

第 11 章 多维标度法

§ 11.1 引言

在实际中往往会碰到这样的问题：有 n 个由多个指标（变量）反映的客体，但反映客体的指标个数是多少不清楚，甚至指标本身是什么也是模糊的，更谈不上直接测量或观察它，仅仅所能知道的是这 n 个客体之间的某种距离（不一定是通常的欧氏距离）或者某种相似性，我们希望仅由这种距离或者相似性给出的信息出发，在较低维的欧氏空间把这 n 个客体（作为几何点）的图形描绘出来，从而尽可能揭示这 n 个客体之间的真实结构关系。这就是多维标度法所要研究的问题，例如有 100 种酒，品尝家可以对每两种酒进行品尝对比，给出一种相近程度的得分（越相近得分越高，相差越远得分越低），我们希望用这些得分数据来了解这 100 种酒之间的结构关系，这样的问题就可以用多维标度法来解决。

一个经典的例子是利用城市之间的距离来绘制地图。

表 11.1.1 列出了通过测量得到的英国 12 个城市之间公路长度的数据。由于公路不是平直的，所以它们还不是城市之间的最短距离，只可以看作是这些城市之间的近似距离（或者说带有误差的距离），我们希望利用这些距离数据画一张平面地图，标出这 12 个城市的位置，使之尽量接近表中所给出的距离数据，从而反映它们的真实地理位置。

表 11.1.1 英国 12 城市之间的公路距离(英里)

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2	244											
3	218	350										
4	284	77	389									
5	197	184	347	242								
6	312	444	94	463	441							
7	215	221	150	236	279	245						
8	469	583	251	598	598	169	380					
9	166	242	116	257	269	210	55	349				
10	212	53	298	72	170	392	188	531	190			
11	253	325	57	340	359	143	117	264	91	273		
12	270	168	284	164	277	378	143	514	173	111	256	

1 = 阿伯瑞斯吹, 2 = 布莱顿, 3 = 卡里斯尔, 4 = 多佛, 5 = 爱塞特,
6 = 格拉斯哥, 7 = 赫尔, 8 = 印威内斯, 9 = 里瑟, 10 = 伦敦, 11 = 纽加塞耳,
12 = 诺利其。

§ 11.2 距离阵和经典解

11.2.1 距离阵

我们这里研究的距离不限于通常的欧氏距离，首先对距离的意义加以拓广，给出如下的距离阵定义。

定义 11.2.1 一个 $n \times n$ 阶矩阵 $D = (d_{ij})$ ，如果满足条件：

(1) $D' = D$,

(2) $d_{ij} \geq 0$, $d_{ii} = 0$, $i, j = 1, \dots, n$,

则称 D 为距离阵， d_{ij} 称为第 i 个点与第 j 个点间的距离。

表 11.1.1 就是一个距离阵。

由上节所述，有了一个距离阵 $D = (d_{ij})$ ，多维标度法的目的就是要确定数 k 并且在 k 维空间 R^k 中求 n 个点 x_1, \dots, x_n ，使得这

n 个点的欧氏距离与距离阵中的相应值在某种意义下尽量接近。即如果用 $\widehat{D}=(\widehat{d}_{ij})$ 记求得的 n 个点的距离, 则要求在某种意义下, \widehat{D} 和 D 尽量接近。在实际中, 为了使求得的结果易于解释, 通常取 $k=1, 2, 3$ 。

下面我们给出多维标度法解的概念和有关性质。

设求得的 n 个点为 x_1, \dots, x_n , 写成矩阵形式

$$X=(x_1, \dots, x_n)',$$

则称 X 为 D 的一个解。在多维标度法中, 形象地称 X 为距离阵 D 的一个拟合构图, 由这 n 个点之间的欧氏距离构成的距离阵称为 D 的拟合距离阵。所谓拟合构图, 其意义是有了这 n 个点的坐标, 可以在 R^k 中画出图来, 使得它们的距离阵 \widehat{D} 和原始的 n 个客体的距离阵 D 接近, 给出原始 n 个客体关系一个有意义的解释。特别地, 如果 $\widehat{D}=D$, 则称 X 为 D 的一个构图。

读者可能已经发现, 由于求解的 n 个点仅仅要求它们的相对欧氏距离和 D 接近, 即只要求它们的相对位置确定而与它们在 R^k 中的绝对位置无关, 所以所求得的解不是唯一的, 根据欧氏距离在正交变换和平移变换下的不变性, 显然有以下结果:

引理11.2.1 如果 $X=(x_1, \dots, x_n)'$ 是 D 的一个多维标度解, 则

$$Y=X\Gamma+1a, \quad (11.2.1)$$

也是 D 的一个解, 而且 Y 与 X 有相同的拟合距离阵, 其中 Γ 为任一正交阵, a 为任一常数向量。

注意, 根据以上距离阵的定义, 并不是任何距离阵 D , 都真实地存在一个欧氏空间 R^k 和其中的 n 个点, 使得 n 个点之间的距离阵等于 D 。于是, 一个距离阵并不一定都有通常距离的含义。为了把有通常意义和没有通常意义的距离阵区别开来, 我们引进欧氏型距离阵和非欧型距离阵的概念, 并给出一个判定定理。

11.2.2 欧氏距离阵及其判定定理

定义11.2.2 对于一个 $n \times n$ 距离阵 $D=(d_{ij})$, 如果存在某个正整数 p 和 R^p 中的 n 个点 x_1, \dots, x_n , 使得

$$d_{ij}^2=(x_i-x_j)'(x_i-x_j), \quad i, j=1, \dots, n, \quad (11.2.2)$$

则称 D 为欧氏的。

为了给出后面的判定定理, 先引进几个记号并证明几个引理,

令

$$\left. \begin{aligned} A &=(a_{ij}), \quad \text{其中 } a_{ij}=-\frac{1}{2}d_{ij}^2, \\ B &=HAH, \quad \text{其中 } H=I-\frac{1}{n}11'. \end{aligned} \right\} \quad (11.2.3)$$

有

引理11.2.2 设 $X=(x_1, \dots, x_n)'$ 为任意 $n \times p$ 数据阵, 则 X , HX , HXG 有相同的距离阵, 其中 G 为正交阵。

证 注意到 $HX=X-1\bar{x}'$, $HXG=HG-1(\bar{x}'G)$, 利用引理 11.2.1 即得证。

引理11.2.3 设 D 是 X 的 (内点) 距离阵, 则

$$B=HAH=(HX)(HX)' \geq 0$$

证 首先证明一般距离阵的一个公式。设 D 为一个一般距离阵, 由于

$$a_{ij}=-\frac{1}{2}d_{ij}^2, \quad (11.2.4)$$

$$\text{则有 } B=HAH=A-\frac{1}{n}A11'-\frac{1}{n}11'A+\frac{1}{n^2}11'A11' \quad (11.2.5)$$

$$\left. \begin{aligned} \bar{a}_{i.} &= \frac{1}{n} \sum_{j=1}^n a_{ij}, \quad i=1, \dots, n, \\ \bar{a}_{.j} &= \frac{1}{n} \sum_{i=1}^n a_{ij}, \quad j=1, \dots, n, \\ \bar{a}_{..} &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n a_{ij}. \end{aligned} \right\} \quad (11.2.6)$$

$$\text{则有 } \frac{1}{n} \mathbf{A} \mathbf{1} = \begin{pmatrix} \bar{a}_{1.} \\ \vdots \\ \bar{a}_{n.} \end{pmatrix}, \quad \frac{1}{n} \mathbf{1}' \mathbf{A} = (\bar{a}_{.1}, \dots, \bar{a}_{.n}),$$

