多维标度法 MDS 及 R 使用

Apocalypse

中国地质大学 (北京)　数理学院

2020 年 10 月 9 日

内容提要

问题引入

问题引入

问题

有 n 个由多个指标反映的客体，但反映这些客体的指标个数未知，甚至指标本身就是模糊的，我们仅知道这 n 个客体间的某种距离 (不一定是通常的欧氏距离) 或某种相似性。

问题引入

问题

有 n 个由多个指标反映的客体，但反映这些客体的指标个数未知，甚至指标本身就是模糊的，我们仅知道这 n 个客体间的某种距离 (不一定是通常的欧氏距离) 或某种相似性。

我们希望仅由这种距离或相似性给出的信息出发，在较低维度的欧氏空间中把这 n 个客体 (作为几何点) 的图形绘制出来，从而尽可能及时地反映这些客体之间真实的结构关系。

问题引入

问题

有 n 个由多个指标反映的客体，但反映这些客体的指标个数未知，甚至指标本身就是模糊的，我们仅知道这 n 个客体间的某种距离 (不一定是通常的欧氏距离) 或某种相似性。

我们希望仅由这种距离或相似性给出的信息出发，在较低维度的欧氏空间中把这 n 个客体 (作为几何点) 的图形绘制出来，从而尽可能及时地反映这些客体之间真实的结构关系。

多维标度 (Multi-Dimensional Scaling, MDS) 分析是以空间分布的形式表现对象之间相似性或亲疏关系的一种多元数据分析方法。其结果主要是**偏好图** (又称多维标度图) 等。

问题引入

问题

有 n 个由多个指标反映的客体，但反映这些客体的指标个数未知，甚至指标本身就是模糊的，我们仅知道这 n 个客体间的某种距离 (不一定是通常的欧氏距离) 或某种相似性。

我们希望仅由这种距离或相似性给出的信息出发，在较低维度的欧氏空间中把这 n 个客体 (作为几何点) 的图形绘制出来，从而尽可能及时地反映这些客体之间真实的结构关系。

多维标度 (Multi-Dimensional Scaling, MDS) 分析是以空间分布的形式表现对象之间相似性或亲疏关系的一种多元数据分析方法。其结果主要是**偏好图** (又称多维标度图) 等。

MDS 分析多见于市场营销领域，近年来在经济管理领域的应用也日趋增多。

MDS 的基本概念

MDS 的基本概念

定义

多维标度法是利用客体间的**相似性数据**去揭示它们之间**空间关系**的统计分析方法。

MDS 的基本概念

定义

多维标度法是利用客体间的**相似性数据**去揭示它们之间**空间关系**的统计分析方法。

根据所分析数据的类型，可将多维标度法分为度量化模型和非度量化模型：

MDS 的基本概念

定义

多维标度法是利用客体间的**相似性数据**去揭示它们之间**空间关系**的统计分析方法。

根据所分析数据的类型，可将多维标度法分为度量化模型和非度量化模型：

度量化模型 (Metric MDS)
若模型所需要的相似性数据是用距离尺度或比率尺度测得的，则称此模型为**度量化模型**。

MDS 的基本概念

定义

多维标度法是利用客体间的**相似性数据**去揭示它们之间**空间关系**的统计分析方法。

根据所分析数据的类型，可将多维标度法分为度量化模型和非度量化模型：

度量化模型 (Metric MDS)
若模型所需要的相似性数据是用距离尺度或比率尺度测得的，则称此模型为**度量化模型**。

非度量化模型 (Nonmetric MDS)
若模型需要顺序量表水平的相似数据，称此模型为**非度量化模型**。

MDS 的基本概念

定义

多维标度法是利用客体间的**相似性数据**去揭示它们之间**空间关系**的统计分析方法。

根据所分析数据的类型，可将多维标度法分为度量化模型和非度量化模型：

度量化模型 (Metric MDS)
若模型所需要的相似性数据是用距离尺度或比率尺度测得的，则称此模型为**度量化模型**。

非度量化模型 (Nonmetric MDS)
若模型需要顺序量表水平的相似数据，称此模型为**非度量化模型**。

基本思想

将高维坐标中的点投影到低维空间中，保持点彼此之间的相似性尽可能不变。

例 1

下表是美国 10 个城市间的公路距离 (并非直线距离)，我们希望在地图上重新标出这 10 个城市，使得它们之间的距离接近表中的距离。

	Atl	Chi	Den	Hou	LA	Mia	NYC	SF	Sea	WDC
Atl	0	587	1212	701	1936	604	748	2139	2182	543
Chi	587	0	920	940	1745	1188	713	1859	1737	597
Den	1212	920	0	879	831	1726	1631	949	1021	1494
Hou	701	940	879	0	1374	968	1420	1645	1891	1220
LA	1936	1745	831	1374	0	2339	2451	347	959	2300
Mia	604	1188	1726	968	2339	0	1092	2594	2734	923
NYC	748	713	1631	1420	2451	1092	0	2571	2408	205
SF	2139	1858	949	1645	347	2594	2571	0	678	2442
Sea	2182	1737	1021	1891	959	2734	2408	678	0	2329
WDC	543	597	1494	1220	2300	923	205	2442	2329	0

例 1

下表是美国 10 个城市间的公路距离 (并非直线距离)，我们希望在地图上重新标出这 10 个城市，使得它们之间的距离接近表中的距离。

	Atl	Chi	Den	Hou	LA	Mia	NYC	SF	Sea	WDC
Atl	0	587	1212	701	1936	604	748	2139	2182	543
Chi	587	0	920	940	1745	1188	713	1859	1737	597
Den	1212	920	0	879	831	1726	1631	949	1021	1494
Hou	701	940	879	0	1374	968	1420	1645	1891	1220
LA	1936	1745	831	1374	0	2339	2451	347	959	2300
Mia	604	1188	1726	968	2339	0	1092	2594	2734	923
NYC	748	713	1631	1420	2451	1092	0	2571	2408	205
SF	2139	1858	949	1645	347	2594	2571	0	678	2442
Sea	2182	1737	1021	1891	959	2734	2408	678	0	2329
WDC	543	597	1494	1220	2300	923	205	2442	2329	0

如果用 $D = (d_{ij})$ 表示上表的矩阵，虽然名义上是距离阵，但不一定是 n 个点的距离，所以下面我们扩展距离矩阵的定义。

MDS 的基本方法

MDS 的基本方法

定义

一个 $n \times n$ 矩阵 $D = (d_{ij})$，若满足 $D^T = D$，$d_{ii} = 0$，$d_{ij} \geqslant 0$，$(i, j = 1, 2, \cdots, n; i \neq j)$，则称 D 为**距离阵**。

MDS 的基本方法

定义

一个 $n \times n$ 矩阵 $D = (d_{ij})$，若满足 $D^T = D$, $d_{ii} = 0$, $d_{ij} \geqslant 0$, $(i, j = 1, 2, \cdots, n; i \neq j)$，则称 D 为**距离阵**。

