



中山大學
SUN YAT-SEN UNIVERSITY

实验设计与数据分析

第十二讲：聚类分析

肖叶，副教授
环境科学与工程学院
中山大学

广州，2023

聚类分析

✿ 基本概念

- 聚类分析的一般过程
- 距离
- 相似系数

✿ 聚类方法

- 系统聚类法
- k均值法
- 模糊聚类法



聚类分析的研究内容

✿ 聚类分析

■ 是指：

- 依照某种准则对个体（样本或变量）进行分类的一种分析方法
- 又称聚簇分析、群分析、类分析
- Cluster Analysis

■ 常用于：

- 各种调查数据的分析处理
- 在环境科学中，常应用于生态系统污染状况的分类，采样点、采样地区污染程度的分类，也常用作数据预处理的手段（例如剔除一些不合理的样本或数据）。



聚类分析的研究内容

✿ 聚类分析的对象

- 处理 p 个变量的 n 组样本数据
- $X_{1i}, X_{2i}, \dots, X_{pi}, i = 1, 2, \dots, n$
- 可以对样本进行聚类
 - 称之为Q型聚类
 - 是本课程要重点讨论的
- 可以对变量进行聚类
 - 称之为R型聚类



聚类分析的研究内容

✿ 聚类分析的常见做法，分三大步：

- 针对要进行聚类的个体，定义一种能够反映个体之间亲疏程度的量
 - 对样本之间可以定义各种距离；对变量之间可以定义各种相似系数
- 以这些量为聚类依据，把一些相似程度大的个体聚合为一类，把另一些彼此间相似程度大的个体聚合为另外一类
 - 是一个动态过程
 - n 个样本，最少可聚成一类，最多可聚成 n 类
- 按聚类过程把不同的分类罗列出来，形成一个由小到大的分类系统
 - 可以用一张图来表示整个聚类过程
 - 聚类图或谱系图，可以直观地表示出个体之间的亲疏关系



聚类分析的研究内容

✿ 聚类分析的基本方法

■ 聚合法：

- 聚类开始时每个个体自成一类，然后将距离最近的类进行合并，使类的数目减少一个；
- 然后再将距离最近的类进行合并，使类的数目又减少一个；
- 如此做下去直至所有的个体都聚合成一类为止
- 例如：系统聚类法

■ 分裂法：

- 与聚合法相反，开始时将全部个体看成一类；
- 然后依据某种准则将其进行逐步分裂，直至分裂成所需要的分类数为止

■ 模糊聚类法：

- 把模糊数学理论应用于聚类分析



距离与相似系数

✿ 样本之间的距离:

- 表征第*i*组样本与第*j*组样本之间的亲疏程度，也称为非类似度，记作 d_{ij} 。

✿ 对于*p*个变量的*n*组样本数据 $(X_{1i}, X_{2i}, \dots, X_{pi}, i = 1, 2, \dots, n)$ 距离有多种定义方法:

- 闵可夫斯基距离 (Minkowski distance)

$$d_{ij} = \left(\sum_{k=1}^p |X_{ki} - X_{kj}|^g \right)^{1/g}$$

- 满足：非负性 $d_{ij} \geq 0$ ；同一性 $d_{ij} = 0$ ，当且仅当 $\mathbf{X}_i = \mathbf{X}_j$ ；对称性 $d_{ij} = d_{ji}$ ；直递性 $d_{ij} \leq d_{ik} + d_{kj}$



距离与相似系数

✿ 样本之间的距离:

- 表征第*i*组样本与第*j*组样本之间的亲疏程度, 也称为非类似度, 记作 d_{ij} 。

✿ 对于*p*个变量的*n*组样本数据 $(X_{1i}, X_{2i}, \dots, X_{pi}, i = 1, 2, \dots, n)$ 距离有多种定义方法:

- 闵可夫斯基距离中 $g = 1$, 绝对值距离 (曼哈顿距离)

$$d_{ij} = \sum_{k=1}^p |X_{ki} - X_{kj}|$$

- 闵可夫斯基距离中 $g = 2$, 欧式距离

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ki} - X_{kj})^2}$$

- 标准化欧式距离

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ki} - X_{kj})^2 / S_k^2}$$

- 切比雪夫距离

$$d_{ij} = \max_{1 \leq k \leq p} |X_{ki} - X_{kj}|$$



距离与相似系数

✿ 相似系数:

