应用多元统计分析

第四章 鉴别分析

第一节 引言

第二节 距离鉴别法

第三节 贝叶斯（Bayes）鉴别法

第四节 费歇（Fisher）鉴别法

第五节 实例分析与计算机实现

第一节　引言

- 在我们的日常生活和工作实践中，经常会遇到鉴别分析问题，即根据历史上划分类别的有关资料和某种最优准则，拟定一种鉴别措施，鉴定一种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，统计了每个患者若干项症状指标数据。目前想利用既有的这些资料找出一种措施，使得对于一种新的病人，当测得这些症状指标数据时，能够鉴定其患有哪种病。又如，在天气预报中，我们有一段较长时间有关某地域每天气象的统计资料（晴阴雨、气温、气压、湿度等），目前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措施。这些问题都能够应用鉴别分析措施予以处理。

- 把此类问题用数学语言来体现，能够论述如下：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，找出一种鉴别函数，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量地域别开来，并对测得一样p项指标（变量）数据的一种新样本，能鉴定这个样本归属于哪一类。

- 鉴别分析内容很丰富，措施诸多。判断分析按鉴别的总体数来区别，有两个总体鉴别分析和多总体鉴别分析；按区别不同总体所用的数学模型来分，有线性鉴别和非线性鉴别；按鉴别时所处理的变量措施不同，有逐渐鉴别和序贯鉴别等。鉴别分析能够从不同角度提出问题，所以有不同的鉴别准则，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、最大约率准则等等，按鉴别准则的不同又提出多种鉴别措施。本章仅简介常用的几种鉴别分析措施：距离鉴别法、Fisher鉴别法、Bayes鉴别法和逐渐鉴别法。

