广告中的数学

在我们的现实生活中，广告无所不在。广告给商家带来了丰厚的利润，广告中蕴藏着诸多学问。以房产销售广告为例，房产开发商为了扩大销售，提高销售量，通常会印制精美的广告分发给大家。虽然买房人的买房行为是随机的，他可能买房，也可能暂时不买，可能买这家开发商的房子，也可能买另一家开发商的房子，但与各开发商的广告投入有一定的关联。一般地，随着广告费用的增加，潜在的购买量会增加，但市场的购买力是有一定限度的。表 9.1 给出了某开发商以往 9 次广告投入及预测的潜在购买力。

表 **9.1** 广告投入与潜在购买力统计（单位：百万元）

广告投入	0.2	0.4	0.5	0.52	0.56	0.65	0.67	0.69	1
购买力	10340	10580	10670	10690	10720	10780	10800	10810	10950

下面从数学角度，通过合理的假设为开发商制定合理的广告策略，并给出单位面积成本 700 元，售价为 4000 元条件下的广告方案。

模型假设

（1）假设单位面积成本为 p_1 元，售价为 p_2 元，忽略其他费用，需求量 r 是随机变量，其概率密度为 $p(r)$。

（2）假设广告投入为 p 百万元，潜在购买力是 p 的函数记作 $s(p)$，实际供应量为 y。

模型建立

开发商制定策略的好坏主要由利润来确定，好的策略应该获得好的利润（平均意义下），为此，必须计算平均销售量 $E(x)$。

$$E(x) = \int_0^y rp(r)dr + \int_y^{+\infty} yp(r)dr$$

上面右边第二项表示当需求量大于等于供应量时，取需求量等于供应量。

因此，利润函数为

$$R(y) = p_2E(x) - p_1y - p$$

利用 $\int_0^{+\infty} p(r)dr = 1$ 得到

$$R(y) = (p_2 - p_1)y - p - p_2\int_0^y (y-r)p(r)dr \qquad (9.1)$$

上式中，第一项表示已售房毛利润，第二项为广告成本，第三项为未售出房的损失。

模型求解

为了获得最大利润，只需对（9.1）式求导并令其为零，设 $R(y)$ 获得最大值时 y 的最优值为 y^*，则

$$\frac{dR(y)}{dy}=(p_2-p_1)-p_2\int_0^y p(r)dr=0$$

因此，y^* 满足关系式

$$\int_0^{y^*} p(r)dr=\frac{p_2-p_1}{p_2} \tag{9.2}$$

通过（9.2）式知道，在广告投入一定的情况下，可以求出最优的供应量，但依赖于需求量的概率分布。为使问题更加明确，增加如下假设：

（3） 假设需求量 r 服从 $U[0, s(p)]$ 分布，即

$$p(r)=\begin{cases}\dfrac{1}{s(p)} & 0\le r\le s(p)\\ 0 & \text{其 他}\end{cases} \tag{9.3}$$

将（9.3）代人（9.2）得到

$$y^*=\frac{p_2-p_1}{p_2}s(p) \tag{9.4}$$

即最优的供应量等于毛利率与由广告费确定的潜在购买力的乘积。将（9.4）式代入（9.1）式，得到最大利润为

$$R(y^*)=\frac{(p_2-p_1)^2}{2p_2}s(p)-p \tag{9.5}$$

对（9.5）式关于 p 求导，得驻点 p^* 满足的方程为

$$s'(p^*)=\frac{2p_2}{(p_2-p_1)^2} \tag{9.6}$$

因此，只要知道了潜在购买力函数，就可以给出最优的广告投入。

下面根据开发商获得的相关数据，来确定潜在购买力函数。通过对表 9.1 数据分析，得知其符合 logistic 型曲线增长率，经拟合得到

$$s(p)=10^5/(9+e^{-2p}) \tag{9.7}$$

记

$$l=\frac{2p_2}{(p_2-p_1)^2}\times 10^{-5}$$

将（9.7）式代入（9.6）式，当 $1-18l>0$ 时，求得

$$p^*=-\frac{1}{2}\ln(1-9l-\sqrt{1-18l})+\frac{1}{2}\ln l \tag{9.8}$$

将 $p_1=0.0007, p_2=0.004$ 代入（9.8）式得到 $p^*=0.49$（百万元）。

定岗定编问题

社会系统中，常常因为职务、地位等的不同，划分出许多的等级，各等级的人数比例称之为等级结构，定岗定编问题即是保持一个稳定合理的等级机构，这类问题在许多单位都可以看到它的缩影。那么等级结构是怎样随时间变化的呢？