- 表征变量之间的亲疏程度，也称为类似度
- 利用相关系数

$$r_{ij} = \frac{\sum_{k=1}^n (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)}{\left[\sum_{k=1}^n (X_{ik} - \bar{X}_i)^2 \sum_{k=1}^n (X_{jk} - \bar{X}_j)^2 \right]^{1/2}}$$

- 利用夹角余弦

$$C_{ij} = \frac{\sum_{k=1}^n X_{ik} X_{jk}}{\left(\sum_{k=1}^n X_{ik}^2 \sum_{k=1}^n X_{jk}^2 \right)^{1/2}}$$



聚类分析

✿ 基本概念

- 聚类分析的一般过程
- 距离
- 相似系数

✿ 聚类方法

- 系统聚类法
- k均值法
- 模糊聚类法



系统聚类法

✿ 系统聚类法：

- 使用最多的聚类分析方法， hierarchical cluster analysis

✿ 基本操作步骤：

- 首先各样本自成一类，对于 n 组样本，就相当于有 n 类；
- 计算各类间的距离，将其中最近的两类进行合并；
- 计算新类与其余各类的距离，再将距离最近的两类合并；
- 重复以上步骤，直到所有的样本都聚为一大类时为止。

✿ 系统聚类过程中，必须要计算：

- 类与类之间的距离 D_{pq}
- 根据 D_{pq} 的定义不同，系统聚类法有多种类型



系统聚类法

✿ 最短距离法 (nearest neighbor method)

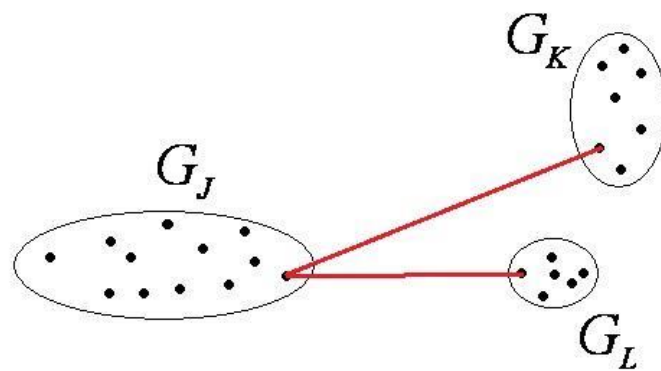
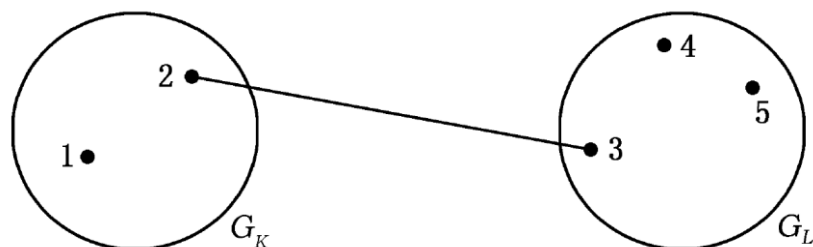
- 将两个类中最近样本的距离定义为两个类之间的距离

- $D_{KL} = \min_{i \in G_K, j \in G_L} d_{ij}$

- 递推公式

- $D_{MJ} = \min\{D_{KJ}, D_{LJ}\}$

- 最短距离法不适合对分离得很差的群体进行聚类。



系统聚类法

✿ 最长距离法 (furthest neighbor method)