第二节 距离鉴别法

一 马氏距离的概念

二 距离鉴别的思想及措施

三 鉴别分析的实质

一、马氏距离的概念

- 设 p R^p $\mathbf{X}=(X_1,X_2,\mathrm{L}\ ,X_p)'$

$\mathbf{Y}=(Y_1,Y_2,\mathrm{L}\ ,Y_p)'$

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

X $N(\mu_1,\sigma^2)$ Y $N(\mu_2,4\sigma^2)$

A X 2σ

Y 3σ A

A X Y

A μ_1 $2\sigma_x$

μ_2 $1.5\sigma_y$ A Y

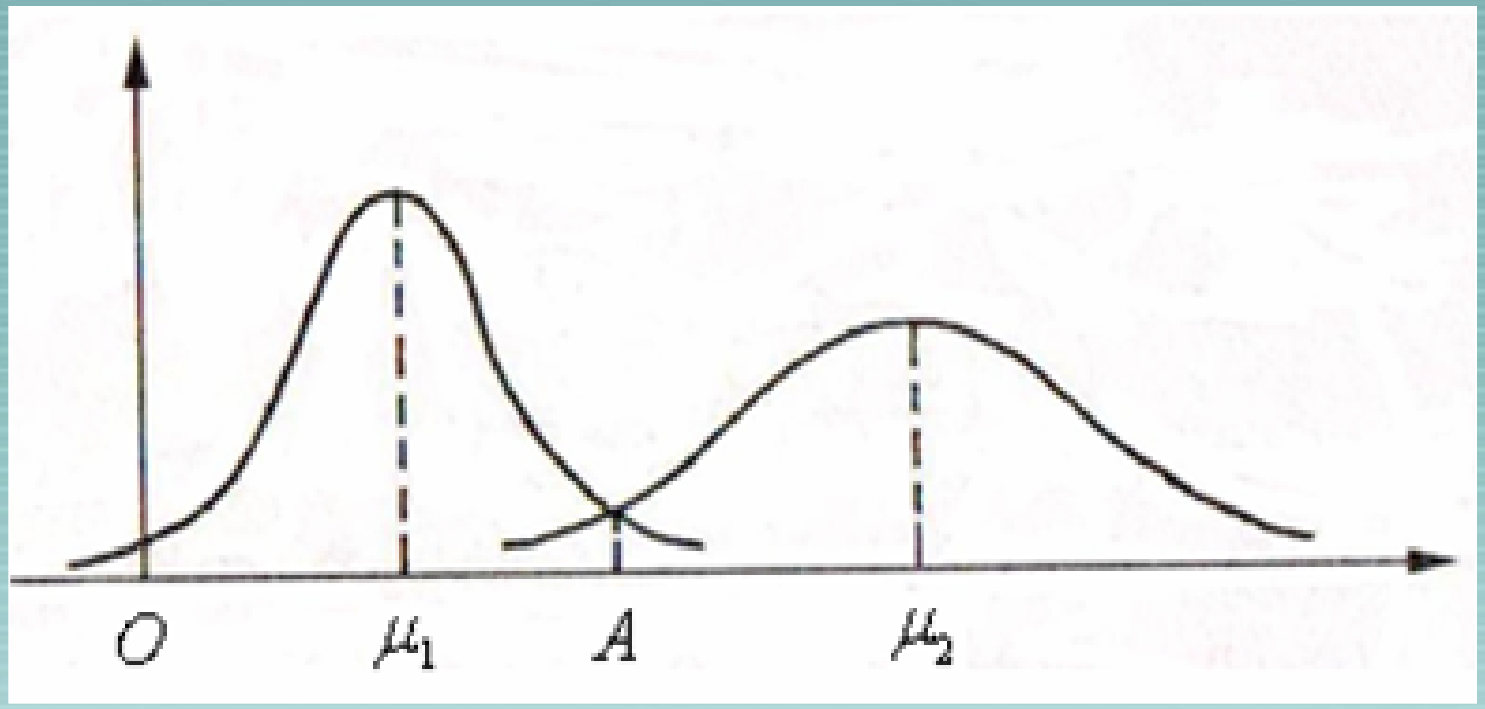


图4.1

第二、设有量度重量和长度的两个变量 X Y

$A(0,5)$ $B(10,0)$ $C(1,0)$ $D(0,10)$

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

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

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

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

- 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 $\mathbf{X}$ $\mathbf{Y}$ $\boldsymbol{\mu}$ $\boldsymbol{\Sigma}(>0)$ G p G $\mathbf{X}$ $\mathbf{Y}$

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

$\mathbf{X}$ G

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

$$\boldsymbol{\Sigma}=\mathbf{I}$$

二、距离鉴别的思想及措施

1、两个总体的距离鉴别问题

- 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，对于一种新的样品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），并按照如下的鉴别规则进行判断

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

- 这个鉴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，假如其值为正，X属于G_2；不然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 $\bar{\boldsymbol{\mu}} = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$

记

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

则判别规则(4.4)式表示为

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

这里称

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

$W(\mathbf{X})$ 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

$\mathbf{X}_1^{(1)}, \cdots, \mathbf{X}_{n_1}^{(1)}$ 是来自总体 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

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

$\boldsymbol{\Sigma}$

$$\hat{\boldsymbol{\Sigma}}=\frac{1}{n_1+n_2-2}(\mathbf{S}_1+\mathbf{S}_2)$$

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

$$\hat{W}(\mathbf{X})=\hat{\boldsymbol{\alpha}}'(\mathbf{X}-\bar{\mathbf{X}})$$

$$\bar{\mathbf{X}}=\frac{1}{2}(\bar{\mathbf{X}}^{(1)}+\bar{\mathbf{X}}^{(2)}) \qquad \hat{\boldsymbol{\alpha}}=\hat{\boldsymbol{\Sigma}}^{-1}(\bar{\mathbf{X}}^{(1)}-\bar{\mathbf{X}}^{(2)})$$

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

- 这里我们应该注意到：

（1） 当 $p=1$ G_1 G_2 $N(\mu_1,\sigma^2)$

$N(\mu_2,\sigma^2)$ μ_1,μ_2,σ^2 $\mu_1<\mu_2$

$$\alpha=\frac{\mu_1-\mu_2}{\sigma^2}<0$$

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

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

（2） 当 $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$

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

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

$\mathbf{X}$

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

2、多种总体的距离鉴别问题

- 问题：设有 k 个总体 $G_1, G_2, \cdots, G_k$，$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_k$，$\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\Sigma}_k$，$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$，$\mathbf{X}$

- $\mathbf{X}$

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

$$= \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} - 2\boldsymbol{\mu}_\alpha' \boldsymbol{\Sigma}^{-1} \mathbf{X} + \boldsymbol{\mu}_\alpha' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_\alpha$$

$$= \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} - 2(\mathbf{I}_\alpha' \mathbf{X} + C_\alpha)$$