等级结构的变化依赖系统内部的等级随时间的转移（即通常所说的职务升降，以及系统内外部的交流（即通常所说的调入、调出、退休、死亡等）。通过数学语言将等级结构随时间变化关系恰当地表示出来，就构成这个问题的数学模型。

模型假设

（1）将一个系统由低向高分成 m 个等级，每隔 s 年进行一次正常的等级调整。

（2）$n_i(k)$ 表示第 k 次调整时第 i 个等级的人数，记 $n(k)=[n_1(k),n_2(k),\cdots,n_m(k)]$，不妨称之为等级结构。$N(k)=\sum_{i=1}^{m} n_i(k)$ 为系统第 k 年的总人数。

（3）记 $a_i(k)=n_i(k)/N(k)$，$a(k)=[a_1(k),a_2(k)\cdots,a_m(k)]$，$a(k)$ 称为等级结构向量。

（4）记 $P_0=[p_{ij}^0]_{k\times k}$，$p_{ij}^0$ 表示每次从等级 i 升到等级 j 的人数占等级 i 中人数的比例；$w=[w_1,w_2,\cdots,w_m]$，w_i 表示每次从等级 i 中退出人数的比例；$r=[r_1,r_2,\cdots,r_m]$，r_i 表示每次调入等级 i 的人数占总调入人数的比例。记 $R(k)$ 为第 k 次调入总人数，且

$$W(k)=\sum_{i=1}^{m} w_i(k)n_i^T(k)。$$

一般地，P_0,w,r 分别称为内部转移矩阵、退出向量、调入向量。为简便起见，不妨假设其与时间无关。

模型建立

根据假设，可以得到 $p_{ij}^0\ge 0,w_i\ge 0,r_i\ge 0$，且

$$\sum_{j=1}^{m} p_{ij}^0+w_i=1,\sum_{i=1}^{m} r_i=1 \qquad (9.9)$$

第 $k+1$ 次的系统总人数满足方程

$$N(k+1)=N(k)+R(k)-W(k) \qquad (9.10)$$

每个等级人数的转移方程为

$$n(k+1)=n(k)P_0+R(k)r \qquad (9.11)$$

从 k 到 $k+1$ 年总人数的增长量记为 $M(k)$，则

$$R(k)=W(k)+M(k)=n(k)w^{T}+M(k) \tag{9.12}$$

将（9.12）代入（ 9.11）得到

$$n(k+1)=n(k)(P_0+w^{T}r)+M(k)r \tag{9.13}$$

记 $P=P_0+w^{T}r$，则 P 也是随机矩阵，（9.13）可以表示为

$$n(k+1)=n(k)P+M(k)r \tag{9.14}$$

通常称（ 9.14）为等级分布基本方程。

假如系统的总人数每年以固定的比例增长，即 $M(k)=\beta N(k)$，则

$$a(k+1)=(1+\beta)^{-1}[a(k)P+\beta r] \tag{9.15}$$

特别地，如果每年进出系统的人数大致相等，即系统总人数 $N(k)$ 保持不变。那么，$M(k)=0$, $\beta=0$，方程（ 9.15）可以简化为

$$a(k+1)=a(k)P \tag{9.16}$$

具有形如（ 9.16）的方程称为马氏链。

对于由（9.14）给出的等级分布基本方程，下面考虑如下问题：给定初始等级结构 $a(0)$，如何确定调入比例 r^*，使等级变化尽快达到或接近给定的理想等级结构 a^*。需要指出的是，如果等级结构 $\bar{a}$ 满足 $\bar{a}=\bar{a}P$，则称等级结构 $\bar{a}$ 为稳定的。