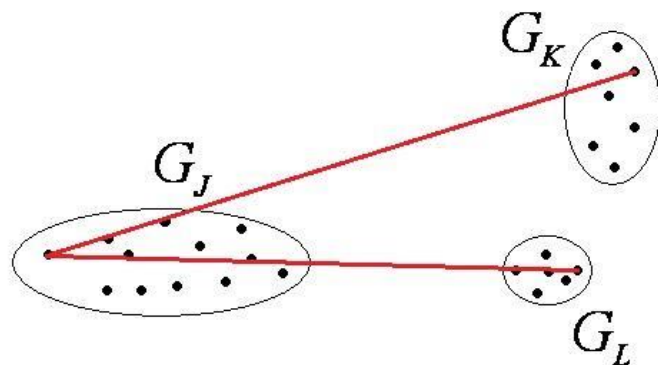
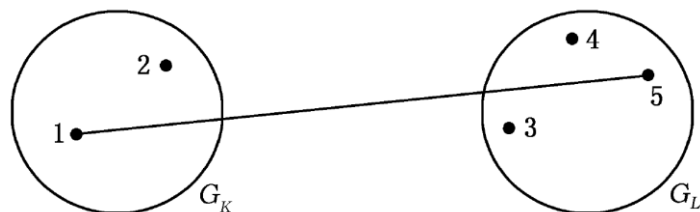
- 将两个类中最远样本之间的距离定义为两个类之间的距离

- $D_{KL} = \max_{i \in G_K, j \in G_L} d_{ij}$

- 递推公式

- $D_{MJ} = \max\{D_{KJ}, D_{LJ}\}$

- 最长距离法容易被异常值严重地扭曲



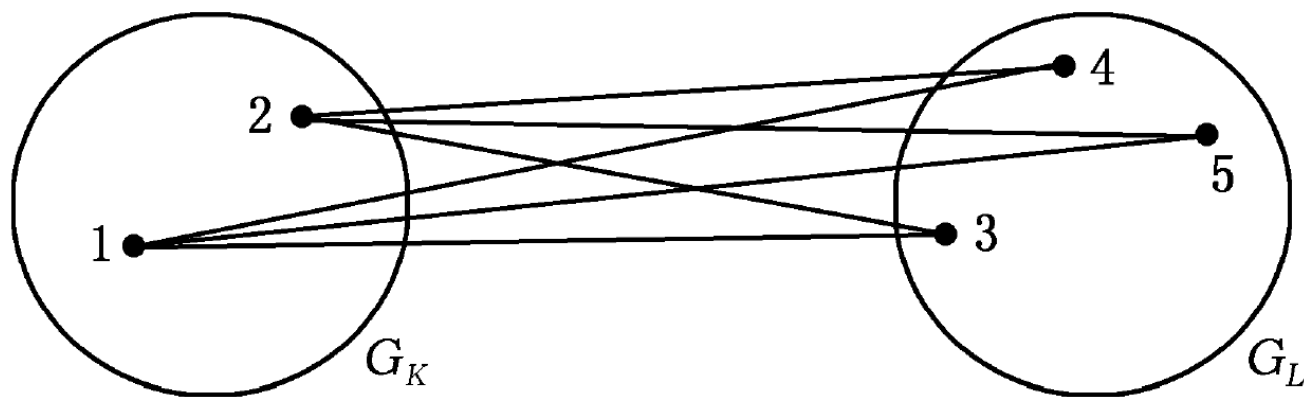
系统聚类法

✿ 类平均法 (group average method)

- 将两个类中样本之间的平均距离定义为两个类之间的距离

- $$D_{KL} = \frac{1}{n_K n_L} \sum_{i \in G_K, j \in G_L} d_{ij}$$

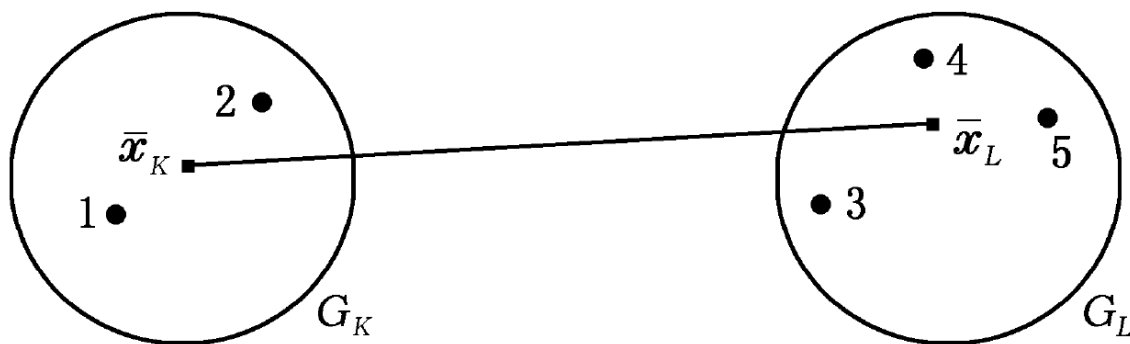
- 类平均法较好地利用了所有样品之间的信息，在很多情况下它被认为是一种比较好的系统聚类法。