$$\frac{1}{n^2} \mathbf{1}' \mathbf{A} \mathbf{1} = \bar{a}_{..}$$

$$\text{和 } \frac{1}{n} \mathbf{A} \mathbf{1} \mathbf{1}' = \begin{pmatrix} \bar{a}_{1.} & \dots & \bar{a}_{1.} \\ \dots & \dots & \dots \\ \bar{a}_{n.} & \dots & \bar{a}_{n.} \end{pmatrix}, \quad \frac{1}{n} \mathbf{1} \mathbf{1}' \mathbf{A} = \begin{pmatrix} \bar{a}_{.1} & \dots & \bar{a}_{.n} \\ \dots & \dots & \dots \\ \bar{a}_{.1} & \dots & \bar{a}_{.n} \end{pmatrix},$$

$$\frac{1}{n^2} \mathbf{1} \mathbf{1}' \mathbf{A} \mathbf{1} \mathbf{1}' = \begin{pmatrix} \bar{a}_{..} & \dots & \bar{a}_{..} \\ \dots & \dots & \dots \\ \bar{a}_{..} & \dots & \bar{a}_{..} \end{pmatrix}.$$

从而由式(11.2.5)有

$$b_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, \quad i, j=1, \dots, n, \quad (11.2.7)$$

这就是一般距离阵D的一个公式。

现设D为X的距离阵, 则有

$$a_{ij} = -\frac{1}{2} d_{ij}^2 = -\frac{1}{2} (\mathbf{x}_i - \mathbf{x}_j)' (\mathbf{x}_i - \mathbf{x}_j). \quad (11.2.8)$$

利用式(11.2.8)和(11.2.6)算出 a_{ij} , $\bar{a}_{i.}$, $\bar{a}_{.j}$, $\bar{a}_{..}$ 关于 \mathbf{x}_i , $i=1, \dots, n$ 的表达式。将结果代入式(11.2.7)得到

$$b_{ij} = (\mathbf{x}_i - \bar{\mathbf{x}})' (\mathbf{x}_j - \bar{\mathbf{x}}), \quad i, j=1, \dots, n, \quad (11.2.9)$$

$$\text{其中 } \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i.$$

将式(11.2.8)用矩阵表示并根据非负定矩阵的性质, 有

$$\mathbf{B} = (\mathbf{H}\mathbf{X})(\mathbf{H}\mathbf{X})' \geq 0, \quad (11.2.10)$$

式中 $\mathbf{H}\mathbf{X} = (\mathbf{x}_1 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}})'$ 。我们把B叫做X的中心化内积阵, 引理证毕。

引理11.2.4 设D为 $n \times n$ 距离阵。如果D满足

$$\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H} \geq 0 \quad (11.2.11)$$

则存在正整数p和一个 $n \times p$ 矩阵 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)'$, 使得D为X的内点距离阵。

证 我们用p记B的秩。由于 $\mathbf{B} \geq 0$, 则由谱分解定理

$$\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H} = \Gamma_1 \Lambda_1 \Gamma_1', \quad (11.2.12)$$

其中 $\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_p)$, $\lambda_1 \geq \dots \geq \lambda_p$ 为B的p个正特征值。 Γ_1 的p个列为对应的p个正交特征向量。取

$$\mathbf{X} = \Gamma_1 \Lambda_1^{1/2}, \quad (11.2.13)$$

这是一个 $n \times p$ 阶矩阵。把X写成

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)' = (\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(p)}),$$

于是有

$$\mathbf{X}' \mathbf{X} = (\Gamma_1 \Lambda_1^{1/2})' (\Gamma_1 \Lambda_1^{1/2}) = \Lambda_1 \quad (11.2.14)$$

$$\text{和 } \mathbf{B} = \mathbf{X}\mathbf{X}' \quad (11.2.15)$$

则 $b_{ij} = \mathbf{x}_i' \mathbf{x}_j$, $i, j=1, \dots, n$ 。

由此求得 \mathbf{x}_i 与 \mathbf{x}_j 两点的距离平方

$$\begin{aligned} (\mathbf{x}_i - \mathbf{x}_j)' (\mathbf{x}_i - \mathbf{x}_j) &= \mathbf{x}_i' \mathbf{x}_i - 2\mathbf{x}_i' \mathbf{x}_j + \mathbf{x}_j' \mathbf{x}_j \\ &= b_{ii} - 2b_{ij} + b_{jj} = a_{ii} - 2a_{ij} + a_{jj} \\ &= -2a_{ij} = d_{ij}^2. \end{aligned} \quad (11.2.16)$$

在上式计算中用到式(11.2.7)和等式 $a_{ij} = a_{ji}$, $\bar{a}_{.i} = \bar{a}_{.i}$, $a_{ii} = -\frac{1}{2} d_{ii}^2 = 0$ 以及 $a_{ij} = -\frac{1}{2} d_{ij}^2$ 。式(11.2.16)意味着式(11.2.13)给出的X以D为它的距离阵, 或者说X是D的一个构图。引理证

毕。

由以上引理不难得到下面的判定定理。

定理 11.2.5 一个 $n \times n$ 距离阵 D 是欧氏的充要条件是 $B = HAH \geq 0$ 。

证 设 D 是欧氏的, 则由定义 11.2.2, 存在一个 $n \times p$ 矩阵 X , 使得 D 是 X 的距离阵。由引理 11.2.3, 有 $B = HAH \geq 0$ 。反之, 如果 $B = HAH \geq 0$, 则由引理 11.2.4, 存在 $n \times p$ 矩阵 X , 而且可取 $X = \Gamma_1 \Lambda_1^{1/2}$ (其中 $\Gamma_1 \Lambda_1$ 由 B 的谱分解 $B = \Gamma_1 \Lambda_1 \Gamma_1'$ 得到), 使得 D 为 X 的距离阵, 从而由定义 11.2.2, D 是欧氏的, 定理得证。

下面给出一个距离阵不是欧氏型的例子。

例 设

$$D = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 4 \\ 2 & 4 & 0 \end{pmatrix}.$$

显然 D 是一个距离阵。由式 (11.2.3), 此时

$$A = \begin{pmatrix} 0 & -\frac{1}{2} & -1 \\ -\frac{1}{2} & 0 & -2 \\ -1 & -2 & 0 \end{pmatrix}.$$

不难算得

$$B = HAH = \frac{1}{18} \begin{pmatrix} 4 & 1 & -5 \\ 1 & 16 & -17 \\ -5 & -17 & 22 \end{pmatrix}.$$

B 的一、二、三阶主子式分别是 $\frac{4}{18}$, $\frac{64}{18^2}$, $-\frac{360}{18^3}$, 故 B 不是非负定的, 因此 D 不是欧氏距离阵。事实上, 在欧氏空间中的三角形必须满足两边之和不小于第三边, 但由 D 给出的距离阵要求在欧氏

空间中找一边长为 4 另外两边长为 1 和 2 的三角形, 这显然是不可能的。

11.2.3 多维标度的经典解

当距离阵 D 是欧氏的时候, 引理 11.2.4 和定理 11.2.5 实际上给出了求 D 的构图 X 的方法。当 D 不是欧氏的情形, 由上面的定理, 显然不存在 D 的构图, 这时只能寻求 D 的拟合构图。为了和构图 X 加以区别, 拟合构图用 \hat{X} 来记。在实际中, 即使 D 是欧氏的从而存在构图, 但如果构图点 x_i 的维数太高, 仍然不实用, 因为失去了直观意义而不便解释。这时一般不求构图 X 而求低维的拟合构图 \hat{X} 。所以在上述两种情形下都需要寻求拟合构图。