系统是否有稳定的等级结构是有条件的，如果存在，则 r 必须满足（ 9.9），且

$$r=\frac{\bar{a}-\bar{a}P_0}{\bar{a}w^{T}} \tag{9.17}$$

保证（ 9.17）成立的充分必要条件是存在非负向量 a（每个分量非负），使

$$a\geq aP_0 \tag{9.18}$$

如果矩阵 $Q=E-P_0$ 可逆，由（ 9.17）得到

$$\bar{a}=\bar{a}w^{T}rQ^{-1} \tag{9.19}$$

令 $e=(1,1,\cdots,1)$，由于 $\bar{a}$ 的各分量之和为 1，即 $\bar{a}e=1$。利用（ 9.19）式得

$$\bar{a}w^{T}=\frac{1}{r(E-P_0)^{-1}e^{T}} \tag{9.20}$$

再将（ 9.20）式代入（ 9.19）式得到

$$\bar{a}=\frac{r(E-P_0)^{-1}}{r(E-P_0)^{-1}e^{T}} \tag{9.21}$$

关于两个等级接近程度的分析

在处理实际问题时，通常会比较两个等级的接近程度，以便确定当前等级的状态。为此，我们引入等级距离的概念。定义两个等级 $a^{(1)},a^{(2)}$ 之间的距离 $D(a^{(1)},a^{(2)})$ 如下：

$$D(a^{(1)}, a^{(2)}) = \left\| a^{(1)} - a^{(2)} \right\|_2 = \sqrt{\sum_{i=1}^{m} \lambda_i (a_i^{(1)} - a_i^{(2)})^2} \qquad (9.22)$$

其中 λ_i 为加权因子，由对各等级的关注程度确定。一个满意的等级分布应该满足如下优化问题：

$$\begin{cases} \min & D(a, \bar{a}) \\ \text{s.t.} & a = a(0)(P_0 + w^T r) \\ & r_i \geq 0, \sum_{i=1}^{m} r_i = 1 \end{cases} \qquad (9.23)$$

由于

$$\bar{a} - a = \bar{a} - a(0)(P_0 + w^T r) = a(0) w^T \left[\frac{\bar{a} - a(0) P_0}{a(0) w^T} - r \right]$$

如果记

$$Y = (Y_1, Y_2, \cdots, Y_m) = \frac{\bar{a} - a(0) P_0}{a(0) w^T} \qquad (9.24)$$

则 $\bar{a} - a$ 与 $Y - r$ 呈正比，（9.23）等价于

$$\begin{cases} \min & \sum_{i=1}^{m} \lambda_i (Y_i - r_i)^2 \\ \text{s.t.} & r_i \geq 0, \sum_{i=1}^{m} r_i = 1 \end{cases} \qquad (9.25)$$

对于上面的优化问题，在某种程度上，它只是条件极值问题，可以用拉格朗日乘子法求解。

分类问题

人以类聚，物以群分。人们认为某一批样品属于同一类，是因为它们之间有相同或相似之处，从指标上来说就是大小比较接近。由于指标往往不只一个，接近程度的衡量标准不一样，结果会有差异。本节借助一个实际问题介绍两类常用的分类方法。

若已知两类蠓共 32 个标本，已由生物专家根据触角长度 x_1 及重量 x_2 的数据分成 A 类和 B 类，具体数据见表 9.4，根据这 32 个样本的特征对未知的 8 个样本进行分类。

表 **9.4(a)** A 类蠓的触角与重量数据

序号	1	2	3	4	5	6	7	8
x_1	8.70	5.00	10.38	10.86	6.560	13.57	13.57	9.89
x_2	32.94	16.64	37.14	46.24	23.08	38.58	42.54	14.02
序号	9	10	11	12	13	14	15	16
x_1	10.98	10.52	9.44	12.18	8.24	16.55	9.59	10.34
x_2	15.59	35.71	26.00	36.90	38.16	37.12	42.90	36.69

表 **9.4(b)** B 类蠓的触角与重量数据

序号	17	18	19	20	21	22	23	24
x_1	27.14	12.78	19.88	19.05	10.37	21.54	13.66	28.49
x_2	23.04	30.15	23.54	16.13	22.28	13.94	19.88	19.71
序号	25	26	27	28	29	30	31	32
x_1	15.16	23.17	21.31	14.46	6.97	19.64	13.93	23.68
x_2	20.00	18.09	26.57	8.75	22.56	25.37	24.38	23.46

未知的 8 个样本为

表 **9.4(c)** 未知类别蠓的触角与重量数据

序号	33	34	35	36	37	38	39	40
x_1	10.12	12.03	11.70	9.23	20.71	21.88	28.66	17.89
x_2	26.22	27.04	15.24	27.66	23.73	24.79	25.64	14.04