系统聚类法

✿ 重心法 (centroid method)

- 将两个类的重心之间的距离定义为两个类之间的距离
- 设类 G_K 和 G_L 的重心 (均值) 分别为 \bar{x}_K 和 \bar{x}_L , 则 G_K 和 G_L 的平方距离定义为
- $D_{KL}^2 = d_{\bar{x}_K \bar{x}_L}^2 = (\bar{x}_K - \bar{x}_L)'(\bar{x}_K - \bar{x}_L)$
- 与其他系统聚类法相比, 重心法在处理异常值方面更稳健, 但是在别的方面一般不如类平均法或离差平方和法的效果好。



系统聚类法

✿ 离差平方和法 (Ward方法)

- G_K 和 G_L 的平方距离定义为

- $$D_{KL}^2 = \frac{n_K n_L}{n_K + n_L} (\bar{x}_K - \bar{x}_L)' (\bar{x}_K - \bar{x}_L)$$

- 离差平方和法使得两个大的类倾向于有较大的距离，因而不易合并；相反，两个小的类却因倾向于有较小的距离而易于合并。这往往符合我们对聚类的实际要求。

- 离差平方和法在许多场合下被认为是一种比较好的系统聚类法，但该方法对异常值很敏感。



最短距离法

✿ 准备工作:

- 规定样本之间的距离计算方法
- 用 G_1, G_2, \dots 表示类, 用 D_{pq} 表示类 G_p 和 G_q 之间的距离

✿ (1) 计算各样本两两距离的对称矩阵, 记为 $D_{(0)}$

- 开始视每个样本分别为一类, 这时显然应有 $D_{pq} = d_{pq}$

✿ (2) 选择距离矩阵 $D_{(0)}$ 中的最小元素

- 不失一般性设为 D_{pq} , 将 G_p 和 G_q 合并成一个新类 G_r , $G_r = \{G_p, G_q\}$

✿ (3) 计算新类 G_r 与其它各类的距离:

- $D_{rk} = \min(d_{ij}) = \min[\min(d_{ij}), \min(d_{ij})] = \min(D_{pk}, D_{qk})$

$$(i \in G_r, j \in G_k) \quad (i \in G_p, j \in G_k) \quad (i \in G_q, j \in G_k)$$



最短距离法

✿ (4) 获得新的距离矩阵

- 将距离阵 $D_{(0)}$ 的第 p , q 行及第 p , q 列删去, 再将 D_{rk} 放到第 p 行和第 p 列上, 得到的矩阵记为 $D_{(1)}$

✿ (5) 重复过程

- 对 $D_{(1)}$ 重复以上步骤, 直到所有样本成为一类为止

✿ 在整个聚类的过程中, 如果在某一步最小元素不止一个, 可以将其同时合并



k均值算法

✿ 定义:

- 根据各类中样本的均值进行聚类划分的一种聚类分析算法

✿ 输入:

- 聚类个数 k , n 个样本数据。

✿ 输出:

- k 个聚类, 并满足:
- 同一聚类中的对象相似度较高;
- 不同聚类中的对象相似度较小;
- 利用各聚类中样本均值所获得的“类中心”来计算聚类相似度。



k均值算法

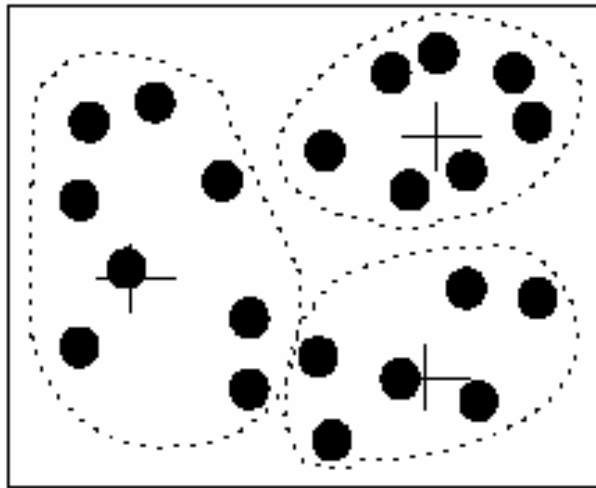
✿ 处理流程:

- 1. 从 n 个样本中任意选择 k 个样本作为初始聚类中心（种子、初始类中心）；
- 2. 根据每个聚类中样本的各变量均值（迭代类中心），计算每个样本与这些类中心的距离；并根据各样本到类中心的最小距离重新对相应样本进行划分；
- 3. 重新计算每个聚类的均值（新的类中心）。
- 4. 循环 2 和 3，直到每个聚类不再发生变化为止，或者达到终止迭代的判据要求为止；

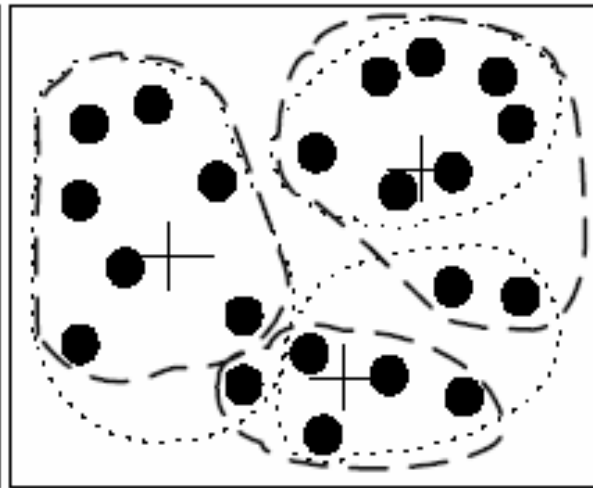


k均值算法

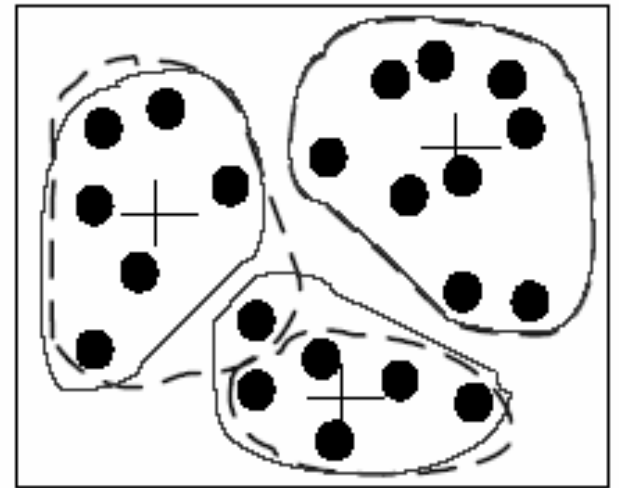
✿ 处理流程示例：（21个样本，聚成3类）



(a)



(b)



(c)

k均值算法

✿ 举例： 8个湖泊， 5个水质变量， 将8个湖泊聚成3类 ($k = 3$)

	叶绿素	TP	TN	高锰酸盐指数	BOD ₅
湖泊	mg/m ³	mg/L	mg/L	mg/L	mg/L
Lake1	4.90	0.04	0.33	6.06	1.52
Lake2	4.91	0.04	0.24	5.40	1.65
Lake3	3.35	0.03	0.87	16.16	1.35
Lake4	38.32	0.19	4.61	22.39	7.51
Lake5	3.32	0.03	0.15	3.92	2.50
Lake6	0.79	0.01	0.84	9.25	2.37
Lake7	33.09	0.17	4.03	9.23	3.34
Lake8	38.31	0.18	0.48	17.03	3.69



k均值算法

✿ 初始类中心:

Initial Cluster Centers

	Cluster		
	1	2	3
叶绿素	38.32	33.09	3.35
总磷	.19	.17	.03
总氮	4.61	4.03	.87
高锰酸盐	22.39	9.23	16.16
生化需氧	7.51	3.34	1.35

- 说明:
- (1) 湖泊4、湖泊7、湖泊3被系统选作初始类中心。
- (2) 系统随机选取“种子”，也可以由用户事先指定。



k均值算法

✿ 聚类结果:

- 说明:
- (1) 聚成3类。
- (2) 第一类: 湖泊4、 8
- (3) 第二类: 湖泊7
- (4) 第三类: 湖泊1、 2、 3、 5、 6
- (5) 每个样本到所属类中心的距离
- $\text{Distance}(7)=0.000$; $\text{Distance}(4)=\text{Distance}(8)=3.885$

Cluster Membership

Case Number	湖泊编号	Cluster	Distance
1	1	3	2.578
2	2	3	3.137
3	3	3	8.029
4	4	1	3.885
5	5	3	4.299
6	6	3	2.942
7	7	2	.000
8	8	1	3.885



k均值算法

✿ 最终类中心:

- 说明:
- (1) 类中心是各类中所包含样本的均值。

- (2) 第1类TN:

$$2.55 = (4.61 + 0.48)/2$$

- (3) 第3类TP:

$$0.03 = (0.04 + 0.04 + 0.03 + 0.03 + 0.01)/5$$

Final Cluster Centers

	Cluster		
	1	2	3
叶绿素	38.32	33.09	3.45
总磷	.19	.17	.03
总氮	2.55	4.03	.49
高锰酸盐	19.71	9.23	8.16
生化需氧	5.60	3.34	1.88



k均值算法

✿ 各类之间的距离:

- 说明:
- (1) 各类之间的距离就是类中心之间的距离。
- (2) 采用的是欧氏距离。

Distances between Final Cluster Centers

Cluster	1	2	3
1		12.018	36.971
2	12.018		29.902
3	36.971	29.902	



k均值算法

✿ 迭代过程:

- 说明:
- (1) 共迭代两次就完成了聚类。
- (2) 在第一次迭代中, 第一、三类中心移动了一定的距离 (例如, 第一类中心由初始设定的湖泊4变成了湖泊4和湖泊8的均值)。
- (3) 第二个类中心始终没有变过 (即湖泊7的值一直是第二类的均值)。

Iteration History^a

Iteration	Change in Cluster Centers		
	1	2	3
1	3.885	.000	8.029
2	.000	.000	.000



模糊聚类法

✿ 模糊聚类法的产生：

- 将模糊数学理论应用于聚类分析

✿ 模糊聚类分析的基本过程：

- (1) 计算样本或变量间的相似系数，建立模糊相似矩阵；
- (2) 利用模糊运算对相似矩阵进行一系列的合成改造，生成模糊等价矩阵；
- (3) 最后根据不同的截取水平 λ 对模糊等价矩阵进行截取分类



模糊聚类法

✱ 设某环境系统有 n 个样本，每个样本用 p 个变量来表征，构成原始数据矩阵 $\mathbf{X} = (x_{ij})_{n \times p}$ 。以样本分类为例，介绍模糊聚类分析的步骤：

- 建立模糊相似矩阵 $\mathbf{R} = (S_{ij})_{n \times n}$ ，其中 S_{ij} 为相似系数，其定义可以有多种形式：

- 夹角余弦：
$$S_{ij} = \frac{\sum_{k=1}^p X_{ik}X_{jk}}{(\sum_{k=1}^p X_{ik}^2 \sum_{k=1}^p X_{jk}^2)^{1/2}}$$

- 由相关系数转化而来：
$$r_{ij} = \frac{\sum_{k=1}^p (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)}{[\sum_{k=1}^p (X_{ik} - \bar{X}_i)^2 \sum_{k=1}^p (X_{jk} - \bar{X}_j)^2]^{1/2}} \quad S_{ij} = 0.5(r_{ij} + 1)$$

