应用多元统计分析

第四章 判别分析

第一节　引言

- 在我们的日常生活和工作实践中，常常会遇到判别分析问题，即根据历史上划分类别的有关资料和某种最优准则，确定一种判别方法，判定一个新的样本归属哪一类。例如，某医院有部分患有肺炎、肝炎、冠心病、糖尿病等病人的资料，记录了每个患者若干项症状指标数据。现在想利用现有的这些资料找出一种方法，使得对于一个新的病人，当测得这些症状指标数据时，能够判定其患有哪种病。又如，在天气预报中，我们有一段较长时间关于某地区每天气象的记录资料（晴阴雨、气温、气压、湿度等），现在想建立一种用连续五天的气象资料来预报第六天是什么天气的方法。这些问题都可以应用判别分析方法予以解决。

- 把这类问题用数学语言来表达，可以叙述如下：设有n个样本，对每个样本测得p项指标（变量）的数据，已知每个样本属于k个类别（或总体）G_1，G_2，…，G_k中的某一类，且它们的分布函数分别为$F_1(x)$，$F_2(x)$，…，$F_k(x)$。我们希望利用这些数据，找出一种判别函数，使得这一函数具有某种最优性质，能把属于不同类别的样本点尽可能地区别开来，并对测得同样p项指标（变量）数据的一个新样本，能判定这个样本归属于哪一类。

- 判别分析内容很丰富，方法很多。判断分析按判别的总体数来区分，有两个总体判别分析和多总体判别分析；按区分不同总体所用的数学模型来分，有线性判别和非线性判别；按判别时所处理的变量方法不同，有逐步判别和序贯判别等。判别分析可以从不同角度提出问题，因此有不同的判别准则，如马氏距离最小准则、Fisher准则、平均损失最小准则、最小平方准则、最大似然准则、最大概率准则等等，按判别准则的不同又提出多种判别方法。本章仅介绍常用的几种判别分析方法：距离判别法、Fisher判别法、Bayes判别法和逐步判别法。

第二节 距离判别法

一 马氏距离的概念

二 距离判别的思想及方法

三 判别分析的实质

一、马氏距离的概念

- 设 p 维欧氏空间 R^p 中的两点 $\mathbf{X}=(X_1,X_2,\cdots,X_p)'$ 和 $\mathbf{Y}=(Y_1,Y_2,\cdots,Y_p)'$，通常我们所说的两点之间的距离，是指欧氏距离，即

$$d^2(\mathbf{X},\mathbf{Y})=(X_1-Y_1)^2+\cdots+(X_p-Y_p)^2 \quad (4.1)$$

在解决实际问题时，特别是针对多元数据的分析问题，欧氏距离就显示出了它的薄弱环节。

第一、设有两个正态总体，$X\sim N(\mu_1,\sigma^2)$ 和 $Y\sim N(\mu_2,4\sigma^2)$，现有一个样品位于如图 4.1 所示的 A 点，距总体 X 的中心 2σ 远，距总体 Y 的中心 3σ 远，那么，A 点处的样品到底离哪一个总体近呢？若按欧氏距离来量度，A 点离总体 X 要比离总体 Y “近一些”。但是，从概率的角度看，A 点位于 μ_1 右侧的 $2\sigma_x$ 处，而位于 μ_2 左侧 $1.5\sigma_y$ 处，应该认为 A 点离总体 Y “近一些”。显然，后一种量度更合理些。