已知的 32 个样本及未知的 8 个样本分布见图 9.5

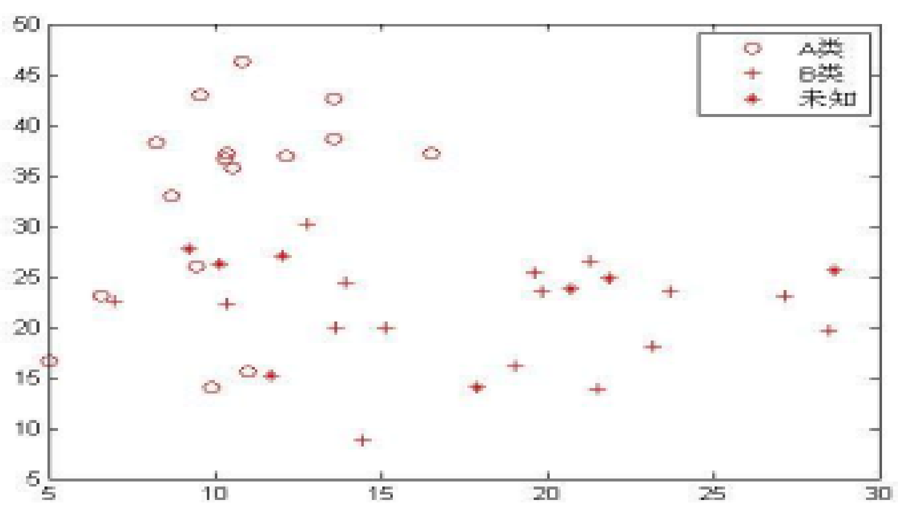

图 **9.5 40** 个样本分布

下面用三种方法来建立模型解决此问题。

模型一　距离判别模型

距离判别法是让指标大小比较接近的属于一类，新样本离谁近就判给谁。当已经给定了一些样本的分类，一般有许多样本属于同一类，以谁作为这类的代表呢？以哪一个样本作代表都不太合适，一个比较恰当的方式是以样本的几何中心为代表。通常的方法是通过计算样本的均值向量及方差。根据样本与均值的接近程度，判断其类型。对于给定的蠓样本，与哪组均值向量越接近，就认为该样本属于此类。考虑 32 个样本得到

A 类的均值向量 $\overline{X_1}=(\overline{x_1},\overline{y_1})^T=(10.40,\ 32.52)^T$，均方差向量为 $\Omega=(2.77,10.22)^T$，B 类的均值向量 $\overline{X_2}=(\overline{x_2},\overline{y_2})^T=(18.21,\ 21.12)^T$，均方差向量为 $\Omega=(6.03,5.18)^T$。

常用的距离判别有欧氏距离与马氏距离判别法。

欧氏距离定义为

$$d_{1i}(X)=d_1(X,\overline{X_i})=\sqrt{(x-\overline{x_i})^2+(y-\overline{y_i})^2},\qquad i=1,2 \quad (9.51)$$

马氏距离定义为

$$d^2_{2i}(X)=d_2(X,\overline{X_i})=(X-\overline{X_i})^T S_i^{-1}(X-\overline{X_i}),\qquad i=1,2 \quad (9.52)$$

其中，$S_i,\ i=1,2$ 表示两类样本的协方差阵。经计算得

$$S_1=\begin{bmatrix}7.66 & 13.33\\ 13.33 & 104.47\end{bmatrix},\qquad S_2=\begin{bmatrix}36.40 & -2.23\\ -2.23 & 26.79\end{bmatrix}$$

