聚类任务

黄晟

huangsheng@cqu.edu.cn

办公室：信息大楼B701

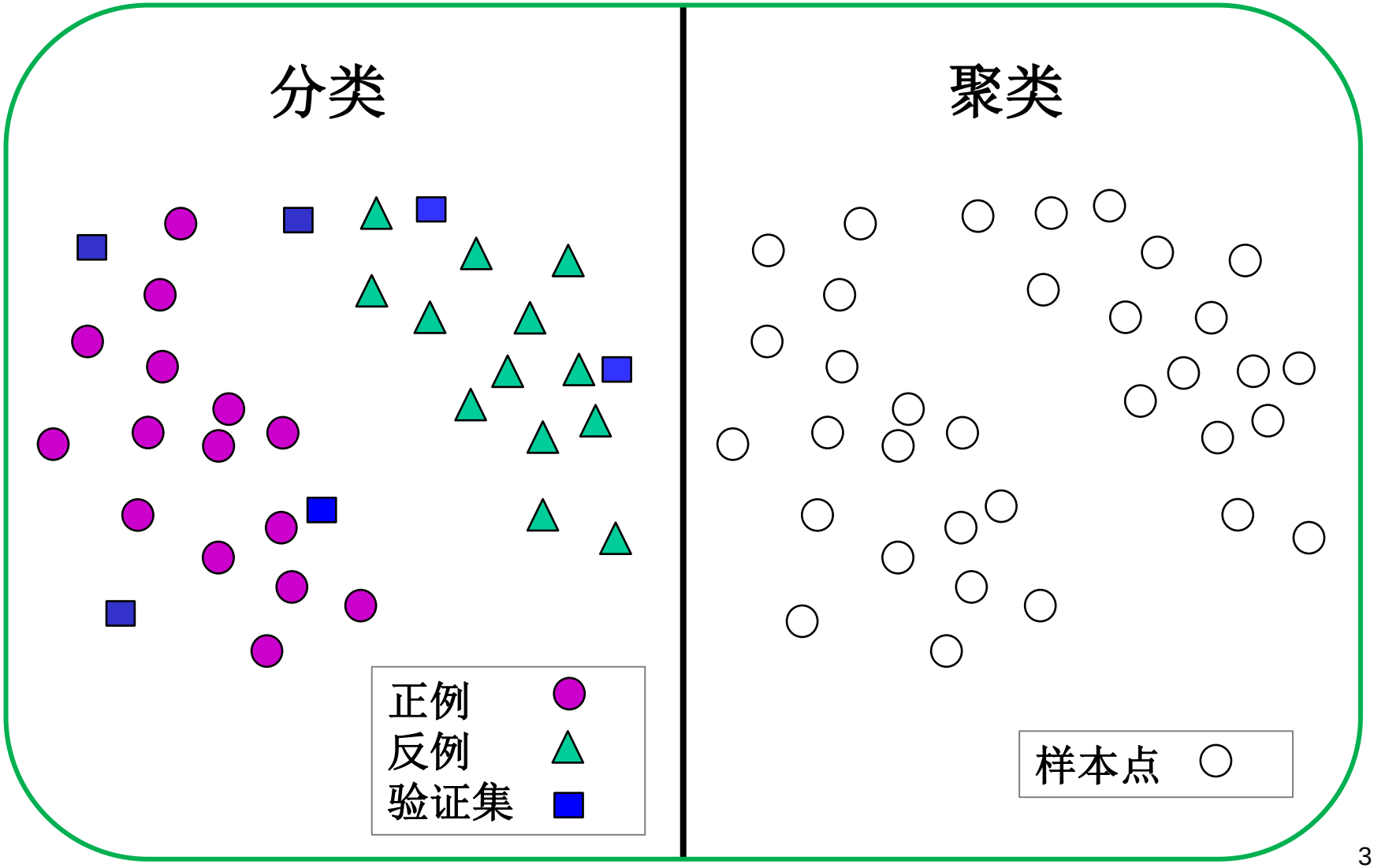
7.1 聚类任务简介

黄晟

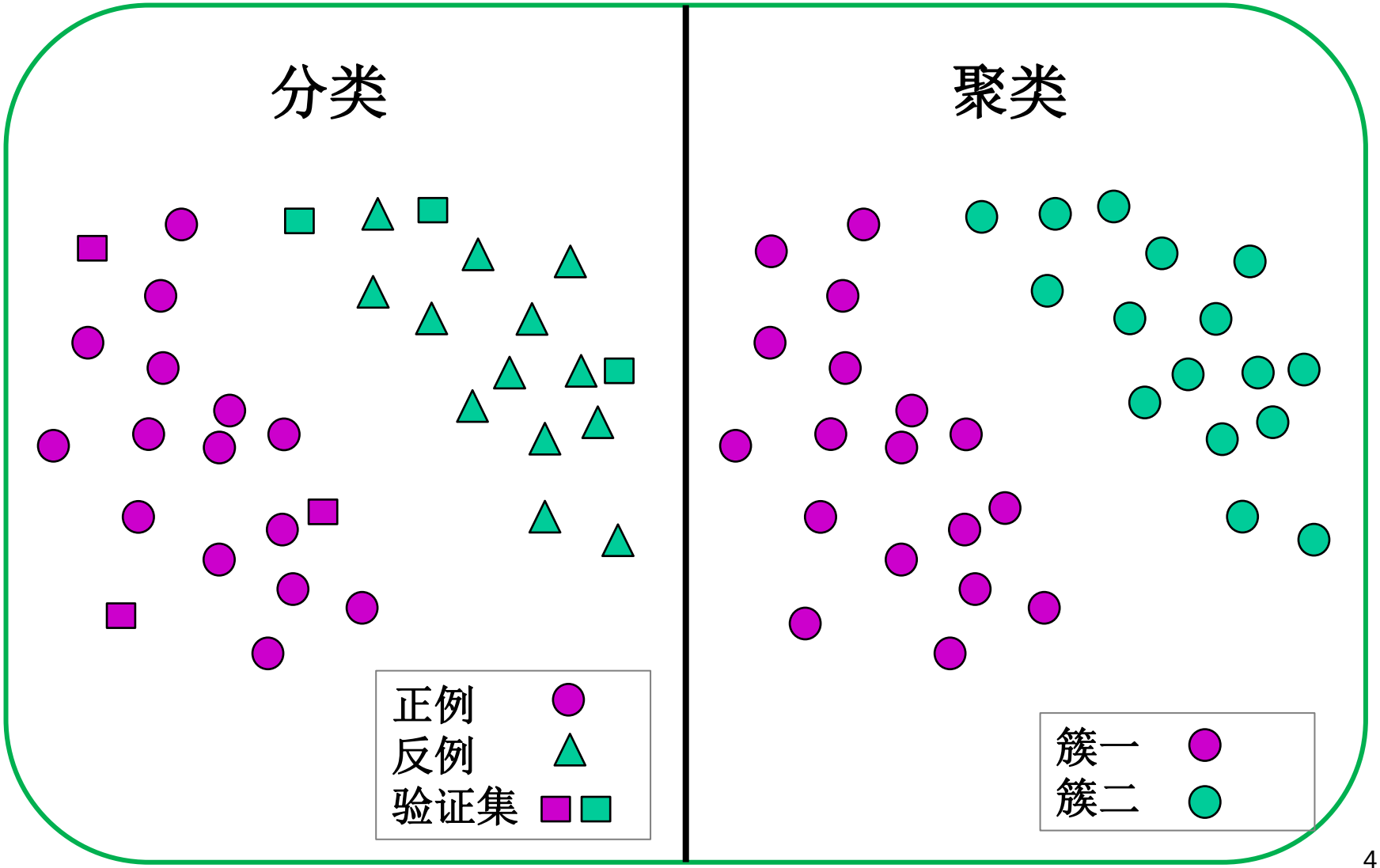
huangsheng@cqu.edu.cn

办公室：信息大楼B701

分类与聚类



分类与聚类



聚类问题

聚类：根据某种相似性，把一组数据划分成若干个簇的过程。

难点一：相似性很难精准定义！

难点二：可能存在的划分太多！

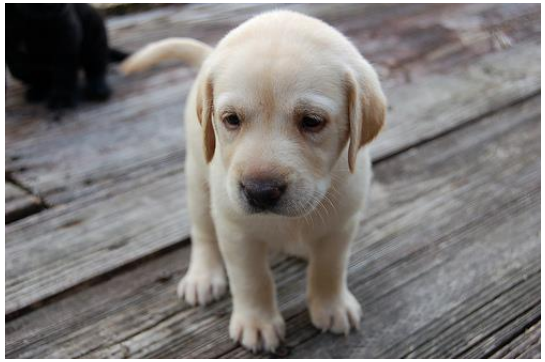
难点三： 若干个簇 = ？

1. 相似性定义？

相似？

相似？

相似？



2. 可能的划分？

假设我们要把 n = 5个不同的数据点放入不相同的 k = 2 个无差别的盒子中，有几种方案？

$$\mathcal{C}_1 = \{1\}, \mathcal{C}_2 = \{2,3,4,5\}$$
$$\mathcal{C}_1 = \{2\}, \mathcal{C}_2 = \{1,3,4,5\}$$
$$\mathcal{C}_1 = \{3\}, \mathcal{C}_2 = \{1,2,4,5\}$$
$$\mathcal{C}_1 = \{4\}, \mathcal{C}_2 = \{1,2,3,5\}$$
$$\mathcal{C}_1 = \{5\}, \mathcal{C}_2 = \{1,2,3,4\}$$

$$\mathcal{C}_1 = \{1,2\}, \mathcal{C}_2 = \{3,4,5\} \quad \mathcal{C}_1 = \{1,3\}, \mathcal{C}_2 = \{2,4,5\}$$