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

- 由（4.8）式，可以取线性判别函数为

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

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

$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_k$ $\boldsymbol{\Sigma}$

$\mathbf{X}_1^{(\alpha)}, \mathrm{L}, \mathbf{X}_{n_\alpha}^{(\alpha)}$ G_α

$\alpha = 1,2,\cdots,k$ $\boldsymbol{\mu}_\alpha$ $\alpha = 1,2,\cdots,k$ $\boldsymbol{\Sigma}$

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

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-k}\sum_{\alpha=1}^{k} \mathbf{S}_\alpha \qquad n = n_1 + n_2 + \cdots + n_k$$

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

- 同样，我们注意到，如果总体 $G_1, G_2, \cdots, G_k$ $\bar{\boldsymbol{\Sigma}}_1, \bar{\boldsymbol{\Sigma}}_2, \cdots, \bar{\boldsymbol{\Sigma}}_k$ $\bar{\mathbf{X}}$

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

$\mathbf{X} \in G_i$ $D^2(\mathbf{X}, G_i) = \min\limits_{1 \le \alpha \le k} D^2(\mathbf{X}, G_\alpha)$

$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_k$ $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\Sigma}_k$ $\boldsymbol{\mu}_\alpha$ $\alpha = 1,2,\cdots,k$

$\boldsymbol{\Sigma}_\alpha$ $\alpha = 1,2,\cdots,k$

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

三、鉴别分析的实质

- 我们懂得，鉴别分析就是希望利用已经测得的变量数据，找出一种鉴别函数，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量地域别开来。为了更清楚的认识鉴别分析的实质，以便能灵活的应用鉴别分析措施处理实际问题，我们有必要了解“划分”这么概念。

- 设R_1，R_2，…，R_k是p维空间R^p的k个子集，假如它们互不相交，且它们的和集为R^p，则称R_1，R_2，…，R_k为R^p的一种划分。

- 在两个总体的距离判别问题中，利用$W(\mathbf{X})=\boldsymbol{\alpha}'(\mathbf{X}-\bar{\boldsymbol{\mu}})$

R^p

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

$\mathbf{X}$ R_1 $\mathbf{X}\in G_1$ R_2 $\mathbf{X}\in G_2$

- 这么我们将会发觉，鉴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构造一种“划分”，这个“划分”就构成了一种鉴别规则。这一思想将在背面的各节中体现的愈加清楚。

例 在企业的考核中，能够根据企业的生产经营情况把企业分为优异企业和一般企业。考核企业经营情况的指标有：

资金利润率=利润总额/资金占用总额

劳动生产率=总产值/职员平均人数

产品净值率=净产值/总产值

三个指标的均值向量和协方差矩阵如下。既有二个企业，观察值分别为

（7.8，39.1，9.6）和（8.1，34.2，6.9），问这两个企业应该属于哪一类？

多元统计

变量	均值向量		协方差矩阵			
	■ 优异	一般				
资金利润率	13.5	5.4	68.39		40.24	21.41
劳动生产率	40.7	29.8	40.24		54.58	11.67
产品净值率	10.7	6.2	21.41		11.67	7.90

$$\Sigma^{-1}=\begin{bmatrix} 0.119337 & -0.02753 & -0.28276 \\ -0.02753 & 0.033129 & 0.025659 \\ -0.28276 & 0.025659 & 0.854988 \end{bmatrix}$$

$$\mu_1-\mu_2=\begin{bmatrix}8.1\\10.9\\4.5\end{bmatrix}\qquad (\mu_1+\mu_2)/2=\begin{bmatrix}9.45\\35.25\\8.45\end{bmatrix}$$

判别函数的系数 $\Sigma^{-1}(\mu_1-\mu_2)=\begin{bmatrix}-0.60581\\0.25362\\1.83679\end{bmatrix}$

判别函数的常数项 $\frac{(\mu_1+\mu_2)'}{2}\Sigma^{-1}(\mu_1-\mu_2)$

$$=\begin{bmatrix}9.45 & 35.25 & 8.45\end{bmatrix}\begin{bmatrix}-0.60581\\0.25362\\1.83679\end{bmatrix}=18.73596$$