对于距离阵 $D = (d_{ij})$，多维标度法的目的是寻找 p 和 $\mathbb{R}^p$ 中的 n 个点 $x_1, x_2, \cdots, x_n$，用 $\widehat{d}_{ij}$ 表示 x_i 与 x_j 的欧氏距离，$\widehat{D} = (\widehat{d}_{ij})$，使得 $\widehat{D}$ 与 D 在某种意义下相近。

MDS 的基本方法

定义

一个 $n \times n$ 矩阵 $D = (d_{ij})$，若满足 $D^T = D$, $d_{ii} = 0$, $d_{ij} \geqslant 0$, $(i, j = 1, 2, \cdots, n; i \neq j)$，则称 D 为**距离阵**。

对于距离阵 $D = (d_{ij})$，多维标度法的目的是寻找 p 和 $\mathbb{R}^p$ 中的 n 个点 $x_1, x_2, \cdots, x_n$，用 $\widehat{d}_{ij}$ 表示 x_i 与 x_j 的欧氏距离，$\widehat{D} = (\widehat{d}_{ij})$，使得 $\widehat{D}$ 与 D 在某种意义下相近。

在实际运用中，常取 $p = 1, 2, 3.$ 将寻找到的 n 个点 $x_1, x_2, \cdots, x_n$ 写成矩阵形式如下：

MDS 的基本方法

定义

一个 $n \times n$ 矩阵 $D = (d_{ij})$，若满足 $D^T = D$，$d_{ii} = 0$，$d_{ij} \geqslant 0$，$(i, j = 1, 2, \cdots, n; i \neq j)$，则称 D 为**距离阵**。

对于距离阵 $D = (d_{ij})$，多维标度法的目的是寻找 p 和 $\mathbb{R}^p$ 中的 n 个点 $x_1, x_2, \cdots, x_n$，用 $\widehat{d}_{ij}$ 表示 x_i 与 x_j 的欧氏距离，$\widehat{D} = (\widehat{d}_{ij})$，使得 $\widehat{D}$ 与 D 在某种意义下相近。

在实际运用中，常取 $p = 1, 2, 3$. 将寻找到的 n 个点 $x_1, x_2, \cdots, x_n$ 写成矩阵形式如下：

$$X = (x_1, x_2, \cdots, x_n)^T \tag{1}$$

MDS 的基本方法

定义

一个 $n\times n$ 矩阵 $D=(d_{ij})$，若满足 $D^T=D$, $d_{ii}=0$, $d_{ij}\geqslant 0$, $(i,\, j=1,\, 2,\, \cdots,\, n;\, i\neq j)$，则称 D 为**距离阵**。

对于距离阵 $D=(d_{ij})$，多维标度法的目的是寻找 p 和 $\mathbb{R}^p$ 中的 n 个点 $x_1, x_2, \cdots, x_n$，用 $\widehat{d}_{ij}$ 表示 x_i 与 x_j 的欧氏距离，$\widehat{D}=(\widehat{d}_{ij})$，使得 $\widehat{D}$ 与 D 在某种意义下相近。

在实际运用中，常取 $p=1,\,2,\,3.$ 将寻找到的 n 个点 $x_1, x_2, \cdots, x_n$ 写成矩阵形式如下：

$$X=(x_1, x_2, \cdots, x_n)^T \tag{1}$$

则称 X 为 D 的一个解 (或称**多维标度解**)。

欧式型距离阵及其判定定理

欧式型距离阵及其判定定理

定义: 欧式型距离阵

一个距离阵 $D = (d_{ij})$ 称为**欧式型**的，若存在某个正整数 p 及 p 维空间 $\mathbb{R}^p$ 中的 n 个点 $x_1, \cdots, x_n$, 使得

欧式型距离阵及其判定定理

定义: 欧式型距离阵

一个距离阵 $D=(d_{ij})$ 称为**欧式型**的，若存在某个正整数 p 及 p 维空间 $\mathbb{R}^p$ 中的 n 个点 $x_1,\cdots,x_n$, 使得

$$d_{ij}^2=(x_i-x_j)'(x_i-x_j), \tag{2}$$

欧式型距离阵及其判定定理

定义: 欧式型距离阵

一个距离阵 $D=(d_{ij})$ 称为**欧式型**的，若存在某个正整数 p 及 p 维空间 $\mathbb{R}^p$ 中的 n 个点 $x_1,\cdots,x_n$, 使得

$$d_{ij}^2=(x_i-x_j)'(x_i-x_j), \tag{2}$$

$$A=(a_{ij}), \quad a_{ij}=-\frac{1}{2}d_{ij}^2, \tag{3}$$

欧式型距离阵及其判定定理

定义: 欧式型距离阵

一个距离阵 $D=(d_{ij})$ 称为**欧式型**的，若存在某个正整数 p 及 p 维空间 $\mathbb{R}^p$ 中的 n 个点 $x_1, \cdots, x_n$, 使得

$$d_{ij}^2 = (x_i - x_j)'(x_i - x_j), \tag{2}$$

$$A = (a_{ij}), \quad a_{ij} = -\frac{1}{2} d_{ij}^2, \tag{3}$$

$$B = H'AH, \; H = I_n - \frac{1}{n} 1_n 1_n'. \tag{4}$$

欧式型距离阵及其判定定理

定义: 欧式型距离阵

一个距离阵 $D=(d_{ij})$ 称为**欧式型**的，若存在某个正整数 p 及 p 维空间 $\mathbb{R}^p$ 中的 n 个点 $x_1,\cdots,x_n$, 使得

$$d_{ij}^2=(x_i-x_j)'(x_i-x_j), \tag{2}$$

$$A=(a_{ij}),\quad a_{ij}=-\frac{1}{2}d_{ij}^2, \tag{3}$$

$$B=H'AH,\ H=I_n-\frac{1}{n}1_n1_n'. \tag{4}$$

定理: 欧式型距离阵判定定理

一个 $n\times n$ 的距离阵 D 是欧式型的充要条件是:$B\geqslant 0$.