$$\mathcal{C}_1 = \{1,4\}, \mathcal{C}_2 = \{2,3,5\} \quad \mathcal{C}_1 = \{1,5\}, \mathcal{C}_2 = \{2,3,4\}$$
$$\mathcal{C}_1 = \{2,3\}, \mathcal{C}_2 = \{1,4,5\} \quad \mathcal{C}_1 = \{2,4\}, \mathcal{C}_2 = \{1,3,5\}$$
$$\mathcal{C}_1 = \{2,5\}, \mathcal{C}_2 = \{1,3,4\} \quad \mathcal{C}_1 = \{3,4\}, \mathcal{C}_2 = \{1,2,5\}$$
$$\mathcal{C}_1 = \{3,5\}, \mathcal{C}_2 = \{1,2,4\} \quad \mathcal{C}_1 = \{4,5\}, \mathcal{C}_2 = \{1,2,3\}$$

2. 可能的划分？

第二类 Stirling 数：集合的一个划分，表示将 n 个不同的元素拆分成 k 个集合的方案数, 记为 $S(n,k)$。

$$S(n,k)=\frac{1}{k!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^n$$

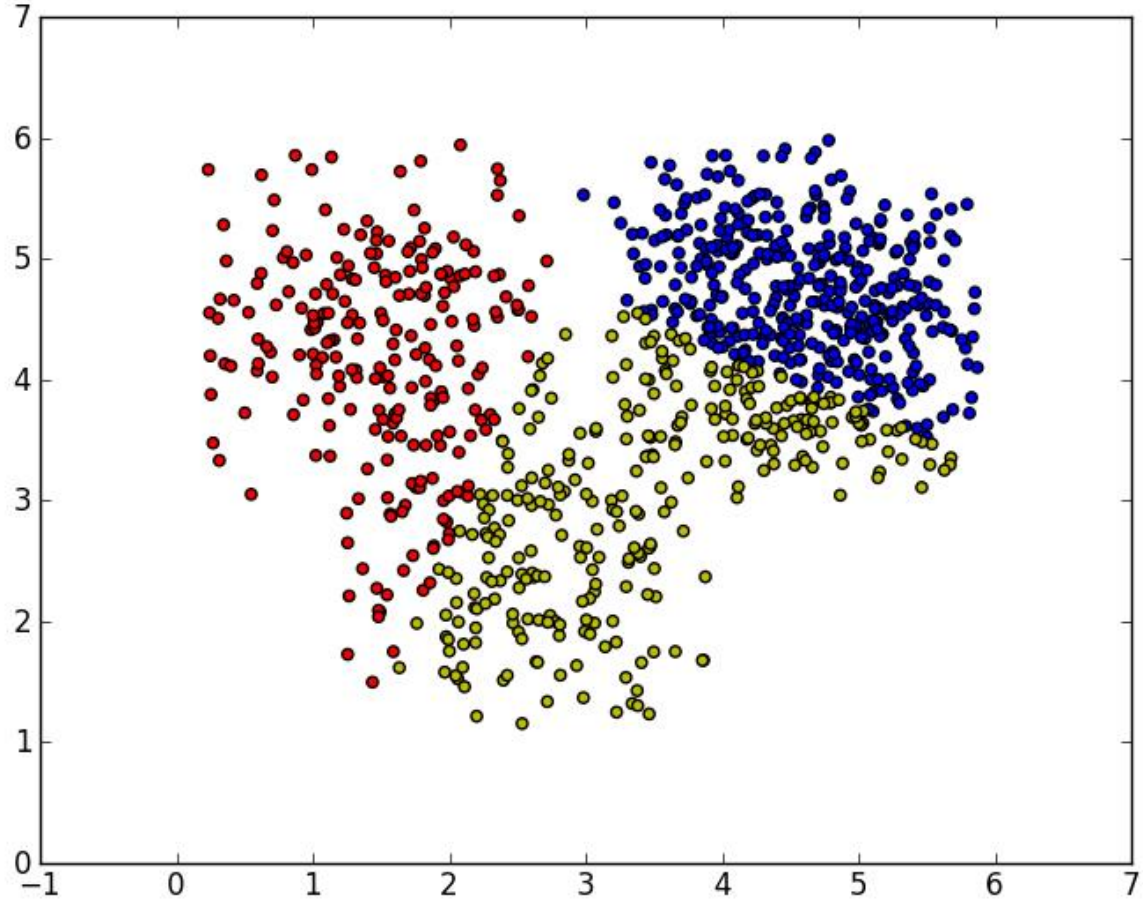
随着集合元素 n 和子集个数 k 的增加，方案数 $S(n,k)$ 呈爆炸式的增长！！！

