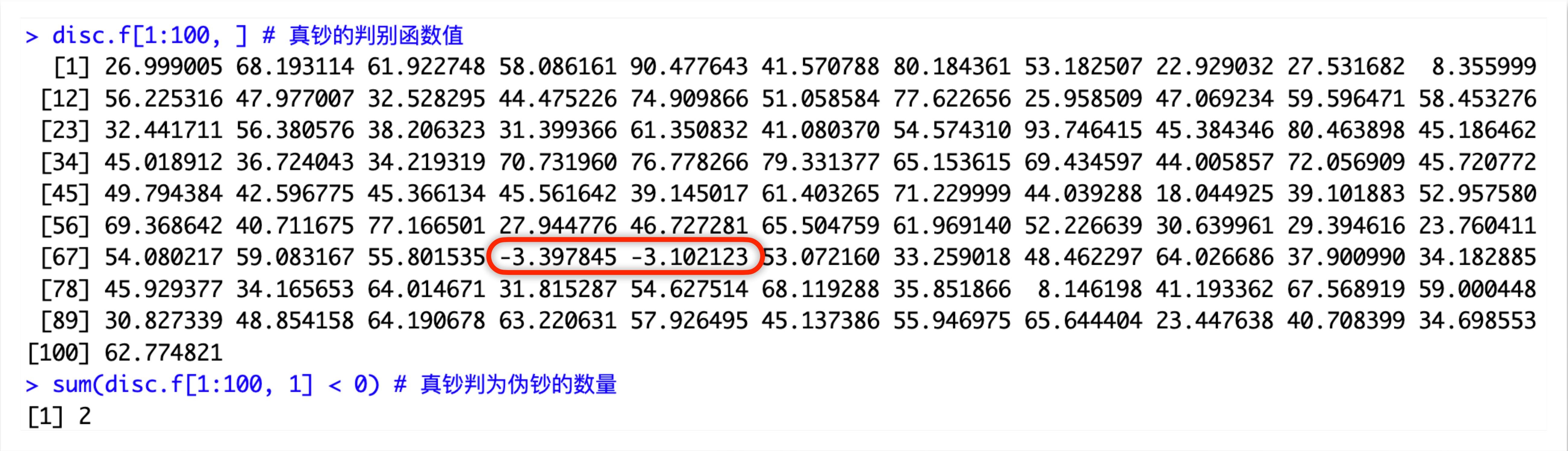
(b) 当 $J = 2$ 时,

$$\boldsymbol{x} \in R_1 \iff \boldsymbol{\alpha}^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}) \geq 0$$

其中

$$\boldsymbol{\alpha} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \quad \boldsymbol{\mu} = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$$

```
data.x = as.matrix(banknote[, 2:7]) # 真伪钞数据集全体
MU = matrix(mu, nrow = dim(data.x)[1], ncol = 6, byrow = TRUE) # 均值向量构造的矩阵
disc.f = (data.x - MU) %*% alpha # 计算每个体的判别函数值：正判为真钞、负判为伪钞
disc.f[1:100, ] # 真钞的判别函数值
sum(disc.f[1:100, 1] < 0) # 真钞判为伪钞的数量
```

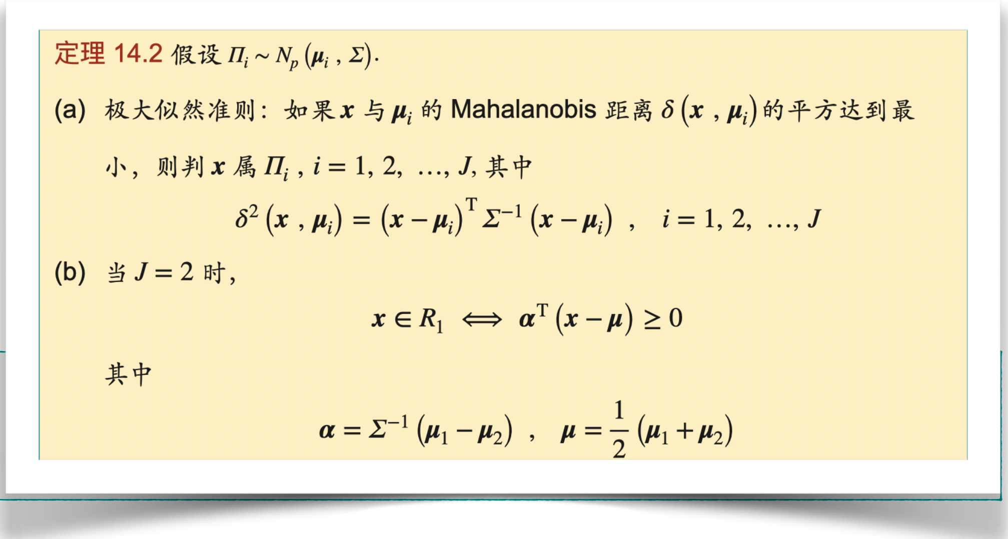


```
> disc.f[1:100, ] # 真钞的判别函数值
  [1] 26.999005 68.193114 61.922748 58.086161 90.477643 41.570788 80.184361 53.182507 22.929032 27.531682  8.355999
 [12] 56.225316 47.977007 32.528295 44.475226 74.909866 51.058584 77.622656 25.958509 47.069234 59.596471 58.453276
 [23] 32.441711 56.380576 38.206323 31.399366 61.350832 41.080370 54.574310 93.746415 45.384346 80.463898 45.186462
 [34] 45.018912 36.724043 34.219319 70.731960 76.778266 79.331377 65.153615 69.434597 44.005857 72.056909 45.720772
 [45] 49.794384 42.596775 45.366134 45.561642 39.145017 61.403265 71.229999 44.039288 18.044925 39.101883 52.957580
 [56] 69.368642 40.711675 77.166501 27.944776 46.727281 65.504759 61.969140 52.226639 30.639961 29.394616 23.760411
 [67] 54.080217 59.083167 55.801535 -3.397845 -3.102123 53.072160 33.259018 48.462297 64.026686 37.900990 34.182885
 [78] 45.929377 34.165653 64.014671 31.815287 54.627514 68.119288 35.851866  8.146198 41.193362 67.568919 59.000448
 [89] 30.827339 48.854158 64.190678 63.220631 57.926495 45.137386 55.946975 65.644404 23.447638 40.708399 34.698553
[100] 62.774821
> sum(disc.f[1:100, 1] < 0) # 真钞判为伪钞的数量
[1] 2
```

Discrimination Rules in Practice 应用当中的判别准则

- Example: 对真伪钞票数据集应用 ML 判别准则.
  - 对真伪钞票数据集所有个体的回判：

```
disc.f[101:200, ] # 伪钞的判别函数值
sum(disc.f[101:200, 1] > 0) # 伪钞判为真钞的数量
```



定理 14.2 假设 $\Pi_i \sim N_p(\boldsymbol{\mu}_i, \Sigma)$.