谱分解定理

谱分解定理

定义：矩阵的谱

矩阵 A 的所有特征值的全体 $\lambda_1, \cdots, \lambda_n$ 称为矩阵 A 的谱，记为 $\lambda(A)$。

谱分解定理

定义：矩阵的谱

矩阵 A 的所有特征值的全体 $\lambda_1, \cdots, \lambda_n$ 称为矩阵 A 的谱，记为 $\lambda(A)$。

定理：谱分解定理

设 $A \geqslant 0$ 为对称矩阵，λ_i, λ_j 是 A 的两个相异特征根，相应的特征向量 ℓ_i 和 ℓ_j 相互正交，则 A 可表示为 $A = T\Lambda T' = \sum_{i=1}^{p} \lambda_i \ell_i \ell_i'$，称为矩阵 A 的**谱分解**。即存在一个正交阵 T，使 $T'AT = \mathrm{diag}(\lambda_1, \lambda_2, \cdots \lambda_p) = \Lambda$，其中 T 的列向量为相应的特征向量。

证明:

证明: (必要性) 设距离阵 D 是欧式型的，则由欧式型距离阵的定义可知，存在 $x_1, \cdots, x_n \in \mathbb{R}^p$，使得

$$d_{ij}^2 = -2a_{ij} = (x_i - x_j)'(x_i - x_j), \tag{5}$$

证明: (必要性) 设距离阵 D 是欧式型的，则由欧式型距离阵的定义可知，存在 $x_1, \cdots, x_n \in \mathbb{R}^p$，使得

$$d_{ij}^2 = -2a_{ij} = (x_i - x_j)'(x_i - x_j), \tag{5}$$

由(4)可得:

$$B = H'AH = A - \frac{1}{n}AJ - \frac{1}{n}JA + \frac{1}{n^2}JAJ, \tag{6}$$

上式中，$J = 1_n 1_n'$。

证明: (必要性) 设距离阵 D 是欧式型的，则由欧式型距离阵的定义可知，存在 $x_1, \cdots, x_n \in \mathbb{R}^p$，使得

$$d_{ij}^2 = -2a_{ij} = (x_i - x_j)'(x_i - x_j), \tag{5}$$

由(4)可得:

$$B = H'AH = A - \frac{1}{n}AJ - \frac{1}{n}JA + \frac{1}{n^2}JAJ, \tag{6}$$

上式中，$J = 1_n 1_n'$。并注意到

$$\frac{1}{n}AJ = \begin{bmatrix} \bar{a}_{1.} \\ \bar{a}_{2.} \\ \vdots \\ \bar{a}_{n.} \end{bmatrix} 1_n', \ \frac{1}{n}JA = 1_n(\bar{a}_{.1}, \bar{a}_{.2}, \cdots, \bar{a}_{.n}), \ \frac{1}{n^2}JAJ = \bar{a}_{..}1_n 1_n', \tag{7}$$

证明: (必要性) 设距离阵 D 是欧式型的，则由欧式型距离阵的定义可知，存在 $x_1, \cdots, x_n \in \mathbb{R}^p$，使得

$$d_{ij}^2 = -2a_{ij} = (x_i - x_j)'(x_i - x_j), \tag{5}$$

由(4)可得:

$$B = H'AH = A - \frac{1}{n}AJ - \frac{1}{n}JA + \frac{1}{n^2}JAJ, \tag{6}$$

上式中，$J = 1_n 1_n'$。并注意到

$$\frac{1}{n}AJ = \begin{bmatrix} \bar{a}_{1.} \\ \bar{a}_{2.} \\ \vdots \\ \bar{a}_{n.} \end{bmatrix} 1_n', \ \frac{1}{n}JA = 1_n(\bar{a}_{.1}, \bar{a}_{.2}, \cdots, \bar{a}_{.n}), \ \frac{1}{n^2}JAJ = \bar{a}_{..}1_n 1_n', \tag{7}$$

其中，

$$\bar{a}_{i.} = \frac{1}{n}\sum_{j=1}^n a_{ij}, \ \bar{a}_{.j} = \frac{1}{n}\sum_{i=1}^n a_{ij}, \ \bar{a}_{..} = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n a_{ij}. \tag{8}$$

将上述各式代入(6)，得到

$$b_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, \tag{9}$$

将上述各式代入(6)，得到

$$b_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, \tag{9}$$

再由(5)式可分别求得 a_{ij}, $\bar{a}_{i.}$, $\bar{a}_{.j}$, $\bar{a}_{..}$，将其代入式(9)，得到

$$b_{ij} = (x_i - \bar{x})'(x_i - \bar{x}) \geqslant 0, \tag{10}$$

其中 $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$,

将上述各式代入(6)，得到

$$b_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, \tag{9}$$

再由(5)式可分别求得 a_{ij}, $\bar{a}_{i.}$, $\bar{a}_{.j}$, $\bar{a}_{..}$，将其代入式(9)，得到

$$b_{ij} = (x_i - \bar{x})'(x_i - \bar{x}) \geqslant 0, \tag{10}$$

其中 $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$, 将上式写成矩阵形式，即得到

$$B = (HX)(HX)' \geqslant 0 \tag{11}$$

必要性得证，下证充分性。

将上述各式代入(6)，得到

$$b_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, \tag{9}$$

再由(5)式可分别求得 a_{ij}, $\bar{a}_{i.}$, $\bar{a}_{.j}$, $\bar{a}_{..}$，将其代入式(9)，得到

$$b_{ij} = (x_i - \bar{x})'(x_i - \bar{x}) \geqslant 0, \tag{10}$$

其中 $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$, 将上式写成矩阵形式，即得到

$$B = (HX)(HX)' \geqslant 0 \tag{11}$$

必要性得证，下证充分性。

(充分性) 记 $p = \mathrm{rank}(B)$, $\lambda_1, \lambda_2, \cdots, \lambda_p$ 为 B 的正特征根，$x_{(1)}, \cdots, x_{(p)}$ 为对应的特征向量。

由已知条件，$B \geqslant 0$，根据**谱分解定理**得到

$$B = H'AH = \Gamma\Lambda\Gamma' \tag{12}$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_p)$, $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ 为 B 的 p 个正特征根 $\Gamma = X\Lambda^{-\frac{1}{2}}$, Γ 的 p 个列为对应的 p 个标准正交化的特征向量。

由已知条件，$B \geqslant 0$，根据**谱分解定理**得到

$$B = H'AH = \mathit{\Gamma}\Lambda\mathit{\Gamma}' \tag{12}$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_p)$, $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ 为 B 的 p 个正特征根
$\mathit{\Gamma} = X\Lambda^{-\frac{1}{2}}$, $\mathit{\Gamma}$ 的 p 个列为对应的 p 个标准正交化的特征向量。取
$X = \mathit{\Gamma}\Lambda^{\frac{1}{2}}$，为一 $n \times p$ 阶矩阵。将 X 写成
$X = (x_1, x_2, \cdots, x_n)' = (x_{(1)}, x_{(2)}, \cdots, x_{(p)})$，于是有

$$X'X = (\mathit{\Gamma}\Lambda^{\frac{1}{2}})'(\mathit{\Gamma}\Lambda^{\frac{1}{2}}) = \Lambda, \ B = XX', \tag{13}$$