下面介绍的多维标度的经典解, 就是具有某种最优性质 (在 § 11.3 中介绍并且证明) 的 D 的拟合构图或者构图。

在介绍求解方法之前, 我们将上面的引理和定理所得的结果稍加剖析会得到一些启发。

1. 上面定理的证明告诉我们, 当 D 是欧氏距离阵时, $X = \Gamma_1 \Lambda_1^{1/2}$ 就是 D 的一个构图。将 X 用 B 的特征值与特征向量具体写出来会更清楚:

$$\begin{array}{l} \text{构} \\ \text{图} \\ \text{点} \\ \text{向量记号} \end{array} \left\{ \begin{array}{l} P_1: x_{11} \quad x_{12} \cdots x_{1p} \\ P_2: x_{21} \quad x_{22} \cdots x_{2p} \\ \vdots \\ P_n: x_{n1} \quad x_{n2} \cdots x_{np} \end{array} \right. \begin{array}{l} \text{特征值} \\ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \\ \text{向量记号} \\ x'_1 \\ x'_2 \\ \vdots \\ x'_p \end{array} \quad (11.2.17)$$

即构图 X 不是别的, 正是由 B 对应于它的正特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ 的特征向量 $x_{(1)} = \sqrt{\lambda_1} Y_{(1)}$, \cdots , $x_{(p)} = \sqrt{\lambda_p} Y_{(p)}$ 所排成的 $n \times p$ 矩阵。而以此个矩阵的行为坐标的 n 个点 P_1, \cdots, P_n 就是这个构图的点。

由式(11.2.12), 又因为 $H1=0$, 故有

$$1'B1=1'HAH1=1'\Gamma_1\Lambda_1\Gamma_1'1=(1'X)(1'X)'=0.$$

从而 $1'X=0$, 并且

$$\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i=\frac{1}{n}X'1=0, \quad (11.2.18)$$

即构图 X 的 n 个点的重心在原点。

另外, 由式(11.2.14), 这 p 个特征向量相互正交, 而且它们的长度平方为

$$x_{(i)}'x_{(i)}=\lambda_i, \quad i=1, \dots, p \quad (11.2.19)$$

2. 由引理11.2.3, 当 D 是欧氏距离阵时, 对于 D 的任一构图 X , B 总是 X 的中心化内积阵, 或者说, D 的任一构图有相同的中心化内积阵。其中 b_{ii} 等于构图点 x_i 到构图重心 \bar{x} 的距离, 而 $b_{ij}/(b_{ii}b_{jj})^{1/2}$ 等于向量 $x_i - \bar{x}$ 与 $x_j - \bar{x}$ 的夹角余弦。

3. 不管 D 是不是欧氏的, B 总有一个 0 特征值和对应于这个特征值的特征向量 1 。显然, 如果 X 为 D 的构图, 则 $(X, 1)$ 仍然是 D 的构图, 即添加的列 1 不起作用。由此得到一点启发, 在构图 X 中, 方差越小的列对构图的作用会越小, 而方差越大的列对构图的作用会越大。

下面我们给出求经典解的步骤:

第 1 步 由距离阵 D 构造矩阵 $A=(a_{ij})=(-\frac{1}{2}d_{ij}^2)$;

第 2 步 作出矩阵 $B=(b_{ij})$, 其中

$$b_{ij}=a_{ij}-\bar{a}_{i.}-\bar{a}_{.j}+\bar{a}_{..};$$

第 3 步 求出 B 的 k 个最大特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$, 和对应的正交特征向量 $x_{(1)}, \dots, x_{(k)}$, 并且满足规格化条件: $x_{(i)}'x_{(i)}=\lambda_i$, $i=1, \dots, k$ 。

注意, 这里关于 k 的选取有两种方法: 一种是事先指定, 例如 $k=1, 2$ 或 3 ; 另一种是考虑前 k 个特征值在全体特征值中所占的比例, 这时需将所有特征值 $\lambda_1 \geq \dots \geq \lambda_n$ 求出。如果 λ_i 都非负, 说明 $B \geq 0$ 从而 D 为欧氏的, 则依据

$$\varphi = \frac{\lambda_1 + \dots + \lambda_k}{\lambda_1 + \dots + \lambda_n} \geq \varphi_0 \quad (11.2.20a)$$

来确定上式成立的最小 k 值, 其中 φ_0 为预先给定的百分数 (即变差贡献比例)。如果 λ_i 中有负值, 表明 D 是非欧氏的, 这时以

$$\varphi = \frac{\lambda_1 + \dots + \lambda_k}{|\lambda_1| + \dots + |\lambda_n|} \geq \varphi_0 \quad (11.2.20b)$$

求出最小的 k 值, 但必须要求 $\lambda_1 \geq \dots \geq \lambda_k > 0$, 否则必须减少 φ_0 的值以减少个数 k 。

第 4 步 将所求得特征向量顺序排成一个 $n \times k$ 矩阵, $\hat{X}=(x_{(1)}, \dots, x_{(k)})$, 则 \hat{X} 就是 D 的一个拟合构图, 或者说 \hat{X} 的行坐标 $x_i=(x_{i1}, \dots, x_{ik})$, $i=1, \dots, n$ 对应的点 P_1, \dots, P_n 是 D 的拟合构图点。我们把这一 k 维拟合构图叫做经典解 k 维拟合构图 (简称经典解)。

为了更好地理解上述求解步骤, 我们来研究两个例子。

例 11.2.1 设 7×7 阶距离阵

$$D = \begin{pmatrix} 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 & 1 \\ & 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 \\ & & 0 & 1 & \sqrt{3} & 2 & 1 \\ & & & 0 & 1 & \sqrt{3} & 1 \\ & & & & 0 & 1 & 1 \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{pmatrix}.$$

根据式(11.2.3)求出矩阵 A , 而且求出

$$\bar{a}_{1.} = \dots = \bar{a}_{6.} = -\frac{13}{14}, \quad \bar{a}_{7.} = -\frac{3}{7},$$

$$\bar{a}_{.i} = \bar{a}_{i.}, \quad i=1, \dots, 7, \quad \bar{a}_{..} = -\frac{6}{7}.$$

再由公式 $b_{ij}=a_{ij}-\bar{a}_{i.}-\bar{a}_{.j}+\bar{a}_{..}$, 求得矩阵 B :

$$B = \frac{1}{2} \begin{pmatrix} 2 & 1 & -1 & -2 & -1 & 1 & 0 \\ 1 & 2 & 1 & -1 & -2 & -1 & 0 \\ -1 & 1 & 2 & 1 & -1 & -2 & 0 \\ -2 & -1 & 1 & 2 & 1 & -1 & 0 \\ -1 & -2 & -1 & 1 & 2 & 1 & 0 \\ 1 & -1 & -2 & -1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

容易看出

$$\begin{aligned} b_{(3)} &= b_{(2)} - b_{(1)}, & b_{(4)} &= -b_{(1)}, & b_{(5)} &= -b_{(2)} \\ b_{(6)} &= b_{(1)} - b_{(2)}, & b_{(7)} &= 0, & b_{(1)} &\text{与 } b_{(2)} \text{ 不相关,} \end{aligned}$$

故B的秩为2. B的特征值容易求得

$$\lambda_1 = 3, \lambda_2 = 3, \lambda_3 = \dots = \lambda_7 = 0.$$

于是D是欧氏的, 而且它的构图可以精确地在 $k=2$ 维空间(即平面)中求出来.