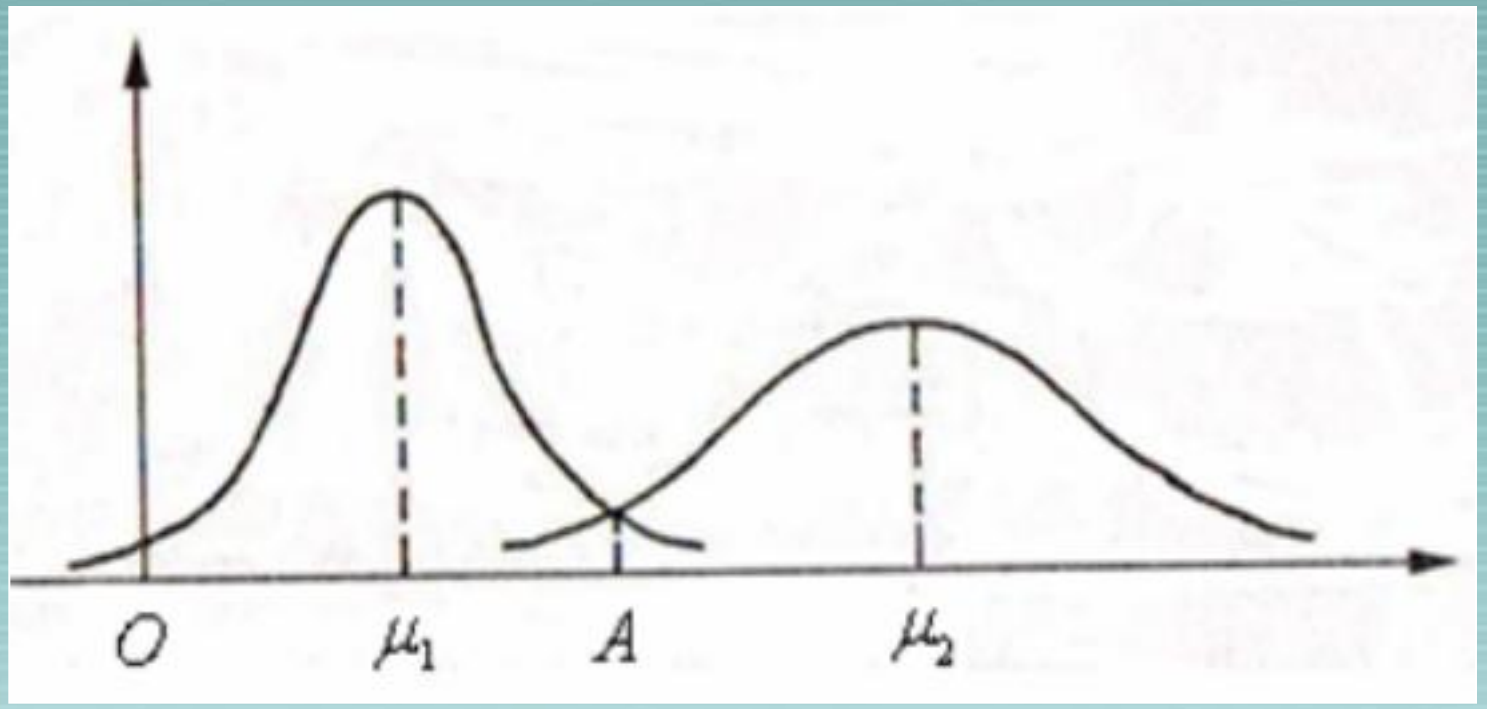


图4.1

第二、设有量度重量和长度的两个变量 X 与 Y，以单位分别为 kg 和 cm 得到样本 $A(0,5)$，$B(10,0)$，$C(1,0)$，$D(0,10)$。今按照欧氏距离计算，有

$$AB=\sqrt{10^2+5^2}=\sqrt{125}\,;$$

$$CD=\sqrt{1^2+10^2}=\sqrt{101}$$

如果我们将长度单位变为 mm，那么，有

$$AB=\sqrt{10^2+50^2}=\sqrt{2600}\,;$$

$$CD=\sqrt{1^2+100^2}=\sqrt{10001}$$

量纲的变化，将影响欧氏距离计算的结果。

- 为此，我们引入一种由印度著名统计学家马哈拉诺比斯（Mahalanobis, 1936）提出的“马氏距离”的概念。
- 设 $\mathbf{X}$ 和 $\mathbf{Y}$ 是来自均值向量为 $\boldsymbol{\mu}$，协方差为 $\boldsymbol{\Sigma}(>0)$ 的总体 G 中的 p 维样本，则总体 G 内两点 $\mathbf{X}$ 与 $\mathbf{Y}$ 之间的马氏距离定义为

$$D^2(\mathbf{X},\mathbf{Y})=(\mathbf{X}-\mathbf{Y})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{Y}) \qquad (4.2)$$

定义点 $\mathbf{X}$ 到总体 G 的马氏距离为

$$D^2(\mathbf{X},G)=(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \qquad (4.3)$$

这里应该注意到，当 $\boldsymbol{\Sigma}=\mathbf{I}$（单位矩阵）时，即为欧氏距离的情形。

二、距离判别的思想及方法

1、两个总体的距离判别问题

- 问题：设有协方差矩阵∑相等的两个总体G_1和G_2，其均值分别是μ_1和μ_2，对于一个新的样品X，要判断它来自哪个总体。
- 一般的想法是计算新样品X到两个总体的马氏距离D^2（X，G_1）和D^2（X，G_2），并按照如下的判别规则进行判断

$$\begin{cases} \mathbf{X} \in G_1, & \text{如果} \quad D^2(\mathbf{X}, G_1) \le D^2(\mathbf{X}, G_2) \\ \mathbf{X} \in G_2, & \text{如果} \quad D^2(\mathbf{X}, G_1) > D^2(\mathbf{X}, G_2) \end{cases} \qquad (4.4)$$

- 这个判别规则的等价描述为：求新样品X到G_1的距离与到G_2的距离之差，如果其值为正，X属于G_2；否则X属于G_1。

- 我们考虑

$$
\begin{aligned}
&D^2(\mathbf{X},G_1)-D^2(\mathbf{X},G_2)\\
&=(\mathbf{X}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}_1)-(\mathbf{X}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}_2)\\
&=\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}-2\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1+\boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1-(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}-2\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2+\boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2)\\
&=2\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)+\boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1-\boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2\\
&=2\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)+(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\
&=-2\left(\mathbf{X}-\frac{\boldsymbol{\mu}_1+\boldsymbol{\mu}_2}{2}\right)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\
&=-2(\mathbf{X}-\bar{\boldsymbol{\mu}})'\boldsymbol{\alpha}=-2\boldsymbol{\alpha}'(\mathbf{X}-\bar{\boldsymbol{\mu}})
\end{aligned}
$$

- 其中 $\bar{\boldsymbol{\mu}}=\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)$ 是两个总体均值的平均值，$\boldsymbol{\alpha}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$，记

$$W(\mathbf{X})=\boldsymbol{\alpha}'(\mathbf{X}-\bar{\boldsymbol{\mu}}) \qquad (4.5)$$

则判别规则（4.4）式可表示为

$$\begin{cases}\mathbf{X}\in G_1, & \text{如果} \quad W(\mathbf{X})\geq 0\\ \mathbf{X}\in G_2, & \text{如果} \quad W(\mathbf{X})<0\end{cases} \qquad (4.6)$$