即 $b_{ij} = x_i'x_j$。

由已知条件，$B \geqslant 0$，根据**谱分解定理**得到

$$B = H'AH = \mathit{\Gamma}\Lambda\mathit{\Gamma}' \tag{12}$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_p)$, $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ 为 B 的 p 个正特征根
$\mathit{\Gamma} = X\Lambda^{-\frac{1}{2}}$, $\mathit{\Gamma}$ 的 p 个列为对应的 p 个标准正交化的特征向量。取
$X = \mathit{\Gamma}\Lambda^{\frac{1}{2}}$，为一 $n \times p$ 阶矩阵。将 X 写成
$X = (x_1, x_2, \cdots, x_n)' = (x_{(1)}, x_{(2)}, \cdots, x_{(p)})$，于是有

$$X'X = (\mathit{\Gamma}\Lambda^{\frac{1}{2}})'(\mathit{\Gamma}\Lambda^{\frac{1}{2}}) = \Lambda,\ B = XX', \tag{13}$$

即 $b_{ij} = x_i'x_j$。由此得到 x_i 与 x_j 两点的距离平方

$$\begin{aligned}(x_i - x_j)'(x_i - x_j) &= x_i'x_i - 2x_i'x_j + x_j'x_j = b_{ii} - 2b_{ij} + b_{jj} &\quad (14)\\ &= a_{ii} - 2a_{ij} + a_{jj} = -2a_{ij} = d_{ij}^2 &\quad (15)\end{aligned}$$

由已知条件，$B \geqslant 0$，根据**谱分解定理**得到

$$B = H'AH = \Gamma\Lambda\Gamma' \tag{12}$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_p)$, $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ 为 B 的 p 个正特征根
$\Gamma = X\Lambda^{-\frac{1}{2}}$, Γ 的 p 个列为对应的 p 个标准正交化的特征向量。取
$X = \Gamma\Lambda^{\frac{1}{2}}$，为一 $n \times p$ 阶矩阵。将 X 写成
$X = (x_1, x_2, \cdots, x_n)' = (x_{(1)}, x_{(2)}, \cdots, x_{(p)})$，于是有

$$X'X = (\Gamma\Lambda^{\frac{1}{2}})'(\Gamma\Lambda^{\frac{1}{2}}) = \Lambda,\ B = XX', \tag{13}$$

即 $b_{ij} = x_i'x_j$。由此得到 x_i 与 x_j 两点的距离平方

$$\begin{aligned}(x_i - x_j)'(x_i - x_j) &= x_i'x_i - 2x_i'x_j + x_j'x_j = b_{ii} - 2b_{ij} + b_{jj} &(14)\\ &= a_{ii} - 2a_{ij} + a_{jj} = -2a_{ij} = d_{ij}^2 &(15)\end{aligned}$$

根据上述推导可知：存在正整数 p 和一个 $n \times p$ 阶矩阵
$X = (x_1, x_2, \cdots, x_n) = \Gamma\Lambda^{\frac{1}{2}}$，使得 X 是 D 的构造点，所以 D 为欧式型。

MDS 古典解的计算步骤 ★

MDS 古典解的计算步骤 ★

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

MDS 古典解的计算步骤 ★

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

令矩阵 $B=(b_{ij})$, 其中 $b_{ij}=a_{ij}-\overline{a}_{i.}-\overline{a}_{.j}+\overline{a}_{..}$;

MDS 古典解的计算步骤 ★

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

令矩阵 $B=(b_{ij})$, 其中 $b_{ij}=a_{ij}-\overline{a}_{i\cdot}-\overline{a}_{\cdot j}+\overline{a}_{\cdot\cdot}$;

求矩阵 B 的特征根 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$, 若无负特征根，表明 $B \geqslant 0$, 从而距离阵 D 是欧式型的; 若有负特征根，距离阵 D 一定不是欧式型的。

MDS 古典解的计算步骤 ★

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

令矩阵 $B=(b_{ij})$, 其中 $b_{ij}=a_{ij}-\overline{a}_{i.}-\overline{a}_{.j}+\overline{a}_{..}$;

求矩阵 B 的特征根 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$, 若无负特征根，表明 $B\geqslant 0$, 从而距离阵 D 是欧式型的; 若有负特征根，距离阵 D 一定不是欧式型的。

此时令

$$a_{1,k}=\frac{\sum\limits_{i=1}^{k}\lambda_i}{\sum\limits_{i=1}^{n}|\lambda_i|},\quad a_{2,k}=\frac{\sum\limits_{i=1}^{k}\lambda_i^2}{\sum\limits_{i=1}^{n}\lambda_i^2},\tag{16}$$

MDS 古典解的计算步骤 ★

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

令矩阵 $B=(b_{ij})$, 其中 $b_{ij}=a_{ij}-\overline{a}_{i\cdot}-\overline{a}_{\cdot j}+\overline{a}_{\cdot\cdot}$;

求矩阵 B 的特征根 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$, 若无负特征根，表明 $B\geqslant 0$, 从而距离阵 D 是欧式型的; 若有负特征根，距离阵 D 一定不是欧式型的。

此时令

$$a_{1,k}=\frac{\sum\limits_{i=1}^{k}\lambda_i}{\sum\limits_{i=1}^{n}|\lambda_i|},\quad a_{2,k}=\frac{\sum\limits_{i=1}^{k}\lambda_i^2}{\sum\limits_{i=1}^{n}\lambda_i^2}, \tag{16}$$

上面两个量相当于主成分分析 (PCA) 中的**累积贡献率**;

MDS 古典解的计算步骤 ★

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

令矩阵 $B=(b_{ij})$, 其中 $b_{ij}=a_{ij}-\bar{a}_{i.}-\bar{a}_{.j}+\bar{a}_{..}$;

求矩阵 B 的特征根 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$, 若无负特征根，表明 $B \geqslant 0$, 从而距离阵 D 是欧式型的; 若有负特征根，距离阵 D 一定不是欧式型的。

此时令

$$a_{1,k}=\frac{\sum\limits_{i=1}^{k}\lambda_i}{\sum\limits_{i=1}^{n}|\lambda_i|}, \quad a_{2,k}=\frac{\sum\limits_{i=1}^{k}\lambda_i^2}{\sum\limits_{i=1}^{n}\lambda_i^2}, \tag{16}$$

上面两个量相当于主成分分析 (PCA) 中的**累积贡献率**;

令 $\hat{X}=(\hat{x}_{(1)},\cdots,\hat{x}_{(k)})$, 则 $\hat{X}$ 的行向量 $x_1,\cdots,x_n$ 即为欲求的古典解。

例 2

设有距离阵 D 如下，计算 MDS 的古典解。

$$D = \begin{bmatrix} 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 & 1 \\ & 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 \\ & & 0 & 1 & \sqrt{3} & 2 & 1 \\ & & & 0 & 1 & \sqrt{3} & 1 \\ & & & & 0 & 1 & 1 \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{bmatrix} \tag{17}$$

例 2

设有距离阵 D 如下，计算 MDS 的古典解。

$$D = \begin{bmatrix} 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 & 1 \\ & 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 \\ & & 0 & 1 & \sqrt{3} & 2 & 1 \\ & & & 0 & 1 & \sqrt{3} & 1 \\ & & & & 0 & 1 & 1 \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{bmatrix} \tag{17}$$