- 线性鉴别函数为：

$$y = -0.60581x_1 + 0.25362x_2 + 1.83679x_3 - 18.73596$$

$$y_1 = -0.60581 \times 7.8 + 0.25362 \times 39.1 + 1.83679 \times 9.6 - 18.73596$$
$$= 4.0892 > 0$$(第一个新企业属于一类

$$y_2 = -0.60581 \times 8.1 + 0.25362 \times 34.2 + 1.83679 \times 6.9 - 18.73596$$
$$= -2.2956 < 0$$(第二个新企业属于二类

- 错判概率：由上面的分析能够看出，马氏距离鉴别法是合理的，但是这并不意谓着不会发生误判。

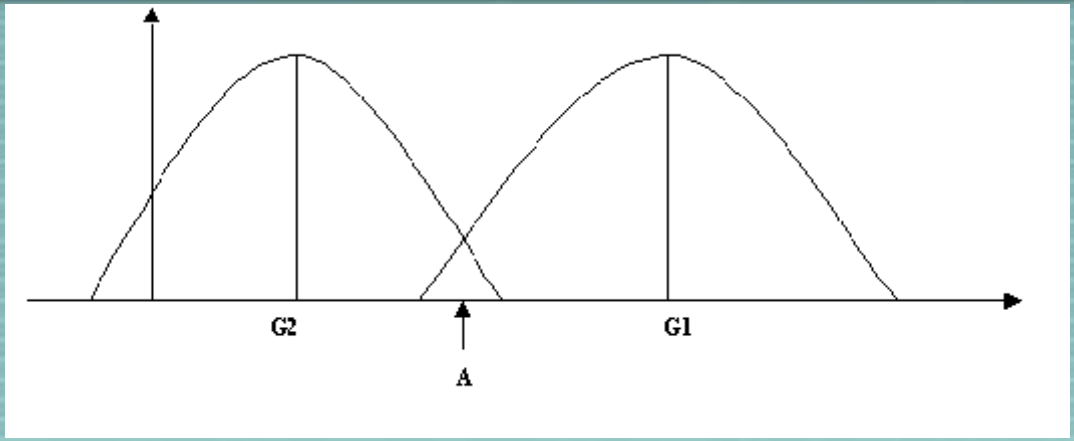
两总体分别服从$N(\mu_1,\sigma^2)$和$N(\mu_2,\sigma^2)$，其线性判别函数为：

$$W(y)=(y-\overline{\mu})\frac{1}{\sigma^2}(\mu_1-\mu_2)$$，其中

$\overline{\mu}=\frac{\mu_1+\mu_2}{2}$。不失一般性设$\mu_1>\mu_2$，这种情况下线性判别函数 W（y）的符号取决于 y>$\overline{\mu}$ 还是 y≤$\overline{\mu}$。

如果 y>$\overline{\mu}$，则$\mathbf{y}\in G_1$

如果 y<$\overline{\mu}$，则$\mathbf{y}\in G_2$。



错判概率：$P(X_2 > \bar{\mu}) = P(X_2 - \mu_2 > \frac{\mu_1 + \mu_2}{2} - \mu_2)$

$= P(X_2 - \mu_2 > \frac{\mu_1 - \mu_2}{2})$

$= 1 - \Phi(\frac{\mu_1 - \mu_2}{2\sigma})$

距离鉴别只要求懂得总体的数字特征，不涉及总体的分布函数，当参数和协方差未知时，就用样本的均值和协方差矩阵来估计。距离鉴别措施简朴实用，但没有考虑到每个总体出现的机会大小，即先验概率，也没有考虑到错判的损失。贝叶斯鉴别法正是为了处理这两个问题提出的鉴别分析措施。