(a) 极大似然准则：如果 $\boldsymbol{x}$ 与 $\boldsymbol{\mu}_i$ 的 Mahalanobis 距离 $\delta(\boldsymbol{x}, \boldsymbol{\mu}_i)$ 的平方达到最小，则判 $\boldsymbol{x}$ 属 Π_i , $i = 1, 2, \ldots, J$, 其中

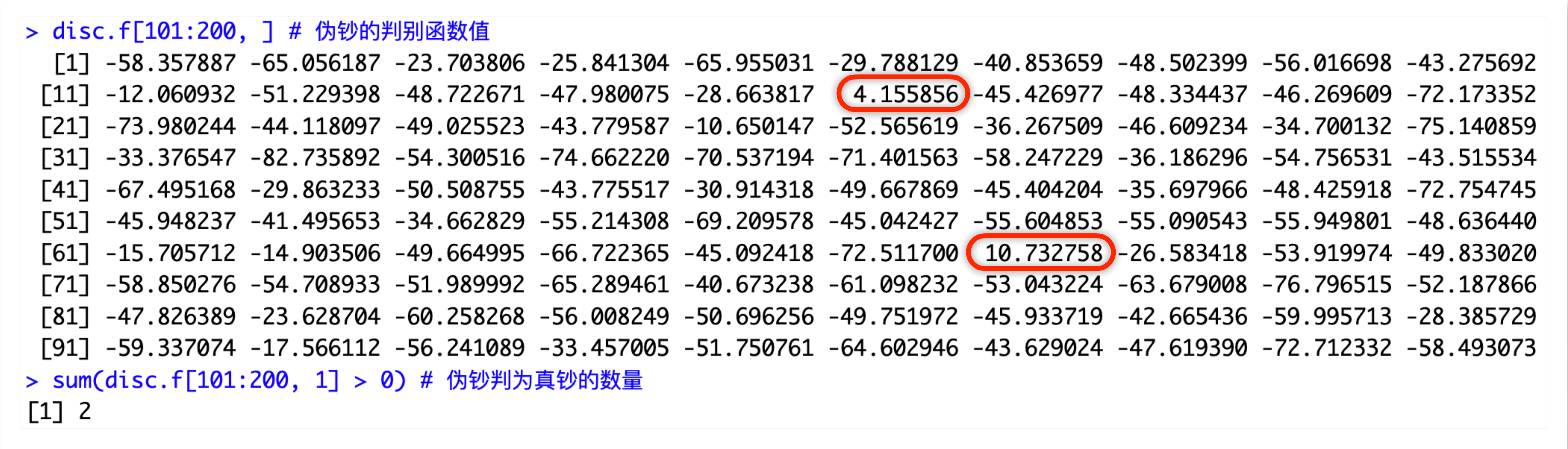
$$\delta^2(\boldsymbol{x}, \boldsymbol{\mu}_i) = (\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i), \quad i = 1, 2, \ldots, J$$

(b) 当 $J = 2$ 时，

$$\boldsymbol{x} \in R_1 \iff \boldsymbol{\alpha}^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}) \geq 0$$

其中

$$\boldsymbol{\alpha} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \quad \boldsymbol{\mu} = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$$



```
> disc.f[101:200, ] # 伪钞的判别函数值
  [1] -58.357887 -65.056187 -23.703806 -25.841304 -65.955031 -29.788129 -40.853659 -48.502399 -56.016698 -43.275692
 [11] -12.060932 -51.229398 -48.722671 -47.980075 -28.663817   4.155856 -45.426977 -48.334437 -46.269609 -72.173352
 [21] -73.980244 -44.118097 -49.025523 -43.779587 -10.650147 -52.565619 -36.267509 -46.609234 -34.700132 -75.140859
 [31] -33.376547 -82.735892 -54.300516 -74.662220 -70.537194 -71.401563 -58.247229 -36.186296 -54.756531 -43.515534
 [41] -67.495168 -29.863233 -50.508755 -43.775517 -30.914318 -49.667869 -45.404204 -35.697966 -48.425918 -72.754745
 [51] -45.948237 -41.495653 -34.662829 -55.214308 -69.209578 -45.042427 -55.604853 -55.090543 -55.949801 -48.636440
 [61] -15.705712 -14.903506 -49.664995 -66.722365 -45.092418 -72.511700  10.732758 -26.583418 -53.919974 -49.833020
 [71] -58.850276 -54.708933 -51.989992 -65.289461 -40.673238 -61.098232 -53.043224 -63.679008 -76.796515 -52.187866
 [81] -47.826389 -23.628704 -60.258268 -56.008249 -50.696256 -49.751972 -45.933719 -42.665436 -59.995713 -28.385729
 [91] -59.337074 -17.566112 -56.241089 -33.457005 -51.750761 -64.602946 -43.629024 -47.619390 -72.712332 -58.493073
> sum(disc.f[101:200, 1] > 0) # 伪钞判为真钞的数量
[1] 2
```

Discrimination Rules in Practice 应用当中的判别准则

- 如果总体的分布与参数均已知，则使用 ML 准则.
  - 假设数据来自多元正态总体 $N_p\left(\boldsymbol{\mu}_j\,,\Sigma\right)$ $, j=1, 2, \ldots, J.$ 当 $J=3$ 时

$$\begin{cases} h_{12}(\boldsymbol{x}) = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^{\mathrm{T}} \mathcal{S}_u^{-1} \left[\boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2)\right] \\ h_{13}(\boldsymbol{x}) = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_3)^{\mathrm{T}} \mathcal{S}_u^{-1} \left[\boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_3)\right] \\ h_{23}(\boldsymbol{x}) = (\bar{\boldsymbol{x}}_2 - \bar{\boldsymbol{x}}_3)^{\mathrm{T}} \mathcal{S}_u^{-1} \left[\boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_2 + \bar{\boldsymbol{x}}_3)\right] \end{cases}$$

  - ML 判别准则：

$$\begin{cases} \text{判 } \boldsymbol{x} \text{ 属 } \Pi_1\,, & \text{当 } h_{12}(\boldsymbol{x}) \geq 0 \text{ 并且 } h_{13}(\boldsymbol{x}) \geq 0 \\ \text{判 } \boldsymbol{x} \text{ 属 } \Pi_2\,, & \text{当 } h_{12}(\boldsymbol{x}) < 0 \text{ 并且 } h_{23}(\boldsymbol{x}) \geq 0 \\ \text{判 } \boldsymbol{x} \text{ 属 } \Pi_3\,, & \text{当 } h_{13}(\boldsymbol{x}) < 0 \text{ 并且 } h_{23}(\boldsymbol{x}) < 0 \end{cases}$$

Discrimination Rules in Practice 应用当中的判别准则

- 误判概率的估计

  - 随机化判别准则 (randomized discriminant rule)：

$$\phi_j(\boldsymbol{x}) = P\left(\boldsymbol{x} \in R_j\right) , \quad j = 1, 2, \ \ldots, J, \quad \text{where} \quad \sum_{j=1}^{J} \phi_j(\boldsymbol{x}) = 1.$$

$$\phi_j(\boldsymbol{x}) = \max_{1\le i\le J} \left\{\phi_i(\boldsymbol{x})\right\} \quad \Longrightarrow \quad \text{判 } \boldsymbol{x} \text{ 属 } \Pi_j$$

  - 用 p_{ij} 表示来自总体 Π_j 的个体 $\boldsymbol{x}$ 但判给总体 Π_i 的概率：

$$p_{ij} = \int \phi_i(\boldsymbol{x}) \cdot f_j(\boldsymbol{x}) d\boldsymbol{x} , \quad i, j = 1, 2, \ \ldots, J .$$

  - 对两个正态总体 $\begin{cases} \Pi_1 : \ N_p\left(\boldsymbol{\mu}_1 , \Sigma\right) \\ \Pi_2 : \ N_p\left(\boldsymbol{\mu}_2 , \Sigma\right) \end{cases}$，我们计算了 ML 准则的误判概率为：

$$p_{12} = p_{21} = \Phi\left(-\frac{1}{2}\delta^2\right)$$

$$\delta^2 = \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)^{\mathrm{T}} \Sigma^{-1} \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)$$