在对于 $\lambda_1 = \lambda_2$ 的二维特征子空间中取一对正交的特征向量

$$\begin{aligned} x'_{(1)} &= (a, a, 0, -a, -a, 0, 0) \\ x'_{(2)} &= (b, -b, -2b, -b, b, 2b, 0) \end{aligned}$$

其中 $a = \frac{\sqrt{3}}{2}$, $b = \frac{1}{2}$, 经验证有 $x'_{(1)}x_{(1)} = x'_{(2)}x_{(2)} = 3$. 故 $X = (x_{(1)}, x_{(2)})$ 就是D的构图. 7个构图点的坐标依次为

$$A = \left(\frac{\sqrt{3}}{2}, \frac{1}{2} \right), \quad B = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2} \right), \quad C = (0, -1),$$

$$D = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2} \right), \quad E = \left(-\frac{\sqrt{3}}{2}, \frac{1}{2} \right),$$

$$F = (0, 1), \quad G = (0, 0),$$

这就是所要求的D的经典解.

不难验证, 这7个点的重心为 $(0, 0)$, 而且这7个点之间的

距离阵等于D. 将这7个点在平面上画出来, 正好A, B, C, D, E, F是边长为1的正六角形的顶点, 重心在G点.

例11.2.3 设有一个场地, 分为七个区域, 如图11.2.1所示. 如果以从一个区域到另一个区域穿过边界线的条数作两个区域的距离, 不难得到这七个区域的距离阵为

$$D = \begin{pmatrix} 0 & 1 & 2 & 2 & 2 & 1 & 1 \\ & 0 & 1 & 2 & 2 & 2 & 1 \\ & & 0 & 1 & 2 & 2 & 1 \\ & & & 0 & 1 & 2 & 1 \\ & & & & 0 & 1 & 1 \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{pmatrix}.$$

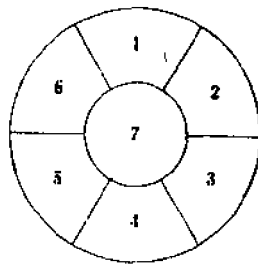


图11.2.1

经计算, $B = HAH$ 的特征值为

$$\begin{aligned} \lambda_1 = \lambda_2 &= \frac{7}{2}, & \lambda_3 = \lambda_4 &= \frac{1}{2}, & \lambda_5 &= 0, & \lambda_6 &= -\frac{1}{7}, \\ \lambda_7 &= -1. \end{aligned}$$

可见这个距离阵D虽然满足三角不等式 $d_{ik} \leq d_{ij} + d_{jk}$, 但仍为非欧氏的.

选择 $k=2$, 对 $\lambda_1 = \lambda_2$ 两个特征值选取对应的两个正交的特征向量 $x_{(1)}, x_{(2)}$, 且满足 $x'_{(1)}x_{(1)} = x'_{(2)}x_{(2)} = \frac{7}{2}$, 利用得到的7个二维点在平面上作图, 把原始图重新构造出来. 恰好其中一点为原点, 其余6个点是正六角形的顶点. 本例的一些中间细节留给读者作为练习.

11.2.4 相似阵情形

有时已知的不是 n 个客体之间的某种距离, 而是已知 n 个客体之间的某种相似性, 即已知的是一个相似矩阵. 关于相似系数的概

念在聚类分析一章介绍过,本节还给出相似系数阵的定义。

定义11.2.3 一个 $n \times n$ 矩阵 $C=(c_{ij})$ 如果满足下面两个条件则称之为相似系数阵(简称相似阵):

$$\left. \begin{array}{l} \textcircled{1} C' = C; \\ \textcircled{2} c_{ii} \leq c_{ij}, i, j = 1, \dots, n. \end{array} \right\} \quad (11.2.21)$$

显然, 对于一个 $n \times p$ 矩阵 X , 如果取 $c_{ij} = x_i' x_j / (\|x_i\| \cdot \|x_j\|)$, 则 $C=(c_{ij})$ 是一个相似阵。

为了应用上面求经典解的方法, 我们把相似阵变换为距离阵。下面定义一种变换。

定义11.2.4 设 C 为一个相似阵, 令

$$d_{ij} = (c_{ij} - 2c_{i.} + c_{.j})^{1/2}, \quad (11.2.22)$$

得到一个距离阵 $D=(d_{ij})$, 我们称变换式(11.2.22)为从相似阵 C 到距离阵 D 的标准变换。

容易验证当 C 满足式(11.2.21)时有 $c_{ii} - 2c_{ij} + c_{jj} \geq 0$, 从而 $d_{ij} \geq 0$, $d_{ii} = 0$ 而且 $d_{ij} = d_{ji}$ 。故通过标准变换得到的 D 确实为一个距离阵。

我们常遇到的相似阵还具有非负定性。根据下面的定理可知, 具有非负定性的相似阵具有特殊的性质。

定理11.2.6 如果 $C \geq 0$, 则由变换式(11.2.22)确定一个欧氏距离阵 D , 而且 D 的中心化内积阵为 $B = HCH$ 。

证 首先注意, 定理中的条件只说 C 为非负定阵, 并未指出 C 是一个相似阵, 所以还得证明由变换式(11.2.22)决定的 D 是距离阵。

由于 $C \geq 0$, 故

$$d_{ii}^2 = c_{ii} - 2c_{ij} + c_{jj} = x' C x \geq 0$$

其中 x 是第 i 个分量取值 1、第 j 个分量取值 -1、其余分量取值 0 的向量 ($i \neq j$)。又显然 $d_{ii}^2 = c_{ii} - 2c_{ii} + c_{ii} = 0$, 因而 D 有定义而且为一距离阵。

设 A, B 由式(11.2.3)定义, 由 $a_{ij} = -\frac{1}{2}d_{ij}^2$ 以及式(11.2.7)和(11.2.6), 我们有

$$\begin{aligned} -2b_{ij} &= d_{ij}^2 - \frac{1}{n} \sum_{i=1}^n d_{i.}^2 - \frac{1}{n} \sum_{j=1}^n d_{.j}^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2 \\ &= c_{ii} - 2c_{ij} + c_{jj} - \frac{1}{n} \sum_{i=1}^n (c_{ii} - 2c_{ij} + c_{jj}) \\ &\quad - \frac{1}{n} \sum_{j=1}^n (c_{ii} - 2c_{ij} + c_{jj}) \\ &\quad + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (c_{ii} - 2c_{ij} + c_{jj}) \\ &= -2c_{ij} + 2\bar{c}_{i.} + 2\bar{c}_{.j} - 2\bar{c}_{..}, \end{aligned}$$

其中 $\bar{c}_{i.}, \bar{c}_{.j}, \bar{c}_{..}$ 与式(11.2.6)的定义类似。从而

$$b_{ij} = c_{ij} - \bar{c}_{i.} - \bar{c}_{.j} + \bar{c}_{..} \quad (11.2.23)$$

根据式(11.2.5)与(11.2.7)的等价性, 同样有式(11.2.23)等价于

$$B = HCH. \quad (11.2.24)$$

由于 $C \geq 0$, 故 $B = HCH \geq 0$ 。

又由定理11.2.5, D 是欧氏距离阵, 而且 D 的中心化内积阵就是 HCH , 从而定理得证。

注1 由以上证明过程可知, 式(11.2.24)只要对用变换式(11.2.22)能得到距离阵 D 的 C 都成立, 不必要求满足 $C \geq 0$ 。因此式(11.2.24)对于任意相似阵 C 都成立。根据这一结果, 当 C 是相似阵时, 不必通过 C 定义 D 求出 A , 然后通过 $B = HAH$ 求出 B , 而可以直接通过式(11.2.24)求出 B 。