欧氏距离判别法可以看作为马氏距离判别法的一个特例，即协方差阵为单位阵。

利用（9.51）式及（9.52）式进行回代，计算 A 类蠓的欧氏距离及马氏距离见表 9.5。

表 **9.5** A 类蠓回代距离统计

序号	1	2	3	4	5	6	7	8
$d^2_{11}(X)$	3.1	281.3	21.3	188.4	103.9	46.8	110.4	342.5
$d^2_{12}(X)$	230.1	194.6	318.0	685.0	139.6	326.4	480.3	119.6
$d^2_{21}(X)$	0.5	4.3	0.3	2.1	2.0	1.3	1.6	3.9
$d^2_{22}(X)$	7.2	5.8	10.7	24.2	3.8	11.7	17.4	4.1
序号	9	10	11	12	13	14	15	16
$d^2_{11}(X)$	286.9	10.19	43.43	22.35	36.48	58.98	108.40	7.39
$d^2_{12}(X)$	82.9	272.0	100.7	285.4	389.8	258.8	548.7	304.4
$d^2_{21}(X)$	4.0	0.1	0.4	0.4	1.7	5.4	1.8	0.2
$d^2_{22}(X)$	2.8	9.1	2.8	9.9	12.9	9.6	19.0	10.2

经检验发现：在欧氏距离意义下序号为 2，8，9 的样本出现误判；在马氏距离意义下序号为 9 的样本出现误判。

利用（9.51）式及（9.52）式进行回代，计算 B 类蠓的欧氏距离及马氏距离见表 9.6。

表 **9.6** B 类蠓回代距离统计

序号	17	18	19	20	21	22	23	24
$d^2_{11}(X)$	370.1	11.3	170.5	343.5	104.9	469.3	170.4	491.3
$d^2_{12}(X)$	83.4	111.0	8.6	25.6	62.8	62.6	22.2	107.7
$d^2_{21}(X)$	54.9	1.3	19.7	21.9	1.3	33.9	5.5	66.8
$d^2_{22}(X)$	2.4	3.6	0.3	0.9	1.7	2.1	0.7	2.9
序号	25	26	27	28	29	30	31	32

$d^2_{11}(X)$	179.4	371.3	154.4	581.5	111.0	136.5	78.7	258.4
$d^2_{12}(X)$	10.6	33.8	39.3	167.1	128.4	20.1	28.9	35.4
$d^2_{21}(X)$	8.3	37.8	23.2	13.8	1.7	17.8	4.1	35.7
$d^2_{22}(X)$	0.3	1.0	1.5	6.3	3.5	0.8	0.8	1.1

经检验发现：在欧氏距离意义下序号为 18，29 的样本出现误判；在马氏距离意义下序号为 18，21，29 样本出现误判。

综上分析发现：对于 32 个已知样本，用欧氏距离判别误判率为 15.6%，用马氏距离判别误判率为 12.5%。易见，距离判别法的误判概率还是比较高的。

将 8 个未知样本数据分别代入（9.51）式及（9.52）式，得到相应的欧氏距离及马氏距离见下表 9.7。

表 **9.7** 未知样本欧氏距离与马氏距离统计

序号	33	34	35	36	37	38	39	40
$d^2_{11}(X)$	39.8	32.7	300.3	25.0	183.6	191.5	380.8	397.6
$d^2_{12}(X)$	91.5	73.2	77.0	123.4	13.1	26.9	129.6	50.2
$d^2_{21}(X)$	0.4	1.2	4.9	0.3	22.7	26.6	61.9	15.3
$d^2_{22}(X)$	2.6	2.2	2.6	3.6	0.5	0.9	4.0	1.9

经比较发现：在欧氏距离意义下序号为 33，34，36 的样本为 A 类，序号为 35，37，38，39，40 的样本为 B 类；在马氏距离意义下到序号为 33，34，36，的样本为 A 类，序号为 35，37，38，39，40 的样本为 B 类。

对于马氏距离判别法，进一步分析可知，如果两类样本的协方差阵相同，则

$$d^2_{21}(X) - d^2_{22}(X) = (X - \bar{X}_1)^T S^{-1}(X - \bar{X}_1) - (X - \bar{X}_2)^T S^{-1}(X - \bar{X}_2)$$

$$= (\bar{X}_2 - \bar{X}_1) S^{-1}\left(X - \frac{\bar{X}_1 + \bar{X}_2}{2}\right)$$

$d_{21}(X) - d_{22}(X) = 0$ 表示平面上的一条直线，它将平面分成两个区域，分别表示两类样本区，位于该直线上的点到两个样本中心距离相等，理论上该直线上的样本点属于无法判别的情况。

如果两类样本的协方差阵不同，则

$$d_{21}(X) - d_{22}(X) = (X - \bar{X}_1)^T S_1^{-1}(X - \bar{X}_1) - (X - \bar{X}_2)^T S_2^{-1}(X - \bar{X}_2)$$

