

模式识别

Pattern Recognition

第4讲 基于谱的聚类与降维

相似图——邻接矩阵

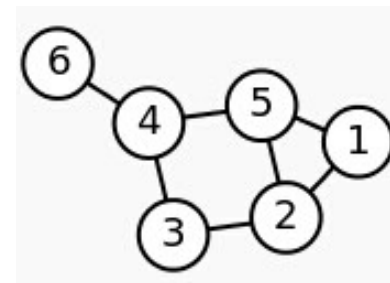
- **相似图**：样本集 $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ 可以表示为相似图
- $G = (V, E)$ 。图可以用邻接矩阵描述：

$$W = \{w_{ij}\}_{i,j=1,\dots,n}$$

- 利用 w_{ij} 直观定义样本i与样本j间的相似关系

- 你可以想到哪些方式?
- 距离、内积?
- 任意两个样本之间的相似度为 $1/n$ ——PCA
- 类内样本相似度为 $1/n_k$ ——LDA

$$\bullet \quad w_{ij} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \sigma), & \mathbf{x}_i \in N_k(\mathbf{x}_j) \vee \mathbf{x}_j \in N_k(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases}$$



$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

拉普拉斯?

1. 微积分中的Laplacian算子——非混合二阶偏导数之和

$$\Delta f = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} \quad \text{例如:} \quad \Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

2. 图像处理中的Laplacian算子——二阶差分

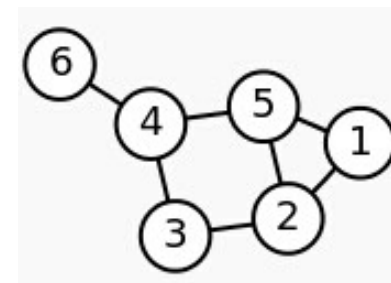
0	-1	0	-1	-1	-1
-1	4	-1	-1	8	-1
0	-1	0	-1	-1	-1

知乎 @owkeyway

3. 图的Laplacian矩阵——邻接矩阵上的二阶差分

相似图的拉普拉斯矩阵

- **相似图**: 样本集 $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ 可以表示为相似图 $G = (V, E)$ 。图 G 也可以用邻接矩阵 $W = \{w_{ij}\}_{i,j=1,\dots,n}$ 描述。



- **节点的度矩阵**:

$$d_i = \sum_{j=1}^n w_{ij}; \quad D = \text{diag}(d_i)$$

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

- **Laplacian矩阵**:

$$L = D - W$$

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

- 随机游走正规化: $L_{rw} = D^{-1}L = I - D^{-1}W$
- 对称正规化: $L_{sym} = D^{-1/2}LD^{-1/2} = I - D^{-1/2}WD^{-1/2}$

Laplacian矩阵的性质与作用

1. Laplacian矩阵：邻接矩阵上的二阶差分和
2. 对于 $\mathbf{y} \in R^n$, y_i 为样本 i 的某维属性，有：

$$\mathbf{y}^t \mathbf{L} \mathbf{y} = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (y_i - y_j)^2$$

$w_{ij}(y_i - y_j)^2$ ：两点间的加权二阶差分

L矩阵的二次型



邻接矩阵上的二阶差分函数

3. 对于 n 个 m 维样本构成的样本矩阵 \mathbf{Y}

$$\begin{aligned} \mathbf{Y} &= [\mathbf{y}_1, \dots, \mathbf{y}_n] \\ &= \begin{bmatrix} \tilde{\mathbf{y}}_1^t \\ \vdots \\ \tilde{\mathbf{y}}_m^t \end{bmatrix} \end{aligned} \quad \Rightarrow \quad \begin{aligned} &\sum_{i,j} w_{ij} \sum_{k=1}^m (y_i^{(k)} - y_j^{(k)})^2 \\ &= \sum_{k=1}^m \sum_{i,j} w_{ij} (y_i^{(k)} - y_j^{(k)})^2 = \sum_{k=1}^m \tilde{\mathbf{y}}_k^t \mathbf{L} \tilde{\mathbf{y}}_k \end{aligned}$$

$\tilde{\mathbf{y}}_k$ ：各样本的第 k 维特征构成的向量， $y_i^{(k)}$ ：样本 i 的第 k 维特征， $k = 1, \dots, m$

样本降维

□降维过程: $\mathbf{y} = f(\mathbf{x})$

$$\mathbf{x} \in R^m, \quad \mathbf{y} \in R^{m'}, \quad m' \ll m$$

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] = \begin{bmatrix} \mathbf{x}_1^t \\ \vdots \\ \tilde{\mathbf{x}}_m^t \end{bmatrix} \longrightarrow \mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n] = \begin{bmatrix} \tilde{\mathbf{y}}_1^t \\ \vdots \\ \tilde{\mathbf{y}}_{m'}^t \end{bmatrix}$$

降维的一般谱方法

- 令 α_{ij} 为样本 \mathbf{x}_i 与 \mathbf{x}_j 之间的相似度，为了保持降维之后样本之间的相似度，可以求解如下的优化问题：

$$\begin{aligned}\min_{\mathbf{Y}} \sum_{i \neq j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 &= \min_{\mathbf{Y}} \sum_{i \neq j} w_{ij} \sum_{k=1}^{m'} \left(y_i^{(k)} - y_j^{(k)} \right)^2 \\ &= \min_{\mathbf{Y}} \sum_{k=1}^{m'} \sum_{i \neq j} w_{ij} \left(y_i^{(k)} - y_j^{(k)} \right)^2 = \min_{\mathbf{Y}} \sum_{k=1}^{m'} \tilde{\mathbf{y}}_k^t \mathbf{L} \tilde{\mathbf{y}}_k\end{aligned}$$

$$\text{约束: } \tilde{\mathbf{y}}_k^t \mathbf{B} \tilde{\mathbf{y}}_k = d, \quad k = 1, \dots, m'$$

$$\text{其中: } \mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n] = \begin{bmatrix} \tilde{\mathbf{y}}_1^t \\ \vdots \\ \tilde{\mathbf{y}}_{m'}^t \end{bmatrix}$$

$$\mathbf{W} = (w_{ij}), \quad \mathbf{D} = \text{diag} \left(\sum_{j=1}^n w_{ij} \right) \quad \mathbf{L} = \mathbf{D} - \mathbf{W}$$

线性降维的优化问题

- f 为线性映射时:

$$\mathbf{y}_i = \mathbf{W}^t \mathbf{x}_i, \quad \tilde{\mathbf{y}}_k = \mathbf{X}^t \mathbf{w}_k$$

其中: \mathbf{x}_i 和 \mathbf{w}_k 分别为 \mathbf{X} 和 \mathbf{W} 的列矢量。

- 优化问题:

$$\min_{\mathbf{W}} \sum_{k=1}^{m'} \tilde{\mathbf{y}}_k^t \mathbf{L} \tilde{\mathbf{y}}_k = \sum_{k=1}^{m'} \mathbf{w}_k^t \mathbf{X} \mathbf{L} \mathbf{X}^t \mathbf{w}_k$$

约束: $\mathbf{w}_k^t \mathbf{X} \mathbf{B} \mathbf{X}^t \mathbf{w}_k = d, \quad k = 1, \dots, m'$

或: $\mathbf{w}_k^t \mathbf{w}_k = d$

- 问题的解: 广义特征值、特征矢量问题

$$(\mathbf{X} \mathbf{L} \mathbf{X}^t) \mathbf{w} = \lambda (\mathbf{X} \mathbf{B} \mathbf{X}^t) \mathbf{w}$$

最小 m' 个特征值对应的特征矢量。

PCA: Principle Component Analysis

- PCA的优化问题: (最大值问题转换最小值问题——丢掉的最少)

$$\min_{\mathbf{W}} \sum_{k=1}^{m'} \mathbf{w}_k^t \mathbf{\Sigma} \mathbf{w}_k$$

$$\text{约束: } \mathbf{w}_k^t \mathbf{w}_k = 1$$

- 协方差矩阵:

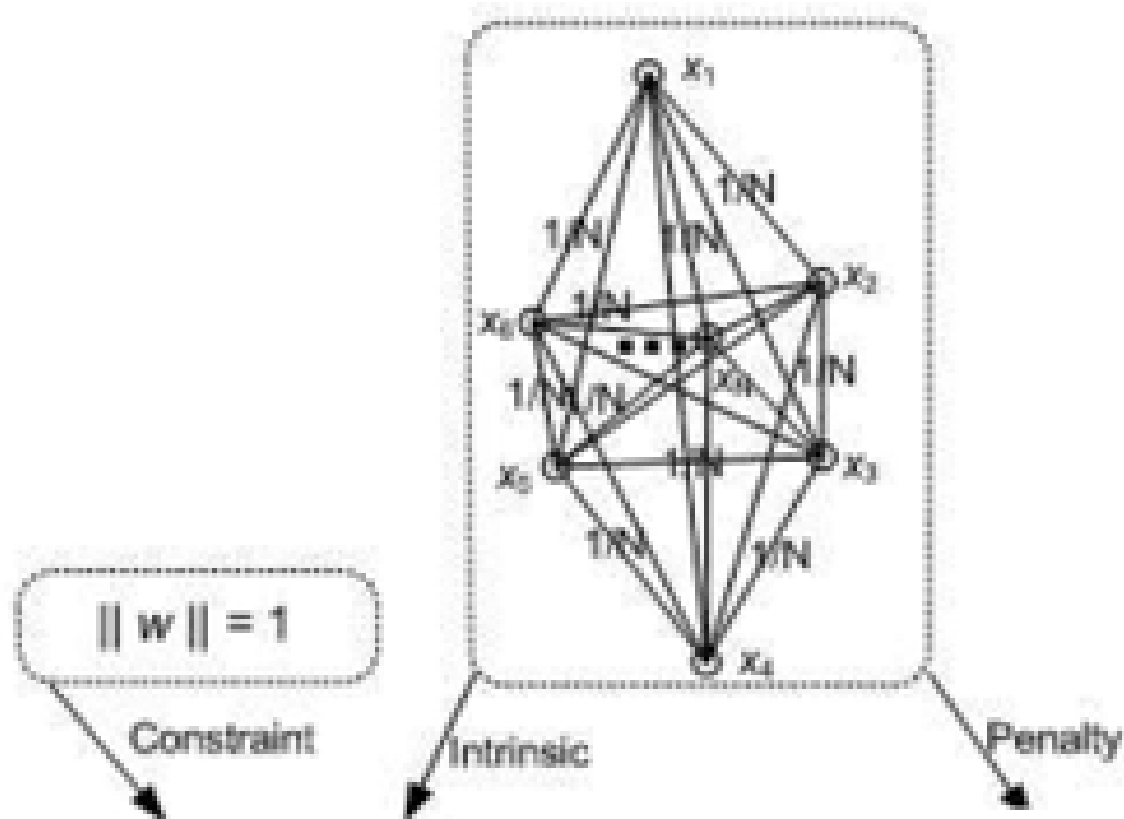
$$\mathbf{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t = \frac{1}{n} \mathbf{X} \left(\mathbf{I} - \frac{1}{n} \mathbf{e} \mathbf{e}^t \right) \mathbf{X}^t$$

- 相似图: Laplacian矩阵 \mathbf{L} 和相似矩阵 \mathbf{A} ($\mathbf{e}=\mathbf{1}$)

$$\mathbf{L} = \mathbf{I} - \frac{1}{n} \mathbf{e} \mathbf{e}^t, \quad \mathbf{A} = \frac{1}{n} \mathbf{e} \mathbf{e}^t$$