注2 按照以上相似阵的定义, 一般非负定阵未必是相似阵, 即形为 $X'X$ 的矩阵未必是相似阵, 其中 X 是 $n \times p$ 阶数据阵。然而

这种矩阵是反映变量(或样本)的某种相似,所以有时在应用中还有必要将相似阵的定义加以推广,例如把能通过变换式(11.2.22)化为距离阵的对称阵C称为相似阵。

我们给出一个用相似阵求经典解的例子。

例11.2.4 表11.2.2是对598个电报收报员的收听调查数据,其中的数据是,连续发两个数字的莫尔斯(Morse)码信号被收听后认为是相同的人数百分比(%).这个矩阵可以用来描述两两数字的莫尔斯码子之间的某种相似性,它就是一个相似阵,记为 $C=(c_{ij})$ 。

表 11.2.2 对598个电报收报员收听信号调查表(%)

	1	2	3	4	5	6	7	8	9	0
1	84									
2	62	89								
3	16	59	86							
4	6	23	38	83						
5	12	8	27	58	90					
6	12	14	33	34	30	86				
7	20	25	17	24	18	65	85			
8	37	25	16	13	10	22	85	88		
9	57	28	9	7	5	8	31	58	91	
0	52	18	9	7	5	18	15	39	79	94

根据标准变换式(11.2.22),令

$$d_{ij} = (c_{ii} - 2c_{ij} + c_{jj})^{1/2},$$

得到距离阵D。由定理11.2.6,求得 $B=H A H=H C H$ 和它的特征值与特征向量。具体结果如下:

$$\lambda_1=187.4, \lambda_2=121.0, \lambda_3=95.4, \lambda_4=55.4, \lambda_5=46.6,$$

$$\lambda_6=31.5, \lambda_7=9.6, \lambda_8=4.5, \lambda_9=0.0, \lambda_{10}=-4.1.$$

可见这个距离阵不是欧氏的。我们从解释距离阵累积变差贡献百分

比 $(\lambda_1 + \lambda_2) / \sum_{i=1}^{10} |\lambda_i| = 56\%$ 选取两个特征值,即取 $k=2$ 来求

D的经典解。对应于 λ_1 和 λ_2 的两个满足规格化条件 $x_{(1)}' x_{(1)} = 187.4, x_{(2)}' x_{(2)} = 121.0$ 的特征向量为:

$$x_{(1)}' = (-4.2, -0.3, 3.7, 5.6, 5.4, 3.8, 0.9, -3.0, -6.2, -5.7),$$

$$x_{(2)}' = (-3.2, -5.8, -4.3, -0.6, 0.0, 4.0, 5.5, 3.6, 0.6, 0.2).$$

把 $P_1=(-4.2, -3.2), P_2=(-0.3, -5.8), \dots, P_9=(-6.2, 0.6), P_{10}=(-5.7, 0.2)$ 十个点画在图11.2.2中,在每个点即数字旁边标有莫尔斯(Morse)码子的信号。由图示结果容易看出 x_1 轴的正向反映“点”信号增加的方向,而 x_2 轴的方向是反映“点”“横”变化情况的。如果我们把 OP_i 看作一个向量,则这些数字信号向量的夹角顺次大约为 45° 。但0与9,4与5除外,这两对数的码子信号离得近,可能是由于仅仅最后一个信号的改变使得对那些缺乏训练收报员的耳朵反应不过来而造成的。这样一张

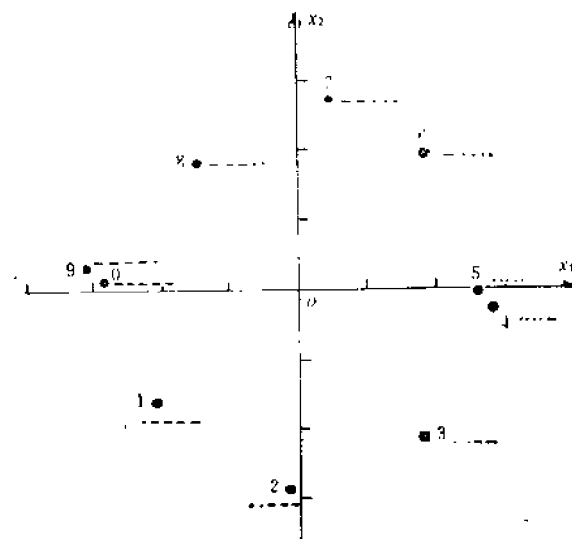


图 11.2.2 莫尔斯(Morse)码子信号数据的经典解

图就大体上反映了这些数字莫尔斯码信号的基本相互关系, 它们大体上以 1, 2, ..., 9, 0 的顺序处在一个圆上。

§ 11.3 经典解的优良性质

11.3.1 经典解和主成分分析的关系

首先回顾一下主成分分析的有关结果。

设 $X(n \times p)$ 是由 R^p 中的 n 个点坐标组成的矩阵, 由第 5 章结果可知, 如果 $nS = X'HX$ 有谱分解

$$nS = X'HX = U\Lambda U',$$

其中 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$, $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ 为 nS 的特征值, U 的列 u_1, \dots, u_p 为对应的标准正交特征向量, 则

$$HXU = (HXu_1, HXu_2, \dots, HXu_p) \quad (11.3.1)$$

的各列为 HX 的主成分, u_1, \dots, u_p 为在 R^p 中的主轴, 而

$$h_i' Xu_1, \dots, h_i' Xu_p \quad (11.3.2)$$

为 HX 的第 i 个行向量的主坐标, 其中 h_i' 表示中心化阵 H 的第 i 行。

另外, 主成分 HXu_1, \dots, HXu_p 是矩阵 $B = (HX)(HX)'$ 的特征向量, 且满足

$$BH Xu_i = \lambda_i H Xu_i, (HXu_i)'(HXu_i) = \lambda_i, \quad i = 1, \dots, p. \quad (11.3.3)$$

特别地, 我们还称

$$HXU_k = (HXu_1, \dots, HXu_k) \quad (11.3.4)$$

为 HX 的 k 维主坐标阵, 其中 $U_k = (u_1, \dots, u_k)$, $k \leq p$ 。

再来对比一下多维标度法经典解的结果。

设 D 为一个 $n \times n$ 欧氏距离阵, 根据定理 11.2.5 和上节中经典解的求法得知:

1. D 存在构图 X (不妨设 X 为 $n \times p$ 阶的), 而且 $B = HAH$

$= HXX'H$ (X 的中心化内积阵)。

2. D 的经典解 k 维拟合构图 X_k 为

$$X_k = \Gamma_k \Lambda_k^{1/2} (\sqrt{\lambda_1} Y_{(1)}, \dots, \sqrt{\lambda_k} Y_{(k)}),$$

其中 $\lambda_1 \geq \dots \geq \lambda_k$ 为 B 的前 k 个特征值, $x_{(i)} = \sqrt{\lambda_i} Y_{(i)}$ 为 B 对应于 λ_i 的特征向量, 还满足规格化条件 $x_{(i)}' x_{(i)} = \lambda_i$, 即

$$Bx_{(i)} = \lambda_i x_{(i)}, \quad x_{(i)}' x_{(i)} = \lambda_i, \quad i = 1, \dots, k. \quad (11.3.5)$$

对照式 (11.3.5) 和 (11.3.4), 则有

$$x_{(i)} = \sqrt{\lambda_i} Y_{(i)} = HXu_i, \quad i = 1, \dots, k, \quad (11.3.6)$$

由此可见, D 的 k 维拟合构图 X_k 正好是 D 的构图 HX 的 k 维主坐标阵。

注意, HX 的 k 维主坐标阵是通过距离阵 D 计算出来的, 而没有直接通过 HX , 所以也可以称 $X_k = \Gamma_k \Lambda_k^{1/2}$ 为 D 的构图的 k 维主坐标阵。