根据上述计算步骤，有 $a_{ij} = -\frac{1}{2}d_{ij}^2$，由此得到

例 2

设有距离阵 D 如下，计算 MDS 的古典解。

$$D = \begin{bmatrix} 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 & 1 \\ & 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 \\ & & 0 & 1 & \sqrt{3} & 2 & 1 \\ & & & 0 & 1 & \sqrt{3} & 1 \\ & & & & 0 & 1 & 1 \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{bmatrix} \tag{17}$$

根据上述计算步骤，有 $a_{ij} = -\frac{1}{2}d_{ij}^2$，由此得到

$$A = \begin{bmatrix} 0 & -\frac{1}{2} & -\frac{3}{2} & -2 & -\frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} \\ & 0 & -\frac{1}{2} & -\frac{3}{2} & -2 & -\frac{3}{2} & -\frac{1}{2} \\ & & 0 & -\frac{1}{2} & -\frac{3}{2} & -2 & -\frac{1}{2} \\ & & & 0 & -\frac{1}{2} & -\frac{3}{2} & -\frac{1}{2} \\ & & & & 0 & -\frac{1}{2} & -\frac{1}{2} \\ & & & & & 0 & -\frac{1}{2} \\ & & & & & & 0 \end{bmatrix} \tag{18}$$

再由 $b_{ij} = a_{ij} - \overline{a}_{i.} - \overline{a}_{.j} + \overline{a}_{..}$，得到

再由 $b_{ij} = a_{ij} - \overline{a}_{i.} - \overline{a}_{.j} + \overline{a}_{..}$，得到

$$2B = \begin{bmatrix} 2 & 1 & -1 & -2 & -1 & 1 & 0 \\ 1 & 2 & 1 & -1 & -2 & -1 & 0 \\ -1 & 1 & 2 & 1 & -1 & -2 & 0 \\ -2 & -1 & 1 & 2 & 1 & -1 & 0 \\ -1 & -2 & -1 & 1 & 2 & 1 & 0 \\ 1 & -1 & -2 & -1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{19}$$

再由 $b_{ij} = a_{ij} - \overline{a}_{i.} - \overline{a}_{.j} + \overline{a}_{..}$，得到

$$2B = \begin{bmatrix} 2 & 1 & -1 & -2 & -1 & 1 & 0 \\ 1 & 2 & 1 & -1 & -2 & -1 & 0 \\ -1 & 1 & 2 & 1 & -1 & -2 & 0 \\ -2 & -1 & 1 & 2 & 1 & -1 & 0 \\ -1 & -2 & -1 & 1 & 2 & 1 & 0 \\ 1 & -1 & -2 & -1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{19}$$

计算矩阵 B 的特征根，并按从大到小排序：$\lambda_1 = \lambda_2 = 3$, $\lambda_3 = \cdots = \lambda_7 = 0$。$B$ 无负特征根，说明 $B \geqslant 0$，所以 D 是欧式型的。

再由 $b_{ij} = a_{ij} - \overline{a}_{i.} - \overline{a}_{.j} + \overline{a}_{..}$，得到

$$2B = \begin{bmatrix} 2 & 1 & -1 & -2 & -1 & 1 & 0 \\ 1 & 2 & 1 & -1 & -2 & -1 & 0 \\ -1 & 1 & 2 & 1 & -1 & -2 & 0 \\ -2 & -1 & 1 & 2 & 1 & -1 & 0 \\ -1 & -2 & -1 & 1 & 2 & 1 & 0 \\ 1 & -1 & -2 & -1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{19}$$

计算矩阵 B 的特征根，并按从大到小排序：$\lambda_1 = \lambda_2 = 3$, $\lambda_3 = \cdots = \lambda_7 = 0$。$B$ 无负特征根，说明 $B \geqslant 0$，所以 D 是欧式型的。计算特征向量，得 $x_{(1)} = (0, -a, -a, 0, a, a, 0)^T$, $a = \frac{1}{2}$, $x_{(2)} = (-2b, -b, b, 2b, b, -b, 0)^T$, $b = \frac{\sqrt{3}}{2}$。

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

$$2B = \begin{bmatrix} 2 & 1 & -1 & -2 & -1 & 1 & 0 \\ 1 & 2 & 1 & -1 & -2 & -1 & 0 \\ -1 & 1 & 2 & 1 & -1 & -2 & 0 \\ -2 & -1 & 1 & 2 & 1 & -1 & 0 \\ -1 & -2 & -1 & 1 & 2 & 1 & 0 \\ 1 & -1 & -2 & -1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{19}$$

计算矩阵 B 的特征根，并按从大到小排序：$\lambda_1 = \lambda_2 = 3$, $\lambda_3 = \cdots = \lambda_7 = 0$。$B$ 无负特征根，说明 $B \geqslant 0$，所以 D 是欧式型的。计算特征向量，得 $x_{(1)} = (0, -a, -a, 0, a, a, 0)^T$, $a = \frac{1}{2}$, $x_{(2)} = (-2b, -b, b, 2b, b, -b, 0)^T$, $b = \frac{\sqrt{3}}{2}$。所以可得七个点的坐标分别为：
$(0, -\sqrt{3})$, $(-\frac{1}{2}, -\frac{\sqrt{3}}{2})$, $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$, $(0, \sqrt{3})$, $(\frac{1}{2}, \frac{\sqrt{3}}{2})$, $(\frac{1}{2}, -\frac{\sqrt{3}}{2})$, $(0, 0)$,

此即 D 的古典解。

MDS 古典解的 R 实现

MDS 古典解的 R 实现

```
cmdscale(d, k = 2, eig = FALSE, add = FALSE, x.ret =
    FALSE, list. = eig || add || x.ret)
```

MDS 古典解的 R 实现

```
cmdscale(d, k = 2, eig = FALSE, add = FALSE, x.ret =
    FALSE, list. = eig || add || x.ret)
```

d 表示进行多维标度分析的距离矩阵，k 表示维度，默认取 2 维.