2. 可能的划分？

帕斯卡三角形：

n=0	1
n=1	0 1
n=2	0 1 1
n=3	0 1 3 1
n=4	0 1 7 6 1
n=5	0 1 15 25 10 1
n=6	0 1 31 90 65 15 1
n=7	0 1 63 301 350 140 21 1
n=8	0 1 127 966 1701 1050 266 28 1
n=9	0 1 255 3025 7770 6951 2646 462 36 1

3. 聚类个数？



聚类问题

聚类：根据某种相似性，把一组数据划分成若干个簇的过程。

难点一：相似性很难精准定义！

--- 各种距离，度量学习。

难点二：可能存在的划分太多！

--- 避免穷举，优化算法。

难点三： 若干个簇 = ？

--- 预先给定，算法自适应。

7.2 K-means 算法

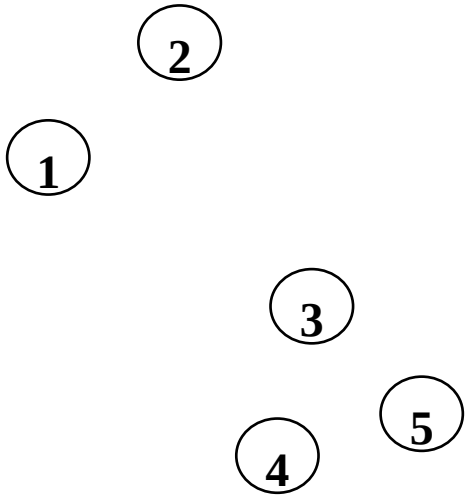
黄晟

huangsheng@cqu.edu.cn

办公室：信息大楼B701

人造例子

如何对下面数据进行聚类？



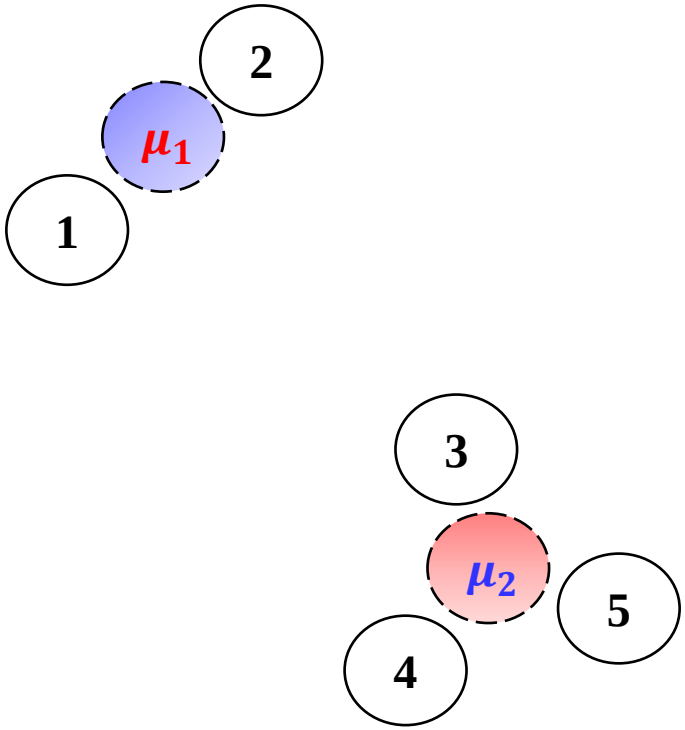
相似性： 欧式距离

簇个数： $\mathbf{k = 2}$

最优化分： $\boldsymbol{C_1 = \{1, 2\}}$, $\boldsymbol{C_2 = \{3, 4, 5\}}$

簇的中心

聚类问题可以通过为每个簇寻找合适的中心来实现。

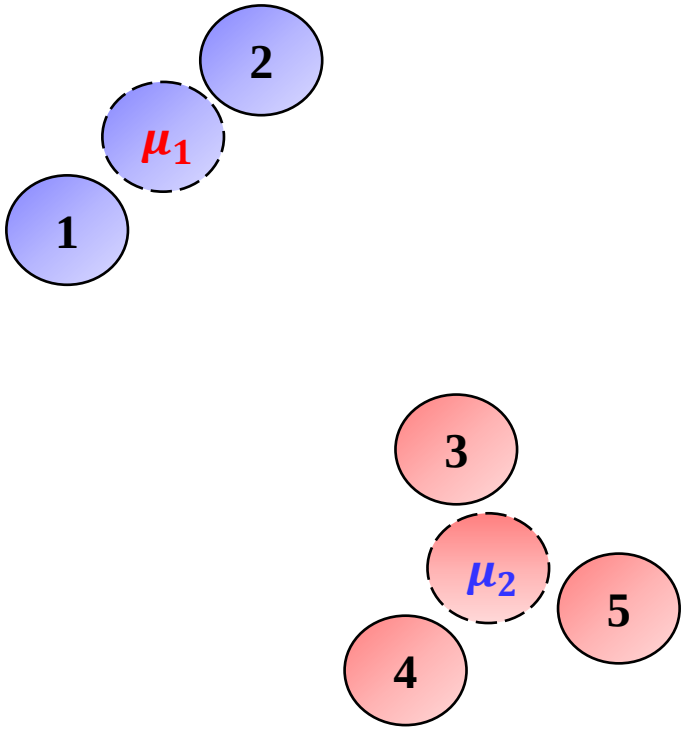


假设每个簇的中心已经找到，可以把所有数据点分配到距离它最近的中心所在的簇。

$$j^* = \underset{j}{\operatorname{argmin}}\ \boldsymbol{dist}(\boldsymbol{x}_i, \boldsymbol{\mu}_j)$$

簇的中心

聚类问题可以通过为每个簇寻找合适的中心来实现。



假设每个簇的中心已经找到，可以把所有数据点分配到距离它最近的中心所在的簇。

$$j^* = \underset{j}{\operatorname{argmin}}\ \boldsymbol{dist}(\boldsymbol{x}_i, \boldsymbol{\mu}_j)$$

K-means

给定数据$\{x_i\}_{i=1,\ldots,n}$以及簇的个数 k， K-means 模型可以写成下面的优化模型：

$$\underset{\mu_j,\, c_j}{\operatorname{argmin}} \sum_{j=1}^{k} \sum_{i \in c_j} \|x_i - \mu_j\|^2$$

变量：

- c_j： 第 j 个簇。
- μ_j： 第 j 个簇的中心。

K-means

目标函数对中心 $\boldsymbol{\mu_j}$ 求偏导，我们有

$$\frac{\partial\left(\sum_{j=1}^{k}\sum_{i\in c_j}\left\|\boldsymbol{x_i}-\boldsymbol{\mu_j}\right\|^2\right)}{\partial\boldsymbol{\mu_j}}=-2\sum_{i\in c_j}(\boldsymbol{x_i}-\boldsymbol{\mu_j})$$