11.3.2 经典解的最优性质和拟合优度

由主成分分析的结果, 对于 R^p 中任意 n 个点 x_1, \dots, x_n , 在所有 k 维子空间的投影中, 以在前 k 个主轴张成的子空间的投影对它们拟合最好。根据前面关于经典解和主成分分析的关系, 可以推想, 经典解 k 维拟合构图也应有相应的最优性质。下面我们就来具体阐述和证明这一点。

在本节, 对于给定的距离阵 D , 用 \hat{X} 记 D 的 k 维拟合构图, 用 \hat{D} 表示 \hat{X} 的内点距离阵。对 D 求多维标度解的目的是使得 \hat{X} 的内点距离阵 \hat{D} 尽可能拟合原始距离阵 D 。

首先考虑 D 为欧氏距离阵的情形。

设 X 是 D 在 R^p 中的一个构图, 不妨设 X 已中心化, 即 $X = HX$ 。又设 $L = (L_1', L_2')$ 为一个 $p \times p$ 阶正交阵, 其中 L_1 是 $p \times k$ 阶的。则 XL_1 表示 X 到由 L_1 的列所张成子空间的投影, $\hat{X} = XL_1$ 就是 D 的一个 k 维拟合构图 (但未必是经典解)。

由于 L 是正交阵, 所以 X 的行之间的距离与 XL 的行之间的距离

离相同。则有

$$d_{ij}^2 = \sum_{r=1}^p (x_{ir} - x_{jr})^2 = \sum_{r=1}^p (x_i' l_{(r)} - x_j' l_{(r)})^2 \quad (11.3.7)$$

其中 $l_{(r)}$ 表示 L 的第 r 列。如果用 \hat{D} 记 XL_1 的行间距离阵, 则

$$\hat{d}_{ij}^2 = \sum_{r=1}^k (x_i' l_{(r)} - x_j' l_{(r)})^2. \quad (11.3.8)$$

由于一个构图到子空间的投影不会增加内点距离, 故 $\hat{d}_{ij}^2 \leq d_{ij}^2$. 于是构图 X 的距离阵 D 与 k 维拟合构图 \hat{X} 的距离阵 \hat{D} 之间的差异可以用

$$\phi = \sum_{i,j=1}^n (d_{ij}^2 - \hat{d}_{ij}^2) \quad (11.3.9)$$

来衡量。

我们把 k 维经典解的最优性质用下列定理来叙述。

定理 11.3.1 设 D 为一个欧氏距离阵, 并设在 R^p 中的构图为 X . 则对任意固定的 k ($1 \leq k \leq p$), 在所有的 k 维拟合构图 $\hat{X} = XL_1$ 中, 以 k 维经典解使式 (11.3.9) 中的 ϕ 值达到最小。

证 由 (11.3.7)–(11.3.9) 式, 有

$$\begin{aligned} \phi &= \sum_{i,j=1}^n \sum_{r=k+1}^p (x_i' l_{(r)} - x_j' l_{(r)})^2 \\ &= \sum_{i,j=1}^n \sum_{r=k+1}^p l_{(r)}' (x_i - x_j)(x_i - x_j)' l_{(r)} \\ &= \text{tr} L_2' \left(\sum_{i,j=1}^n (x_i - x_j)(x_i - x_j)' \right) L_2. \end{aligned}$$

$$\begin{aligned} \text{由于 } \sum_{i,j=1}^n (x_i - x_j)(x_i - x_j)' &= 2n \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' \\ &= 2 \sum_{i=1}^n (x_i - \bar{x}) \sum_{i=1}^n (x_i - \bar{x})' \end{aligned}$$

$$= 2nX'HX = 2n^2S,$$

$$\text{故 } \phi = 2n^2 \text{tr}(L_2' SL_2). \quad (11.3.10)$$

利用谱分解 $nS = U\Lambda U'$, 其中 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$, 并且 $\lambda_1 \geq \dots \geq \lambda_p$, 则有

$$\phi = 2n \text{tr}(F_2' \Lambda F_2)$$

其中 $F_2 = U' L_2$ 为列正交阵, 即 $F_2' F_2 = I_{p-k}$. 注意到

$$\text{tr}(F_2' \Lambda F_2) = \sum_{i=1}^{p-k} f_{(i)}' \Lambda f_{(i)}, \quad (11.3.11)$$

由二次型极值性质容易得证 ϕ 的极小值是

$$2n(\lambda_{k+1} + \dots + \lambda_p), \quad (11.3.12)$$

而且是在 $F_2 = (0, I_{p-k})'$ 上达到。根据 L_2 与 F_2 的关系, 也就是 ϕ 在 $L_2 = UF_2 = (u_{k+1}, \dots, u_p)$ 上达到最小。换句话说, 当取 $L_1 = (u_1, \dots, u_k)$ 时, 拟合构图 $\hat{X} = XL_1$ 的距离阵 \hat{D} 与 D 的差异 ϕ 达到最小。 $\hat{X} = (Xu_1, \dots, Xu_k)$ 正好是 k 维经典解, 这就是要证的。

当 D 是非欧氏距离阵时, 由于 D 不存在构图 X , 故用式 (11.3.9) 作为拟合标准不合适。这时我们选用对矩阵 $B = HAH$ 的拟合作为拟合优度准则。设 \hat{X} 是 D 的一个拟合构图, 用 $\hat{B} = H\hat{X}X'H$ 记它的中心化内积阵, 则用

$$\psi = \sum_{i,j=1}^n (b_{ij} - \hat{b}_{ij})^2 = \text{tr}(B - \hat{B})^2 \quad (11.3.13)$$

作为拟合优度的衡量标准。下面的定理告诉我们, 在这一准则下, 经典解具有最优性质。同时得出当 D 是欧氏距离阵时, 在两种标准下的最优解一致。

定理 11.3.2 设 D 为一个距离阵 (未必是欧氏的), 则对任意固定的 k , 在所有的 k 维拟合构图 \hat{X} 中以经典解使式 (11.3.13) 的 ψ 值达到最小。

证 由谱分解定理

$$B = \Gamma \Lambda \Gamma', \quad (11.3.14)$$

其中 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_1 \geq \dots \geq \lambda_n$ 为 B 的特征值 (可能有负值)。又以 $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_n \geq 0$ 记 \hat{B} 的特征值, 其中 $\hat{B} = H\hat{X}\hat{X}'H$ 。

我们考虑 $\Gamma' \hat{B} \Gamma$ 的谱分解:

$$\Gamma' \hat{B} \Gamma = G \hat{\Lambda} G' \quad (11.3.15)$$

其中 $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_n)$, G 为正交阵。于是

$$\begin{aligned} \psi &= \text{tr}(\hat{B} - \hat{B})^2 = \text{tr} \Gamma' (\hat{B} - \hat{B}) \Gamma \Gamma' (\hat{B} - \hat{B}) \Gamma \\ &= \text{tr}(\Lambda - G \hat{\Lambda} G')^2. \end{aligned} \quad (11.3.16)$$

根据下面的引理 11.3.3, 当 $G=I$ 时上式 ψ 达到最小值

$$\psi = \sum_{i=1}^n (\lambda_i - \hat{\lambda}_i)^2. \quad (11.3.17)$$

由于考虑的拟合构图 \hat{X} 在一个 k 维空间中, 故中心化内积阵 $\hat{B} = H\hat{X}\hat{X}'H$ 至多有 k 个非零特征值, 而且是非负值。于是只有当取

$$\hat{\lambda}_i = \begin{cases} \lambda_i, & i=1, \dots, t, \\ 0, & i=t+1, \dots, n \end{cases} \quad (11.3.18)$$