PCA的相似图

- PCA的相似图为全连接图，任意两个样本之间的相似度为 $1/n$ 。



LDA: Linear Discriminant Analysis

- LDA优化的问题:

$$\min_{\mathbf{w}} \sum_{k=1}^{m'} \mathbf{w}_k^t \mathbf{S}_w \mathbf{w}_k$$

$$\text{约束: } \mathbf{w}_k^t \mathbf{S}_b \mathbf{w}_k = d$$

- 类内散布矩阵:

$$\mathbf{S}_w = \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}^{c_i}) (\mathbf{x}_i - \boldsymbol{\mu}^{c_i})^t = \mathbf{X} \left(\mathbf{I} - \sum_{k=1}^c \frac{1}{n_k} \mathbf{e}_k \mathbf{e}_k^t \right) \mathbf{X}^t$$

其中: c_i 为 \mathbf{x}_i 的类别, c 为类别数, n_k 为第 k 类样本数, \mathbf{e}_k 为 n 维指示矢量:

$$e_{ki} = \begin{cases} 1, & \mathbf{x}_i \in \omega_k \\ 0, & \text{otherwise} \end{cases}$$

LDA的优化问题

- 类间散布矩阵:

$$\mathbf{S}_b = \sum_{k=1}^c n_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^t = n\boldsymbol{\Sigma} - \mathbf{S}_w$$

因此约束相当于: $\mathbf{w}_k^t \boldsymbol{\Sigma} \mathbf{w}_k = d$

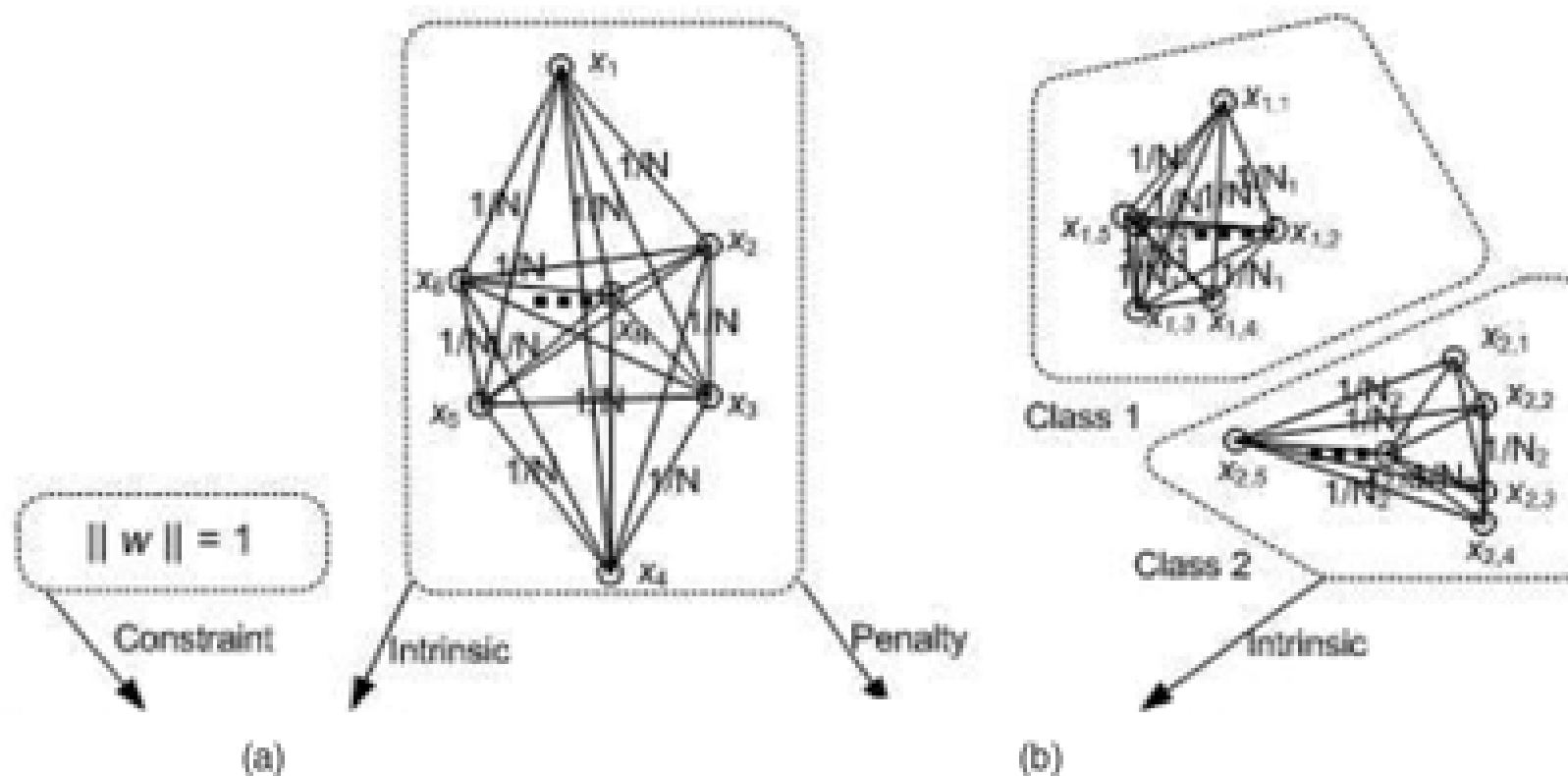
- 相似图:

$$\mathbf{L} = \mathbf{I} - \sum_{k=1}^c \frac{1}{n_k} \mathbf{e}_k \mathbf{e}_k^t$$
$$\mathbf{W} = \sum_{k=1}^c \frac{1}{n_k} \mathbf{e}_k \mathbf{e}_k^t, \quad \mathbf{B} = \boldsymbol{\Sigma}$$

\mathbf{W} 实际上是分块矩阵。

LDA的相似图

- 相似图**：类内样本之间全连接，权重为 $1/n_c$ ，类间样本无连接。约束图为所有样本的全连接，权重为 $1/n$ 。



LE: Laplacian Eigenmap 拉普拉斯映射

LPP: Locality Preserving Projection 局部保留投影

- 分别采用k近邻或 ϵ 近邻方式定义相似矩阵:

- **LE:**
$$w_{ij} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \sigma), & \mathbf{x}_i \in N_k(\mathbf{x}_j) \vee \mathbf{x}_j \in N_k(\mathbf{x}_i) \\ 0, & otherwise \end{cases}$$

- **LPP:**
$$w_{ij} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \sigma), & \|\mathbf{x}_i - \mathbf{x}_j\|^2 < \epsilon \\ 0, & otherwise \end{cases}$$

MDS: Multidimensional Scaling 多维尺度变换

- **MDS的目标**: 给定样本之间距离（相似度）的条件下，在低维空间对样本进行表示，力图保证样本之间的距离不变。
- **MDS的优化问题**: 给定距离矩阵 $D = (d_{ij})_{n \times n}$
$$\min_{\mathbf{y}_1, \dots, \mathbf{y}_n} \sum_{i < j} (\|\mathbf{y}_i - \mathbf{y}_j\| - d_{ij})^2$$
- 求解方法很多，一种是直接采用梯度法优化 $\mathbf{y}_1, \dots, \mathbf{y}_n$ 。

NMDS: Non-metric Multidimensional Scaling

- NMDS算法

NMDS

- 根据距离矩阵 \mathbf{D} 构造内积矩阵 \mathbf{S} ;
- 计算 \mathbf{S} 的特征值和特征矢量, 保留最大的 m' 个特征值(矩阵) $\mathbf{\Lambda}$ 和特征矢量(矩阵) $\mathbf{\Gamma}$;
- 构造映射之后的样本矩阵 \mathbf{Y} (相差一个旋转和尺度):

$$\mathbf{S} = \mathbf{Y}^t \mathbf{Y} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^t$$

$$\mathbf{Y} = \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{\Gamma}^t$$

NMDS与PCA的一致性

- Y的奇异值分解:

$$\mathbf{Y} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Gamma}^t$$

- 内积矩阵和协方差矩阵:

$$\mathbf{S} = \mathbf{Y}^t\mathbf{Y} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^t$$

$$\mathbf{\Sigma} = \mathbf{Y}\mathbf{Y}^t = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^t$$

- 两个矩阵的特征值相同, 特征矢量不同, 分别是Y的左奇异矢量和右奇异矢量。

距离矩阵构造内积矩阵

- 距离矩阵 $D = (d_{ij})_{n \times n}$, 内积矩阵 $S = (s_{ij})_{n \times n}$

$$s_{ij} = \mathbf{x}_i^t \mathbf{x}_j$$

$$\begin{aligned} d_{ij} &= (\mathbf{x}_i - \mathbf{x}_j)^t (\mathbf{x}_i - \mathbf{x}_j) \\ &= \mathbf{x}_i^t \mathbf{x}_i + \mathbf{x}_j^t \mathbf{x}_j - 2\mathbf{x}_i^t \mathbf{x}_j \\ &= \mathbf{x}_i^t \mathbf{x}_i + \mathbf{x}_j^t \mathbf{x}_j - 2s_{ij} \\ &= s_{ii} + s_{jj} - 2s_{ij} \end{aligned}$$

- S可以唯一的构造D, 而D不可能唯一构造S。平移变换下, 距离矩阵不变, 而内积矩阵不同。
- 限定坐标原点为样本的中心: $\bar{\mathbf{x}} = 0$

距离矩阵构造内积矩阵

- 利用关系:

$$\begin{aligned} & \frac{1}{n} \sum_j d_{ij} + \frac{1}{n} \sum_i d_{ij} - \frac{1}{n^2} \sum_{i,j} d_{ij} \\ &= \mathbf{x}_i^t \mathbf{x}_i + \frac{1}{n} \sum_j \mathbf{x}_j^t \mathbf{x}_j + \mathbf{x}_j^t \mathbf{x}_j + \frac{1}{n} \sum_i \mathbf{x}_i^t \mathbf{x}_i - \frac{1}{n} \sum_i \mathbf{x}_i^t \mathbf{x}_i - \frac{1}{n} \sum_j \mathbf{x}_j^t \mathbf{x}_j \\ & \quad - \frac{2}{n^2} \sum_{i,j} \mathbf{x}_i^t \mathbf{x}_j = \mathbf{x}_i^t \mathbf{x}_i + \mathbf{x}_j^t \mathbf{x}_j \end{aligned}$$

- 有:

$$\begin{aligned} s_{ij} &= -\frac{1}{2} d_{ij} + \frac{1}{2n} \sum_j d_{ij} + \frac{1}{2n} \sum_i d_{ij} - \frac{1}{2n^2} \sum_{i,j} d_{ij} \\ &= -\frac{1}{2} \left(d_{ij} - \frac{1}{n} (\mathbf{D}\mathbf{e})_i - \frac{1}{n} (\mathbf{e}^t \mathbf{D})_j + \frac{1}{n^2} \mathbf{e}^t \mathbf{D} \mathbf{e} \right) \end{aligned}$$

距离矩阵构造内积矩阵

- 矩阵形式:

$$\mathbf{S} = -\frac{1}{2} \left(\mathbf{D} - \frac{1}{n} \mathbf{D} \mathbf{e} \mathbf{e}^t - \frac{1}{n} \mathbf{e} \mathbf{e}^t \mathbf{D} + \frac{1}{n^2} \mathbf{e} \mathbf{e}^t \mathbf{D} \mathbf{e} \mathbf{e}^t \right)$$

$$= -\frac{1}{2} \left(\mathbf{I} - \frac{1}{n} \mathbf{e} \mathbf{e}^t \right) \mathbf{D} \left(\mathbf{I} - \frac{1}{n} \mathbf{e} \mathbf{e}^t \right)$$

NMDS (Non-metric multidimensional scaling)