即

$$\boldsymbol{\mu_j}=\frac{1}{|c_j|}\sum_{i\in c_j}\boldsymbol{x_i}$$

因此，上述模型通常被称作 **K-means**。

K-means

K-means 模型：

$$\underset{\mathcal{C}_j}{\operatorname{argmin}} \sum_{j=1}^{k} \sum_{i \in \mathcal{C}_j} \left\| \boldsymbol{x_i} - \frac{\mathbf{1}}{|\mathcal{C}_j|} \sum_{i \in \mathcal{C}_j} \boldsymbol{x_i} \right\|^{\mathbf{2}}$$

可能的划分数：

$$\mathrm{S}(n,k) = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n$$

非凸组合优化问题，NP-难！！！

非凸优化问题

求解非凸组合优化问题的两种常见方法：

- 启发式方法 (Heuristic method)：一个基于直观或经验构造的算法，在可接受的时间内下给出待解决组合优化问题每一个实例的一个可行解，该可行解与最优解的偏离程度一般不能被预计。
- 松弛方法 (Relaxation method)：对组合优化问题进行适当的松弛，将其转化为多项式时间内可解的优化问题，松弛后问题的解不是原组合优化问题的解，需要适当的后处理。

Lloyd 算法

给定数据$\{x_i\}_{i=1,\dots,n}$以及簇的个数 k：

初始化：随机选取 k 个簇的中心 $\{\mu_j\}_{j=1,\dots,k}$

重复下面迭代过程直到收敛：

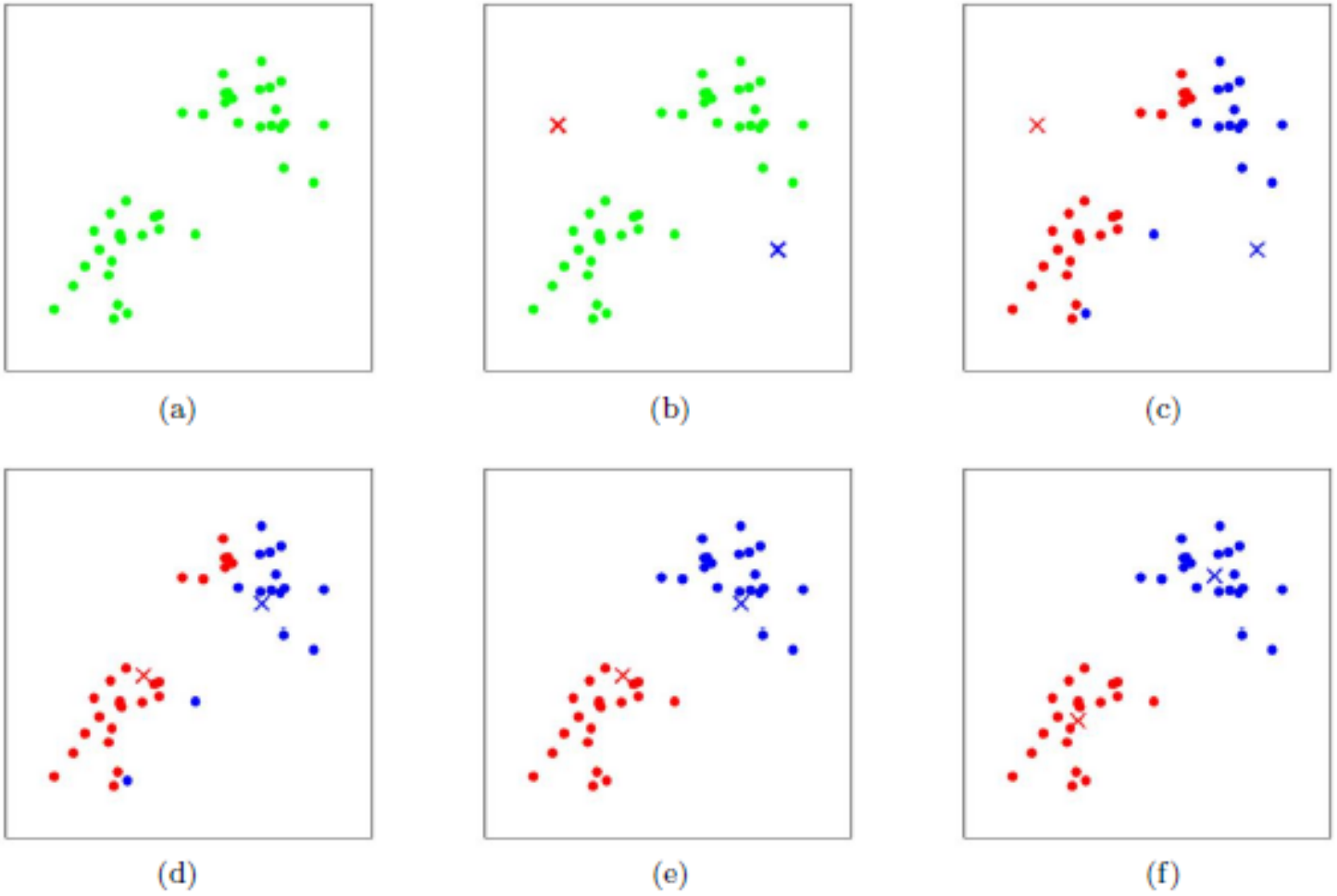
- 划分步骤：对于每一个数据点 x_i，计算其应该属于的簇

$$\operatorname*{argmin}_j \|x_i - \mu_j\|_2^2$$

- 更新步骤：重新计算每个簇的中心

$$\mu_j = \frac{1}{|c_j|}\sum_{i \in c_j} x_i$$

Lloyd 算法



Lloyd 算法

优势：

- Lloyd 算法属于EM算法（期望最大化），可以保证收敛到K-means问题的局部最优解。
- Lloyd 算法的速度快，计算复杂度为O(nk)。
- Lloyd 算法思想简单，容易实现，可拓展性强。

劣势：

- 簇的个数 k 需要预先给定。
- 聚类结果依赖于初值的选取。

K-means++

K-means++是一种初始化方法，目的是改进K-means算法对初值的影响。

Step 1：随机选取一个簇的中心 μ_1

Step 2: 选择离已有中心最远的数据点作为新的中心

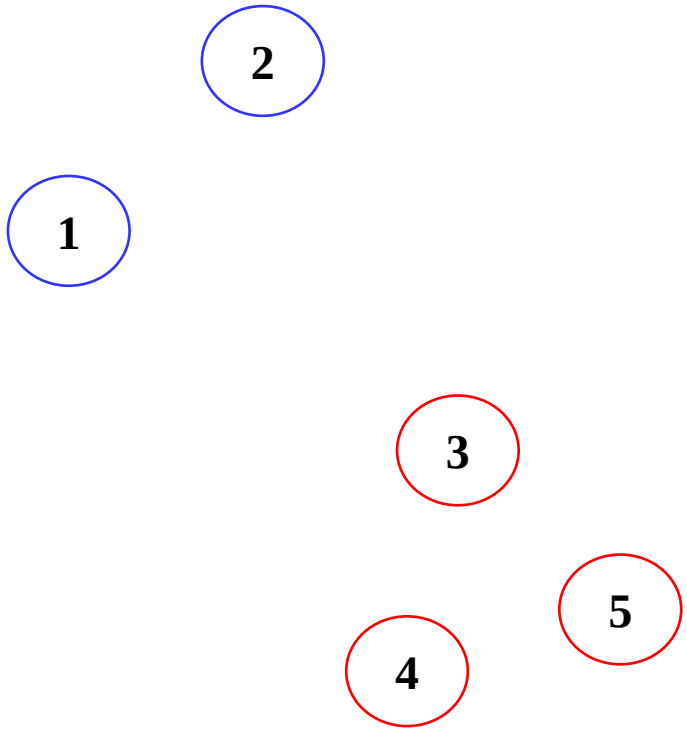
Step 3: 重复Step 2直到选出 k 个中心

Step 4: 运行K-means算法

示性矩阵

示性矩阵（**Indicator matrix**）：

$$\widetilde{H} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \in \mathbb{R}^{n \times k}$$