MDS 古典解的 R 实现

```
cmdscale(d, k = 2, eig = FALSE, add = FALSE, x.ret =
    FALSE, list. = eig || add || x.ret)
```

d 表示进行多维标度分析的距离矩阵，k 表示维度，默认取 2 维. 下面应用 R 中自带的函数 **cmdscale()**，对上述例题进行 MDS 古典解的计算。

MDS 古典解的 R 实现

```
cmdscale(d, k = 2, eig = FALSE, add = FALSE, x.ret =
    FALSE, list. = eig || add || x.ret)
```

d 表示进行多维标度分析的距离矩阵，k 表示维度，默认取 2 维. 下面应用 R 中自带的函数 **cmdscale()**，对上述例题进行 MDS 古典解的计算。

```
D <- matrix(c(0,1,sqrt(3),2,sqrt(3),1,1,
              1,0,1,sqrt(3),2,sqrt(3),1,
              sqrt(3),1,0,1,sqrt(3),2,1,
              2,sqrt(3),1,0,1,sqrt(3),1,
              sqrt(3),2,sqrt(3),1,0,1,1,
              1,sqrt(3),2,sqrt(3),1,0,1,
              1,1,1,1,1,1,0), nrow = 7)
result <- round(cmdscale(D), 3); result
plot(result)
```

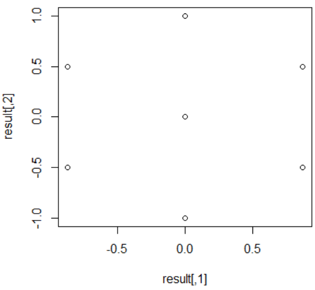
运行结果：

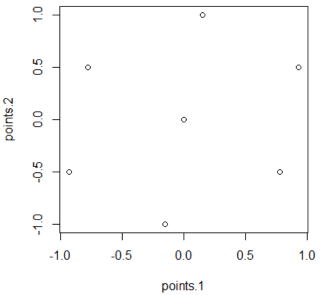
运行结果:

```
      [,1]  [,2]
[1,]  0.000  1.0
[2,] -0.866  0.5
[3,] -0.866 -0.5
[4,]  0.000 -1.0
[5,]  0.866 -0.5
[6,]  0.866  0.5
[7,]  0.000  0.0
```

运行结果：

```
     [,1]  [,2]
[1,]  0.000  1.0
[2,] -0.866  0.5
[3,] -0.866 -0.5
[4,]  0.000 -1.0
[5,]  0.866 -0.5
[6,]  0.866  0.5
[7,]  0.000  0.0
```





另一种方法也可以得出同样的结论，根据前面定理证明的方法。下面具体讲解。

另一种方法也可以得出同样的结论，根据前面定理证明的方法。下面具体讲解。

```
calc_mds <- function(D) {
    dim <- nrow(D)
    A <- -1/2*D^2
    onen <- matrix(rep(1, dim), nrow=dim)
    H <- diag(dim) - 1/dim * onen %*% t(onen)
    B <- t(H) %*% A %*% H
    val <- eigen(B, symmetric=F)$values
    vec <- eigen(B, symmetric=F)$vectors
    MDS <- vec[,1:2] %*% diag(sqrt(val[1:2]));
    data.frame(points = round(MDS, 3))
}
MDS <- calc_mds(D)
plot(MDS)
```

例 3

考虑例 1 中美国 10 个城市的距离阵，计算其 MDS 古典解。

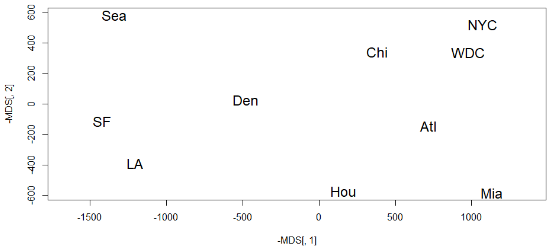
例 3

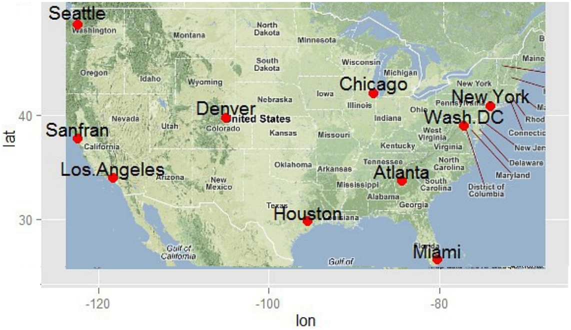
考虑例 1 中美国 10 个城市的距离阵，计算其 MDS 古典解。

```
X <- openxlsx::read.xlsx("mvstats4.xlsx", sheet="d12.1"
    , rowNames = T)
B <- calc_mds(as.matrix(X))
# 自编函数计算累积贡献率
calc_acr <- function(lambda, k) {
    a1k <- sum(lambda[1:k]) / sum(abs(lambda))
    a2k <- sum(lambda[1:k]^2) / sum(lambda^2)
    list(a1k=round(a1k, 3), a2k=round(a2k, 3))
}
calc_acr(eigen(B)$values, 2)
MDS <- round(cmdscale(X), 2)# 计算古典解
# 图形绘制
plot(-MDS[, 1], -MDS[, 2], type = "n", asp = 1)
text(-MDS[, 1], -MDS[, 2], labels = rownames(X), cex
    =1.5)
```

上述代码运行后可绘制如下图形，对比发现 MDS 准确度较高。

上述代码运行后可绘制如下图形，对比发现 MDS 准确度较高。





古典解的优良性

由欧式型距离阵的判定定理可知，距离阵 D 在 k 维实数空间中拟合构造点的古典解就是 X 的 k 维主坐标。

古典解的优良性

由欧式型距离阵的判定定理可知，距离阵 D 在 k 维实数空间中拟合构造点的古典解就是 X 的 k 维主坐标。

定理 2

X 的 k 维主坐标是将 X 中心化后 n 个样本的前 k 个主成分的值。

非度量方法

度量化模型

古典解 (基于主成分分析的思想):

$$d_{ij} = \widehat{d}_{ij} + \varepsilon_{ij}$$

非度量方法

度量化模型
古典解 (基于主成分分析的思想)：

$$d_{ij} = \widehat{d}_{ij} + \varepsilon_{ij}$$

式中，$\widehat{d}_{ij}$ 是拟合 d_{ij} 的值，ε_{ij} 为误差。

非度量方法

度量化模型
古典解 (基于主成分分析的思想)：

$$d_{ij} = \widehat{d}_{ij} + \varepsilon_{ij}$$

式中，$\widehat{d}_{ij}$ 是拟合 d_{ij} 的值，ε_{ij} 为误差。

非度量化模型
非古典解：

$$d_{ij} = f\left(\widehat{d}_{ij} + \varepsilon_{ij}\right)$$

非度量方法

度量化模型
古典解 (基于主成分分析的思想):

$$d_{ij} = \widehat{d}_{ij} + \varepsilon_{ij}$$

式中，$\widehat{d}_{ij}$ 是拟合 d_{ij} 的值，ε_{ij} 为误差。

非度量化模型
非古典解:

$$d_{ij} = f\left(\widehat{d}_{ij} + \varepsilon_{ij}\right)$$

式中，f 为一个未知的单调增加的函数。此时只能用 $|d_{ij}|$ 的秩来构造 $\widehat{d}_{ij}$.