非度量多维尺度分析

- 内积矩阵**S**构造相似矩阵**W**:

$$w_{ij} = \begin{cases} s_{ij}, & i \neq j \\ 0, & i = j \end{cases}$$

- S**的行求和:

$$\begin{aligned} \sum_j s_{ij} &= \sum_j \left(-\frac{1}{2} \left(\mathbf{I} - \frac{1}{n} \mathbf{e} \mathbf{e}^t \right) \mathbf{D} \left(\mathbf{I} - \frac{1}{n} \mathbf{e} \mathbf{e}^t \right) \right)_{ij} \\ &= \frac{1}{2} \sum_j \left(-d_{ij} - \frac{1}{n} \sum_{i'} d_{i'j} + \frac{1}{n} \sum_{j'} d_{ij'} - \frac{1}{n^2} \sum_{i',j'} d_{i'j'} \right) \\ &= \left(-\frac{1}{2} \sum_j d_{ij} + \frac{1}{2n} \sum_{j,j'} d_{ij'} \right) + \left(\frac{1}{2n} \sum_{i',j} d_{i',j} - \frac{1}{2n^2} \sum_{i',j'} d_{i',j'} \right) = 0 \end{aligned}$$

NMDS的相似图

- NMDS相似图上的优化问题:

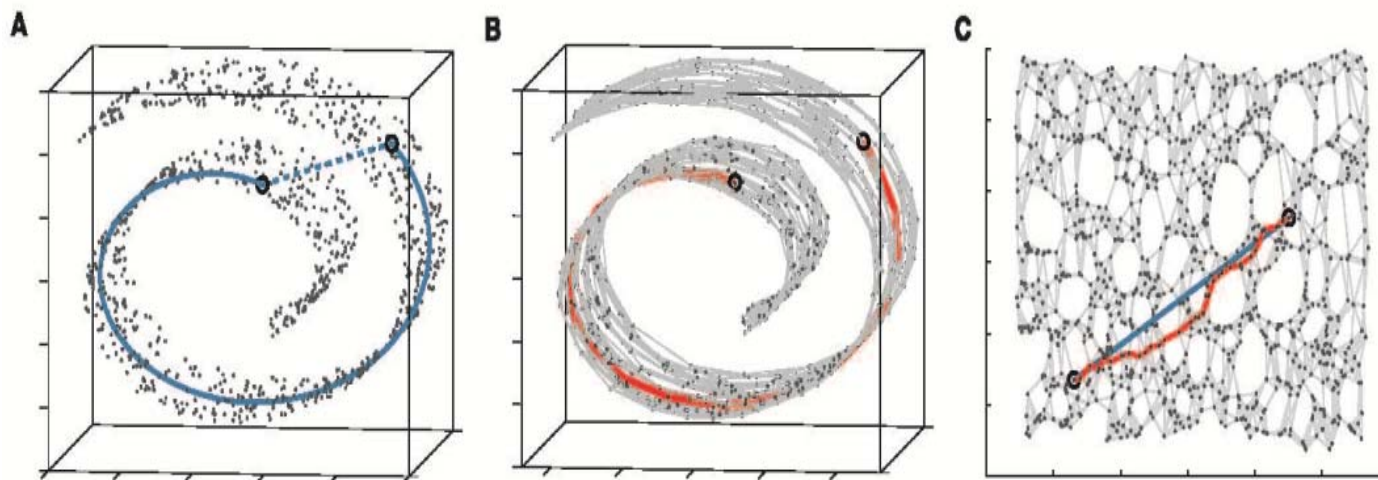
$$\arg \max_{\mathbf{y}^t \mathbf{y} = \lambda} \mathbf{y}^t \mathbf{S} \mathbf{y} = \arg \max_{\mathbf{y}^t \mathbf{y} = \lambda} \mathbf{y}^t (\mathbf{W} - \mathbf{D}) \mathbf{y} = \arg \min_{\mathbf{y}^t \mathbf{y} = \lambda} \mathbf{y}^t (\mathbf{D} - \mathbf{W}) \mathbf{y}$$

其中矩阵 $\mathbf{D} = (d_{ij})$:

$$d_{ij} = \begin{cases} -s_{ij} = \sum_j w_{ij}, & i = j \\ 0, & i \neq j \end{cases}$$

ISOMAP等距特征映射

- 欧氏距离与测地距离：



- **ISOMAP的思想：** 根据两个样本的测地距离构造相似图，采用MDS完成向低维空间的映射。

ISOMAP算法

- **ISOMAP算法:**

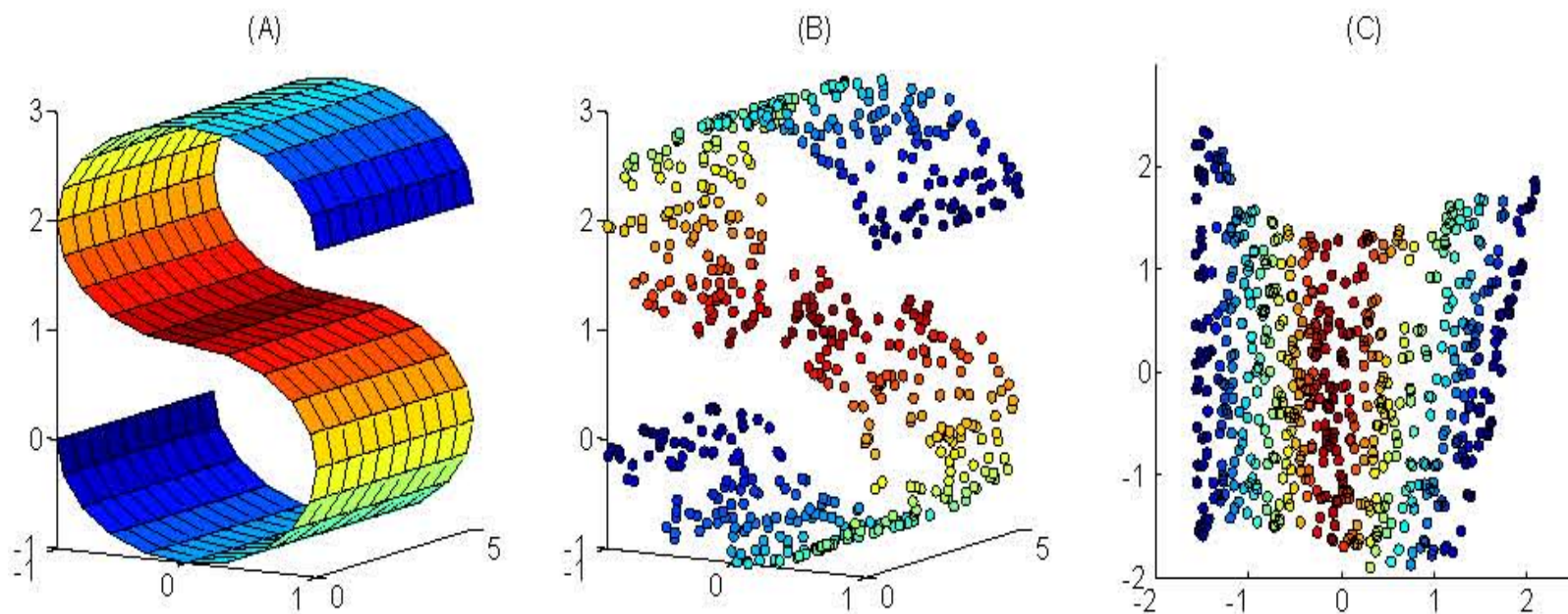
1. 构造样本的 ε 近邻或k近邻图，连接近邻样本节点，计算相连节点之间的（欧氏）距离 $d_G(i, j)$ ，不相连节点之间 $d_G(i, j) = \infty$;
2. 计算测地距离：（Dijkstra算法）

$$d_G(i, j) = \min_{k=1, \dots, n} (d_G(i, j), d_G(i, k) + d_G(k, j))$$

3. $D = (d_G(i, j))_{n \times n}$ ，用NMDS计算低维空间映射。

LLE: Locally Linear Embedding

- **LLE的思想:** 样本分布在一个嵌入于高维空间中的非线性流形，连续非线性流形的局部可以用一个线性流形逼近。



LLE算法

- LLE算法分成两个步骤完成：

1. 在局部线性流形上计算每个样本的线性表示；
2. 保持局部线性流形的拓扑不变性，向低维空间映射样本点。

- **LLE算法**

1. 计算每个样本点的 K 近邻，求解由 K 个近邻样本点线性表示当前样本点的最优组合系数：

① 由 \mathbf{x}_i 的 K 近邻 $\mathbf{x}^1, \dots, \mathbf{x}^K$ 计算局部协方差矩阵 Σ_i ：

$$\Sigma_i^{jk} = (\mathbf{x}_i - \mathbf{x}^j)^t (\mathbf{x}_i - \mathbf{x}^k)$$

LLE算法

- ② 对应 $\mathbf{x}^1, \dots, \mathbf{x}^K$ 的线性组合系数（如 Σ_i 奇异，则正则化）

$$\mathbf{w}_i = \frac{\Sigma_i^{-1} \mathbf{e}}{\mathbf{e}^t \Sigma_i^{-1} \mathbf{e}}$$

- ③ 扩展 \mathbf{w}_i 至 n 维矢量，非 K 近邻样本处置0；
④ 以 \mathbf{w}_i 为行矢量构成局部线性组合系数矩阵 \mathbf{W} 。

2. 计算局部保持的低维映射

- ① 由 \mathbf{W} 计算矩阵 \mathbf{M} ：

$$\mathbf{M} = (\mathbf{I} - \mathbf{W})^t (\mathbf{I} - \mathbf{W})$$

- ② 计算 \mathbf{M} 的最小 $d + 1$ 个特征值和特征矢量，以第2至 $d + 1$ 个特征矢量作为行矢量构造矩阵 \mathbf{Y} ；
③ \mathbf{Y} 的列矢量即为 n 个样本的LLE映射结果。

局部重构系数矩阵

- 令 \mathbf{x} 的 K 个近邻为： $\mathbf{x}_1, \dots, \mathbf{x}_K$ ，求解如下优化问题：

$$\min_{\mathbf{w}} \left\| \mathbf{x} - \sum_{j=1}^K w_j \mathbf{x}_j \right\|^2$$

约束： $\sum_{j=1}^K w_j = 1$

- 写成矩阵形式：

$$\left\| \mathbf{x} - \sum_{j=1}^K w_j \mathbf{x}_j \right\|^2 = \left\| \sum_{j=1}^K w_j (\mathbf{x} - \mathbf{x}_j) \right\|^2 = \sum_{j,k=1}^K w_j w_k (\mathbf{x} - \mathbf{x}_j)^t (\mathbf{x} - \mathbf{x}_k)$$

- 优化问题转化为：

$$\min_{\mathbf{w}} \mathbf{w}^t \Sigma \mathbf{w}$$

约束：

$$\mathbf{w}^t \mathbf{e} = 1$$

局部重构系数矩阵

- 构造拉格朗日函数:

$$L(\mathbf{w}) = \mathbf{w}^t \Sigma \mathbf{w} - \lambda(\mathbf{w}^t \mathbf{e} - 1)$$

$$\frac{\partial L(\mathbf{w})}{\partial \mathbf{w}} = 2\Sigma \mathbf{w} - \lambda \mathbf{e} = \mathbf{0}, \quad \mathbf{w} = \frac{\lambda}{2} \Sigma^{-1} \mathbf{e}$$

- 代入约束 $\mathbf{w}^t \mathbf{e} = 1$, 得到:

$$\lambda = \frac{2}{\mathbf{e}^t \Sigma^{-1} \mathbf{e}}$$

- 因此:

$$\mathbf{w} = \frac{\Sigma^{-1} \mathbf{e}}{\mathbf{e}^t \Sigma^{-1} \mathbf{e}}, \quad w_{ij} = \frac{\sum_{k=1}^K \Sigma_{jk}^{-1}}{\sum_{m,n=1}^K \Sigma_{mn}^{-1}}$$