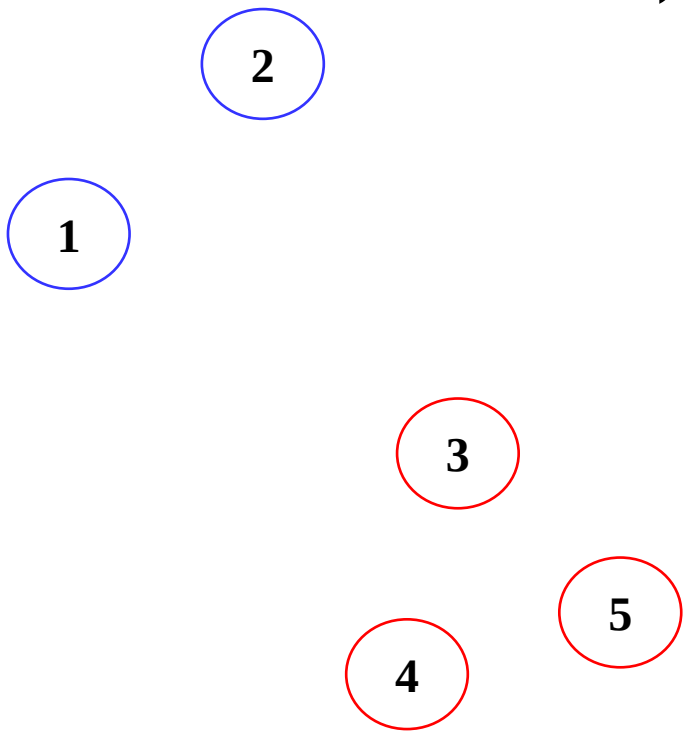
其中：

$$\widetilde{H_{ij}} = 1 \Leftrightarrow i \in \mathcal{C}_j$$

示性矩阵

标准化示性矩阵 (Normalized Indicator matrix)

$$H = \begin{pmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{pmatrix} \in \mathbb{R}^{n \times k}$$



其中：

$$H_{ij} = 1/\sqrt{|\mathcal{C}_j|} \iff i \in \mathcal{C}_j$$

示性矩阵

- $H^T H = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{pmatrix} = \boldsymbol{I}$
- $HH^T \mathbf{1}_n = \begin{pmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{pmatrix}^{\boldsymbol{T}} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \mathbf{1}_n$
- $H \geq 0$

示性矩阵

所有标准化示性矩阵都满足：

$$H \geq 0 \qquad H^T H = I \qquad HH^T \mathbf{1}_n = \mathbf{1}_n$$

事实上，满足上述三个条件的矩阵一定是标准化示性矩阵。

目标函数 I

令 $X = [x_1, x_2, \dots, x_5]$,

$$XHH^T = [x_1, x_2, \dots, x_5] \begin{pmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{pmatrix}^T = [\mu_1, \mu_1, \mu_2, \mu_2, \mu_2]$$

其中，$\mu_1 = (x_1 + x_2)/2$, $\mu_2 = (x_3 + x_4 + x_5)/3$.

$$\sum_{j=1}^{k} \sum_{i \in C_j} \|x_i - \mu_j\|^2 = \sum_{i=1}^{2} \|x_i - \mu_1\|^2 + \sum_{i=3}^{5} \|x_i - \mu_2\|^2 = \|X - XHH^T\|_F^2$$

目标函数 II

如果正交约束 $H^T H = I$ 成立，我们有

$$\|X - XHH^T\|_F^2$$

$$= \mathbf{Tr}\left((X - XHH^T)(X - XHH^T)^T\right)$$

$$= \mathbf{Tr}\left((X - XHH^T)(X^T - HH^T X^T)\right)$$

$$= \mathbf{Tr}(XX^T) + \mathbf{Tr}(XH[\boldsymbol{H^T H}]H^T X^T)$$

$$- \mathbf{Tr}(XHH^T X^T) - \mathbf{Tr}(X^T XHH^T)$$

$$= \mathbf{Tr}(XX^T) - \mathbf{Tr}(XHH^T X^T)$$

$$= \mathbf{Tr}(XX^T) - \mathbf{Tr}(H^T X^T XH)$$

K-means模型矩阵形式

K-means 模型：

$$\underset{\boldsymbol{\mu}_j, \boldsymbol{\mathcal{C}}_j}{\operatorname{argmin}} \sum_{j=1}^{k} \sum_{i \in \boldsymbol{\mathcal{C}}_j} \left\| \boldsymbol{x}_i - \boldsymbol{\mu}_j \right\|^2$$

等价于

$$\underset{H \in \mathcal{H}}{\operatorname{argmin}} \left\| X - XHH^T \right\|_F^2$$

或者

$$\underset{H \in \mathcal{H}}{\operatorname{argmax}} \ \mathbf{Tr}\left(H^T X^T X H\right)$$

$$\boldsymbol{\mathcal{H}} = \left\{ H \in \mathbb{R}^{n \times k} \middle| H \geq 0, H^T H = I, HH^T \mathbf{1}_n = \mathbf{1}_n \right\}$$

Kernel K-means

K-means 模型：

$$\underset{H\in\mathcal{H}}{\text{argmax}}\ \mathbf{Tr}\left(H^T X^T X H\right)$$

$$\mathcal{H}=\left\{H\in\mathbb{R}^{n\times k}\middle| H\geq 0, H^T H=I, HH^T\mathbf{1}_n=\mathbf{1}_n\right\}$$

Kernel K-means 模型：

$$\underset{H\in\mathcal{H}}{\text{argmax}}\ \mathbf{Tr}\left(H^T K H\right)$$

其中K是合适的核矩阵。

7.3 模糊C聚类

K-means

给定数据$\{x_i\}_{i=1,\ldots,n}$以及簇的个数k， K-means模型可以写成下面的优化模型：

$$\underset{\mu_j, C_j}{\operatorname{argmin}} \sum_{j=1}^{k} \sum_{i \in C_j} \|x_i - \mu_j\|^2$$