这里称 $W(\mathbf{X})$ 为两总体距离判别的判别函数，由于它是 $\mathbf{X}$ 的线性函数，故又称为线性判别函数，$\boldsymbol{\alpha}$ 称为判别系数。

- 在实际应用中，总体的均值和协方差矩阵一般是未知的，可由样本均值和样本协方差矩阵分别进行估计。设 $\mathbf{X}_1^{(1)},\cdots,\mathbf{X}_{n_1}^{(1)}$ 来自总体 G_1 的样本，$\mathbf{X}_1^{(2)},\cdots,\mathbf{X}_{n_2}^{(2)}$ 是来自总体 G_2 的样本，$\boldsymbol{\mu}_1$ 和 $\boldsymbol{\mu}_2$ 的一个无偏估计分别为

$$\bar{\mathbf{X}}^{(1)}=\frac{1}{n_1}\sum_{i=1}^{n_1}\mathbf{X}_i^{(1)} \quad \text{和} \quad \bar{\mathbf{X}}^{(2)}=\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbf{X}_i^{(2)}$$

Σ 的一个联合无偏估计为 $$\hat{\Sigma}=\frac{1}{n_1+n_2-2}(\mathbf{S}_1+\mathbf{S}_2)$$

这里 $$\mathbf{S}_\alpha=\sum_{i=1}^{n_\alpha}(\mathbf{X}_i^{(\alpha)}-\bar{\mathbf{X}}^{(\alpha)})(\mathbf{X}_i^{(\alpha)}-\bar{\mathbf{X}}^{(\alpha)})', \quad \alpha=1,2$$

- 此时，两总体距离判别的判别函数为 $\hat{W}(\mathbf{X})=\hat{\boldsymbol{\alpha}}'(\mathbf{X}-\bar{\mathbf{X}})$

其中 $\bar{\mathbf{X}}=\frac{1}{2}(\bar{\mathbf{X}}^{(1)}+\bar{\mathbf{X}}^{(2)})$，$\hat{\boldsymbol{\alpha}}=\hat{\Sigma}^{-1}(\bar{\mathbf{X}}^{(1)}-\bar{\mathbf{X}}^{(2)})$。这样，判别规则为

$$\begin{cases}\mathbf{X}\in G_1, & \text{如果} \quad \hat{W}(\mathbf{X})\geq 0\\ \mathbf{X}\in G_2, & \text{如果} \quad \hat{W}(\mathbf{X})<0\end{cases} \qquad (4.7)$$

- 这里我们应该注意到：

（1） 当 $p=1$，G_1 和 G_2 的分布分别为 $N(\mu_1,\sigma^2)$ 和 $N(\mu_2,\sigma^2)$ 时，μ_1,μ_2,σ^2 均为已知，且 $\mu_1<\mu_2$，则判别系数为 $\alpha=\dfrac{\mu_1-\mu_2}{\sigma^2}<0$，判别函数为

$$W(x)=\alpha(x-\overline{\mu})$$

判别规则为

$$\begin{cases} x\in G_1, & \text{如果} \quad x\le\overline{\mu} \\ x\in G_2, & \text{如果} \quad x>\overline{\mu} \end{cases}$$

（2） 当$\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$，$\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$时，我们采用（4.4）式作为判别规则的形式。选择判别函数为

$$W^*(\mathbf{X}) = D^2(\mathbf{X}, G_1) - D^2(\mathbf{X}, G_2)$$

$$= (\mathbf{X}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_1^{-1}(\mathbf{X}-\boldsymbol{\mu}_1) - (\mathbf{X}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\mathbf{X}-\boldsymbol{\mu}_2)$$

它是$\mathbf{X}$的二次函数，相应的判别规则为

$$\begin{cases} \mathbf{X} \in G_1, & \text{如果} \quad W^*(\mathbf{X}) \leq 0 \\ \mathbf{X} \in G_2, & \text{如果} \quad W^*(\mathbf{X}) > 0 \end{cases}$$

2、多个总体的距离判别问题

- 问题：设有 k 个总体 $G_1, G_2, \cdots, G_k$，其均值和协方差矩阵分别是 $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\Sigma}_k$，而且 $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$。对于一个新的样品 $\mathbf{X}$，要判断它来自哪个总体。
- 该问题与两个总体的距离判别问题的解决思想一样。计算新样品 $\mathbf{X}$ 到每一个总体的距离，即

$$
\begin{aligned}
D^2(\mathbf{X}, G_\alpha) &= (\mathbf{X}-\boldsymbol{\mu}_\alpha)'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}_\alpha) \\
&= \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} - 2\boldsymbol{\mu}_\alpha'\boldsymbol{\Sigma}^{-1}\mathbf{X} + \boldsymbol{\mu}_\alpha'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_\alpha \\
&= \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} - 2(\mathbf{I}_\alpha'\mathbf{X} + C_\alpha)
\end{aligned}
\qquad (4.8)
$$

这里 $\mathbf{I}_\alpha = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_\alpha$，$C_\alpha = -\frac{1}{2}\boldsymbol{\mu}_\alpha'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_\alpha$，$\alpha = 1, 2, \cdots, k$。

- 由（4.8）式，可以取线性判别函数为

$$W_\alpha(\mathbf{X}) = \mathbf{I}'_\alpha \mathbf{X} + C_\alpha, \quad \alpha = 1,2,\cdots,k$$