Discrimination Rules in Practice 应用当中的判别准则

- 误判概率的估计
  - 将其中的参数用样本估计量替换即可得到误判概率的估计为

$$\hat{p}_{12} = \hat{p}_{21} = \Phi\left(-\frac{1}{2}\hat{\delta}^2\right)$$

$$\hat{\delta}^2 = \left(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2\right)^{\mathrm{T}} \mathcal{S}_u^{-1} \left(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2\right)$$

  - 对两个正态总体 $\begin{cases} \Pi_1 : & N_p\left(\boldsymbol{\mu}_1 , \Sigma\right) \\ \Pi_2 : & N_p\left(\boldsymbol{\mu}_2 , \Sigma\right) \end{cases}$ ，我们计算了 ML 准则的误判概率为：

$$p_{12} = p_{21} = \Phi\left(-\frac{1}{2}\delta^2\right)$$

$$\delta^2 = \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)^{\mathrm{T}} \Sigma^{-1} \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)$$

Discrimination Rules in Practice 应用当中的判别准则

- 误判概率的估计
  - 误判概率还可以通过回判方法 (re-substitution method) 估计：
    - 将初始数据 $\boldsymbol{x}_i$, $i = 1, 2, \ldots, n$ 利用判别准则回判给 $\Pi_1, \Pi_2, \ldots, \Pi_J$.

      n_{ij} ： 属于 Π_j 但回判给 Π_i 的数量

      $$\Longrightarrow \quad \hat{p}_{ij} = \frac{n_{ij}}{n_j}, \quad i, j = 1, 2, \ldots, J$$

    - 称矩阵 $\left(\hat{p}_{ij}\right)$ 为混淆矩阵 (confusion matrix).

Discrimination Rules in Practice 应用当中的判别准则

- Example: 对真伪钞票数据集应用 ML 判别准则.
  - 混淆矩阵为：

		观测值	
		真钞 Π_1	伪钞 Π_2
回判结果	Π_1	98	2
	Π_2	2	98

  - 定义明显错误率 (apparent error rate, or APER) 为误判观测值所占的比例 (百分比):

$$\text{APER} = \frac{4}{200} \times 100\,\% = 2\,\%$$

Discrimination Rules in Practice 应用当中的判别准则

- Fisher 线性判别函数
  - 思想：使投影 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$ 构成很好的分类来确定判别准则.
  - 初始观测数据：

$$\mathscr{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \end{pmatrix}$$

  - 投影后的数据：

$$\mathscr{Y} = \mathscr{X}\boldsymbol{a} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} a_1x_{11} + a_2x_{12} + \cdots + a_px_{1p} \\ a_1x_{21} + a_2x_{22} + \cdots + a_px_{2p} \\ \vdots \\ a_1x_{n1} + a_2x_{n2} + \cdots + a_px_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_1 \\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Discrimination Rules in Practice 应用当中的判别准则

- Fisher 线性判别函数
  - 投影后数据 $\mathscr{Y}$ 的总平方和：

$$\mathscr{H} = \mathscr{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \mathscr{Y}^{\mathrm{T}}\mathscr{H}\mathscr{Y} = \boldsymbol{a}^{\mathrm{T}}\mathscr{X}^{\mathrm{T}}\mathscr{H}\mathscr{X}\boldsymbol{a} = \boldsymbol{a}^{\mathrm{T}}\mathscr{T}\boldsymbol{a}$$

$$\mathscr{T} = \mathscr{X}^{\mathrm{T}}\mathscr{H}\mathscr{X}$$

  - 假设有来自 J 个总体 $\Pi_1, \Pi_2, \ldots, \Pi_J$ 的观测矩阵 $\mathscr{X}_1, \mathscr{X}_2, \ldots, \mathscr{X}_J$.
  - Fisher 判别：寻找线性组合 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$，使得组间平方和 (between-group-sum of squares) 与组内平方和 (within-group-sum of squares) 之比达到最大.

Discrimination Rules in Practice 应用当中的判别准则

- Fisher 线性判别函数
  - 投影后数据 $\mathscr{Y}$ 的总平方和：

$$\mathscr{H}=\mathscr{I}-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\mathscr{Y}^{\mathrm{T}}\mathscr{H}\mathscr{Y}=\boldsymbol{a}^{\mathrm{T}}\,\mathscr{X}^{\mathrm{T}}\,\mathscr{H}\,\mathscr{X}\,\boldsymbol{a}=\boldsymbol{a}^{\mathrm{T}}\mathscr{T}\boldsymbol{a}$$

$$\mathscr{T}=\mathscr{X}^{\mathrm{T}}\mathscr{H}\mathscr{X}$$

  - 假设有来自 J 个总体 Π_1，Π_2，…，Π_J 的观测矩阵 $\mathscr{X}_1$，$\mathscr{X}_2$，…，$\mathscr{X}_J$.
  - 组内平方和：

投影 ： $\mathscr{X}_k\boldsymbol{a}$

$$\mathscr{Y}_1=\mathscr{X}_1\boldsymbol{a}\,,\ \mathscr{Y}_2=\mathscr{X}_2\boldsymbol{a}\,,\ \dots\,,\ \mathscr{Y}_J=\mathscr{X}_J\boldsymbol{a}$$

各组内的平方和： $\mathscr{Y}_1^{\mathrm{T}}\mathscr{H}_1\mathscr{Y}_1$，$\mathscr{Y}_2^{\mathrm{T}}\mathscr{H}_2\mathscr{Y}_2$，…，$\mathscr{Y}_J^{\mathrm{T}}\mathscr{H}_J\mathscr{Y}_J$

$$\sum_{k=1}^{J}\mathscr{Y}_k^{\mathrm{T}}\mathscr{H}_k\mathscr{Y}_k=\sum_{k=1}^{J}\boldsymbol{a}^{\mathrm{T}}\mathscr{X}_k^{\mathrm{T}}\mathscr{H}_k\mathscr{X}_k\boldsymbol{a}=\boldsymbol{a}^{\mathrm{T}}\left(\sum_{k=1}^{J}\mathscr{X}_k^{\mathrm{T}}\mathscr{H}_k\mathscr{X}_k\right)\boldsymbol{a}=\boldsymbol{a}^{\mathrm{T}}\mathscr{W}\boldsymbol{a}$$

Discrimination Rules in Practice 应用当中的判别准则

- Fisher 线性判别函数

各组的均值：$\bar{\boldsymbol{x}}_1$, $\bar{\boldsymbol{x}}_2$, … , $\bar{\boldsymbol{x}}_J$;　总均值：$\bar{\boldsymbol{x}} = \dfrac{n_1\bar{\boldsymbol{x}}_1 + n_2\bar{\boldsymbol{x}}_2 + \cdots + n_J\bar{\boldsymbol{x}}_J}{n_1 + n_2 + \cdots + n_J}$

  - 投影后数据 $\mathscr{Y}$ 的总平方和：

$$\mathscr{H} = \mathscr{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$$

$$\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \mathscr{Y}^{\mathrm{T}}\mathscr{H}\mathscr{Y} = \boldsymbol{a}^{\mathrm{T}}\,\boxed{\mathscr{X}^{\mathrm{T}}\,\mathscr{H}\,\mathscr{X}}\,\boldsymbol{a} = \boldsymbol{a}^{\mathrm{T}}\mathscr{T}\boldsymbol{a}$$

$$\mathscr{T} = \mathscr{X}^{\mathrm{T}}\mathscr{H}\mathscr{X}$$

  - 假设有来自 J 个总体 Π_1 , Π_2 , … , Π_J 的观测矩阵 $\mathscr{X}_1$, $\mathscr{X}_2$, … , $\mathscr{X}_J$.