- 由距离转化而来：

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ki} - X_{kj})^2} \quad S_{ij} = \frac{1}{1 + d_{ij}} \quad \text{或者} \quad S_{ij} = 1 - \frac{d_{ij}}{\max_{1 \leq i, j \leq n} \{d_{ij}\}}$$



模糊聚类法

✿ 模糊聚类分析的步骤:

- (2) 创建模糊等价矩阵 \mathbf{R}^*
- 设 \mathbf{R} 为一个 n 阶模糊相似矩阵, 对 \mathbf{R} 进行模糊复合运算 (模糊矩阵积), 记为 $\mathbf{R} \circ \mathbf{R} = \mathbf{R}^2$
- 利用模糊矩阵积运算, 依次计算 $\mathbf{R}^2 \circ \mathbf{R}^2 = \mathbf{R}^4$, $\mathbf{R}^4 \circ \mathbf{R}^4 = \mathbf{R}^8$,
- 直到第 k 步使得 $\mathbf{R}^{2k} \circ \mathbf{R}^{2k} = \mathbf{R}^{4k} = \mathbf{R}^{2k} = \mathbf{R}^* = (s_{ij}^*)_{n \times n}$ 为止, 则 \mathbf{R}^* 就是模糊等价矩阵
- 模糊矩阵积运算法则为:

$$\mathbf{R} \circ \mathbf{R} = \mathbf{R}^2 = (s_{ij}^2) \quad s_{ij}^2 = \vee \{s_{ik} \wedge s_{kj}\}$$

- \vee 为并运算, \wedge 为交运算, 分别取较大者和较小者

$$s_{ij}^2 = \max\{\min\{s_{i1}, s_{1j}\}, \min\{s_{i2}, s_{2j}\}, \dots, \min\{s_{in}, s_{nj}\}\}$$



模糊聚类法

✿ 模糊聚类分析的步骤:

- (3) 选取截取水平 λ ($0 < \lambda < 1$), 对样本进行模糊聚类
- 依次由大到小选取一系列的 λ ($0 < \lambda < 1$), 对矩阵 \mathbf{R}^* 进行截取, 得到截

取矩阵: $\mathbf{R}_\lambda^* = (r_{ij})_{n \times n}$

- 截取原则为:

$$r_{ij} = \begin{cases} 1 & s_{ij}^* \geq \lambda \\ 0 & s_{ij}^* < \lambda \end{cases}$$

- 将 \mathbf{R}_λ^* 中相同行所对应的样本归为一类, 并由此画出聚类图
 - λ 越大, 分组越多; λ 越小, 分组越少



模糊聚类法

✿ 举例：下表为某地区6个采样点的SO₂、NO_x、TSP (mg/m³)的平均监测值，试对这6个采样点进行模糊聚类。（提示：可以将欧氏距离作为样本之间的距离）

	1#	2#	3#	4#	5#	6#
SO ₂	0.227	0.147	0.227	0.316	0.145	0.392
NO _x	0.180	0.213	0.253	0.213	0.220	0.280
TSP	1.214	0.678	1.871	0.974	0.754	1.221



模糊聚类法

✿ 解：（1）计算6个采样点两两之间的欧氏距离矩阵D

$$D = \begin{pmatrix} 0 & 0.543 & 0.661 & 0.258 & 0.469 & 0.193 \\ 0.543 & 0 & 1.196 & 0.341 & 0.076 & 0.599 \\ 0.661 & 1.196 & 0 & 0.902 & 1.120 & 0.671 \\ 0.258 & 0.341 & 0.902 & 0 & 0.279 & 0.267 \\ 0.469 & 0.076 & 1.120 & 0.279 & 0 & 0.532 \\ 0.193 & 0.599 & 0.671 & 0.267 & 0.532 & 0 \end{pmatrix}$$

✿ （2）计算相似系数，得到模糊相似矩阵R

$$s_{ij} = \frac{1}{1 + d_{ij}} \quad R = \begin{pmatrix} 1 & 0.648 & 0.602 & 0.795 & 0.681 & 0.838 \\ 0.648 & 1 & 0.455 & 0.746 & 0.929 & 0.625 \\ 0.602 & 0.455 & 1 & 0.526 & 0.472 & 0.598 \\ 0.795 & 0.746 & 0.526 & 1 & 0.781 & 0.789 \\ 0.681 & 0.929 & 0.472 & 0.781 & 1 & 0.653 \\ 0.838 & 0.625 & 0.598 & 0.789 & 0.653 & 1 \end{pmatrix}$$