$d^2_{21}(X) - d^2_{22}(X) = 0$ 表示平面上的一条曲线，它将平面分成两个区域，分别表示两类样本区，位于该曲线上的点到两个样本中心距离相等，理论上该曲线上的样本点属于无法判别的情况。

我们在前面的回代中已经发现，马氏距离判别法会出现误判，那么怎样估计误判率呢？为此，不妨简化假设蒙的触角长度与重量服从正态分布，即

$$X_i \sim N(\bar{X}_i, S_i), \quad i = 1, 2$$

记

$$D=\{X \mid X-\frac{\overline{X}_1+\overline{X}_2}{2}\geq 0,\ X=(x_1,x_2)^T\}$$

则 A 类被误判为 B 类的概率为随机变量 X_1 落入区域 D 内的概率，类似地可以计算将 B 类误判为 A 类的概率。

模型二 Fisher 判别法

多变量的判别分析有多个指标，它们对于判别样本属于哪一类一般都有影响，但影响程度一般不完全相同，总会有些指标影响大，有些指标影响程度小，因此按主要差异来进行判别将会有比较好的效果。通常指标间有一定的关联性，因此主要差异不一定是某个指标，而是某些指标的某种线性组合，在这个方向上，样本点最容易区分。Fisher 判别法就是这一思想的某种体现。

借助多元统计中方差的思想，一般可以分类的样本，应该是类与类之间方差很大，而各类之间却靠得狠近，方差较小。可否用一个指标来表征这些特征呢？因为方差总是非负的，故可以用组间方差除以组内方差的商来衡量。当这个商很小时，一定是组间方差小，组内方差大，对应于不易分类的情况，反之，就比较容易分类了。Fisher 判别法就是寻找一个最恰当的方向，使在这个方向上，组间方差与组内方差的商最大。下面我们给出其一般的提法：

若已知 k 个 p 维类别 $X_i, i=1,2,\cdots,k$，它们的均值向量为 $\overline{X}_i$，假如第 i 类有已知样本集 $G_i: x_i^j, j=1,2,\cdots,m_i, i=1,2,\cdots,k$，令

$$S_i=\sum_{j=1}^{m_i}(x_i^j-\overline{X}_i)(x_i^j-\overline{X}_i)^T,\qquad S=\sum_{i=1}^{k}S_i$$

$$\overline{X}=\frac{1}{k}\sum_{i=1}^{k}\overline{X}_i,\qquad Y=\sum_{i=1}^{k}m_i(\overline{X}_i-\overline{X})(\overline{X}_i-\overline{X})^T$$

在给定的方向 $\underline{a}$ 上样本的组间方差为

$$G=\sum_{i=1}^{k}m_i(a^T\overline{X}_i-a^T\overline{X})(a^T\overline{X}_i-a^T\overline{X})^T=a^TYa$$

在给定的方向 $\underline{a}$ 上样本组内方差为

$$E=\sum_{i=1}^{k}\sum_{j=1}^{m_i}(a^Tx_i^j-a^T\overline{X}_i)(a^Tx_i^j-a^T\overline{X}_i)^T=a^TSa$$

记

$$L(a)=\frac{G}{E}=\frac{a^TYa}{a^TSa} \tag{9.53}$$

则所求的方向为使（9.53）式取最大值的方向。

对于本例

$$\overline{X}=\begin{bmatrix}14.3\\26.8\end{bmatrix},\quad S_1=\begin{bmatrix}7.66 & 13.33\\13.33 & 104.47\end{bmatrix},\quad S_2=\begin{bmatrix}36.40 & -2.23\\-2.23 & 26.79\end{bmatrix}$$

$$S=16\begin{bmatrix}2.75 & 0.69\\0.69 & 8.2\end{bmatrix},\quad Y=16\begin{bmatrix}29.73 & -43.95\\-43.95 & 64.98\end{bmatrix}$$

$$L(a)=\frac{(29.73a_1^2-87.9a_1a_2+64.98a_2^2)}{2.75a_1^2+1.38a_1a_2+8.2a_2^2}$$