时式 (11.3.17) 达到最小, 其中 $t = \min(k, v)$, 而 v 为 $\lambda_i > 0$ 的最大下标号。

因此, 使式 (11.3.16) 达到最小的 \hat{B} 正好取 $\hat{B} = \Gamma_1 \Lambda_1 \Gamma_1'$, 其中 $\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_t)$, $\Gamma_1 = (\Gamma_{(1)}, \dots, \Gamma_{(t)})$, 而使得中心化内积阵为 $\Gamma_1 \Lambda_1 \Gamma_1'$ 的 t 维拟合构图 $\hat{X} = \Gamma_1 \Lambda_1^{1/2}$, 这正好是 D 的 t 维经典解, ψ 的最小值为

$$\psi = \lambda_{t+1}^2 + \dots + \lambda_n^2. \quad (11.3.19)$$

定理证毕。

引理 11.3.3 设 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, 其中 $\lambda_1 \geq \dots \geq \lambda_n$ 为实数。又设 $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_n)$, 其中 $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_n \geq 0$, 则对所有正交阵 G ,

$$\text{tr}(\Lambda - G \hat{\Lambda} G')^2 \quad (11.3.20)$$

在 $G=I$ 上达到最小值。

证 由于

$$\text{tr}(\Lambda - G \hat{\Lambda} G')^2 = \text{tr} \Lambda^2 + \text{tr} \hat{\Lambda}^2 - 2 \text{tr}(\Lambda G \hat{\Lambda} G')$$

故我们只要来考虑 $\text{tr}(\Lambda G \hat{\Lambda} G')$ 的最大值。

利用二次型极值的性质, 有

$$\begin{aligned} \text{tr}(\Lambda G \hat{\Lambda} G') &= \text{tr}((\Lambda g_{(1)}, \dots, \Lambda g_{(n)}) \begin{pmatrix} \hat{\lambda}_1 g_{(1)}' \\ \vdots \\ \hat{\lambda}_n g_{(n)}' \end{pmatrix}) \\ &= \text{tr} \left(\sum_{i=1}^n \Lambda g_{(i)} \cdot \hat{\lambda}_i g_{(i)}' \right) = \sum_{i=1}^n \hat{\lambda}_i g_{(i)}' \Lambda g_{(i)} \\ &\leq \max_{g_{(1)}} \hat{\lambda}_1 g_{(1)}' \Lambda g_{(1)} + \max_{g_{(2)} \perp g_{(1)}} \hat{\lambda}_2 g_{(2)}' \Lambda g_{(2)} \\ &\quad + \dots + \max_{g_{(n-1)} \perp g_{(i)}, i=1, \dots, n-1} \hat{\lambda}_n g_{(n)}' \Lambda g_{(n)} \\ &= \hat{\lambda}_1 \lambda_1 + \hat{\lambda}_2 \lambda_2 + \dots + \hat{\lambda}_n \lambda_n, \end{aligned}$$

其中 $g_{(i)}$, $i=1, \dots, n$ 为单位正交向量, 而且当取 $G=I$ 时可

达到上面的最大值, 即 $\text{tr}(\Lambda G \hat{\Lambda} G') = \text{tr}(\Lambda \hat{\Lambda}) = \sum_{i=1}^n \hat{\lambda}_i \lambda_i$, 引理

证毕。

§ 11.4 非度量方法

11.4.1 引言

在前面介绍的多维标度求解方法中, 我们是从观测到的 n 个客体之间的距离阵 D 出发, 在某个 k 维空间求一个拟合构图 \hat{X} , 使得 \hat{X} 的内点距离阵 \hat{D} 和 D 尽量接近。所以, 这种方法实际上隐含着下面的假设模型:

$$d_{ij} = \hat{d}_{ij} + e_{ij} \quad (11.4.1)$$

即观测到的距离 d_{ij} 等于在某 k 维空间的实际距离 \hat{d}_{ij} 加上误差

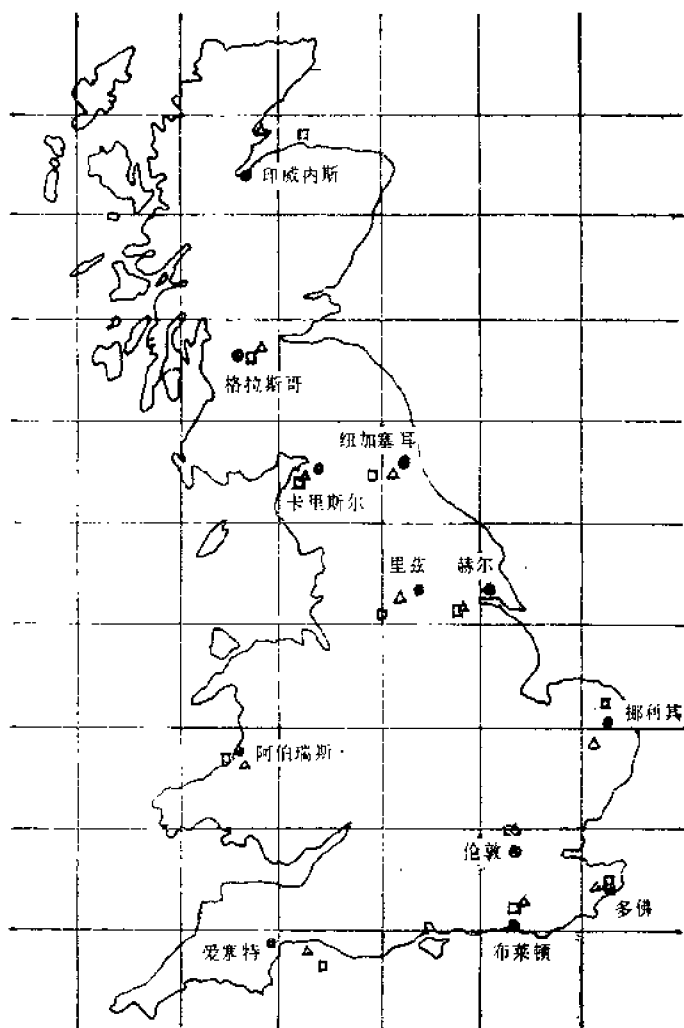


图 11.2.3 表 11.1.1 数据的多维标度解: •, 原始点; △, 经典解; □, 舍帕得——克卢斯卡解

e_{ij} , 其中 e_{ij} 可能是由于测量过程或其它干扰因素所造成的, 由此导致求解方法是使 \hat{d}_{ij} 与 d_{ij} 尽可能接近。

但在实际中, 对 n 个客体所能观测到的可能既不是它们之间的距离也不是相似系数, 而只是它们之间某种差异程度的次序。确切一点说, 例如对其中的两对客体 i 和 j , s 和 t , 每对之间都有差异, 但具体差异是多少都难以用数值来表示, 只知道 i 和 j 的差异要比 s 和 t 的差异大。这样对于 n 个客体的 $\frac{1}{2}n(n-1)$ 对之间的差异程度可以排一个次序:

$$d_{i_1 j_1} \leq \dots \leq d_{i_m j_m}, \quad m = \frac{1}{2}n(n-1), \quad (11.4.2)$$