相应的判别规则为

$$\mathbf{X} \in G_i \quad \text{如果} \quad W_i(\mathbf{X}) = \max_{1\le\alpha\le k}(\mathbf{I}'_\alpha \mathbf{X} + C_\alpha) \qquad (4.9)$$

针对实际问题，当$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_k$和$\boldsymbol{\Sigma}$均未知时，可以通过相应的样本值来替代。设$\mathbf{X}_1^{(\alpha)}, \cdots, \mathbf{X}_{n_\alpha}^{(\alpha)}$是来自总体$G_\alpha$中的样本（$\alpha = 1,2,\cdots,k$），则$\boldsymbol{\mu}_\alpha$（$\alpha = 1,2,\cdots,k$）和$\boldsymbol{\Sigma}$可估计为

$$\bar{\mathbf{X}}^{(\alpha)} = \frac{1}{n_\alpha}\sum_{i=1}^{n_\alpha} \mathbf{X}_i^{(\alpha)}, \qquad \alpha = 1,2,\cdots,k$$

和 $\hat{\boldsymbol{\Sigma}} = \dfrac{1}{n-k}\sum_{\alpha=1}^{k} \mathbf{S}_\alpha$ ， 其中 $n = n_1 + n_2 + \cdots + n_k$

$$\mathbf{S}_\alpha = \sum_{i=1}^{n_\alpha} (\mathbf{X}_i^{(\alpha)} - \bar{\mathbf{X}}^{(\alpha)})(\mathbf{X}_i^{(\alpha)} - \bar{\mathbf{X}}^{(\alpha)})'，\quad \alpha = 1,2,\cdots,k$$

- 同样，我们注意到，如果总体 $G_1, G_2, \cdots, G_k$ 的协方差矩阵分别是 $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\Sigma}_k$，而且它们不全相等，则计算 $\mathbf{X}$ 到各总体的马氏距离，即

$$D^2(\mathbf{X}, G_\alpha) = (\mathbf{X} - \boldsymbol{\mu}_\alpha)' \boldsymbol{\Sigma}_\alpha^{-1} (\mathbf{X} - \boldsymbol{\mu}_\alpha) \quad \alpha = 1,2,\cdots,k$$

则判别规则为

$$\mathbf{X} \in G_i \quad \text{如果} \quad D^2(\mathbf{X}, G_i) = \min_{1 \le \alpha \le k} D^2(\mathbf{X}, G_\alpha) \qquad (4.10)$$

当 $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\Sigma}_k$ 均未知时，$\boldsymbol{\mu}_\alpha$（$\alpha = 1,2,\cdots,k$）的估计同前，$\boldsymbol{\Sigma}_\alpha$（$\alpha = 1,2,\cdots,k$）的估计为

$$\hat{\boldsymbol{\Sigma}}_\alpha = \frac{1}{n_\alpha - 1} \mathbf{S}_\alpha，\qquad \alpha = 1,2,\cdots,k$$

三、判别分析的实质

- 我们知道，判别分析就是希望利用已经测得的变量数据，找出一种判别函数，使得这一函数具有某种最优性质，能把属于不同类别的样本点尽可能地区别开来。为了更清楚的认识判别分析的实质，以便能灵活的应用判别分析方法解决实际问题，我们有必要了解“划分”这样概念。

- 设R_1，R_2，…，R_k是p维空间R^p的k个子集，如果它们互不相交，且它们的和集为R^p，则称R_1，R_2，…，R_k为R^p的一个划分。

- 在两个总体的距离判别问题中，利用 $W(\mathbf{X})=\boldsymbol{\alpha}'(\mathbf{X}-\bar{\boldsymbol{\mu}})$ 可以得到空间 R^p 的一个划分

$$\begin{cases} R_1=\{\mathbf{X}:W(\mathbf{X})\geq 0\} \\ R_2=\{\mathbf{X}:W(\mathbf{X})<0\} \end{cases} \qquad (4.11)$$

新的样品 $\mathbf{X}$ 落入 R_1 推断 $\mathbf{X}\in G_1$，落入 R_2 推断 $\mathbf{X}\in G_2$。

- 这样我们将会发现，判别分析问题实质上就是在某种意义上，以最优的性质对p维空间R^p构造一个“划分”，这个“划分”就构成了一个判别规则。这一思想将在后面的各节中体现的更加清楚。

第三节 贝叶斯（Bayes）判别法

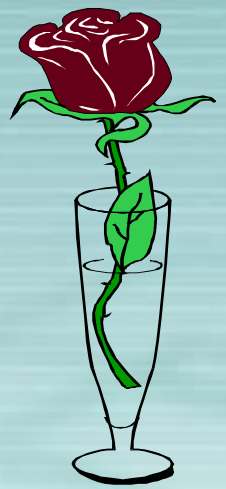
一 Bayes判别的基本思想

二 Bayes判别的基本方法

- 从上节看距离判别法虽然简单，便于使用。但是该方法也有它明显的不足之处。

  第一，判别方法与总体各自出现的概率的大小无关；

  第二，判别方法与错判之后所造成的损失无关。Bayes判别法就是为了解决这些问题而提出的一种判别方法。



一、Bayes判别的基本思想

- 问题：设有k个总体$G_1, G_2, \cdots, G_k$，其各自的分布密度函数$f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_k(\mathbf{x})$互不相同的，假设$k$个总体各自出现的概率分别为$q_1, q_2, \cdots, q_k$（先验概率），$q_i \geq 0$，$\sum_{i=1}^{k} q_i = 1$。