令 $t=\dfrac{a_1}{a_2}$，记

$$L(t)=\frac{29.73t^2-87.9t+64.98}{2.75t^2+1.38t+8.2} \tag{9.54}$$

借助求导运算得到 $t=-1.94$ 时，$L(t)$ 取得最大值，此时 $a=-1.94a_2$。

最优的方向向量为 $\bar{a}=[0.8889,\ -0.4582]^T$。

记

$$\sigma_i^2=\frac{1}{m_i-1}\bar{a}^T S_i \bar{a},\qquad i=1,2$$

取两类样本的分界点为 $\dfrac{\sigma_2\overline{X_1}+\sigma_1\overline{X_2}}{\sigma_1+\sigma_2}$，其在该最优方向上的投影为

$$\bar{a}\frac{\sigma_2\overline{X_1}+\sigma_1\overline{X_2}}{\sigma_1+\sigma_2}=0.8889\times13.49-0.4582\times27.69=-0.694。$$

经过计算得到 40 个样本点在 $\bar{a}=[0.8889,\ -0.4582]^T$ 上的投影见表 9.8。

表 **9.8 40** 个样本投影数据

序号	1	2	3	4	5	6	7	8
投影	-7.4	-3.2	-7.8	-11.5	- 4.7	-5.6	-7.4	-2.4
序号	9	10	11	12	13	14	15	16
投影	2.6	-7.0	-3.5	-6.1	-10.2	-2.3	-11.1	-7.6
序号	17	18	19	20	21	22	23	24
投影	13.6	-2.5	6.9	9.5	-1.0	12.8	3.0	16.3
序号	25	26	27	28	29	30	31	32
投影	4.3	12.3	6.8	8.4	-4.1	5.8	1.2	10.3
序号	33	34	35	36	37	38	39	40
投影	-3.0	-1.7	3.4	-4.5	7.5	8.1	13.7	9.5

经比较发现：对于已知的 32 个样本，序号 8，9 误判为 B 类；序号为 18，21，29 误判为 A 类。误判率为 15.6%。用此判别法对未知的 8 个样本分类结果为序号为 33，34，36 的样本为 A 类，其他的为 B 类。

模型三　Bayes 判别法

从前面介绍的两种方法中我们可以看出都存在不同程度的误判。Bayes 判别法的思想是：分类判断需要考虑因错判带来的损失。下面介绍仅考虑两个总体的 Bayes 判别法。

假设考虑的两个总体 X_1, X_2 分别具有密度函数 $f_1(x)$，$f_2(x)$，记 Ω 为样本空间，R_1 表示要判给 X_1 的全体，$R_2=\Omega-R_1$ 为判给 X_2 的全体。记总体 X_1, X_2 的先验概率为 p_1, p_2，通常根据假设或先验统计得到。据此得到

一个来自总体 X_1 的样本被判给 X_2 的条件概率为

$$p(1\mid 2)=\int_{R_2} f_1(x)dx$$

一个来自总体 X_2 的样本被判给 X_1 的条件概率为

$$p(2\mid 1)=\int_{R_1} f_2(x)dx$$

且

$$p(\text{正确判入}X_1) = p(1|1)p_1, \quad p(\text{错误判入}X_1) = p(1|2)p_1$$

$$p(\text{正确判入}X_2) = p(2|2)p_2, \quad p(\text{错误判入}X_1) = p(2|1)p_2$$

以 $C(i|j)$ 表示来自 X_i 的样本被判给 X_j 所造成的损失。对任一判别法则，平均误判损失为

$$L(R_1) = C(2|1)p(2|1)p_2 + C(1|2)p(1|2)p_1$$

$$= C(2|1)p_2\int_{R_1} f_2(x)dx + C(1|2)p_1\int_{R_2} f_1(x)dx$$

$$= C(2|1)p_2\int_{R_1} f_2(x)dx + C(1|2)p_1\left(1 - \int_{R_1} f_1(x)dx\right)$$

$$= C(1|2)p_1 + \int_{R_1}[C(2|1)p_2 f_2(x) - C(1|2)p_1 f_1(x)]dx \qquad (9.55)$$

一个合理的判别法应使误判损失最小。

记

$$R_1 = \left\{x \mid \frac{f_1(x)}{f_2(x)} \geq \frac{C(2|1)p_2}{C(1|2)p_1}\right\}$$

在 R_1 上，（9.55）式第二项中的被积函数是负的，因此（9.55）式取得极小值。