缺陷：

- 每个点只能属于一个簇！
- 每个点对中心的贡献一致！

Fuzzy c-means

给定数据$\{x_i\}_{i=1,\dots,n}$以及簇的个数 k， K-means 模型可以写成下面的优化模型：

$$\underset{w_{ij},\mu_j}{\operatorname{argmin}} \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ij}^{m}\left\|x_i-\mu_j\right\|^2$$

$$\text{s.t.} \quad \sum_{j=1}^{k} w_{ij}=1, \qquad i=1,\dots,n$$

变量：

- $w_{ij}\in[0,1]$: 第 i 个点属于第j个簇的概率
- μ_j： 第 j 个簇的中心。

Hard clustering vs Soft clustering

K-means:

$$\underset{w_{ij},\mu_j}{\operatorname{argmin}} \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ij}\left\|x_i-\mu_j\right\|^2$$

$$\text{s.t.} \sum_{j=1}^{k} w_{ij}=1, \qquad w_{ij}\in\{0,1\}$$

Fuzzy c-means:

$$\underset{w_{ij},\mu_j}{\operatorname{argmin}} \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ij}^{m}\left\|x_i-\mu_j\right\|^2$$

$$\text{s.t.} \sum_{j=1}^{k} w_{ij}=1, \qquad w_{ij}\in[0,1]$$

Fuzzy c-means

Fuzzy c-means：

$$\underset{w_{ij},\mu_j}{\operatorname{argmin}} \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ij}^{m}\left\|x_i-\mu_j\right\|^2$$

$$\text{s.t.} \quad \sum_{j=1}^{k} w_{ij}=1, \qquad i=1,\dots,n$$

拉格朗日函数：

$$\mathcal{L}=\sum_{j=1}^{k}\sum_{i=1}^{n} w_{ij}^{m}\left\|x_i-\mu_j\right\|^2+\sum_{i=1}^{n}\lambda_i\left(\sum_{j=1}^{k} w_{ij}-1\right)$$

极小问题 $\min_{w,\mu} \mathcal{L}(w, \mu, \lambda)$

$$\min_{w,\mu} \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ij}^{m}\left\|x_i - \mu_j\right\|^2 + \sum_{i=1}^{n}\lambda_i\left(\sum_{j=1}^{k} w_{ij} - 1\right)$$

$$\frac{\partial \mathcal{L}}{\partial \mu_j} = -2\sum_{i=1}^{n} w_{ij}^{m}(x_i - \mu_j)$$

$$\mu_j = \frac{\sum_{i=1}^{n} w_{ij}^{m} x_i}{\sum_{i=1}^{n} w_{ij}^{m}}$$

极小问题 $\min_{w,\mu} \mathcal{L}(w, \mu, \lambda)$

$$\min_{w,\mu} \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ij}^{m}\left\|x_i - \mu_j\right\|^2 + \sum_{i=1}^{n}\lambda_i\left(\sum_{j=1}^{k} w_{ij} - 1\right)$$

$$\frac{\partial \mathcal{L}}{\partial w_{ij}} = m w_{ij}^{m-1}\left\|x_i - \mu_j\right\|^2 + \lambda_i$$

$$w_{ij} = \left(-\frac{\lambda_i}{m}\right)^{\frac{1}{m-1}} \frac{1}{\left\|x_i - \mu_j\right\|^{\frac{2}{m-1}}} = \left(-\frac{\lambda_i}{m}\right)^{\frac{1}{m-1}} \left\|x_i - \mu_j\right\|^{-\frac{2}{m-1}}$$

Fuzzy c-means Model

$$w_{ij} = (-\frac{\lambda_i}{m})^{\frac{1}{m-1}} \left\| x_i - \mu_j \right\|^{-\frac{2}{m-1}}$$

对 i 求和有，

$$1 = \sum_{j=1}^{k} w_{ij} = (-\frac{\lambda_i}{m})^{\frac{1}{m-1}} \sum_{j=1}^{k} \left\| x_i - \mu_j \right\|^{-\frac{2}{m-1}}$$

即

$$(-\frac{\lambda_i}{m})^{\frac{1}{m-1}} = 1 / \sum_{j=1}^{k} \left\| x_i - \mu_j \right\|^{-\frac{2}{m-1}}$$

因此，

$$w_{ij} = \frac{\left\| x_i - \mu_j \right\|^{-\frac{2}{m-1}}}{\sum_{j=1}^{k} \left\| x_i - \mu_j \right\|^{-\frac{2}{m-1}}}$$

Fuzzy c-means

Fuzzy c-means:

$$\underset{w_{ij},\mu_j}{\operatorname{argmin}} \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ij}^{m}\left\|x_i-\mu_j\right\|^2$$

$$\text{s.t.} \quad \sum_{j=1}^{k} w_{ij}=1, \qquad i=1,\dots,n$$

EM算法

$$w_{ij}=\frac{\left\|x_i-\mu_j\right\|^{-\frac{2}{m-1}}}{\sum_{j=1}^{k}\left\|x_i-\mu_j\right\|^{-\frac{2}{m-1}}}, \qquad \mu_j=\frac{\sum_{i=1}^{n} w_{ij}^{m} x_i}{\sum_{i=1}^{n} w_{ij}^{m}}$$

Fuzzy c-means

EM算法

$$w_{ij} = \frac{\left\|x_i - \mu_j\right\|^{-\frac{2}{m-1}}}{\sum_{j=1}^{k}\left\|x_i - \mu_j\right\|^{-\frac{2}{m-1}}}, \quad \mu_j = \frac{\sum_{i=1}^{n} w_{ij}^{m} x_i}{\sum_{i=1}^{n} w_{ij}^{m}}$$