其中 $d_{i,j}$ 表示客体 i 和 j 之间的差异, 数学上, 可以赋予每个 $d_{i,j}$ 一个数值, 但数值大小本身没有什么意义, 仅仅为了标明式 (11.4.2) 中的次序而用的。我们希望仅从 n 个客体之间的这种差异顺序出发找出一个拟合构图 \hat{X} 来反映这 n 个客体之间的结构关系。这就是非度量多维标度所要解决的问题。这时, 一般是用构图 \hat{X} 的内点距离 (反映差异) 的大小次序来拟合原始客体之间的差异大小次序, 即用在此意义上拟合的好坏来衡量构图的好坏。那么, 这时的模型假设为

$$d_{ij} = f(\hat{d}_{ij} + e_{ij}) \quad (11.4.3)$$

才比较合适, 其中 f 是未知的单增函数。自然可以想象, 如果求得的构图 \hat{X} , 其内点间距离的次序和原始客体之间的差异次序完全吻合, 则认为这样的拟合构图是最好的。当然, 如果把一个距离阵中的元素排成一个次序, 把它当作客体间差异程度的次序, 那么也可以把一个度量标度问题化为一个非度量标度问题。事实上, 当距离阵中的数据不可靠, 而对距离大小的顺序是可靠的时候, 用模型 (11.4.3) 比起用模型 (11.4.1) 来考虑可能更真实一些。

举一个非度量问题的例子。在例 11.1.1 的问题中如果不给出表 11.1.1 的距离数据, 而仅知道关于英国 12 个城市的如下信息:

最短的路程是从布莱顿到伦敦,
其次最短的路程是从赫尔到里兹,
.....

最长的路程是从多佛到印威内斯。

我们试图利用这些信息来绘制英国12个城市的地理位置图,则这就是一个典型的非度量多维标度问题。

11.4.2 舍帕得-克卢斯卡算法

目前求解非度量标度问题已有一些方法,下面仅介绍比较流行的一种,这是由舍帕得 (Shepard) (1962)和克卢斯卡 (Kruskal) (1964)*提出来的。步骤如下:

1. 给定 n 个客体之间差异程度的次序

$$d_{i_1 j_1} \leq d_{i_2 j_2} \leq \dots \leq d_{i_m j_m}, \quad m = \frac{1}{2}n(n-1), \quad (11.4.4)$$

其中 $(i_1, j_1), \dots, (i_m, j_m)$ 记所有不同下标的配对,不妨设 $i_r < j_r$, 赋给 $d_{i_r j_r}, r = 1, \dots, m$ 数值以后的对称矩阵 (对角线元素为 0) 记为 D , 并称之为**差异矩阵**。如果另有一个差异阵 $D^* = (d_{ij}^*)$, 满足

$$d_{i_1 j_1}^* \leq d_{i_2 j_2}^* \leq \dots \leq d_{i_m j_m}^*, \quad m = \frac{1}{2}n(n-1), \quad (11.4.5)$$

即与 D 有相同的次序, 则称 D^* 和 D 是**次序等价的**。我们把所有与 D 次序等价的差异阵组成的集合记为 \mathcal{D} 。

2. 设 $\hat{X} (n \times k)$ 是在 R^k 中的一个拟合构图, 并有内点距离阵 $\hat{D} = (\hat{d}_{ij})$, 定义 \hat{X} 与 \mathcal{D} 的 (平方) 距离为:

$$S^2(\hat{X}) = \min_{\hat{D} \in \mathcal{D}} \sum_{i < j} (d_{ij} - \hat{d}_{ij})^2 / \sum_{i < j} \hat{d}_{ij}^2, \quad (11.4.6)$$

其中的最小值是在与 D 次序等价的类 \mathcal{D} 上取的。我们把达到 (11.4.6) 最小值的 \hat{D}^* 叫做在 \mathcal{D} 中 (或关于 D) 对于 \hat{D} 的**最小二乘单调回归**。于是 $S(\hat{X})$ 表示 \hat{X} 的内点距离阵 \hat{D} 的次序与 D 的次序协调一致的程度。如果 $S(\hat{X}) = 0$, 则说 \hat{D} 精确地与 D 匹配, 这时显然有 $\hat{D} \in \mathcal{D}$ 。

注意, 式 (11.4.6) 中用 $\sum_{i < j} \hat{d}_{ij}^2$ 作分母是为了使定义的 距离 (平方) 标准化, 同时还使得在标度变换 $y_r = cx_r, r = 1, \dots, n, c \neq 0$ 以及正交平移变换 $y_r = \Gamma x_r + b, r = 1, \dots, n$ 下是不变的, 其中 Γ 是正交阵, b 为常数向量。

3. 对于每个维数 k , 我们称达到最小距离的拟合构图叫做**最优 k 维拟合构图**。用

$$S_k = \min_{\hat{X} (n \times k)} S(\hat{X})$$

记这一最小距离。

4. 为选择合适的维数, 我们计算 S_1, S_2, \dots 的值。由于 S_k 是 k 的单减函数 (这点容易从式 (11.4.7) 看出), 当 S_k 减小稳定在一个低值时, 则停止计算, 例如 $k = k_0$ 。克卢斯卡给出一种经验的评价 S_k 优劣的尺度, 一般认为 $S_k \geq 20\%$ 时是差的, 在 $S_k = 10\%$ 左右时是可用的, 当 $S_k \leq 5\%$ 时是好的, 当 $S_k = 0$ 时是最理想的。

注 1 最优拟合构图的求法可用克卢斯卡提出的迭代法来进行, 这种迭代法是用来求局部最优的急剧下降法, 迭代的初值可以任意给定一个构图, 也可以取经典解 (在经典解可求的情况下)。但在实际中还没有办法判定所求得的结果是局部最优还是整体最优。

注 2 上述非度量法是用一个差异阵。由于差异是反映客体之间的一种不相似性, 所以很容易把这一方法用于相似阵的情形,

*Shepard, R.N. (1962a). The Analysis of proximities: Multidimensional Scaling with an unknown Distance Function I. *Psychometrika*, 27, 125-139.

—— (1962b). The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function II. *Psychometrika*, 27, 219-240.

Kruskal, J.B. (1964). Non-metric Multidimensional Scaling. *Psychometrika*, 29, 1-27, 115-129.

只要简单地改变一下不等式的方向就可以了。

例11.4.1 利用11.4.2中关于英国12个城市间路程长短顺序的信息绘制这12个城市的地理位置图。舍帕得-克卢斯卡解标在图11.2.3中。

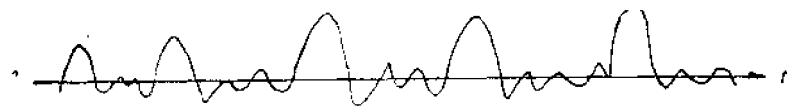
第 12 章 周期分析

§ 12.1 问题的提出

给出一组数据

$$X(1), X(2), \dots, X(n) \quad (12.1.1)$$

或波形 $X(t)$;



为了分析这组数据或曲线的性质，常常需要了解它的周期现象。所谓周期现象就是相隔一个固定时间段重复出现的现象。在实际观测到的数据中出现像周期函数那样的严格周期现象，即满足条件

$$X(t+1)=X(t) \quad (0 \leq t \leq n-1)$$

的周期函数是极少的。我们并不是去研究这样的周期现象，而是去分析那种相隔一个固定时间段出现近似相同值的周期现象。如果数据 $X(t)$ ($t=1, 2, \dots, n$)用时间 l 长去截段，各段是非常相似的，那么我们就说 $X(t)$ 具有明显的 l 的周期。但是如何去衡量截段后各段的相似程度呢？需要给出一个客观标准，这就是本章所要研究的问题。

§ 12.2 周期分析方法

对已给的数据 $X(t)$ ($t=1, \dots, n$)以 l 长分段，一共分