  假设已知若将本来属于G_i总体的样品错判到总体G_j时造成的损失为$C(j \mid i)$，$i, j = 1,2,\cdots,k$。在这样的情形下，对于新的样品$\mathbf{X}$判断其来自哪个总体。

- 下面我们对这一问题进行分析。首先应该清楚$C(i|i)=0$、$C(j|i)\geq 0$，对于任意的$i,j=1,2,\cdots,k$成立。设k个总体$G_1,G_2,\cdots,G_k$相应的p维样本空间为$R_1,R_2,\cdots,R_k$，即为一个划分，故我们可以简记一个判别规则为$R=(R_1,R_2,\cdots,R_k)$。从描述平均损失的角度出发，如果原来属于总体G_i且分布密度为$f_i(\mathbf{x})$的样品，正好取值落入了R_j，我们就将会错判为属于G_j。

- 故在规则 R 下，将属于 G_i 的样品错判为 G_j 的概率为

$$P(j \mid i, R) = \int_{R_j} f_i(\mathbf{x}) d\mathbf{x} \qquad i, j = 1,2,\cdots,k \qquad i \neq j$$

如果实属 G_i 的样品，错判到其它总体 $G_1,\cdots,G_{i-1},G_{i+1}\cdots,G_k$ 所造成的损失为 $C(1 \mid i),\cdots,C(i-1 \mid i),C(i+1 \mid i)\cdots,C(k \mid i)$，则这种判别规则 R 对总体 G_i 而言，样品错判后所造成的平均损失为

$$r(i \mid R) = \sum_{j=1}^{k} [C(j \mid i) P(j \mid i, R)] \qquad i = 1,2,\cdots,k$$

其中 $C(i \mid i) = 0$

- 由于 k 个总体 $G_1, G_2, \cdots, G_k$ 出现的先验概率分别为 $q_1, q_2, \cdots, q_k$，则用规则 R 来进行判别所造成的总平均损失为

$$g(R) = \sum_{i=1}^{k} q_i r(i, R)$$

$$= \sum_{i=1}^{k} q_i \sum_{j=1}^{k} C(j \mid i) P(j \mid i, R) \qquad (4.12)$$

所谓 **Bayes** 判别法则，就是要选择 $R_1, R_2, \cdots, R_k$，使得(**4.12**)式表示的总平均损失 $g(R)$ 达到极小。

二、Bayes判别的基本方法

- 设每一个总体G_i的分布密度为$f_i(\mathbf{x})$，$i=1,2,\cdots,k$，来自总体G_i的样品$\mathbf{X}$被错判为来自总体G_j（$i,j=1,2,\cdots,k$）时所造成的损失记为$C(j|i)$，并且$C(i|i)=0$。那么，对于判别规则$R=(R_1,R_2,\cdots,R_k)$产生的误判概率记为$P(j|i,R)$，有

$$P(j|i,R)=\int_{R_j} f_i(\mathbf{x})d\mathbf{x}$$

- 如果已知样品X来自总体G_i的先验概率为q_i，$i=1,2,\cdots,k$，则在规则R下，由（4.12）式知，误判的总平均损失为

$$g(R)=\sum_{i=1}^{k}q_i\sum_{j=1}^{k}C(j\mid i)P(j\mid i,R)$$

$$=\sum_{i=1}^{k}q_i\sum_{j=1}^{k}C(j\mid i)\int_{R_j}f_i(\mathbf{x})d\mathbf{x}$$

$$=\sum_{j=1}^{k}\int_{R_j}(\sum_{i=1}^{k}q_iC(j\mid i)f_i(\mathbf{x}))d\mathbf{x} \qquad (4.13)$$

令 $\sum_{i=1}^{k}q_iC(j\mid i)f_i(\mathbf{x})=h_j(\mathbf{x})$，那么，（4.13）式为

$$g(R)=\sum_{j=1}^{k}\int_{R_j}h_j(\mathbf{x})d\mathbf{x}$$

- 如果空间 R^p 有另一种划分 $R^* = (R_1^*, R_2^*, \cdots, R_k^*)$，则它的总平均损失为

$$g(R^*) = \sum_{j=1}^{k} \int_{R_j^*} h_j(\mathbf{x}) d\mathbf{x}$$

那么，在两种划分下的总平均损失之差为

$$g(R) - g(R^*) = \sum_{i=1}^{k} \sum_{j=1}^{k} \int_{R_i \cap R_j^*} [h_i(\mathbf{x}) - h_j(\mathbf{x})] d\mathbf{x} \qquad (4.14)$$

由 R_i 的定义，在 R_i 上 $h_i(\mathbf{x}) \le h_j(\mathbf{x})$ 对一切 j 成立，故(4.14)式小于或等于零，这说明 $R_1, R_2, \cdots, R_k$ 确能使总平均损失达到极小，它是 **Bayes** 判别的解。

- 这样，我们以 **Bayes** 判别的思想得到的划分 $R=(R_1,R_2,\cdots,R_k)$ 为

$$R_i=\{\mathbf{x}\,|\,h_i(\mathbf{x})=\min_{1\le j\le k}h_j(\mathbf{x})\}\qquad i=1,2,\cdots,k\qquad(4.15)$$