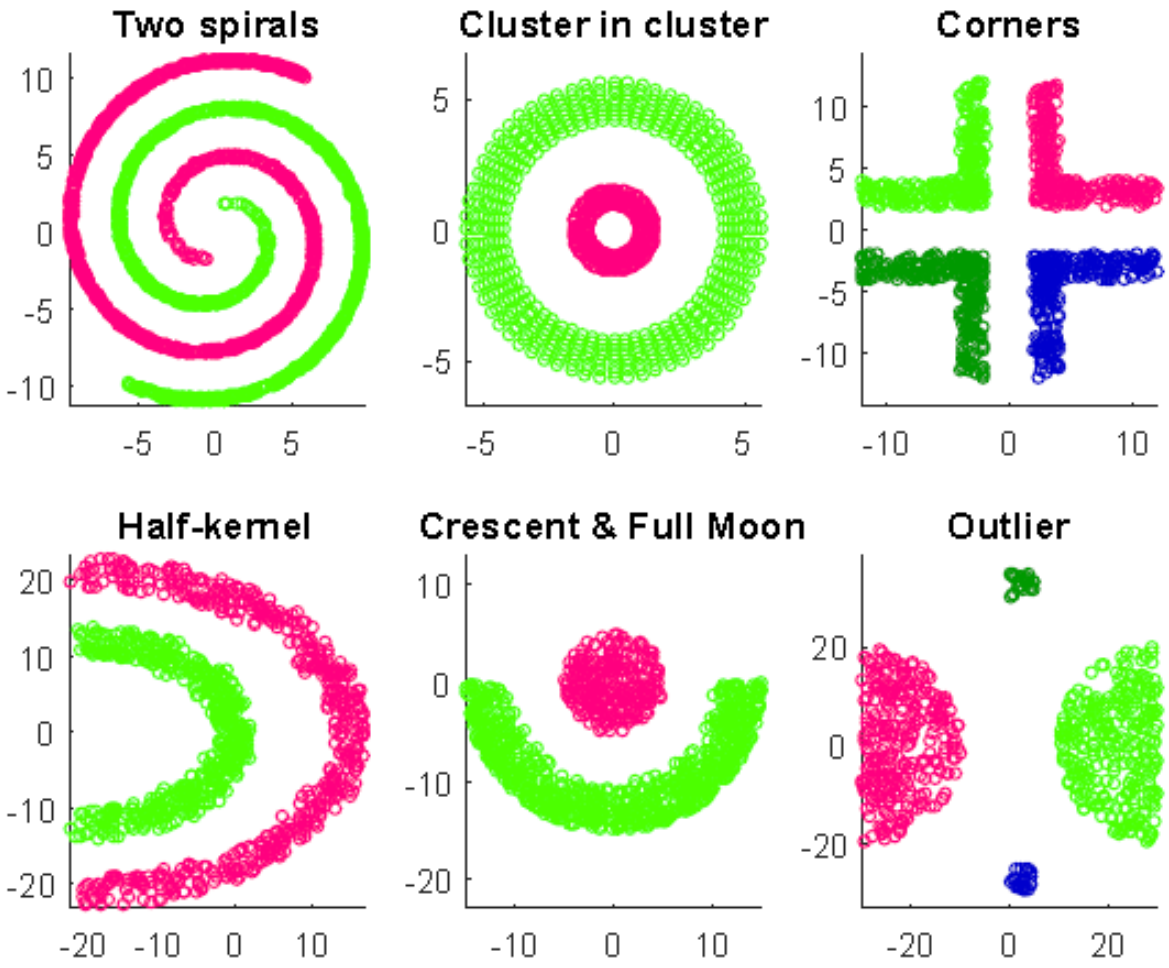
7.4 谱聚类

黄晟

huangsheng@cqu.edu.cn

办公室：信息大楼B701

Non-globular Clustering



Non-globular Clustering

解决方案：核化 K-means ?

$$\underset{\boldsymbol{\mu}_j, \boldsymbol{C}_j}{\operatorname{argmin}} \sum_{j=1}^{k} \sum_{x \in \boldsymbol{C}_j} \|\boldsymbol{x} - \boldsymbol{\mu}_j\|^2$$

- K-means 模型仅和数据点 $\boldsymbol{x}$ 有关！
- Lloyd 算法对核矩阵选取中心？

结论：K- means 模型 + Lloyd 算法不能直接推广！

Non-globular Clustering

矩阵形式 **K-means** 模型：

$$\text{argmax} \quad \mathbf{Tr}\left(H^T X^T X H\right)$$

$$\text{s.t.} \quad H^T H = I, H \geq 0, H H^T \mathbf{1}_n = \mathbf{1}_n$$

- **K-means** 模型和数据矩阵$X^T X$有关！！！
- 求解 **K-means** 和 **Kernel K-means** 是一样的。
- 近似解法: proximal alternating linearized minimization (PALM)

谱聚类（Spectral Clustering）

谱 （Spectral）：矩阵的特征值！

谱聚类： 根据图论，把聚类问题转化为求解拉普拉斯矩阵的特征值问题！

优势：理论高深，实现简单，可适用于各种形状的数据！

图论基础知识 I

通常用 $G(V,E)$表示一个图，V 中的元素称为节点，E 中的元素称为边。

若节点 v_i 和 v_j 之间有边相连，可以给其对应的边 e_{ij} 赋予一个非负的权重 $w_{ij}>0$。反之，若 $w_{ij}=0$，则说明节点 v_i 和 v_j 之间没有边相连。

节点 v_i 的度 d_i：$d_i=\sum_{j=1}^{n} w_{ij}$。

邻接矩阵（Adjacency Matrix）

构造邻接矩阵 W 的方法：

- ε – 邻域：连接所有距离小于 ε 的点。
- k – 近邻：把每个点与它最近的 k 个点相连。

（可能非对称，对称化或相互最近邻）

- 全连接图：$w_{ij} = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$

（其中 σ 控制邻域的宽度）

拉普拉斯矩阵

给定邻接矩阵 W，构造拉普拉斯矩阵 L：

$$L = D - W$$

其中 $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ 是所有数据点的度矩阵。

Normalized 拉普拉斯矩阵：

$$L_{rw} = D^{-1}L = I - D^{-1}W$$

$$L_{sym} = D^{-1/2}LD^{-1/2} = I - D^{-1/2}WD^{-1/2}$$

拉普拉斯矩阵

拉普拉斯矩阵 $L = D - W$满足下列性质：

- 对于任意的向量 f

$$f^T L f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2$$

- L 是对称半正定矩阵
- L 的最小特征值是0，其对应的特征向量是 1
- L 有 n 个非负的实特征值，$0 = \lambda_1 \leq \cdots \leq \lambda_n$

图论基础知识 II

若$A \subset V$是全部节点集的一个子集，记 $\overline{A} = V \backslash A$ 是A的补集。

$$A \cap \overline{A} = \emptyset, \quad A \cup \overline{A} = V$$

我们称$A_1, \ldots, A_k$构成图V的一个划分，如果

$$A_i \cap A_j = \emptyset, \quad A_1 \cup \cdots \cup A_k = V$$

切割图问题 I

聚类：把 n 个数据点划分为 k 个簇，使得相同簇中的数据点相似性高，不同簇之间的数据点相似性低。

切割图：把 n 个节点的图切割成 k 个连通子图，使得相似点之间的边所对应的权值大，不相似点之间的边所对应的权值小。

因此，可以用切割图的方法实现聚类！

切割图问题 II

给定两个不相交的子集 $A, B \subset V$，我们可以定义：

$$\mathrm{cut}(A, B) = \sum_{i \in A, j \in B} w_{ij}$$

给定 k 个不相交的子集 $A_1, \ldots, A_k$，同样可以定义：

$$\mathrm{cut}(A_1, \ldots, A_k) = \sum_{i=1}^{k} \mathrm{cut}\left(A_i, \overline{A_i}\right)$$

切割图问题 III

最小化图切方法可能会导致不平衡的结果，即 $A_1, \ldots, A_k$ 中点的个数极度不平衡。

我们需要引入关于点集A的大小的平衡性条件：

$$|A| := A\text{中节点的个数}$$

$$\mathrm{vol}(A) := \sum_{i \in A} d_i$$

切割图问题 IV

Ratio Cut:

$$\text{Ratiocut}(A_1, \dots, A_k) = \sum_{i=1}^{k} \frac{\text{cut}(A_i, \overline{A_i})}{|A_i|}$$

Normalized Cut:

$$\text{Ncut}(A_1, \dots, A_k) = \sum_{i=1}^{k} \frac{\text{cut}(A_i, \overline{A_i})}{\text{vol}(A_i)}$$