局部拓扑保持的低维映射

- **映射的思想**：映射之前和映射之后的样本具有同样的邻域关系， K 个近邻相同，线性组合系数 \mathbf{w} 相同。
- **优化问题**：最小均方误差

$$\min_{\mathbf{y}_1, \dots, \mathbf{y}_n} \sum_{i=1}^n \left\| \mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_j \right\|^2$$

- 优化问题的解：对于任意的平移、旋转和尺度缩放是不稳定的，需要引入约束条件：

$$\sum_i \mathbf{y}_i = \mathbf{0}, \quad \frac{1}{n} \sum_i \mathbf{y}_i \mathbf{y}_i^t = \mathbf{I}$$

局部拓扑保持的低维映射

- 由于:

$$\begin{aligned}\sum_{i=1}^n \left\| \mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_j \right\|^2 &= \sum_i \left(\mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_j \right)^t \left(\mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_j \right) \\ &= \sum_i \left(\mathbf{y}_i^t \mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_j^t \mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_i^t \mathbf{y}_j + \sum_{j,k} w_{ij} w_{ik} \mathbf{y}_j^t \mathbf{y}_k \right) \\ &= \sum_{i,j} \left(\delta_{ij} - w_{ij} - w_{ji} + \sum_k w_{ki} w_{kj} \right) \mathbf{y}_i^t \mathbf{y}_j\end{aligned}$$

- 定义矩阵:

$$\mathbf{M} = \mathbf{I} - \mathbf{W} - \mathbf{W}^t + \mathbf{W}^t \mathbf{W} = (\mathbf{I} - \mathbf{W})^t (\mathbf{I} - \mathbf{W})$$

- 有:

$$\sum_{i=1}^n \left\| \mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_j \right\|^2 = \sum_{i,j} \mathbf{M}_{ij} \mathbf{y}_i^t \mathbf{y}_j$$

局部拓扑保持的低维映射

- 定义变换之后的样本矩阵:

$$\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)_{d \times n} = \begin{pmatrix} \mathbf{Y}_1^t \\ \vdots \\ \mathbf{Y}_d^t \end{pmatrix}_{d \times n}$$

- 优化目标变为:

$$\sum_{i,j} \mathbf{M}_{ij} \mathbf{y}_i^t \mathbf{y}_j = \sum_{i,j} \sum_{k=1}^d \mathbf{M}_{ij} y_{ik} y_{jk} = \sum_{k=1}^d \mathbf{Y}_k^t \mathbf{M} \mathbf{Y}_k$$

- 约束变为:

$$\sum_i \mathbf{y}_i = \mathbf{0} \Leftrightarrow \mathbf{Y}_k^t \mathbf{e} = \mathbf{0}$$
$$\frac{1}{n} \mathbf{Y} \mathbf{Y}^t = \frac{1}{n} \sum_i \mathbf{y}_i \mathbf{y}_i^t = \mathbf{I} \Leftrightarrow \mathbf{Y}_i^t \mathbf{Y}_j = \begin{cases} n, & i = j \\ 0, & i \neq j \end{cases}$$

局部拓扑保持的低维映射

- 重写优化问题:

$$\min_{\mathbf{Y}_1, \dots, \mathbf{Y}_d} \sum_{k=1}^d \mathbf{Y}_k^t \mathbf{M} \mathbf{Y}_k$$

约束:

$$\mathbf{Y}_k^t \mathbf{e} = 0, \quad k = 1, \dots, d$$

$$\mathbf{Y}_i^t \mathbf{Y}_j = \begin{cases} n, & i = j \\ 0, & i \neq j \end{cases}, \quad i, j = 1, \dots, d$$

- 问题的解: $\mathbf{Y}_1, \dots, \mathbf{Y}_d$ 是矩阵 \mathbf{M} 的特征矢量。

局部拓扑保持的低维映射

- 矩阵 \mathbf{M} 是一个Laplacian矩阵:

$$\begin{aligned}\sum_j M_{ij} &= \sum_j \left(\delta_{ij} - w_{ij} - w_{ji} + \sum_k w_{ki} w_{kj} \right) \\ &= 1 - 1 - \sum_j w_{ji} + \sum_{j,k} w_{ki} w_{kj} = 0\end{aligned}$$

- 因此: \mathbf{M} 的最小特征值为0, 对应特征矢量为 \mathbf{e} , 不满足约束 $\mathbf{Y}_k^t \mathbf{e} = 0$ 。
- 而 \mathbf{M} 为对称矩阵, 因此其它特征矢量与 \mathbf{e} 正交。优化问题的解为第2至 $d + 1$ 个最小特征矢量。

LLE的相似图

- **相似矩阵**：矩阵 \mathbf{M} 是Laplacian矩阵，显然就可以由 \mathbf{M} 构造相似矩阵 Δ

$$\mathbf{M} = \mathbf{I} - \mathbf{W} - \mathbf{W}^t + \mathbf{W}^t \mathbf{W}$$

$$\Delta = \mathbf{W} + \mathbf{W}^t - \mathbf{W}^t \mathbf{W}$$

- 样本 i 和 j 之间的相似度为：

$$\delta_{ij} = w_{ij} + w_{ji} - \sum_k w_{ki} w_{kj}$$

谱方法小结

基于图的几何解释



邻接矩阵上的代价函数



基于L矩阵的优化问题



基于谱的求解方法

相似/邻接矩阵 $\mathbf{W} = (w_{ij})$

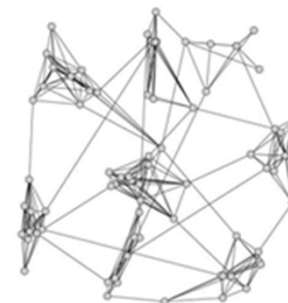
$$\mathbf{D} = \text{diag} \left(\sum_{j=1}^n w_{ij} \right), \quad \mathbf{L} = \mathbf{D} - \mathbf{W}$$

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

$$\min \sum_k \mathbf{f}_k^t \mathbf{L} \mathbf{f}_k \quad s.t. \mathbf{f}_k^t \mathbf{B} \mathbf{f}_k = d, \dots$$

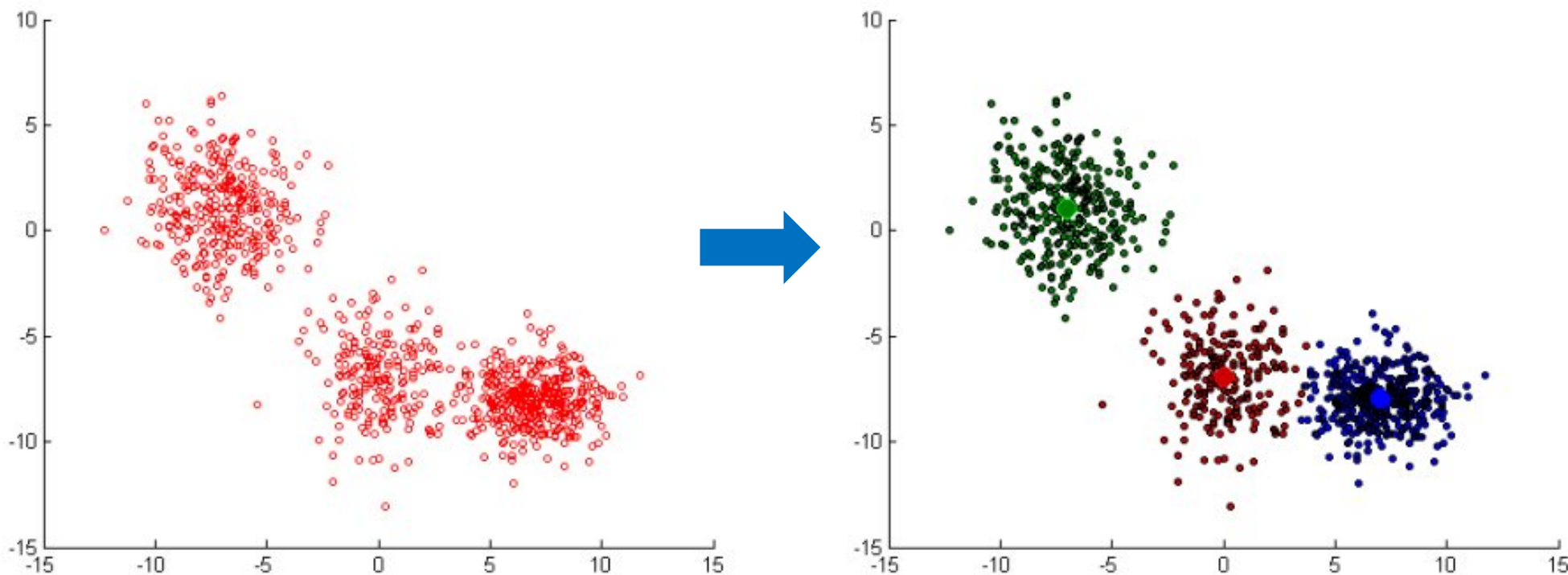
拉格朗日乘子法、广义瑞利商

\mathbf{L} 的特征值、特征向量

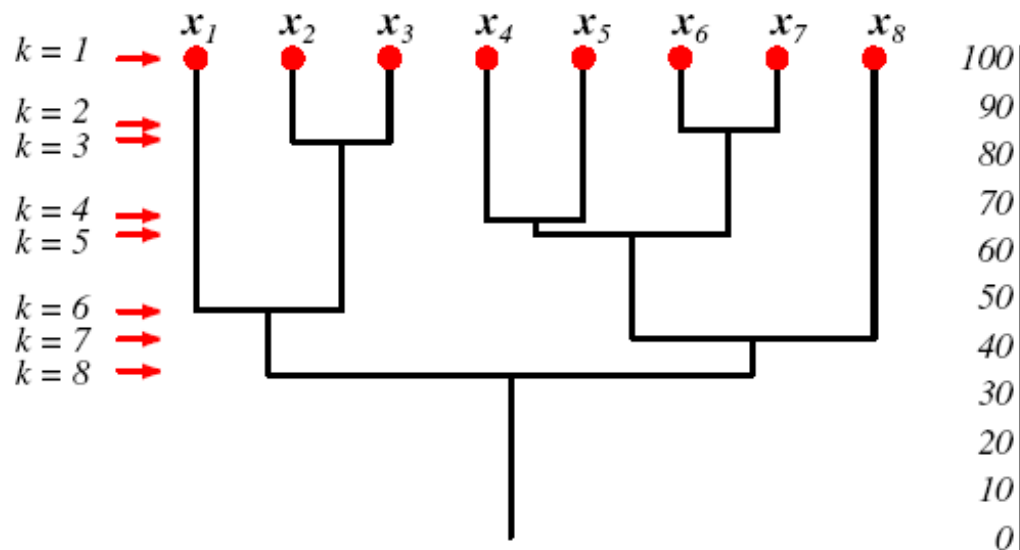


聚类和聚类分析

- **聚类**：是将数据分类到不同的类或者簇(Cluster)的过程，使得同一个簇中的对象具有最大的**相似性**，不同簇间的对象具有最大的**相异性**。



Connectivity based clustering



Hierarchical Clustering

- 初始化：每个样本作为一个单独的聚类， $C_i = \{\mathbf{x}_i\}$, $i = 1, \dots, n$ 。
- 循环，直到所有样本属于一个聚类为止：

□ 寻找当前聚类中最相近的两个聚类：

$$d(C_i, C_j) = \min_{r,s} d(C_r, C_s)$$

□ 删除聚类 C_i 和 C_j ，增加新的聚类 $C_q = C_i \cup C_j$ 。

- 输出：样本的合并过程，形成层次化谱系。

类与类之间相似性度量

- 最短距离:

$$D_{ij} = \min \left(d \left(\mathbf{x}_l^{(i)}, \mathbf{x}_k^{(j)} \right) \right)$$