具体说来，当抽取了一个未知总体的样本值 $\mathbf{X}$，要判断它属于哪个总体，只要前计算出 k 个按先验分布加权的误判平均损失

$$h_j(\mathbf{x})=\sum_{i=1}^{k}q_iC(j\,|\,i)f_i(\mathbf{x}))\qquad j=1,2,\cdots,k\qquad(4.16)$$

然后再比较这 k 个误判平均损失 $h_1(\mathbf{x}),h_2(\mathbf{x}),\cdots,h_k(\mathbf{x})$ 的大小，选取其中最小的，则判定样品 $\mathbf{X}$ 来自该总体。

- 这里我们看一个特殊情形，当$k=2$时，由（4.16）式得

$$h_1(\mathbf{x})=q_2C(1|2)f_2(\mathbf{x}))\qquad h_2(\mathbf{x})=q_1C(2|1)f_1(\mathbf{x}))$$

从而

$$R_1=\{\mathbf{x}\,|\,q_2C(1|2)f_2(\mathbf{x})\le q_1C(2|1)f_1(\mathbf{x})\}$$

$$R_2=\{\mathbf{x}\,|\,q_2C(1|2)f_2(\mathbf{x})> q_1C(2|1)f_1(\mathbf{x})\}$$

若令

$$V(\mathbf{x})=\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})},\qquad d=\frac{q_2C(1|2)}{q_1C(2|1)}$$

则判别规则可表示为

$$\begin{cases}\mathbf{x}\in G_1, & \text{当}V(\mathbf{x})\ge d\\ \mathbf{x}\in G_2, & \text{当}V(\mathbf{x})<d\end{cases}\qquad(4.17)$$

- 如果在此，$f_1(\mathbf{x})$ 与 $f_2(\mathbf{x})$ 分别为 $N(\boldsymbol{\mu}_1,\boldsymbol{\Sigma})$ 和 $N(\boldsymbol{\mu}_2,\boldsymbol{\Sigma})$，那么

$$V(\mathbf{x})=\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}$$

$$=\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)+\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\right\}$$

$$=\exp\left\{[\mathbf{x}-(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)/2]'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right\}$$

$$=\exp W(\mathbf{x})$$

其中 $W(\mathbf{x})$ 由（4.5）所定义。于是，判定样品 $\mathbf{X}$ 来自该总体时，判别规则（4.17）成

$$\begin{cases}\mathbf{X}\in G_1, & \text{如果} \quad W(\mathbf{X})\geq \ln d\\ \mathbf{X}\in G_2, & \text{如果} \quad W(\mathbf{X})< \ln d\end{cases} \qquad (4.18)$$

对比判别规则（4.6），唯一的差别仅在于阈值点，（4.6）用 0 作为阈值点，而这里用 $\ln d$。当 $q_1=q_2$，$C(1|2)=C(2|1)$ 时，$d=1$，$\ln d=0$，则（4.6）与（4.18）完全一致。

第四节 费歇（Fisher）判别法

一 Fisher判别的基本思想

二 Fisher判别函数的构造

三 线性判别函数的求法

多元统计

- Fisher判别法是1936年提出来的，该方法的主要思想是通过将多维数据投影到某个方向上，投影的原则是将总体与总体之间尽可能的放开，然后再选择合适的判别规则，将新的样品进行分类判别。

一、Fisher判别的基本思想

- 从k个总体中抽取具有p个指标的样品观测数据，借助方差分析的思想构造一个线性判别函数

$$U(\mathbf{X}) = u_1X_1 + u_2X_2 + \cdots + u_pX_p = \mathbf{u}'\mathbf{X} \qquad (4.19)$$

其中系数$\mathbf{u} = (u_1, u_2, \cdots, u_p)'$确定的原则是使得总体之间区别最大，而使每个总体内部的离差最小。有了线性判别函数后，对于一个新的样品，将它的p个指标值代入线性判别函数（4.19）式中求出$U(\mathbf{X})$值，然后根据判别一定的规则，就可以判别新的样品属于哪个总体。

二、Fisher判别函数的构造

1、针对两个总体的情形

- 假设有两个总体G_1, G_2，其均值分别为$\boldsymbol{\mu}_1$和$\boldsymbol{\mu}_2$，协方差矩阵为$\boldsymbol{\Sigma}_1$和$\boldsymbol{\Sigma}_2$。当$\mathbf{X} \in G_i$时，我们可以求出$\mathbf{u}'\mathbf{X}$的均值和方差，即

$$E(\mathbf{u}'\mathbf{X}) = E(\mathbf{u}'\mathbf{X} \mid G_i) = \mathbf{u}'E(\mathbf{X} \mid G_i) = \mathbf{u}'\boldsymbol{\mu}_i \triangleq \overline{\mu}_i, \quad i = 1,2$$

$$D(\mathbf{u}'\mathbf{X}) = D(\mathbf{u}'\mathbf{X} \mid G_i) = \mathbf{u}'D(\mathbf{X} \mid G_i)\mathbf{u} = \mathbf{u}'\boldsymbol{\Sigma}_i\mathbf{u} \triangleq \sigma_i^2, \quad i = 1,2$$