  - 组间平方和：

投影 ：$\mathscr{X}_k\boldsymbol{a}$

$$\mathscr{Y}_1 = \mathscr{X}_1\boldsymbol{a}\ ,\ \mathscr{Y}_2 = \mathscr{X}_2\boldsymbol{a}\ ,\ \ldots\ ,\ \mathscr{Y}_J = \mathscr{X}_J\boldsymbol{a}$$

各组的均值：$\bar{\boldsymbol{y}}_1$, $\bar{\boldsymbol{y}}_2$, … , $\bar{\boldsymbol{y}}_J$;　总均值：$\bar{\boldsymbol{y}} = \dfrac{n_1\bar{\boldsymbol{y}}_1 + n_2\bar{\boldsymbol{y}}_2 + \cdots + n_J\bar{\boldsymbol{y}}_J}{n_1 + n_2 + \cdots + n_J}$

$$\sum_{k=1}^{J} n_k\left(\bar{\boldsymbol{y}}_k - \bar{\boldsymbol{y}}\right)^2 = \sum_{k=1}^{J} n_k\left[\boldsymbol{a}^{\mathrm{T}}\left(\bar{\boldsymbol{x}}_k - \bar{\boldsymbol{x}}\right)\right]^2 = \boldsymbol{a}^{\mathrm{T}}\left\{\sum_{k=1}^{J}\left[n_k\left(\bar{\boldsymbol{x}}_k - \bar{\boldsymbol{x}}\right)\left(\bar{\boldsymbol{x}}_k - \bar{\boldsymbol{x}}\right)^{\mathrm{T}}\right]\right\}\boldsymbol{a} = \boldsymbol{a}^{\mathrm{T}}\mathscr{B}\boldsymbol{a}$$

Discrimination Rules in Practice 应用当中的判别准则

- Fisher 线性判别函数
  - 投影后数据 $\mathscr{Y}$ 的总平方和：

$$\mathscr{H} = \mathscr{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \mathscr{Y}^{\mathrm{T}}\mathscr{H}\mathscr{Y} = \boldsymbol{a}^{\mathrm{T}}\,\boxed{\mathscr{X}^{\mathrm{T}}\,\mathscr{H}\,\mathscr{X}}\,\boldsymbol{a} = \boldsymbol{a}^{\mathrm{T}}\mathscr{T}\boldsymbol{a}$$

$$\mathscr{T} = \mathscr{X}^{\mathrm{T}}\mathscr{H}\mathscr{X}$$

$$\mathscr{X} = \begin{pmatrix} \mathscr{X}_1 \\ \mathscr{X}_2 \\ \vdots \\ \mathscr{X}_J \end{pmatrix}$$

  - 假设有来自 J 个总体 $\Pi_1, \Pi_2, \ldots, \Pi_J$ 的观测矩阵 $\mathscr{X}_1, \mathscr{X}_2, \ldots, \mathscr{X}_J$.
  - 总平方和：

投影：$\mathscr{X}_k\boldsymbol{a}$

$$\mathscr{Y}_1 = \mathscr{X}_1\boldsymbol{a},\ \mathscr{Y}_2 = \mathscr{X}_2\boldsymbol{a},\ \ldots,\ \mathscr{Y}_J = \mathscr{X}_J\boldsymbol{a}$$

$$\mathscr{Y} = \begin{pmatrix} \mathscr{Y}_1 \\ \mathscr{Y}_2 \\ \vdots \\ \mathscr{Y}_J \end{pmatrix}$$

$$\mathscr{Y}^{\mathrm{T}}\mathscr{H}\mathscr{Y} = \boldsymbol{a}^{\mathrm{T}}\mathscr{T}\boldsymbol{a} = \boldsymbol{a}^{\mathrm{T}}\mathscr{W}\boldsymbol{a} + \boldsymbol{a}^{\mathrm{T}}\mathscr{B}\boldsymbol{a}$$

总平方和 = 组内平方和 + 组间平方和

Discrimination Rules in Practice 应用当中的判别准则

- Fisher 线性判别函数

  - 投影后数据 $\mathscr{Y}$ 的总平方和：

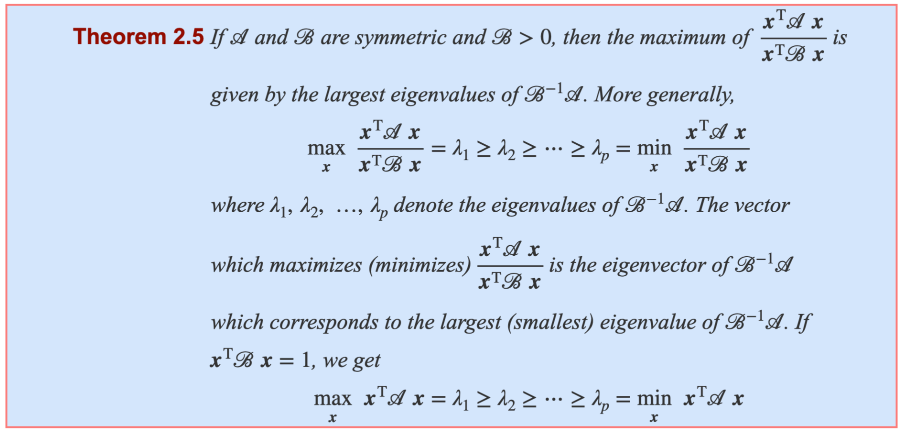
$$\mathscr{H} = \mathscr{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \mathscr{Y}^{\mathrm{T}}\mathscr{H}\mathscr{Y} = \boldsymbol{a}^{\mathrm{T}}\,\mathscr{X}^{\mathrm{T}}\,\mathscr{H}\,\mathscr{X}\,\boldsymbol{a} = \boldsymbol{a}^{\mathrm{T}}\mathscr{T}\boldsymbol{a}$$

$$\mathscr{T} = \mathscr{X}^{\mathrm{T}}\mathscr{H}\mathscr{X}$$

  - 假设有来自 J 个总体 Π_1 , Π_2 , … , Π_J 的观测矩阵 $\mathscr{X}_1$, $\mathscr{X}_2$, … , $\mathscr{X}_J$.
  - Fisher 判别：寻找线性组合 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$，使得组间平方和 (between-group-sum of squares) 与组内平方和 (within-group-sum of squares) 之比达到最大.

求 $\boldsymbol{a}$ 使得 $\dfrac{\boldsymbol{a}^{\mathrm{T}}\mathscr{B}\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\mathscr{W}\boldsymbol{a}}$ 达到最大.



Theorem 2.5 *If $\mathscr{A}$ and $\mathscr{B}$ are symmetric and $\mathscr{B} > 0$, then the maximum of $\dfrac{\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\mathscr{B}\,\boldsymbol{x}}$ is given by the largest eigenvalues of $\mathscr{B}^{-1}\mathscr{A}$. More generally,*

$$\max_{\boldsymbol{x}} \frac{\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\mathscr{B}\,\boldsymbol{x}} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p = \min_{\boldsymbol{x}} \frac{\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\mathscr{B}\,\boldsymbol{x}}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ denote the eigenvalues of $\mathscr{B}^{-1}\mathscr{A}$. The vector which maximizes (minimizes) $\dfrac{\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\mathscr{B}\,\boldsymbol{x}}$ is the eigenvector of $\mathscr{B}^{-1}\mathscr{A}$ which corresponds to the largest (smallest) eigenvalue of $\mathscr{B}^{-1}\mathscr{A}$. If $\boldsymbol{x}^{\mathrm{T}}\mathscr{B}\,\boldsymbol{x} = 1$, we get

$$\max_{\boldsymbol{x}}\ \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p = \min_{\boldsymbol{x}}\ \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}$$

Discrimination Rules in Practice 应用当中的判别准则

- Fisher 线性判别函数

> **定理 14.4** 使得 $\dfrac{\boldsymbol{a}^{\mathrm{T}}\mathscr{B}\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\mathscr{W}\boldsymbol{a}}$ 达到最大的向量 $\boldsymbol{a}$ 是矩阵 $\mathscr{W}^{-1}\mathscr{B}$ 的最大特征值对应的特征向量.