为第*i*类中所有样本与第*j*类中所有样本之间距离的最小值

- 最长距离:

$$D_{ij} = \max \left(d \left(\mathbf{x}_l^{(i)}, \mathbf{x}_k^{(j)} \right) \right)$$

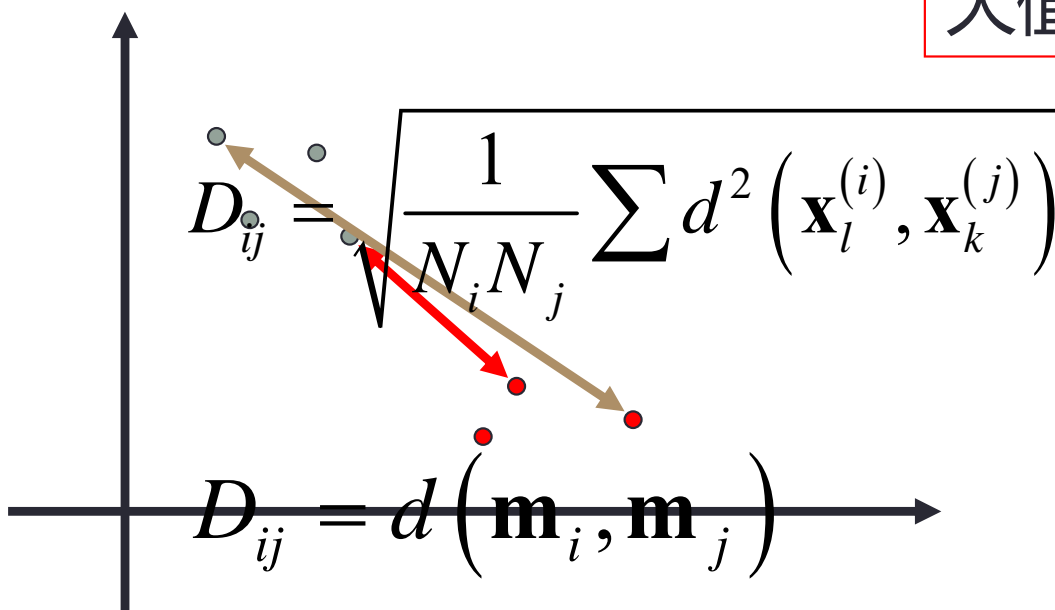
为第*i*类中所有样本与第*j*类中所有样本之间距离的最大值

- 平均距离:

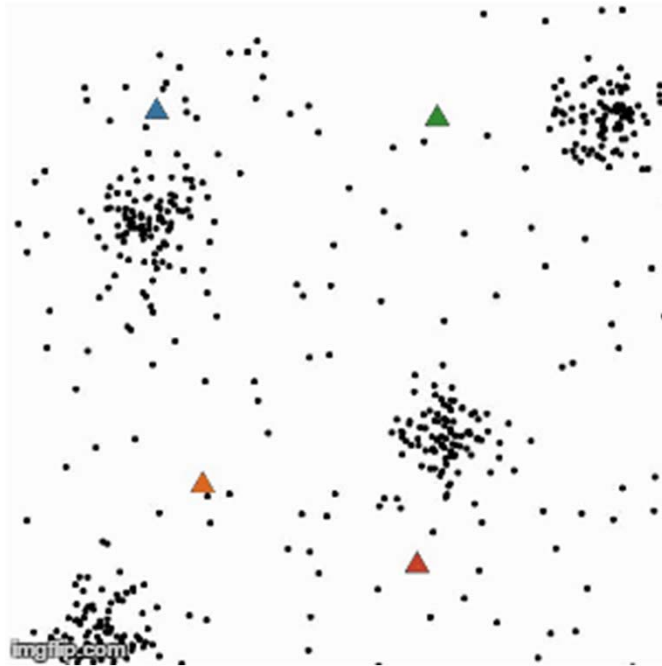
$$D_{ij} = \sqrt{\frac{1}{N_i N_j} \sum d^2 \left(\mathbf{x}_l^{(i)}, \mathbf{x}_k^{(j)} \right)}$$

- 平均样本:

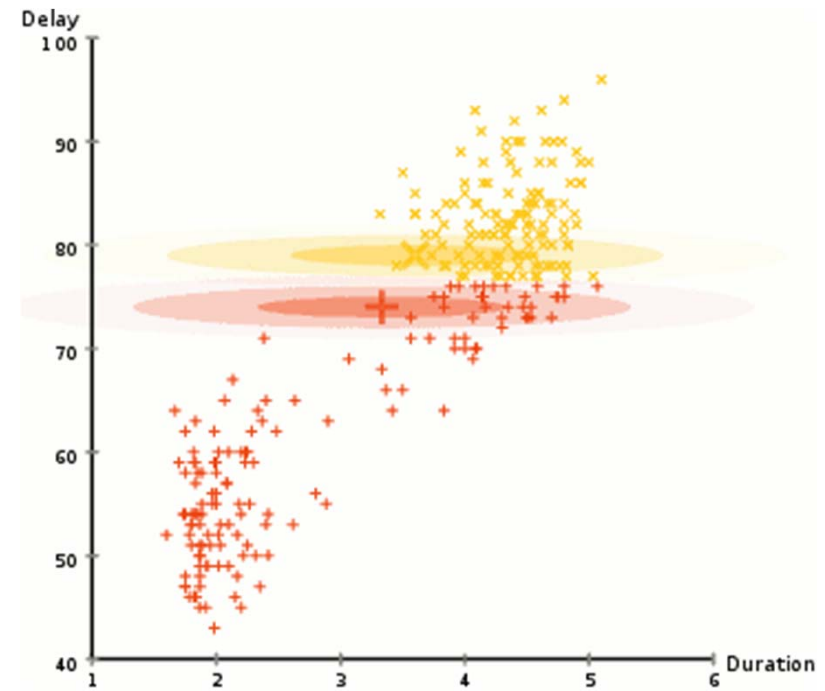
$$D_{ij} = d \left(\mathbf{m}_i, \mathbf{m}_j \right)$$



Centroid-based clustering



K-均值



混合高斯

Centroid-based clustering

K-means Clustering

- 初始化：随机选择 K 个聚类均值 \mathbf{m}_j , $j = 1, \dots, K$;
- 循环，直到 K 个均值都不再变化为止：

- $C_j = \emptyset$, $j = 1, \dots, K$;

- for $i = 1$ to n

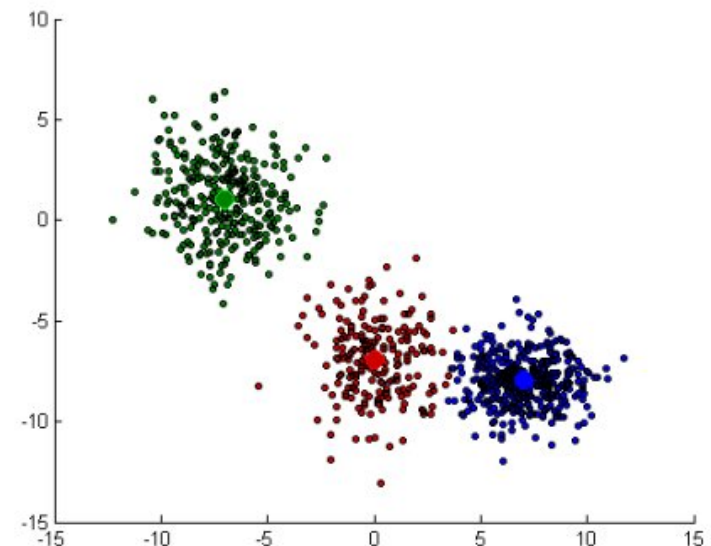
- $$k = \operatorname{argmin}_{1 \leq j \leq K} \|\mathbf{x}_i - \mathbf{m}_j\|, \quad C_k = C_k \cup \{\mathbf{x}_i\}$$

- end for

- 更新 K 个聚类的均值：

- $$\mathbf{m}_j = \frac{1}{n_j} \sum_{\mathbf{x} \in C_j} \mathbf{x}, \quad j = 1, \dots, K$$

- 输出：聚类 $\{C_1, \dots, C_K\}$
-



Distribution-based clustering

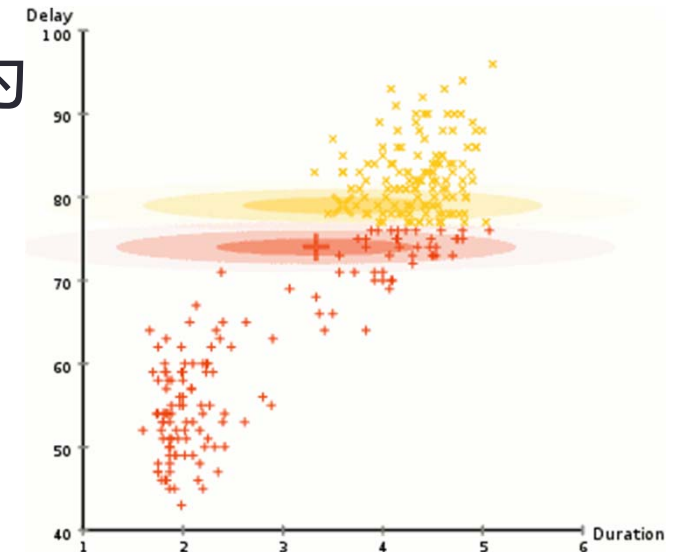
- **样本的产生过程**: 以概率 $P(\omega_j)$ 决定其所属类别 ω_j , 根据概率密度 $p(\mathbf{x}|\omega_j, \boldsymbol{\theta}_j)$ 生成一个具体的样本 \mathbf{x} 。

- **混合密度模型**: 样本 \mathbf{x} 的产生概率密度为

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{j=1}^K p(\mathbf{x}|\omega_j, \boldsymbol{\theta}_j)P(\omega_j)$$

- **混合高斯模型**:

$$p(\mathbf{x}) = \sum_{i=1}^K a_i N(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$



- **参数估计**: EM算法(Expectation-Maximization)

高斯混合模型

复杂的概率密度函数：可以由多个简单的密度函数混合构成：

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^K a_k p_k(\mathbf{x}|\boldsymbol{\theta}_k), \quad a_k > 0, \quad \sum_{k=1}^K a_k = 1$$