在求线性判别函数时，尽量使得总体之间差异大，也就是要求$\mathbf{u}'\boldsymbol{\mu}_1 - \mathbf{u}'\boldsymbol{\mu}_2$尽可能的大，即$\overline{\mu}_1 - \overline{\mu}_2$变大；同时要求每一个总体内的离差平方和最小，即$\sigma_1^2 + \sigma_2^2$，则我们可以建立一个目标函数

$$\Phi(\mathbf{u}) = \frac{(\overline{\mu}_1 - \overline{\mu}_2)}{\sigma_1^2 + \sigma_2^2} \qquad (4.20)$$

这样，我们就将问题转化为，寻找$\mathbf{u}$使得目标函数$\Phi(\mathbf{u})$达到最大。从而可以构造出所要求的线性判别函数。

2、针对多个总体的情形

- 假设有 k 个总体 $G_1, G_2, \cdots, G_k$，其均值和协方差矩阵分别为 $\boldsymbol{\mu}_i$ 和 $\boldsymbol{\Sigma}_i\ (>0)$（$i=1,2,\cdots,k$）。同样，我们考虑线性判别函数 $\mathbf{u}'\mathbf{X}$，在 $\mathbf{X}\in G_i$ 的条件下，有

$$E(\mathbf{u}'\mathbf{X})=E(\mathbf{u}'\mathbf{X}\,|\,G_i)=\mathbf{u}'E(\mathbf{X}\,|\,G_i)=\mathbf{u}'\boldsymbol{\mu}_i \qquad i=1,2,\cdots,k$$

$$D(\mathbf{u}'\mathbf{X})=D(\mathbf{u}'\mathbf{X}\,|\,G_i)=\mathbf{u}'D(\mathbf{X}\,|\,G_i)\mathbf{u}=\mathbf{u}'\boldsymbol{\Sigma}_i\mathbf{u} \qquad i=1,2,\cdots,k$$

令

$$b=\sum_{i=1}^{k}(\mathbf{u}'\boldsymbol{\mu}_i-\mathbf{u}'\overline{\boldsymbol{\mu}})^2$$

$$e=\sum_{i=1}^{k}\mathbf{u}'\boldsymbol{\Sigma}_i\mathbf{u}=\mathbf{u}'(\sum_{i=1}^{k}\boldsymbol{\Sigma}_i)\mathbf{u}=\mathbf{u}'\mathbf{E}\mathbf{u}$$

- 其中 $\bar{\boldsymbol{\mu}}=\frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\mu}_i$，$\mathbf{E}=\sum_{i=1}^{k}\boldsymbol{\Sigma}_i$。这里 b 相当于一元方差分析中的组间差 e 相当于组内差，应用方差分析的思想，选择 $\mathbf{u}$ 使得目标函数

$$\Phi(\mathbf{u})=\frac{b}{e} \qquad (4.21)$$

达到极大。

这里我们应该说明的是，如果我们得到线性判别函数 $\mathbf{u}'\mathbf{X}$，对于一个新的样品 $\mathbf{X}$ 可以这样构造一个判别规则，如果

$$\left|\mathbf{u}'\mathbf{X}-\mathbf{u}'\boldsymbol{\mu}_j\right|=\min_{1\le i\le k}\left|\mathbf{u}'\mathbf{X}-\mathbf{u}'\boldsymbol{\mu}_i\right| \qquad (4.22)$$

则判定 $\mathbf{X}$ 来自总体 G_j。

三、线性判别函数的求法

- 针对多个总体的情形，我们讨论使目标函数（4.21）式达到极大的求法。设 $\mathbf{X}$ 为 p 维空间的样品，那么 $\bar{\boldsymbol{\mu}} = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\mu}_i = \frac{1}{k}\mathbf{M}'\mathbf{1}$

其中

$$\mathbf{M} = \begin{pmatrix} \mu_{11} & \mu_{21} & \cdots & \mu_{p1} \\ \mu_{12} & \mu_{22} & \cdots & \mu_{p1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{1k} & \mu_{1k} & \cdots & \mu_{pk} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1' \\ \boldsymbol{\mu}_2' \\ \cdots \\ \boldsymbol{\mu}_k' \end{pmatrix} \qquad \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{pmatrix}$$

注意到

$$\mathbf{M}'\mathbf{M} = \begin{pmatrix} \boldsymbol{\mu}_1 & \boldsymbol{\mu}_2 & \cdots & \boldsymbol{\mu}_k \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_1' \\ \boldsymbol{\mu}_2' \\ \cdots \\ \boldsymbol{\mu}_k' \end{pmatrix} = \sum_{i=1}^{k}\boldsymbol{\mu}_i\boldsymbol{\mu}_i'$$

- 从而

$$
\begin{aligned}
b &= \sum_{i=1}^{k}(\mathbf{u}'\boldsymbol{\mu}_i - \mathbf{u}'\bar{\boldsymbol{\mu}})^2 \\
&= \mathbf{u}'\sum_{i=1}^{k}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\mathbf{u} \\
&= \mathbf{u}'[\sum_{i=1}^{k}\boldsymbol{\mu}_i\boldsymbol{\mu}_i' - k\bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}']\mathbf{u} \\
&= \mathbf{u}'(\mathbf{M}'\mathbf{M} - \frac{1}{k}\mathbf{M}'\mathbf{1}\mathbf{1}'\mathbf{M})\mathbf{u} \\
&= \mathbf{u}'\mathbf{M}'(\mathbf{I} - \frac{1}{k}\mathbf{J})\mathbf{M}\mathbf{u} \\
&= \mathbf{u}'\mathbf{B}\mathbf{u}
\end{aligned}
$$