- Fisher 判别准则：将 $\boldsymbol{x}$ 判给投影后 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$ 最接近的那一组.

$$\text{如果 } k = \arg\min_{1\le i\le J} \left|\boldsymbol{a}^{\mathrm{T}}\left(\boldsymbol{x}-\overline{\boldsymbol{x}}_i\right)\right|,\ \text{则将 } \boldsymbol{x} \text{ 判给 } \Pi_k\,.$$

Discrimination Rules in Practice 应用当中的判别准则

- Fisher 线性判别函数
  - 当 $J = 2$ 时, $\begin{cases} \Pi_1 : \mathscr{X}_1 \ (n_1 \times p) \\ \Pi_2 : \mathscr{X}_2 \ (n_2 \times p) \end{cases}$

$$\bar{\boldsymbol{x}} = \frac{n_1 \bar{\boldsymbol{x}}_1 + n_2 \bar{\boldsymbol{x}}_2}{n_1 + n_2}$$

$$\mathscr{B} = \sum_{k=1}^{J} \left[n_k (\bar{\boldsymbol{x}}_k - \bar{\boldsymbol{x}}) (\bar{\boldsymbol{x}}_k - \bar{\boldsymbol{x}})^{\mathrm{T}} \right] = n_1 (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}) (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}})^{\mathrm{T}} + n_2 (\bar{\boldsymbol{x}}_2 - \bar{\boldsymbol{x}}) (\bar{\boldsymbol{x}}_2 - \bar{\boldsymbol{x}})^{\mathrm{T}}$$

$$= n_1 \left(\bar{\boldsymbol{x}}_1 - \frac{n_1 \bar{\boldsymbol{x}}_1 + n_2 \bar{\boldsymbol{x}}_2}{n_1 + n_2} \right) \left(\bar{\boldsymbol{x}}_1 - \frac{n_1 \bar{\boldsymbol{x}}_1 + n_2 \bar{\boldsymbol{x}}_2}{n_1 + n_2} \right)^{\mathrm{T}} + n_2 \left(\bar{\boldsymbol{x}}_2 - \frac{n_1 \bar{\boldsymbol{x}}_1 + n_2 \bar{\boldsymbol{x}}_2}{n_1 + n_2} \right) \left(\bar{\boldsymbol{x}}_2 - \frac{n_1 \bar{\boldsymbol{x}}_1 + n_2 \bar{\boldsymbol{x}}_2}{n_1 + n_2} \right)^{\mathrm{T}}$$

$$= \frac{n_1 n_2^2}{(n_1 + n_2)^2} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^{\mathrm{T}} + \frac{n_2 n_1^2}{(n_1 + n_2)^2} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^{\mathrm{T}}$$

$$= \frac{n_1 n_2}{n_1 + n_2} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^{\mathrm{T}} = \frac{n_1 n_2}{n} \boldsymbol{d} \boldsymbol{d}^{\mathrm{T}}$$

$$\boldsymbol{d} = \bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2$$

$$n = n_1 + n_2$$

Discrimination Rules in Practice 应用当中的判别准则

- Fisher 线性判别函数
  - 当 $J = 2$ 时, $\begin{cases} \Pi_1 : \mathscr{X}_1 \ (n_1 \times p) \\ \Pi_2 : \mathscr{X}_2 \ (n_2 \times p) \end{cases}$
  - 可以证明：矩阵 $\mathscr{W}^{-1}\mathscr{B}$ 只有一个特征值 $\mathrm{tr}\left(\mathscr{W}^{-1}\mathscr{B}\right) = \frac{n_1 n_2}{n}\boldsymbol{d}^{\mathrm{T}}\mathscr{W}^{-1}\boldsymbol{d}$.
  - 对应的特征向量为 $\boldsymbol{a} = \mathscr{W}^{-1}\boldsymbol{d}$.
  - 相应的判别准则为：

$$\text{如果 } \boldsymbol{a}^{\mathrm{T}}\left[\boldsymbol{x} - \frac{1}{2}\left(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2\right)\right] > 0, \ \text{则判 } \boldsymbol{x} \longrightarrow \Pi_1$$

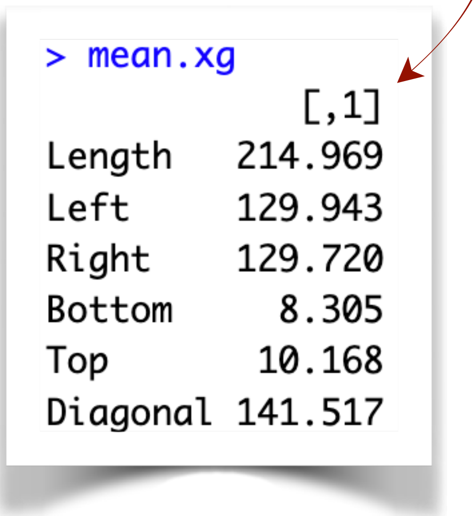
$$\text{如果 } \boldsymbol{a}^{\mathrm{T}}\left[\boldsymbol{x} - \frac{1}{2}\left(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2\right)\right] \leq 0, \ \text{则判 } \boldsymbol{x} \longrightarrow \Pi_2$$

Discrimination Rules in Practice 应用当中的判别准则

- **Example**: 讨论真伪钞票数据集的 Fisher 判别.

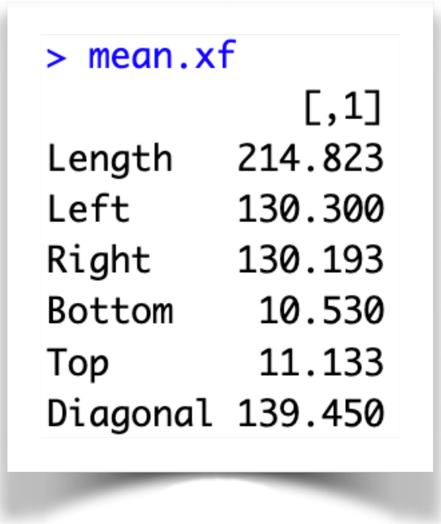
```
rm(list = ls(all = TRUE))
graphics.off()
library(mclust)
data(banknote)
xg = as.matrix(banknote[1:100, 2:7]) # 真钞数据
xf = as.matrix(banknote[101:200, 2:7]) # 伪钞数据
mean.xg = as.matrix(apply(xg, 2, mean)) # 真钞数据的均值
mean.xf = as.matrix(apply(xf, 2, mean)) # 伪钞数据的均值
```

$\bar{\boldsymbol{x}}_g =$



```
> mean.xg
            [,1]
Length   214.969
Left     129.943
Right    129.720
Bottom     8.305
Top       10.168
Diagonal 141.517
```

$\bar{\boldsymbol{x}}_f =$



```
> mean.xf
            [,1]
Length   214.823
Left     130.300
Right    130.193
Bottom    10.530
Top       11.133
Diagonal 139.450
```

- Fisher 判别：寻找线性组合 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$，使得组间平方和与组内平方和之比

$$\frac{\boldsymbol{a}^{\mathrm{T}}\mathscr{B}\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\mathscr{W}\boldsymbol{a}}$$

达到最大.