模糊聚类法

✿ (3) 根据模糊复合运算法则，创建模糊等价矩阵 R^*

$$R = \begin{pmatrix} 1 & 0.648 & 0.602 & 0.795 & 0.681 & 0.838 \\ 0.648 & 1 & 0.455 & 0.746 & 0.929 & 0.625 \\ 0.602 & 0.455 & 1 & 0.526 & 0.472 & 0.598 \\ 0.795 & 0.746 & 0.526 & 1 & 0.781 & 0.789 \\ 0.681 & 0.929 & 0.472 & 0.781 & 1 & 0.653 \\ 0.838 & 0.625 & 0.598 & 0.789 & 0.653 & 1 \end{pmatrix}$$
$$R \circ R = R^2 = \begin{pmatrix} 1 & 0.746 & 0.602 & 0.795 & 0.781 & 0.838 \\ 0.746 & 1 & 0.602 & 0.781 & 0.929 & 0.746 \\ 0.602 & 0.602 & 1 & 0.602 & 0.602 & 0.602 \\ 0.795 & 0.781 & 0.602 & 1 & 0.781 & 0.795 \\ 0.781 & 0.929 & 0.602 & 0.781 & 1 & 0.781 \\ 0.838 & 0.746 & 0.602 & 0.795 & 0.781 & 1 \end{pmatrix}$$
$$R^2 \circ R^2 = R^4 = \begin{pmatrix} 1 & 0.781 & 0.602 & 0.795 & 0.781 & 0.838 \\ 0.781 & 1 & 0.602 & 0.781 & 0.929 & 0.781 \\ 0.602 & 0.602 & 1 & 0.602 & 0.602 & 0.602 \\ 0.795 & 0.781 & 0.602 & 1 & 0.781 & 0.795 \\ 0.781 & 0.929 & 0.602 & 0.781 & 1 & 0.781 \\ 0.838 & 0.781 & 0.602 & 0.795 & 0.781 & 1 \end{pmatrix}$$

$$R^4 \circ R^4 = R^8 = R^4 = R^*$$



模糊聚类法

✿ (4) 进行模糊等价聚类

- 分别取 $\lambda = 0.929, 0.838, 0.795, 0.781, 0.602$, \mathbf{R}^* 中元素大于或等于 λ 者改为1, 否则为0, 得到不同的 λ 截取矩阵。

$$R_{0.929}^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R_{0.838}^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- 从 $R_{0.929}^*$ 知, 在 $\lambda = 0.929$ 水平上, 样本分为5类: $\{1\#$, $\{2\#, 5\#$, $\{3\#$, $\{4\#$, $\{6\#$ };
- 从 $R_{0.838}^*$ 知, 在 $\lambda = 0.838$ 水平上, 样本分为4类: $\{1\#, 6\#$, $\{2\#, 5\#$, $\{3\#$, $\{4\#$;



模糊聚类法

✿ (4) 进行模糊等价聚类

- 分别取 $\lambda = 0.929, 0.838, 0.795, 0.781, 0.602$, R^* 中元素大于或等于 λ 者改为1, 否则为0, 得到不同的 λ 截取矩阵。

$$R_{0.795}^* = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \quad R_{0.781}^* = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \quad R_{0.602}^* = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

- 从 $R_{0.795}^*$ 知, 在 $\lambda = 0.795$ 水平上, 样本分为3类: $\{1\#, 4\#, 6\#\}$, $\{2\#, 5\#\}$, $\{3\#\}$;
- 从 $R_{0.781}^*$ 知, 在 $\lambda = 0.781$ 水平上, 样本分为2类: $\{1\#, 4\#, 6\#, 2\#, 5\#\}$, $\{3\#\}$;
- 从 $R_{0.602}^*$ 知, 在 $\lambda = 0.602$ 水平上, 所有样本聚为一类。





中山大學
SUN YAT-SEN UNIVERSITY

谢谢!