Shepard-Kruskal 算法

已知相似系数矩阵 $D=(d_{ij})$，将其非对角元素由从小到大顺序排列起来: $d_{i_1j_1} \leqslant d_{i_2j_2} \leqslant \cdots \leqslant d_{i_mj_m}$, $m=\frac{1}{2}n(n-1)$, $i_l < j_l$, $l=1, 2, \cdots, m$;

Shepard-Kruskal 算法

已知相似系数矩阵 $D=(d_{ij})$，将其非对角元素由从小到大顺序排列起来: $d_{i_1j_1}\leqslant d_{i_2j_2}\leqslant\cdots\leqslant d_{i_mj_m}$, $m=\frac{1}{2}n(n-1)$, $i_l<j_l$, $l=1,\ 2,\ \cdots,\ m$;
设 $\widehat{X}_{n\times k}$ 是 k 维拟合构造点，相应的距离阵 $\widehat{D}=(\widehat{d}_{ij})$, 令

Shepard-Kruskal 算法

已知相似系数矩阵 $D=(d_{ij})$，将其非对角元素由从小到大顺序排列起来: $d_{i_1j_1}\leqslant d_{i_2j_2}\leqslant\cdots\leqslant d_{i_mj_m}$, $m=\frac{1}{2}n(n-1)$, $i_l<j_l$, $l=1,\ 2,\ \cdots,\ m$;

设 $\widehat{X}_{n\times k}$ 是 k 维拟合构造点，相应的距离阵 $\widehat{D}=(\widehat{d}_{ij})$, 令

$$S^2(\widehat{X})=\frac{\min\sum\limits_{i<j}(d^*_{ij}-\widehat{d}_{ij})^2}{\sum\limits_{i<j}d^2_{ij}}; \tag{20}$$

Shepard-Kruskal 算法

已知相似系数矩阵 $D=(d_{ij})$，将其非对角元素由从小到大顺序排列起来: $d_{i_1j_1}\leqslant d_{i_2j_2}\leqslant\cdots\leqslant d_{i_mj_m}$, $m=\frac{1}{2}n(n-1)$, $i_l<j_l$, $l=1,\,2,\,\cdots,\,m$;

设 $\widehat{X}_{n\times k}$ 是 k 维拟合构造点，相应的距离阵 $\widehat{D}=(\widehat{d}_{ij})$, 令

$$S^2(\widehat{X})=\frac{\min\sum\limits_{i<j}(d_{ij}^*-\widehat{d}_{ij})^2}{\sum\limits_{i<j}d_{ij}^2}; \tag{20}$$

若 k 固定，且能存在一个 $\widehat{X}_0$，使得 $S(\widehat{X}_0)=\min\limits_{\widehat{X}_{n\times k}}S(\widehat{X})\equiv S_k$, 则称 $\widehat{X}_0$ 为 k 维最佳拟合构造点;

Shepard-Kruskal 算法

已知相似系数矩阵 $D=(d_{ij})$，将其非对角元素由从小到大顺序排列起来：$d_{i_1j_1}\leqslant d_{i_2j_2}\leqslant\cdots\leqslant d_{i_mj_m}$, $m=\frac{1}{2}n(n-1)$, $i_l<j_l$, $l=1,\ 2,\ \cdots,\ m$;

设 $\widehat{X}_{n\times k}$ 是 k 维拟合构造点，相应的距离阵 $\widehat{D}=(\widehat{d}_{ij})$, 令

$$S^2(\widehat{X})=\frac{\min\sum\limits_{i<j}(d_{ij}^*-\widehat{d}_{ij})^2}{\sum\limits_{i<j}d_{ij}^2};\tag{20}$$

若 k 固定，且能存在一个 $\widehat{X}_0$，使得 $S(\widehat{X}_0)=\min\limits_{\widehat{X}_{n\times k}}S(\widehat{X})\equiv S_k$, 则称 $\widehat{X}_0$ 为 k 维最佳拟合构造点;

由于 S_k(称为压力指数 stress) 是 k 的单调下降序列，取 k 使 S_k 适当小。

Shepard-Kruskal 算法

已知相似系数矩阵 $D=(d_{ij})$，将其非对角元素由从小到大顺序排列起来: $d_{i_1j_1}\leqslant d_{i_2j_2}\leqslant\cdots\leqslant d_{i_mj_m}$, $m=\frac{1}{2}n(n-1)$, $i_l<j_l$, $l=1,\ 2,\ \cdots,\ m$;

设 $\widehat{X}_{n\times k}$ 是 k 维拟合构造点，相应的距离阵 $\widehat{D}=(\widehat{d}_{ij})$, 令

$$S^2(\widehat{X})=\frac{\min\sum\limits_{i<j}(d^*_{ij}-\widehat{d}_{ij})^2}{\sum\limits_{i<j}d^2_{ij}}; \tag{20}$$

若 k 固定，且能存在一个 $\widehat{X}_0$，使得 $S(\widehat{X}_0)=\min\limits_{\widehat{X}_{n\times k}}S(\widehat{X})\equiv S_k$, 则称 $\widehat{X}_0$ 为 k 维最佳拟合构造点;

由于 S_k(称为压力指数 stress) 是 k 的单调下降序列，取 k 使 S_k 适当小。

例如 $S_k<5\%$ 最好，$5\%\leqslant S_k\leqslant 10\%$ 次之，$S_k>10\%$ 较差。

MDS 非度量方法的 R 实现

MDS 非度量方法的 R 实现

```
# 需要导入MASS包
isoMDS(d, y = cmdscale(d, k), k = 2, maxit = 50, trace
    = TRUE, tol = 1e-3, p = 2)
Shepard(d, x, p = 2)
```

MDS 非度量方法的 R 实现

```r
# 需要导入MASS包
isoMDS(d, y = cmdscale(d, k), k = 2, maxit = 50, trace
    = TRUE, tol = 1e-3, p = 2)
Shepard(d, x, p = 2)
```

d 表示进行多维标度分析的距离矩阵，k 表示维度，默认取 2 维.

MDS 非度量方法的 R 实现

```
# 需要导入MASS包
isoMDS(d, y = cmdscale(d, k), k = 2, maxit = 50, trace
    = TRUE, tol = 1e-3, p = 2)
Shepard(d, x, p = 2)
```

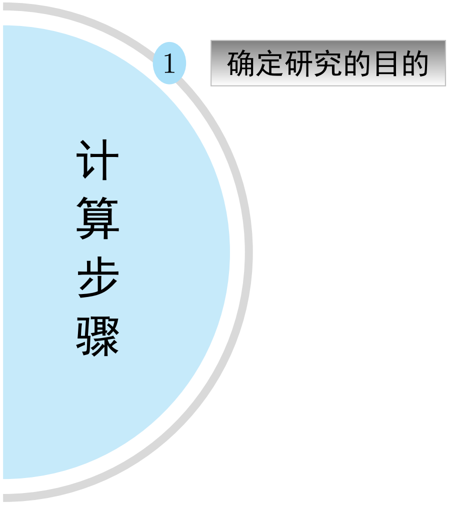
d 表示进行多维标度分析的距离矩阵，k 表示维度，默认取 2 维.