这里，$\mathbf{B} = \mathbf{M}'(\mathbf{I} - \frac{1}{k}\mathbf{J})\mathbf{M}$，$\mathbf{I}_{p\times p}$ 为 $p \times p$ 的单位阵，$\mathbf{J} = \begin{pmatrix} 1 & \cdots & 1 \\ & \ddots & \\ 1 & \cdots & 1 \end{pmatrix}$。

即有 $\Phi(\mathbf{u}) = \dfrac{\mathbf{u}'\mathbf{B}\mathbf{u}}{\mathbf{u}'\mathbf{E}\mathbf{u}}$ （4.23）求使得（4.23）式达到极大的 $\mathbf{u}$。

- 为了确保解的唯一性，不妨设$\mathbf{u}'\mathbf{E}\mathbf{u}=1$，这样问题转化为，在$\mathbf{u}'\mathbf{E}\mathbf{u}=1$的条件下，求$\mathbf{u}$使得$\mathbf{u}'\mathbf{B}\mathbf{u}$式达到极大。

考虑目标函数　$\varphi(\mathbf{u})=\mathbf{u}'\mathbf{B}\mathbf{u}-\lambda(\mathbf{u}'\mathbf{E}\mathbf{u}-1)$　（4.24）

对（4.24）式求导，有

$$\begin{cases}\dfrac{\partial\varphi}{\partial\mathbf{u}}=2(\mathbf{B}-\lambda\mathbf{E})\mathbf{u}=0 & (4.25)\\ \dfrac{\partial\varphi}{\partial\lambda}=\mathbf{u}'\mathbf{E}\mathbf{u}-1=0 & (4.26)\end{cases}$$

对（4.25）式两边同乘$\mathbf{u}'$，有　$\mathbf{u}'\mathbf{B}\mathbf{u}=\lambda\mathbf{u}'\mathbf{E}\mathbf{u}=\lambda$

从而，$\mathbf{u}'\mathbf{B}\mathbf{u}$的极大值为$\lambda$。再用$\mathbf{E}^{-1}$左乘（4.25）式，有

$$(\mathbf{E}^{-1}\mathbf{B}-\lambda\mathbf{I})\mathbf{u}=0 \qquad (4.27)$$

由（4.27）式说明λ为$\mathbf{E}^{-1}\mathbf{B}$特征值，$\mathbf{u}$为$\mathbf{E}^{-1}\mathbf{B}$的特征向量。在此最大特征值所对应的特征向量$\mathbf{u}=(u_1,u_2,\cdots,u_p)'$为我们所求结果。

- 这里值得注意的是，本书有几处利用极值原理求极值时，只给出了不要条件的数学推导，而有关充分条件的论证省略了，因为在实际问题中，往往根据问题本身的性质就能肯定有最大值（或最小值），如果所求的驻点只有一个，这时就不需要根据极值存在的充分条件判定它是极大还是极小而就能肯定这唯一的驻点就是所求的最大值（或最小值）。为了避免用较多的数学知识或数学上的推导，这里不追求数学上的完整性。
- 在解决实际问题时，当总体参数未知，需要通过样本来估计，我们仅对 $k=2$ 的情形加以说明。设样本分别为 $\mathbf{X}_1^{(1)},\mathbf{X}_2^{(1)},\cdots\mathbf{X}_{n_1}^{(1)}$ 和 $\mathbf{X}_1^{(2)},\mathbf{X}_2^{(2)},\cdots\mathbf{X}_{n_2}^{(2)}$，则

$$\bar{\mathbf{X}}=\frac{n_1\bar{\mathbf{X}}^{(1)}+n_2\bar{\mathbf{X}}^{(2)}}{n_1+n_2}$$

$$\bar{\mathbf{X}}^{(1)}-\bar{\mathbf{X}}=\frac{n_2}{n_1+n_2}(\bar{\mathbf{X}}^{(1)}-\bar{\mathbf{X}}^{(2)})$$

$$\bar{\mathbf{X}}^{(2)}-\bar{\mathbf{X}}=\frac{n_1}{n_1+n_2}(\bar{\mathbf{X}}^{(2)}-\bar{\mathbf{X}}^{(1)})$$

那么 $$\hat{\mathbf{B}}=n_1(\bar{\mathbf{X}}^{(1)}-\bar{\mathbf{X}})(\bar{\mathbf{X}}^{(1)}-\bar{\mathbf{X}})'+n_2(\bar{\mathbf{X}}^{(2)}-\bar{\mathbf{X}})(\bar{\mathbf{X}}^{(2)}-\bar{\mathbf{X}})'$$

$$=\frac{n_1n_2}{n_1+n_2}(\bar{\mathbf{X}}^{(1)}-\bar{\mathbf{X}}^{(2)})(\bar{\mathbf{X}}^{(1)}-\bar{\mathbf{X}}^{(2)})'$$

当$\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\cdots,\boldsymbol{\mu}_k$和$\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2,\cdots,\boldsymbol{\Sigma}_k$均未知时，$\boldsymbol{\mu}_\alpha$（$\alpha=1,2,\cdots,k$）的估计同前，$\boldsymbol{\Sigma}_\alpha$（$\alpha=1,2,\cdots,k$）的估计为

$$\hat{\boldsymbol{\Sigma}}_\alpha=\frac{1}{n_\alpha-1}\mathbf{S}_\alpha,\qquad \alpha=1,2,\cdots,k$$

本章结束