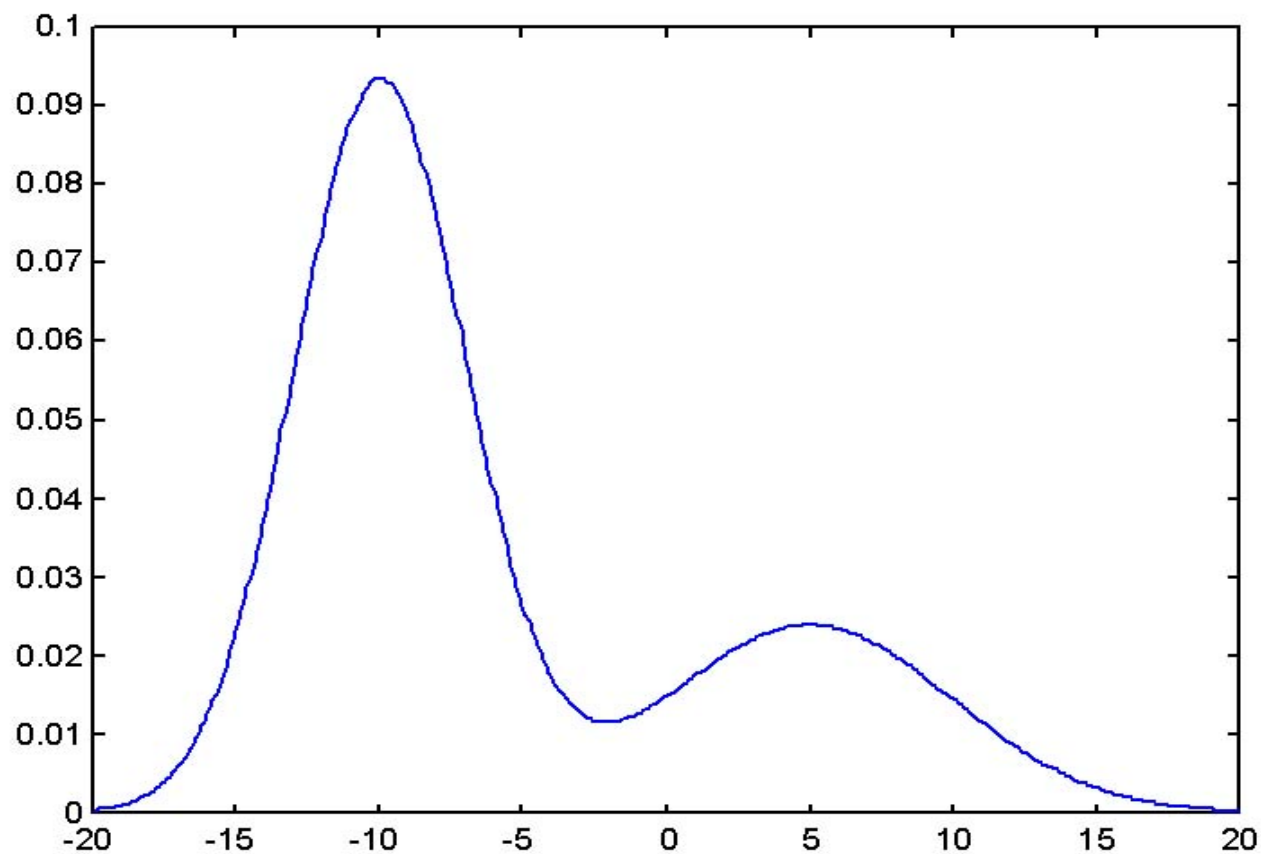
高斯混合模型(GMM, Gauss Mixture Model): 由多个高斯密度函数混合，用于逼近复杂的概率密度

$$p(\mathbf{x}) = \sum_{k=1}^K a_k N(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

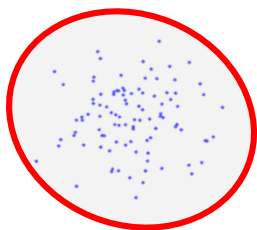
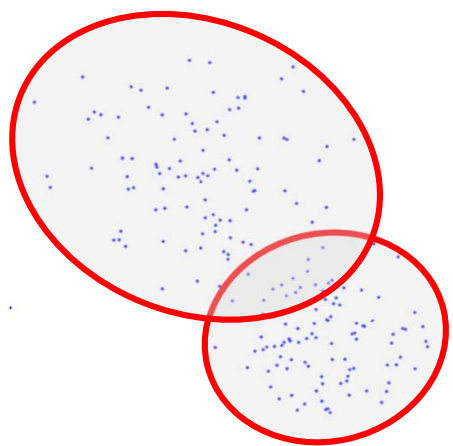
一定条件下，GMM能够以任意精度逼近任意概率密度函数！

两个高斯函数的混合

$$p(x) = 0.7N(-10, 2) + 0.3N(5, 3)$$



GMM模型的样本产生、参数估计



$$p(\mathbf{x}) = \sum_{k=1}^3 a_k N(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

样本产生:

- 1) 以组合系数 a_k 作为先验概率, 随机选择一个高斯分量 k
- 2) 利用该分量 $N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ 产生样本

参数估计:

- 1) 样本 \mathbf{x}_i 是由哪个分量产生的, 记为 $y_i \in \{1, 2, \dots, K\}$
- 2) 高斯分量的先验概率 及参数 $\boldsymbol{\theta}_k = \{a_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$

- 参数估计:
- 1) 样本 \mathbf{x}_i 是由哪个分量产生的, 记为 $y_i \in \{1, 2, \dots, K\}$
 - 2) 高斯分量的先验概率及参数 $\boldsymbol{\theta}_k = \{a_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$

已知 $\boldsymbol{\theta}_k$, 估计 y_i :

计算分量 k 产生 \mathbf{x}_n 的概率

$$p(\mathbf{x}_i | \boldsymbol{\theta}_k) = N(\mathbf{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

根据最大后验概率估计 y_i

$$y_i = \arg \max_k \alpha_k p(\mathbf{x}_i | \boldsymbol{\theta}_k)$$

已知 y_i , 估计 $\boldsymbol{\theta}_k$: 单个高斯分量参数估计

$$\alpha_k = \frac{1}{n} \sum_{i=1}^n I(y_i = k) \quad \text{示性函数} = \begin{cases} 1, & y_i = k \\ 0, & y_i \neq k \end{cases}$$

$$\boldsymbol{\mu}_k = \sum_{i=1}^n I(y_i = k) \mathbf{x}_i / \sum_{i=1}^n I(y_i = k)$$

$$\boldsymbol{\Sigma}_k = \sum_{i=1}^n I(y_i = k) (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^t / \sum_{i=1}^n I(y_i = k)$$

两者都未知时: 先随机初始化高斯分量参数 $\boldsymbol{\theta}^0$, 再迭代优化

根据第 t 次迭代得到的参数 $\boldsymbol{\theta}^t$, 估计样本集的产生分量 \mathbf{y}^t

根据 \mathbf{y}^t 估计新参数 $\boldsymbol{\theta}^{t+1}$

- 参数估计:
- 1) 样本 \mathbf{x}_i 是由哪个分量产生的, 记为 $y_i \in \{1, 2, \dots, K\}$
 - 2) 高斯分量的先验概率及参数 $\boldsymbol{\theta}_k = \{a_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$

两者都未知时: 先随机初始化高斯分量参数 $\boldsymbol{\theta}^0$, 再迭代优化

根据第t次迭代得到的参数 $\boldsymbol{\theta}^t$, 估计样本集的产生分量 \mathbf{y}^t

根据 \mathbf{y}^t 估计新参数 $\boldsymbol{\theta}^{t+1}$

当GMM的混合系数相等、协方差矩阵为相同对角阵时——K均值聚类

示性函数改进: \mathbf{y}^t 是一个不准确估计, 采用概率替代示性函数更恰当:

$$\begin{aligned} I(y_i = k) &= P(y_i = k) \\ &= a_k N(\mathbf{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) / \sum_{j=1}^K a_j N(\mathbf{x}_i; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \end{aligned}$$

- 参数估计:
- 1) 样本 \mathbf{x}_i 是由哪个分量产生的, 记为 $y_i \in \{1, 2, \dots, K\}$
 - 2) 高斯分量的先验概率及参数 $\boldsymbol{\theta}_k = \{a_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$

GMM参数估计: 先随机初始化高斯分量参数 $\boldsymbol{\theta}^0$, 再迭代优化

根据第t次迭代得到的参数 $\boldsymbol{\theta}^t$, 估计样本集的产生分量 \mathbf{y}^t

根据 \mathbf{y}^t 估计新参数 $\boldsymbol{\theta}^{t+1}$

已知 $\boldsymbol{\theta}_k$, 估计 y_i :

计算 y_i 由分量 k 产生的概率

$$P(y_i = k) = \frac{a_k N(\mathbf{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K a_j N(\mathbf{x}_i; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

已知 y_i , 估计 $\boldsymbol{\theta}_k$: 单个高斯分量参数估计

$$\alpha_k = \frac{1}{n} \sum_{i=1}^n P(y_i = k)$$

$$\boldsymbol{\mu}_k = \sum_{i=1}^n P(y_i = k) \mathbf{x}_i \bigg/ \sum_{i=1}^n P(y_i = k)$$

$$\boldsymbol{\Sigma}_k = \sum_{i=1}^n P(y_i = k) (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^t \bigg/ \sum_{i=1}^n P(y_i = k)$$

- 参数估计:**
- 1) 样本 \mathbf{x}_i 是由哪个分量产生的, 记为 $y_i \in \{1, 2, \dots, K\}$
 - 2) 高斯分量的先验概率及参数 $\boldsymbol{\theta}_k = \{a_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$

GMM学习算法

- 设置模型中高斯分量的个数 K , 随机初始参数 $\boldsymbol{\theta}$
设置收敛精度 η
- 循环: $t \leftarrow t + 1$
 - 计算训练样本由各分量产生的概率 $P(y_i = k)$;
 - 重新估计参数 $\boldsymbol{\theta}_i$;
 - 计算似然函数值 $L(\boldsymbol{\theta}) = \prod_{i=1}^n p(\mathbf{x}_i | \boldsymbol{\theta})$;
- 直到满足收敛条件: $L_t(\boldsymbol{\theta}) - L_{t-1}(\boldsymbol{\theta}) < \eta$

K 的选择: 越大拟合能力越强, 结构风险越大

参数初始化:

$$a_k > 0, \quad \sum_{k=1}^K a_k = 1$$

$\boldsymbol{\Sigma}_k$ 为对称正定矩阵

收敛条件: 似然函数变化小于阈值 η

计算稳定性:

对数阈上计算, 克服概率过小

样本过少时, 约束为对角阵

当协方差矩阵为奇异阵时, 叠加小对角阵

期望最大化算法 (Expectation Maximization EM)

样本集合由两部分构成 $D=\{X,Y\}$ ，其中 X 已知， Y 未知

$$l(\theta) = \ln p(D|\theta) = \ln p(X,Y|\theta)$$

Y 未知，无法优化；考虑 Y 所有可能情况下的对数似然函数：

$$Q(\theta) = E_Y [\ln p(X,Y|\theta)] = \int \ln p(X,Y|\theta) p(Y) dY$$

$p(Y)$ 仍然未知，首先设置 θ 的猜测值 θ^g ，在已知 X ， θ^g 的条件下估计 $p(Y|X,\theta^g)$ 替代 $p(Y)$ ：

E步: $Q(\theta;\theta^g) = \int \ln p(X,Y|\theta) p(Y|X,\theta^g) dY$

用 $Q(\theta;\theta^g)$ 替代对数似然函数进行优化

；表示这是关于 θ 的函数
 θ^g 是一个相关的固定值

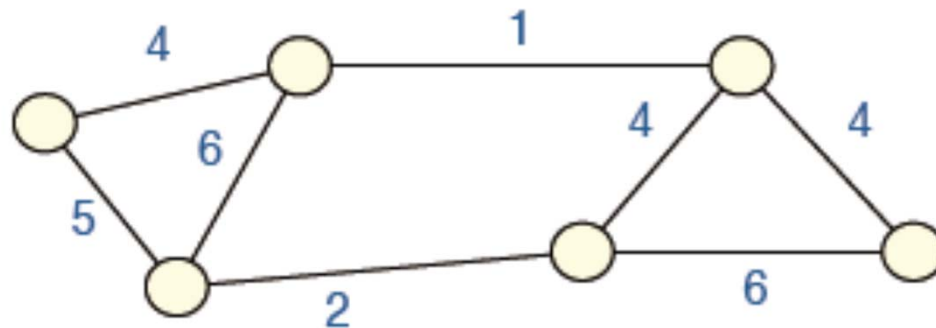
M步: $\theta^* = \arg \max_{\theta} Q(\theta;\theta^g)$