Shepard() 函数用于比较 MDS 结果的好坏。作图后，折线越趋近于一条平滑的斜线表明 MDS 降维的效果越好。

多维标度法的计算过程小结

计
算
步
骤

多维标度法的计算过程小结



多维标度法的计算过程小结

计算步骤

1 确定研究的目的

2 选择样品和变量

多维标度法的计算过程小结

计
算
步
骤

1 确定研究的目的

2 选择样品和变量

3 计算样品间的距离矩阵

多维标度法的计算过程小结

计算步骤

1. 确定研究的目的
2. 选择样品和变量
3. 计算样品间的距离矩阵
4. 分析样品间的距离矩阵

多维标度法的计算过程小结

计算步骤

1. 确定研究的目的
2. 选择样品和变量
3. 计算样品间的距离矩阵
4. 分析样品间的距离矩阵
5. 计算距离矩阵的古典解

多维标度法的计算过程小结

计算步骤

1. 确定研究的目的
2. 选择样品和变量
3. 计算样品间的距离矩阵
4. 分析样品间的距离矩阵
5. 计算距离矩阵的古典解
6. 检验模型的拟合结果

多维标度法的计算过程

多维标度法的计算过程

例 4：广东省各地区农村经济发展状况分析

地区	农业产值	林业产值	牧业产值	企业人数	企业总产值	利润总额
广州市	97.84	1.28	38.86	141.98	2089.55	121.07
深圳市	11.2	0.66	12.59	156.52	418.16	50.12
珠海市	5.67	0.11	3.6	17.39	360.58	10.58
汕头市	29.87	0.57	17.26	52.45	673.74	24.07
佛山市	52.39	0.29	32.14	90.77	1649.81	62.74
韶关市	47.82	4.47	18.44	27.91	144.51	16.14
河源市	33.57	3.1	12.84	12.62	51.25	4.73
梅州市	57.1	2.74	28.02	44.12	226.65	19.75

多维标度法的计算过程

多维标度法的计算过程

R 程序如下：

多维标度法的计算过程

R 程序如下：

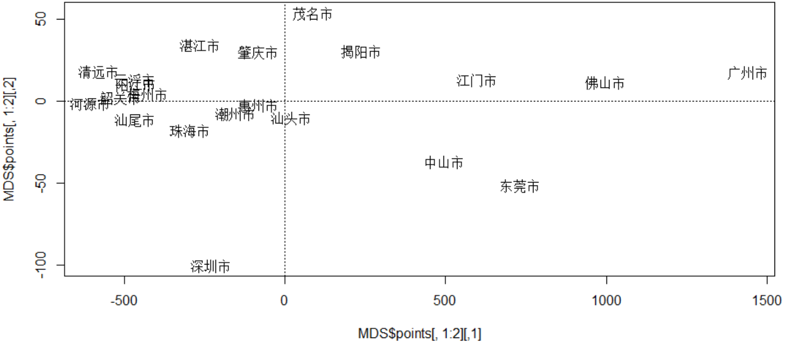
```
data <- openxlsx::read.xlsx("mvstats4.xlsx", sheet="d12
    .4", rowNames=T)
# 生成距离阵(欧氏距离)
D <- dist(data, method = "minkowski", p=2)
# 计算非度量解
MDS <- MASS::isoMDS(D); MDS
plot(MDS$points[, 1:2], type = "n")
abline(h=0, v=0, lty=3)
text(MDS$points[, 1:2], adj=.5, labels=rownames(data))
plot(MASS::Shepard(D, MASS::isoMDS(D)$points))
```

多维标度法的计算过程

结果及分析

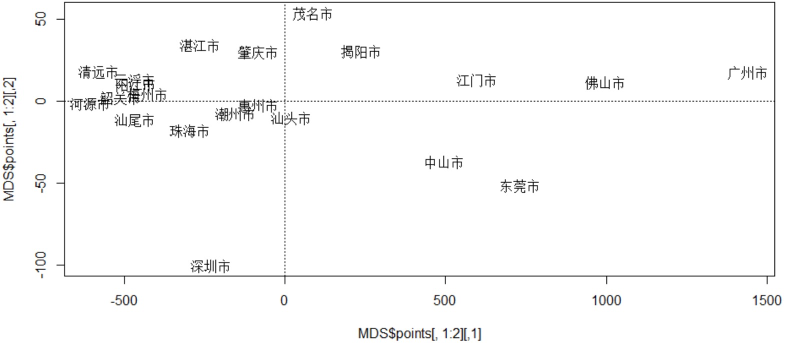
多维标度法的计算过程

结果及分析



多维标度法的计算过程

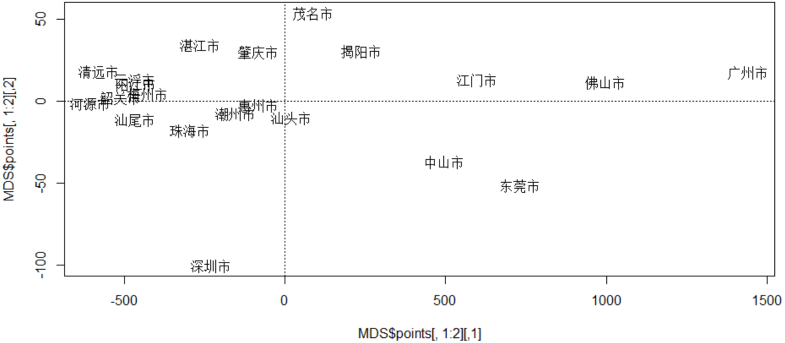
结果及分析



在综合排名中，广州市处于总排名的第一名，佛山市排在第二名，而深圳市明显大大落后；

多维标度法的计算过程

结果及分析



在综合排名中，广州市处于总排名的第一名，佛山市排在第二名，而深圳市明显大大落后；

茂名市、中山市、珠海市和江门市在农、林、牧业产值中表现优越。

本章小结——思维导图

本章小结——思维导图

- **多维标度法 MDS及R实现**
  - MDS的概念和方法
    - MDS的基本概念
    - MDS的基本方法
  - MDS的古典解(基于PCA)
    - 欧式型距离阵及其判定定理
    - 多维标度法的古典解`cmdscale()`
    - 古典解的优良性
  - MDS非度量方法
    - 非度量方法的思想
    - 非度量方法的算法`isoMDS()`
  - MDS的计算过程
    - 一、确定研究的目的
    - 二、选择样品与变量
    - 三、计算样本间的距离矩阵
    - 四、选择适当的求解方法
    - 五、选择适当的维数
    - 六、检验模型的拟合效果

参考文献

- 王斌会.
  多元统计分析及 R 语言建模 (第四版).
  暨南大学出版社, 2016.
- Warren S. Torgerson.
  Theory and Methods of Scaling.
  New York: John Wiley, 1958.
- 程其襄, 张奠宙.
  实变函数与泛函分析基础 (第三版).
  高等教育出版社, 2010, pages 299–301.

The End ✌

欢迎老师同学们批评指正!