Discrimination Rules in Practice 应用当中的判别准则

- Example: 讨论真伪钞票数据集的 Fisher 判别.
  - Fisher 判别准则: $\begin{cases} \text{当 } \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x}-\overline{\boldsymbol{x}}) > 0 \text{ 时，判 } \boldsymbol{x} \text{ 是真钞} \\ \text{当 } \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x}-\overline{\boldsymbol{x}}) \leq 0 \text{ 时，判 } \boldsymbol{x} \text{ 是伪钞} \end{cases}$

$$\Longrightarrow \quad \boldsymbol{a} = \mathscr{W}^{-1}\left(\overline{\boldsymbol{x}}_g - \overline{\boldsymbol{x}}_f\right) = \begin{pmatrix} 0.000 \\ 0.029 \\ -0.029 \\ -0.039 \\ -0.041 \\ 0.054 \end{pmatrix}$$

$$\begin{cases} \left(y_g\right)_i = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_i \\ \left(y_f\right)_i = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_{100+i} \end{cases}, \qquad i = 1, 2, \ldots, 100. \quad \Longrightarrow \quad \sum_{i=1}^{100}\left[\left(y_g\right)_i - \overline{y}_g\right]^2 + \sum_{i=1}^{100}\left[\left(y_f\right)_i - \overline{y}_f\right]^2 = \boldsymbol{a}^{\mathrm{T}}\mathscr{W}\boldsymbol{a}$$

$$100\left[\left(\overline{y}_g - \overline{y}\right)^2 + \left(\overline{y}_f - \overline{y}\right)^2\right] = \boldsymbol{a}^{\mathrm{T}}\mathscr{B}\boldsymbol{a} \quad \Longleftarrow \quad \overline{y} = \frac{1}{2}\left(\overline{y}_g + \overline{y}_f\right)$$

  - Fisher 判别：寻找线性组合 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$，使得组间平方和与组内平方和之比

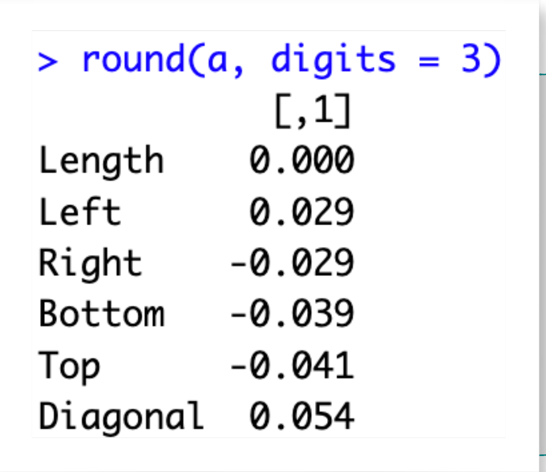
$$\frac{\boldsymbol{a}^{\mathrm{T}}\mathscr{B}\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\mathscr{W}\boldsymbol{a}}$$ 达到最大.

Discrimination Rules in Practice 应用当中的判别准则

- Example: 讨论真伪钞票数据集的 Fisher 判别.

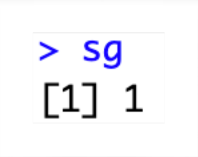
```
m = (mean.xg + mean.xf)/2 # 两个总体均值的平均
w = 100 * (cov(xg) + cov(xf)) # 组内平方和对应的矩阵
d = mean.xg - mean.xf # 均值差
a = solve(w) %*% d # 确定线性组合（投影）的系数
round(a, digits = 3)
```

$a =$

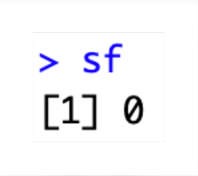


```
> round(a, digits = 3)
            [,1]
Length     0.000
Left       0.029
Right     -0.029
Bottom    -0.039
Top       -0.041
Diagonal   0.054
```

```
yg = as.matrix(xg - matrix(m, nrow = 100, ncol = 6, byrow = T)) %*% a # 计算真钞数据的投影值
yf = as.matrix(xf - matrix(m, nrow = 100, ncol = 6, byrow = T)) %*% a # 计算伪钞数据的投影值

xgtest = yg
sg = sum(xgtest < 0) # 真钞判归伪钞的数量
sg

xftest = yf
sf = sum(xftest > 0) # 伪钞判归真钞的数量
sf
```



```
> sg
[1] 1
```