θ^* 是假设了 θ^g 之后的一个改进值

Graph-based clustering

- **相似图**：样本集 $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ 可以表示为相似图 $G = (V, E)$ 。连接两个样本的边 e_{ij} 赋予权重 w_{ij} 表示两个样本之间的相似度。图 G 也可以用邻接矩阵 $W = \{w_{ij}\}_{i,j=1,\dots,n}$ 描述。
- **无交叠子集 $A, B \subset V$ 之间的连接**：

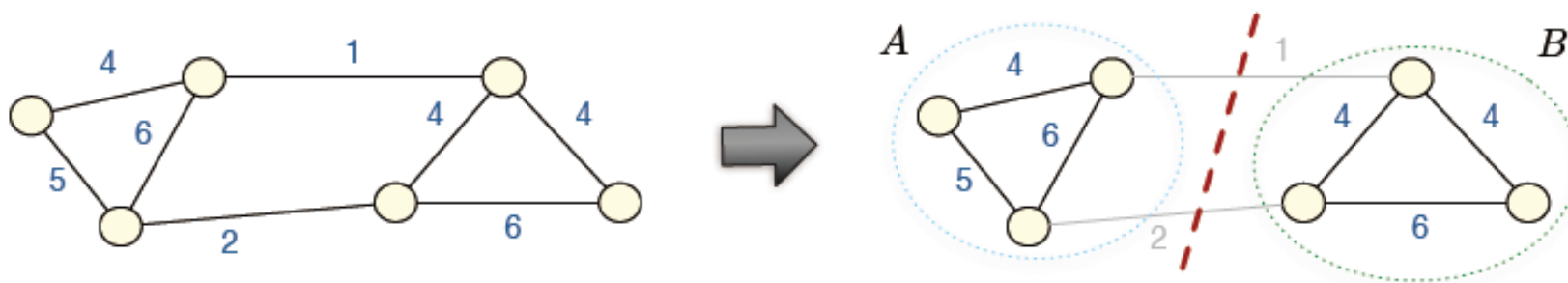
$$W(A, B) = \sum_{i \in A, j \in B} w_{ij}$$



最小割(mincut)

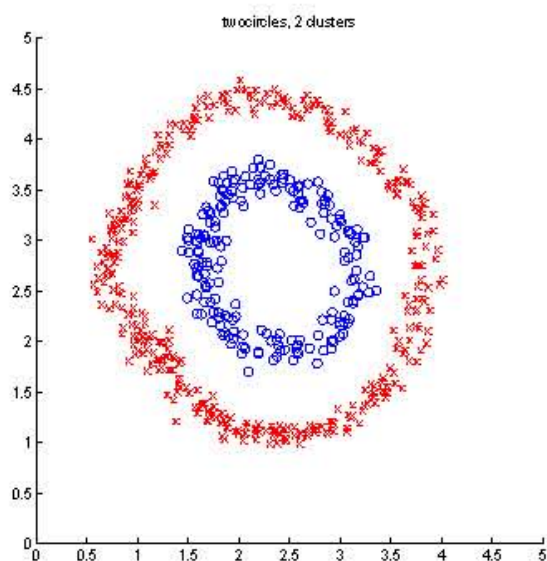
- **最小割**: 在相似图上对样本集的聚类可以看作是将图的节点划分为k个子集 A_1, \dots, A_k , 使得子集之间的连接权重最小

$$\min_{A_1, \dots, A_k} \text{Cut}(A_1, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k W(A_i, \bar{A}_i)$$

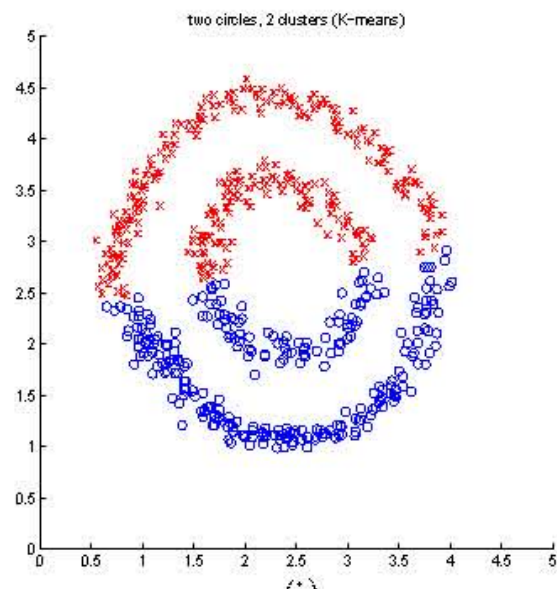


- 最小割 \rightarrow 最大流算法

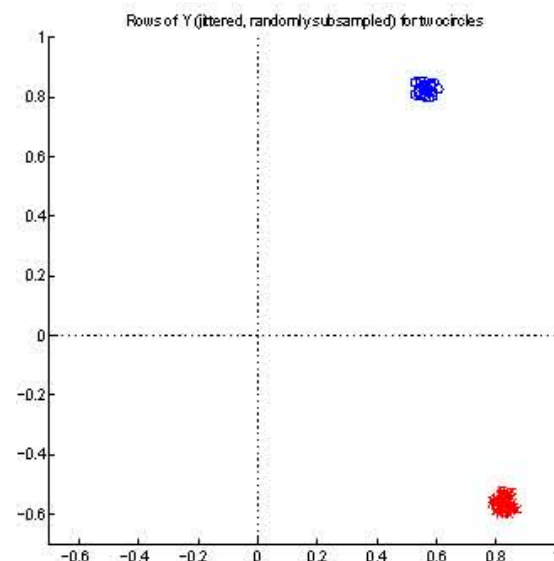
谱聚类示例



原样本分布



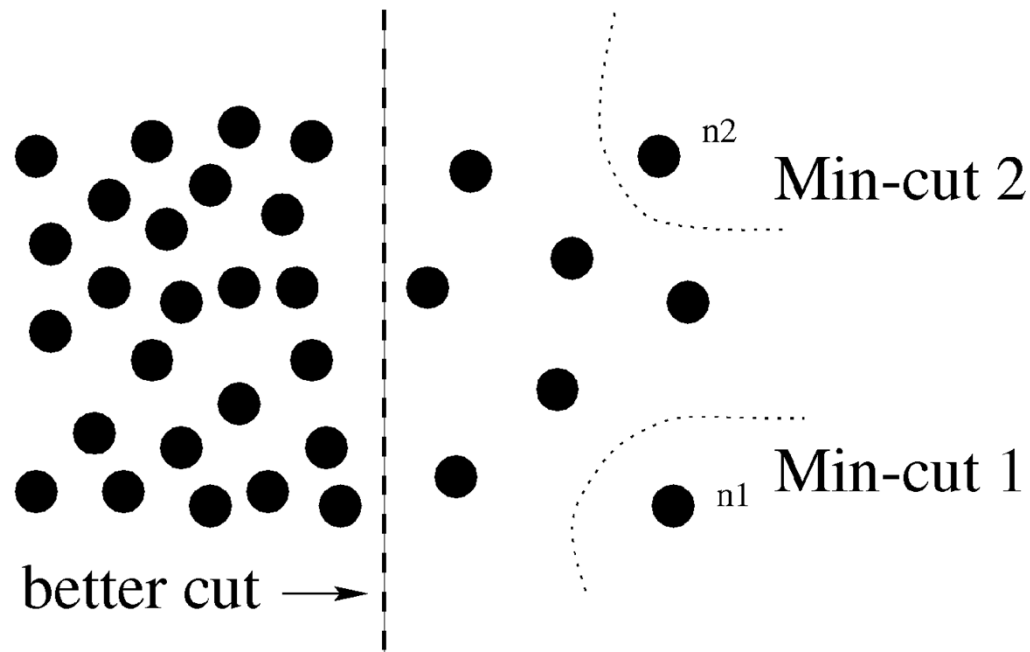
K均值聚类



特征值矩阵的行矢量

Normalized Cut

- MinCut算法在实践应用中很容易造成将单个样本划分为一个子集的现象。



Normalized Cut

- **规格化**: 用节点子集的大小规格化子集之间的连接。
- **节点的度**:

$$d_i = \sum_{j=1}^n w_{ij}$$

- **节点子集的大小**: $A \subset V$

$|A|$ = A中节点的个数

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

Normalized Cut

- Normalized Cut:**

$$RatioCut(A_1, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{|A_i|} = \sum_{i=1}^k \frac{Cut(A_i, \bar{A}_i)}{|A_i|}$$

个数归一化

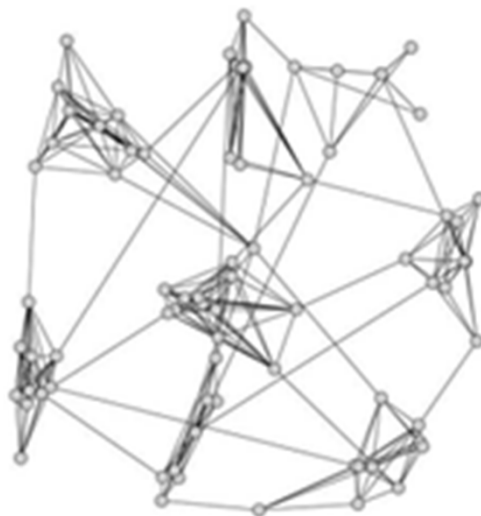
$$NCut(A_1, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{vol(A_i)} = \sum_{i=1}^k \frac{Cut(A_i, \bar{A}_i)}{vol(A_i)}$$

“度” 归一化

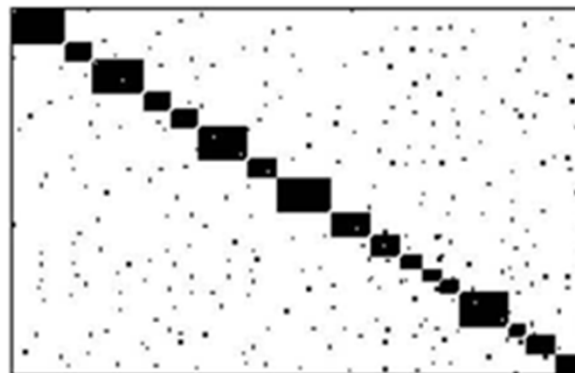
- Normalized Cut的求解是一个NP完全问题。

相似图和邻接矩阵

相似图:



邻接矩阵:



Laplacian矩阵的性质

1. 对任意矢量 \mathbf{f} , 成立:

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

证明:

$$\begin{aligned} \mathbf{f}^t L \mathbf{f} &= \mathbf{f}^t D \mathbf{f} - \mathbf{f}^t W \mathbf{f} = \sum_{i=1}^n d_i f_i^2 - \sum_{i,j=1}^n f_i f_j w_{ij} \\ &= \frac{1}{2} \left(\sum_{i=1}^n d_i f_i^2 - 2 \sum_{i,j=1}^n f_i f_j w_{ij} + \sum_{j=1}^n d_j f_j^2 \right) \\ &= \frac{1}{2} \left(\sum_{i,j=1}^n w_{ij} f_i^2 - 2 \sum_{i,j=1}^n f_i f_j w_{ij} + \sum_{i,j=1}^n w_{ij} f_j^2 \right) \\ &= \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2 \end{aligned}$$

Laplacian矩阵的性质

2. L 为对称的半正定矩阵

证明： D 和 W 为对称矩阵， L 为对称矩阵，实对称矩阵为半正定矩阵。

3. L 的最小特征值为0，对应的特征矢量为1

证明： $(D - W)\mathbf{1} = \mathbf{0}$ ，因此0为特征值， $\mathbf{1}$ 为特征矢量。半正定矩阵，所以0为最小特征值。

RatioCut的近似谱求解: $K=2$

- **定义:** n 维矢量 f (指示矢量)

$$f_i = \begin{cases} \sqrt{|\bar{A}|/|A|}, & v_i \in A \\ -\sqrt{|A|/|\bar{A}|}, & v_i \in \bar{A} \end{cases}$$