Ratio Cut 和 Ncut 都是NP 难的问题！ 松弛！

Ratio Cut

令 $h_{ij} = \begin{cases} 1/\sqrt{|A_j|} & if\ i \in A_j \\ 0 & otherwise \end{cases}$ ⇒ $H^T H = I$

$$h_l^T L h_l = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(h_{il} - h_{jl})^2 = \frac{\sum_{i\in A_l, j\in \overline{A_l}} w_{ij}}{|A_l|}$$

$$= \frac{\mathrm{cut}(A_l, \overline{A_l})}{|A_l|}$$

因此，

$$\mathrm{Ratiocut}(A_1, \ldots, A_k) = \sum_{l=1}^{k} h_l^T L h_l = \mathbf{Tr}(H^T L H)$$

Normalized Cut

令 $h_{ij} = \begin{cases} 1/\sqrt{\mathrm{vol}(\boldsymbol{A_i})} & if\ i \in A_j \\ 0 & otherwise \end{cases}$ ⇒ $H^T DH = I$

$$h_l^T L h_l = \frac{1}{2}\sum_{i,j=1}^{n} \boldsymbol{w_{ij}}(\boldsymbol{h_{il}} - \boldsymbol{h_{jl}})^2 = \frac{\sum_{i\in A_l, j\in \overline{A_l}}^{n} \boldsymbol{w_{ij}}}{\mathrm{vol}(\boldsymbol{A_l})}$$

$$= \frac{\mathrm{cut}(\boldsymbol{A_l}, \overline{\boldsymbol{A_l}})}{\mathrm{vol}(\boldsymbol{A_l})}$$

因此，

$$\mathrm{Ncut}(\boldsymbol{A_1}, \ldots, \boldsymbol{A_k}) = \sum_{l=1}^{k} h_l^T L h_l = \mathbf{Tr}(H^T L H)$$

Ratio Cut and Ncut

Ratio Cut:

$$\min_{\boldsymbol{H}} \quad \mathbf{Tr}\left(H^T L H\right) \quad \text{s.t.} \quad h_{ij} = \begin{cases} 1/\sqrt{|\boldsymbol{A_i}|} & if\ i \in A_j \\ 0 & otherwise \end{cases}$$

Ratio Cut 是一个特殊的Kernel K-means模型

Normalized Cut:

$$\min_{\boldsymbol{H}} \quad \mathbf{Tr}\left(H^T L H\right) \ \text{s.t.} \quad h_{ij} = \begin{cases} 1/\sqrt{\mathrm{vol}(\boldsymbol{A_i})} & if\ i \in A_j \\ 0 & otherwise \end{cases}$$

谱聚类（Spectral Clustering）

Relaxed Ratio Cut:

$$\min_{H} \quad \frac{1}{2}\mathbf{Tr}\left(H^T L H\right) \qquad \text{s.t.} \qquad H^T H = I$$

Relaxed Normalized Cut:

令 $U = D^{1/2}H$, 则 $U^T U = H^T D^{1/2} D^{1/2} H = H^T D H = I$

$$\mathbf{Tr}\left(H^T L H\right) = \mathbf{Tr}\left(U^T D^{-1/2} L D^{-1/2} U\right)$$

$$\min_{U} \quad \frac{1}{2}\mathbf{Tr}\left(U^T D^{-1/2} L D^{-1/2} U\right) \qquad \text{s.t} \qquad U^T U = I$$

拉格朗日乘子法

考虑下面瑞丽商(Rayleigh quotient)问题：

$$\min_{x} \quad x^T L x \qquad \text{s.t} \qquad x^T x = 1$$

拉格朗日函数：

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = x^T L x - \boldsymbol{\lambda}(x^T x - 1)$$

$\boldsymbol{\mathcal{L}}$ 对变量 $\boldsymbol{x}$ 求偏导等于0：

$$\frac{\partial \mathcal{L}(x,\lambda)}{\partial x} = 2\boldsymbol{L}\boldsymbol{x} - 2\boldsymbol{\lambda}\boldsymbol{x} \quad \Rightarrow \quad \boldsymbol{L}\boldsymbol{x} = \boldsymbol{\lambda}\boldsymbol{x}$$

说明 $\boldsymbol{L}$ 的所有特征对$(\boldsymbol{\lambda}, \boldsymbol{x})$都满足$\boldsymbol{\mathcal{L}}$ 对变量 $\boldsymbol{x}$一阶导数等于0。

谱聚类（Spectral Clustering）

此时，拉格朗日函数：

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}) = x^T(\boldsymbol{\lambda}x) - \boldsymbol{\lambda}\left(x^T x - 1\right) = \boldsymbol{\lambda}$$

极小问题：

$$\min_{x} \ \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda})$$

等价于

$$\min_{\boldsymbol{\lambda}} \quad \boldsymbol{\lambda} \qquad \mathbf{s.t} \qquad \boldsymbol{Lx} = \boldsymbol{\lambda x}$$

因此 $\boldsymbol{\lambda}$ 是 $\boldsymbol{L}$ 最小的特征值，$\boldsymbol{x}$是其对应的特征向量。

注意，此时极大问题不需要再求解。

谱聚类（Spectral Clustering）

输入：数据矩阵 X，簇个数 k。

- 构建邻接矩阵 W 。
- 计算拉普拉斯矩阵 L 。
- 计算 L 最小 k 个特征值对应的特征向量 U 。
- 对 U 进行 K-means 聚类。

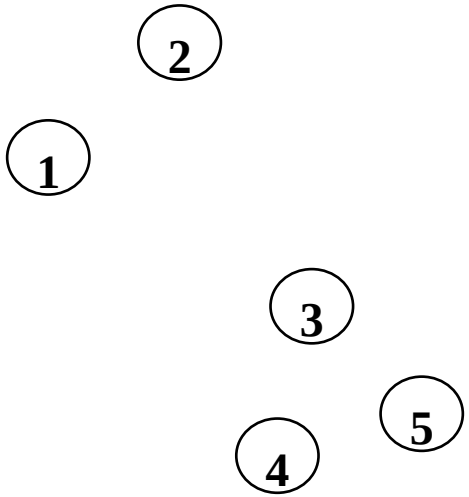
输出：k 个簇。

拉普拉斯矩阵的谱 vs 图的连同分支

定理：假设 $G(V,E)$ 是一个图，L 是它的拉普拉斯矩阵。那么 $G(V,E)$ 的连通分支$A_1, A_2, \dots, A_k$的个数等于 L 的零特征值的重数。并且，零特征值的特征空间由示性向量$\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \dots, \mathbf{1}_{A_k}$张成。

人造例子

如何对下面数据进行聚类？



相似性： 欧式距离

簇个数： $\mathbf{k=2}$

最优化分： $\boldsymbol{C_1=\{1,2\}}$, $\boldsymbol{C_2=\{3,4,5\}}$

人造例子

邻接矩阵与度矩阵：

$$W = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

人造例子

拉普拉斯矩阵：

$$L = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix} - \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

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

人造例子

特征向量：

$$Lu_1 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 & -1 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

且

$$Lu_2 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 & -1 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

人造例子

聚类：

$$U=(u_1,u_2)^{\mathrm{T}}=\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix}$$

前2个、后3个分别重合！

问题？

1. 相比直接进行 K-means 聚类，谱聚类有什么不同？

   相当于先做了一次特征提取，再聚类！

2. 谱聚类背后的机理是什么？

   图论中的切割图问题！