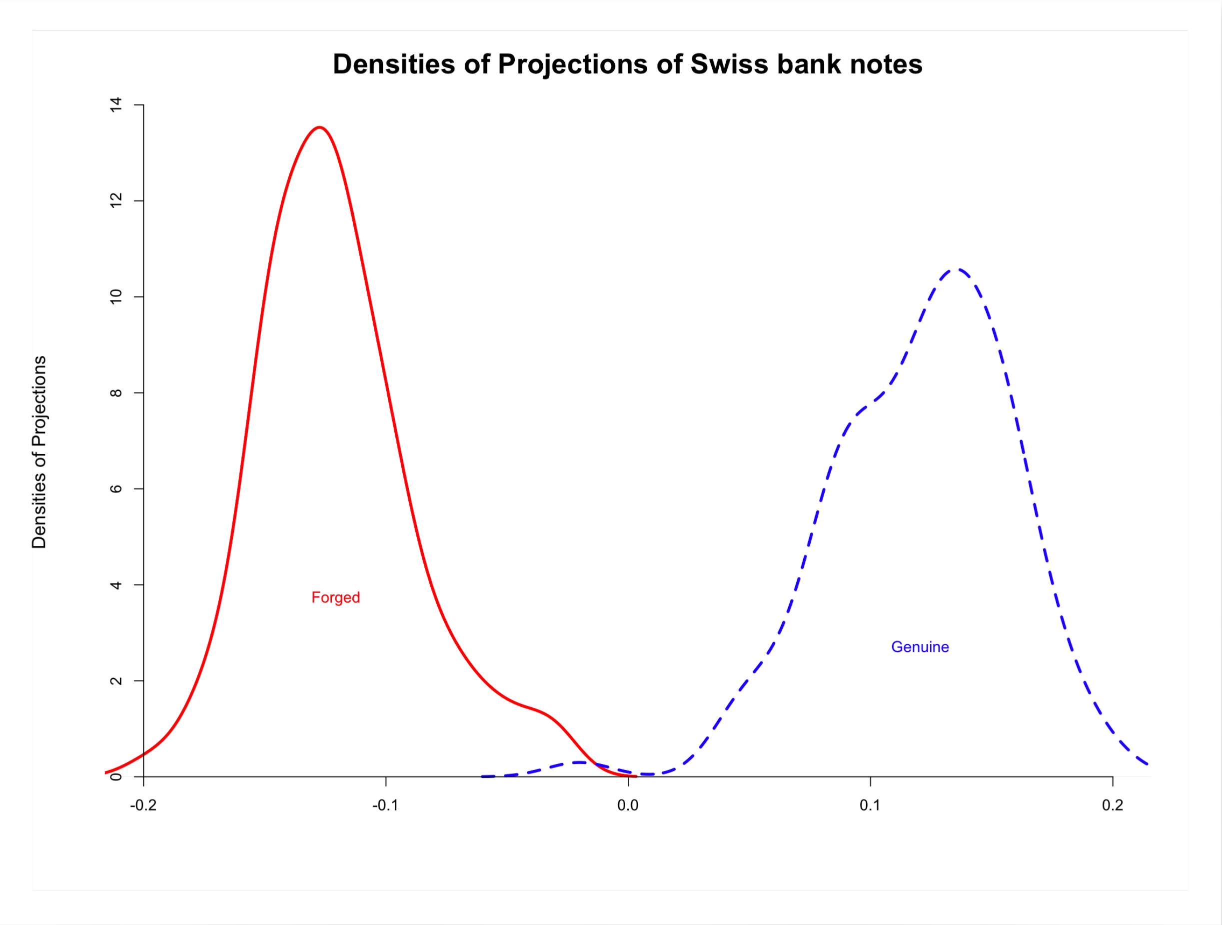


```
> sf
[1] 0
```

Discrimination Rules in Practice 应用当中的判别准则

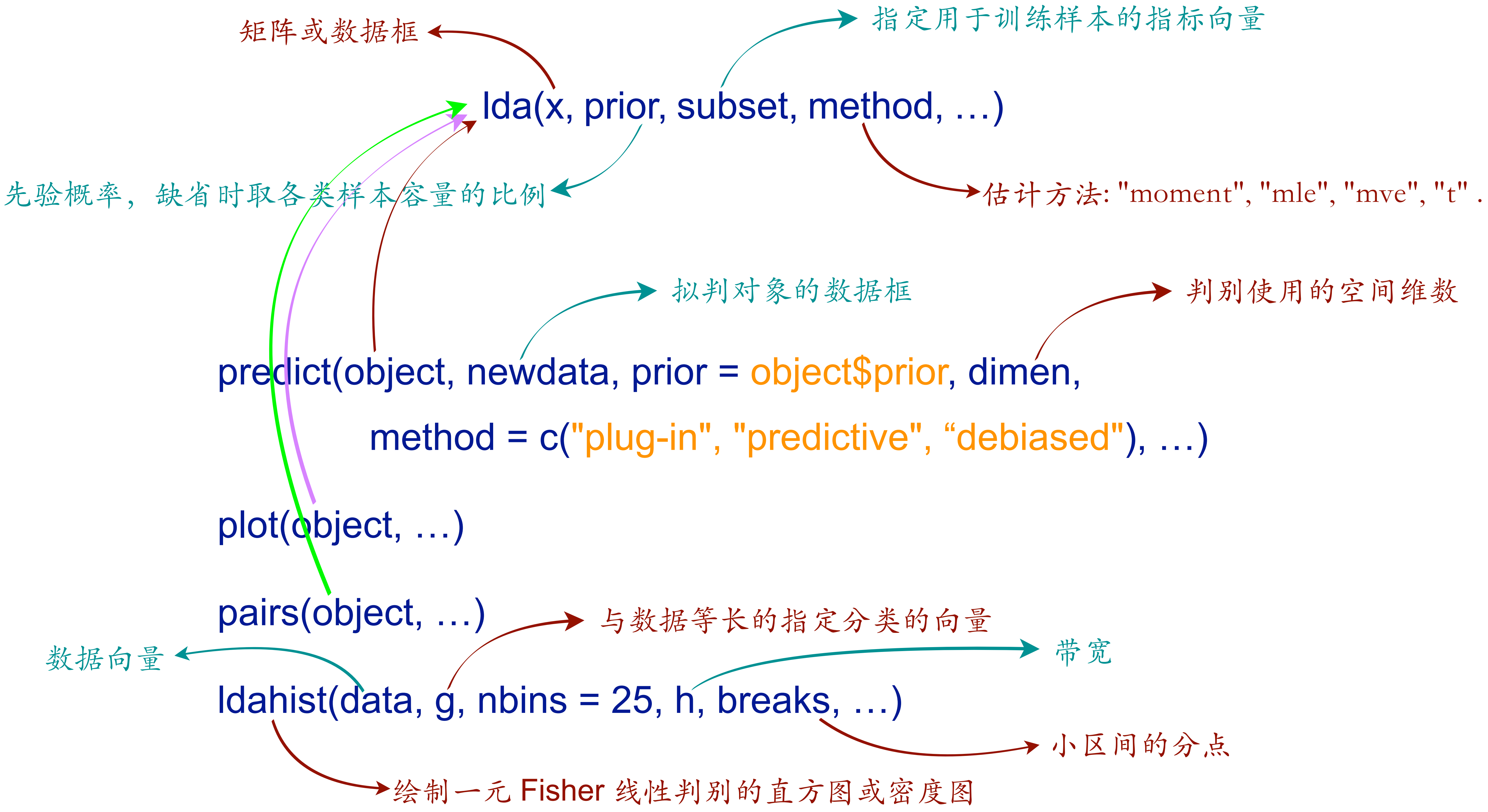
- **Example**: 讨论真伪钞票数据集的 Fisher 判别.

```
# 绘制两个总体投影值核密度估计曲线
plot(ff, lwd = 3, col = "red", xlab = "", ylab = "Densities of Projections", main = "Densities of Projections of Swiss bank notes",
    axes = FALSE, xlim = c(-0.2, 0.2), cex.lab = 1.2, cex.axis = 1.2, cex.main = 1.8)
lines(fg, lwd = 3, col = "blue", lty = 2)
text(mean(yf), 3.72, "Forged", col = "red")
text(mean(yg), 2.72, "Genuine", col = "blue")
axis(1, pos = 0)
axis(2, pos = -0.2)
```



ℜ 中的判别分析

- 线性判别分析 (Linear Discriminant Analysis)



矩阵或数据框

指定用于训练样本的指标向量

lda(x, prior, subset, method, …)

先验概率，缺省时取各类样本容量的比例

估计方法: "moment", "mle", "mve", "t" .

拟判对象的数据框

判别使用的空间维数

predict(object, newdata, prior = object$prior, dimen,

method = c("plug-in", "predictive", “debiased"), …)

plot(object, …)

pairs(object, …)

与数据等长的指定分类的向量

数据向量

带宽

ldahist(data, g, nbins = 25, h, breaks, …)

小区间的分点

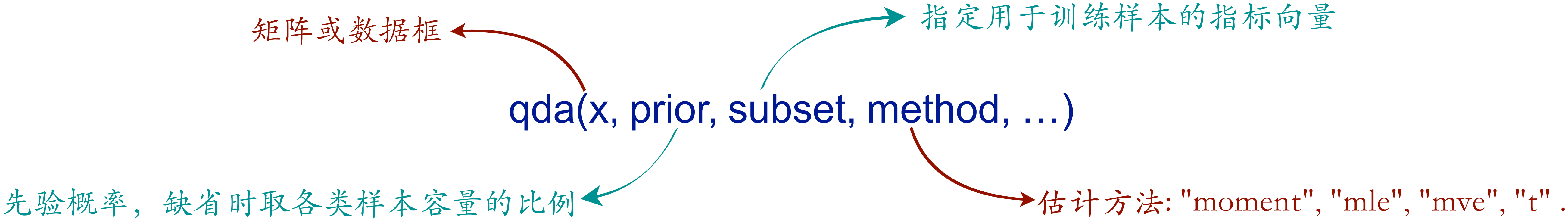
绘制一元 Fisher 线性判别的直方图或密度图

iris 数据集的判别分析

```
library(MASS)
x = iris[, 1:4] # 读入数据
g = gl(n = 3, k = 50) # 生成有三类、每类 50 个对象的分类因子
x = cbind(x, Type = g) # 将分类因子 g 添加到数据中
attach(x)
x.lda = lda(Type ~ Petal.Length + Petal.Width + Sepal.Length + Sepal.Width) # 方差相等下的 Bayes 判别
x.lda # 显示结果
detach(x)
plot(x.lda, dimen = 1) # 绘制一个线性判别函数的图形
colors = rep(1:3, times = c(50, 50, 50))
plot(x.lda, cex = 1.5, dimen = 2, col = colors) # 绘制两个线性判别函数的图形
x.pred = predict(x.lda) # 回判
x.pred$class # 显示回判的结果
x.pred$posterior # 后验概率
x.pred$x # 线性判别函数 (投影) 的取值
x.alloc = cbind(Type = g, x.pred$x, Allocations = x.pred$class) # 真实类型、判别类型对照 (含判别函数)
x.alloc
x.confu = table(g, x.pred$class) # 混淆矩阵
x.confu
x.prob = prop.table(x.confu) # 混淆矩阵对应的错判概率
x.prob
sum(diag(x.prob)) # 判对率
1 - sum(diag(x.prob)) # 错判率
x.alloc = as.data.frame(x.alloc)
ldahist(x.alloc$LD1, g, type = "b") # 第一线性判别函数(投影)值的直方图与核密度
ldahist(x.alloc$LD2, g, type = "b") # 第二线性判别函数(投影)值的直方图与核密度
pairs(x.lda, cex = 1.5, col = colors) # 线性判别的二维散点图
```