第三节 贝叶斯（Bayes）鉴别法

一 Bayes鉴别的基本思想

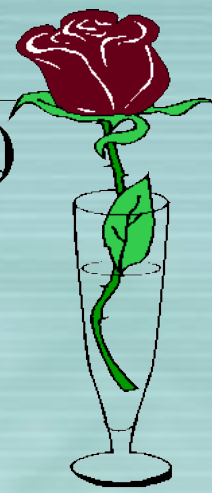
二 Bayes鉴别的基本措施

办公室新来了一种雇员小王，小王是好人还是坏人大家都在猜测。按人们主观意识，一种人是好人或坏人的概率均为0.5。坏人总是要做坏事，好人总是做好事，偶尔也会做一件坏事，一般好人做好事的概率为0.9，坏人做好事的概率为0.2，一天，小王做了一件好事，小王是好人的概率有多大，你目前把小王判为何种人。

$$P(\text{好人}/\text{做好事})$$

$$= \frac{P(\text{好人})P(\text{做好事}/\text{好人})}{P(\text{好人})P(\text{做好事}/\text{好人}) + P(\text{坏人})P(\text{做好事}/\text{坏人})}$$

$$= \frac{0.5\times 0.9}{0.5\times 0.9 + 0.5\times 0.2} = 0.82$$



$$P(\text{坏人}/\text{做好事})$$

$$=\frac{P(\text{坏人})P(\text{做好事}/\text{坏人})}{P(\text{好人})P(\text{做好事}/\text{好人})+P(\text{坏人})P(\text{做好事}/\text{坏人})}$$

$$=\frac{0.5\times0.2}{0.5\times0.9+0.5\times0.2}=0.18$$

距离鉴别简朴直观，很实用，但是距离鉴别的措施把总体等同看待，没有考虑到总体会以不同的概率（先验概率）出现，也没有考虑误判之后所造成的损失的差别。

一种好的鉴别措施，既要考虑到各个总体出现的先验概率，又要考虑到错判造成的损失，Bayes鉴别就具有这些优点，其鉴别效果愈加理想，应用也更广泛。

- 贝叶斯公式是一种我们熟知的公式

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\Sigma P(A \mid B_i)P(B_i)}$$

- 贝叶斯鉴别

在各总体的概率分布及先验概率已知的前提下，分别计算待判对象属于各总体的后验概率，并以最大后验概率相应的总体来作为待判对象的所属总体。

一、Bayes鉴别的基本思想

- 问题：设有 k 个总体 $G_1, G_2, \cdots, G_k$，

  $f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_k(\mathbf{x})$ 分别是 k 个总体的概率密度函数，

  $q_1, q_2, \cdots, q_k$ 是 k 个总体的先验概率，$q_i \geq 0$，$\sum_{i=1}^{k} q_i = 1$，

  G_i 表示第 i 个总体，G_j 表示第 j 个总体，

  $C(j \mid i)$ $i, j = 1, 2, \cdots, k$ 表示将来自 G_i 的样品 $\mathbf{X}$ 错判为 G_j 时所造成的损失。

- 下面我们对这一问题进行分析。首先应该清楚$C(i|i)=0$

$C(j|i)\geq 0$ $i,j=1,2,\square,k$ k

$G_1,G_2,\square,G_k$ p $R_1,R_2,\square,R_k$

$\bar{R}=(R_1,R_2,\square,R_k)$

G_i $f_i(\mathbf{x})$ R_j

G_j

- 故在规则 R G_i G_j

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

G_i $G_1, \square, G_{i-1}, G_{i+1} \square, G_k$

$C(1 \mid i), \square, C(i-1 \mid i), C(i+1 \mid i) \square, C(k \mid i)$

R G_i

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

$$C(i \mid i) = 0$$

- 由于 k $G_1, G_2, \square, G_k$ $q_1, q_2, \square, q_k$ R

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

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

$R_1, R_2, \square, R_k$

$g(R)$

二、Bayes鉴别的基本措施

- 设每一个总体 G_i $f_i(\mathbf{x})$ $i=1,2,\square,k$

  G_i $\mathbf{X}$ G_j $i,j=1,2,\square,k$

  $C(j\mid i)$ $C(i\mid i)=0$

  $R=(R_1,R_2,\square,R_k)$

  $P(j\mid i,R)$

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

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

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

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

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

$$\sum_{i=1}^{k}q_iC(j\mid i)f_i(\mathbf{x})=h_j(\mathbf{x})$$

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

- 如果空间 R^p $R^* = (R_1^*, R_2^*, \square, R_k^*)$

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

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

R_i R_i $h_i(\mathbf{x}) \le h_j(\mathbf{x})$ j

$R_1, R_2, \square, R_k$

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

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

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

$h_1(\mathbf{x}),h_2(\mathbf{x}),\square,h_k(\mathbf{x})$