ℜ 中的判别分析

- 二次判别分析 (Quadratic Discriminant Analysis)



矩阵或数据框

指定用于训练样本的指标向量

qda(x, prior, subset, method, …)

先验概率，缺省时取各类样本容量的比例

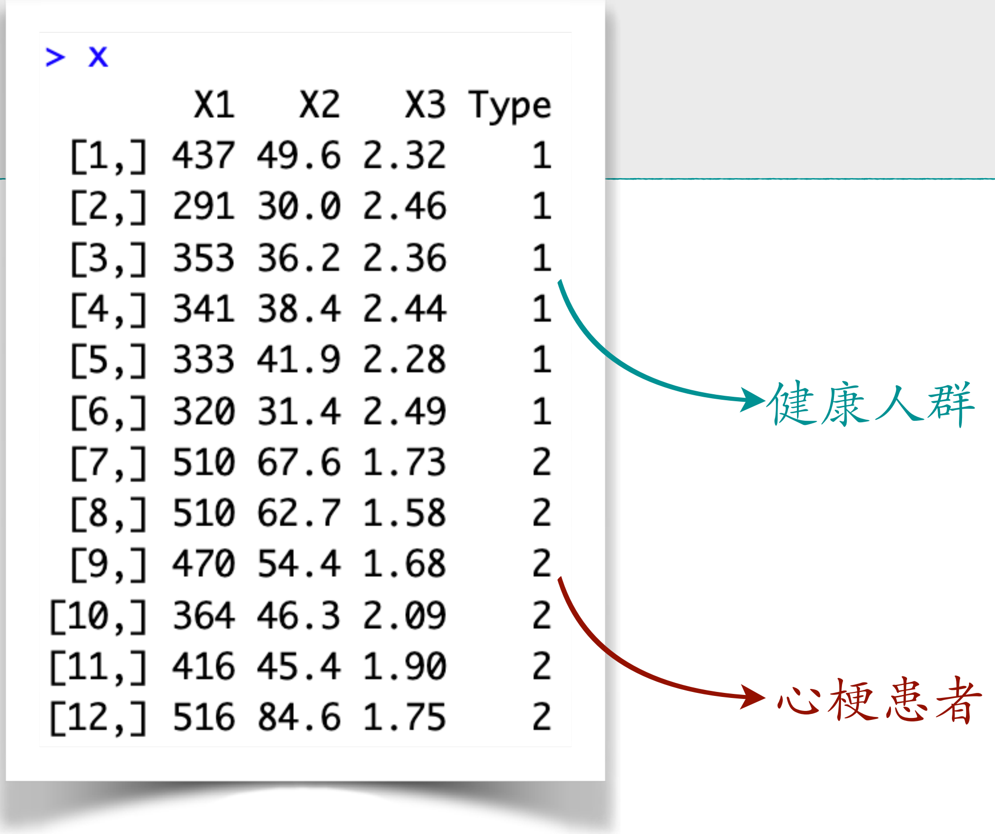
估计方法: "moment", "mle", "mve", "t" .

iris 数据集的判别分析

```
rm(list = ls(all = TRUE))
graphics.off()
options(digits = 3)

library(MASS)
x = iris[, 1:4] # 读入数据

g = gl(n = 3, k = 50) # 生成有三类、每类 50 个对象的分类因子
x = cbind(x, Type = g) # 将分类因子 g 添加到数据中

attach(x)
x.qda = qda(Type ~ Petal.Length + Petal.Width + Sepal.Length + Sepal.Width) # 方差不等条件下的二次 Bayes 判别
x.qda # 显示结果
detach(x)

x.qpred = predict(x.qda) # 回判
x.qpred$class # 回判结果
round(x.qpred$posterior, digits = 2) # 回判的后验概率

x.qalloc = cbind(Type = g, Allocations = x.qpred$class) # 真实类型、判别类型对照
x.qconfu = table(g, x.qpred$class) # 混淆矩阵
x.qconfu
x.qprob = prop.table(x.qconfu) # 混淆矩阵对应的错判概率
x.qprob
sum(diag(x.qprob)) # 判对率
1 - sum(diag(x.qprob)) # 错判率
```

健康人群与心梗患者的判别分析

```
rm(list = ls(all = TRUE))
graphics.off()
library(MASS)
library(readr)
setwd("~/Desktop/2023_Applied Multivariate Statistical Analysis/R Codes with data/Data")
x = read_csv("myocardial.csv") # 读入数据
x = as.data.frame(x[, 2:5])
x$Type = as.factor(x$Type)
x
```



```
> x
       X1   X2   X3 Type
 [1,] 437 49.6 2.32    1
 [2,] 291 30.0 2.46    1
 [3,] 353 36.2 2.36    1
 [4,] 341 38.4 2.44    1
 [5,] 333 41.9 2.28    1
 [6,] 320 31.4 2.49    1
 [7,] 510 67.6 1.73    2
 [8,] 510 62.7 1.58    2
 [9,] 470 54.4 1.68    2
[10,] 364 46.3 2.09    2
[11,] 416 45.4 1.90    2
[12,] 516 84.6 1.75    2
```

- 判断三项指标 $x = \begin{pmatrix} 420.50 \\ 32.42 \\ 1.98 \end{pmatrix}$ 的一人属于健康人群还是心梗患者.

健康人群与心梗患者的判别分析

```
attach(x)
x.lda = lda(Type ~ X1 + X2 + X3) # 方差相等条件下的 Bayes 判别
x.lda # 显示判别结果
plot(x.lda, dimen = 1, type = "b") # 绘制线性判别函数的图形

x.pred = predict(x.lda) # 回判
x.pred$class # 显示回判的结果
round(x.pred$posterior, digits = 2) # 后验概率
x.pred$x # 线性判别函数 (投影) 的取值

x.alloc = cbind(x, x.pred$x, Allocations = x.pred$class) # 真实类型、判别类型对照 (含判别函数)
x.alloc

x.confu = table(x$Type, x.pred$class) # 混淆矩阵
x.confu

x.prob = prop.table(x.confu) # 混淆矩阵对应的错判概率
x.prob

sum(diag(x.prob)) # 判对率
1 - sum(diag(x.prob)) # 错判率

x.new = data.frame(X1 = c(420.50), X2 = c(32.42), X3 = c(1.98)) # 新的观测数据
x.new
x.new.pred = predict(x.lda, newdata = x.new) # 利用上述判别分析模型的预测
x.new.pred # 查看结果
x.new.pred$posterior # 查看后验概率：属于第2类的概率 0.78，判此人为心梗患者
detach(x)
```