- **f 与RatioCut的关系:**

$$\begin{aligned} \mathbf{f}^t L \mathbf{f} &= \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2 \\ &= \frac{1}{2} \sum_{i \in A, j \in \bar{A}} w_{ij} \left(\sqrt{\frac{|\bar{A}|}{|A|}} + \sqrt{\frac{|A|}{|\bar{A}|}} \right)^2 + \frac{1}{2} \sum_{i \in \bar{A}, j \in A} w_{ij} \left(-\sqrt{\frac{|\bar{A}|}{|A|}} - \sqrt{\frac{|A|}{|\bar{A}|}} \right)^2 \\ &= \text{cut}(A, \bar{A}) \left(\frac{|\bar{A}|}{|A|} + \frac{|A|}{|\bar{A}|} + 2 \right) \\ &= \text{cut}(A, \bar{A}) \left(\frac{|A| + |\bar{A}|}{|A|} + \frac{|A| + |\bar{A}|}{|\bar{A}|} \right) \\ &= |V| \text{RatioCut}(A, \bar{A}) \end{aligned}$$

RatioCut的近似谱求解: $K=2$

- \mathbf{f} 与矢量 $\mathbf{1}$ 正交:

$$\sum_{i=1}^n f_i = \sum_{i \in A} \sqrt{\frac{|\bar{A}|}{|A|}} - \sum_{i \in \bar{A}} \sqrt{\frac{|A|}{|\bar{A}|}} = |A| \sqrt{\frac{|\bar{A}|}{|A|}} - |\bar{A}| \sqrt{\frac{|A|}{|\bar{A}|}} = 0$$

即:

$$\sum_{i=1}^n f_i = \mathbf{f}^t \mathbf{1} = 0$$

- \mathbf{f} 的长度平方为 n :

$$\|\mathbf{f}\|^2 = \sum_{i=1}^n f_i^2 = |A| \frac{|\bar{A}|}{|A|} + |\bar{A}| \frac{|A|}{|\bar{A}|} = |\bar{A}| + |A| = n$$

RatioCut的优化问题

- 严格的优化问题:

$$\min_{A \subset V} \mathbf{f}^t L \mathbf{f}$$

约束: $\mathbf{f}^t \mathbf{1} = 0$

$$\sum_{i=1}^n f_i^2 = n$$

$$f_i = \begin{cases} \sqrt{|\bar{A}|/|A|}, & v_i \in A \\ -\sqrt{|A|/|\bar{A}|}, & v_i \in \bar{A} \end{cases}$$

- 仍然是一个NP问题。

近似的RatioCut的优化问题

- **近似的优化问题**: 放松对 \mathbf{f} 中元素的离散性约束

$$\min_{\mathbf{f} \in R^n} \mathbf{f}' L \mathbf{f}$$

$$\text{约束: } \mathbf{f} \perp \mathbf{1}, \|\mathbf{f}\| = \sqrt{n}$$

- **问题的解**: 对应 L 第2小特征值的特征矢量

- **证明**:

1. 不考虑正交约束, 问题变成Rayleigh商的优化, 解是 L 的最小特征值对应的特征矢量;
2. 最小特征值对应特征矢量为 $\mathbf{1}$, 不满足正交条件, 第2小特征值对应特征矢量满足正交条件 (L 为实对称矩阵);

k=2 示例

- 将19个样本分成2个聚类。

$$\begin{array}{llll} x_1=(0,0)^t, & x_2=(1,0)^t, & x_3=(0,1)^t, & x_4=(1,1)^t, \\ x_5=(2,1)^t, & x_6=(1,2)^t, & x_7=(2,2)^t, & x_8=(3,2)^t, \\ x_9=(6,6)^t, & x_{10}=(7,6)^t, & x_{11}=(8,6)^t, & x_{12}=(7,7)^t, \\ x_{13}=(8,7)^t, & x_{14}=(9,7)^t, & x_{15}=(7,8)^t, & x_{16}=(8,8)^t, \\ x_{17}=(9,8)^t, & x_{18}=(8,9)^t, & x_{19}=(9,9)^t & \end{array}$$

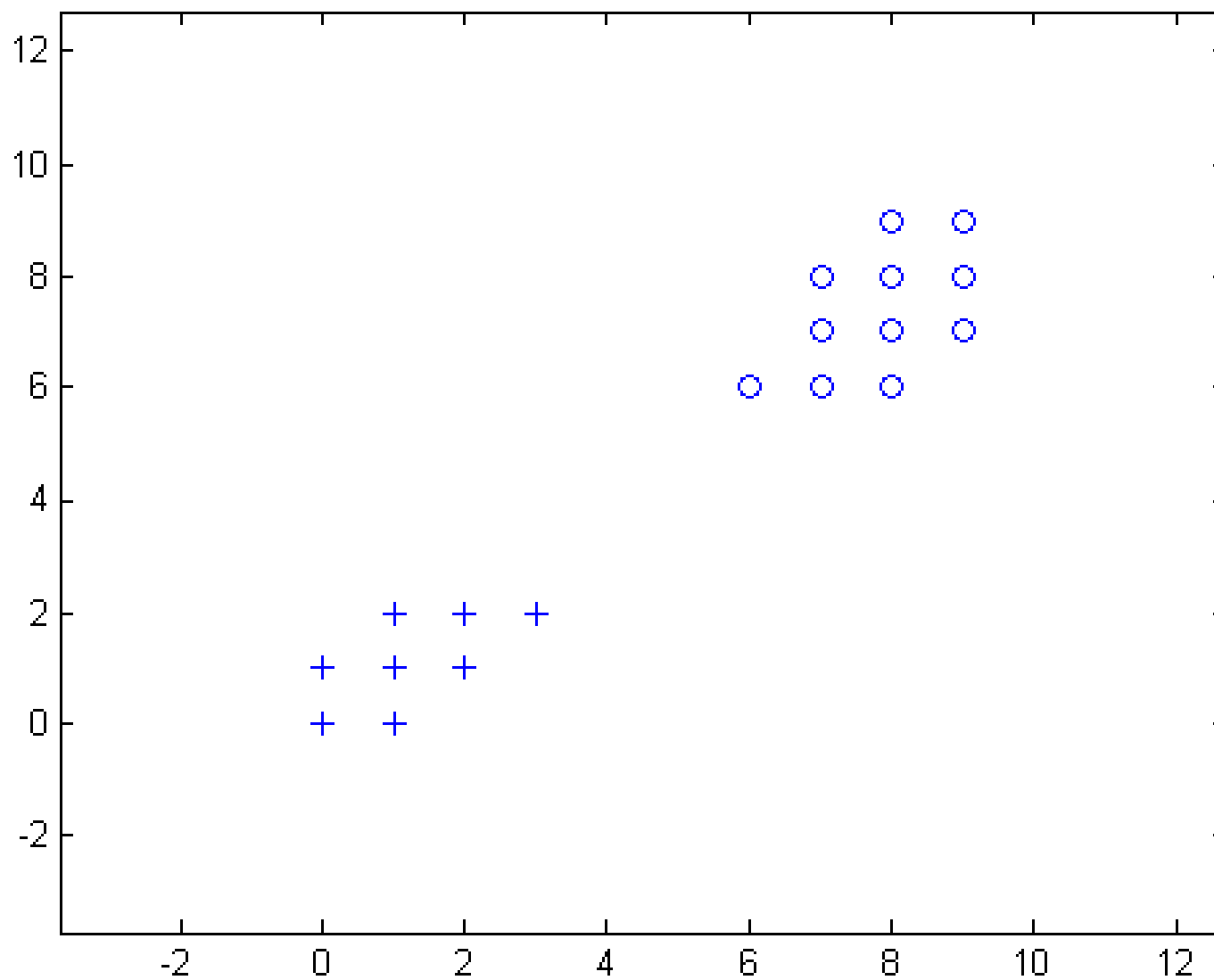
- 特征值

特征值前2个特征值对应特征矢量

0.0000
0.0682
4.3510
5.1267
5.4904
5.9142
5.9461
6.3080
6.4175
6.4826
6.7696
6.9957
7.3704
7.6983
7.7789
7.9342
8.3716
8.6444
8.8704

-0.2294	+0.2740
-0.2294	+0.2728
-0.2294	+0.2731
-0.2294	+0.2715
-0.2294	+0.2694
-0.2294	+0.2699
-0.2294	+0.2655
-0.2294	+0.2553
-0.2294	-0.1838
-0.2294	-0.1920
-0.2294	-0.1954
-0.2294	-0.1953
-0.2294	-0.1968
-0.2294	-0.1978
-0.2294	-0.1969
-0.2294	-0.1977
-0.2294	-0.1984
-0.2294	-0.1985
-0.2294	-0.1991

聚类结果



RatioCut的近似谱求解: $K > 2$

- **定义:** K 个指示矢量 $\mathbf{h}_k = (h_{1,k}, \dots, h_{n,K})^t$:

- $$h_{i,k} = \begin{cases} \frac{1}{\sqrt{|A_k|}}, & v_i \in A_k \\ 0, & \text{otherwise} \end{cases}$$

- 与RatioCut的关系:

$$\begin{aligned} \mathbf{h}_k^t L \mathbf{h}_k &= \frac{1}{2} \sum_{i,j=1}^n w_{ij} (h_{i,k} - h_{j,k})^2 \\ &= \frac{1}{2} \sum_{\substack{i \in A_k \\ j \in A_k}} w_{ij} \left(\frac{1}{\sqrt{|A_k|}} \right)^2 + \frac{1}{2} \sum_{\substack{i \in A_k \\ j \notin A_k}} w_{ij} \left(-\frac{1}{\sqrt{|A_k|}} \right)^2 \\ &= \frac{1}{|A_k|} \text{Cut}(A_k, \bar{A}_k) \end{aligned}$$

$$\text{RatioCut}(A_1, \dots, A_K) = \sum_{j=1}^K \frac{1}{|A_j|} \text{Cut}(A_j, \bar{A}_j) = \sum_{j=1}^K \mathbf{h}_j^t L \mathbf{h}_j = \text{Tr}(H^t L H)$$

例: 五个样本, 3个指示向量

$$H = \begin{pmatrix} \frac{1}{\sqrt{2}} & & \\ \frac{1}{\sqrt{2}} & & \\ & \frac{1}{\sqrt{2}} & \\ & \frac{1}{\sqrt{2}} & \\ & & 1 \end{pmatrix}$$

将样本从特征空间映射到3维
“指示”空间?

RatioCut的优化问题

- 严格的优化问题:

$$\min_{A_1, \dots, A_k} \sum_{j=1}^k \mathbf{h}_j^t L \mathbf{h}_j$$

约束: $\mathbf{h}_i^t \mathbf{h}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \Rightarrow H^t H = I$

$$h_{i,j} = \begin{cases} 1/\sqrt{|A_j|}, & v_i \in A_j \\ 0, & \text{otherwise} \end{cases}$$

- 仍然是NP问题。

$$H^t H = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & & & \\ & & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \\ & & & & 1 \\ & & & & & \frac{1}{\sqrt{2}} \\ & & & & & \frac{1}{\sqrt{2}} \\ & & & & & & 1 \end{pmatrix}$$

近似的RatioCut的优化问题

- 1, 放松对 \mathbf{h} 的离散性约束, 求解连续标签

$$\min_{\mathbf{h}_1, \dots, \mathbf{h}_k} \sum_{j=1}^k \mathbf{h}_j^t L \mathbf{h}_j$$

$$\text{约束: } \mathbf{h}_i^t \mathbf{h}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

- \mathbf{h} 为最小K个特征值对应特征矢量, 构成H矩阵

- 2, 标签离散化
H每行是一个样本的k维指示标签 (接近离散最优解的连续量)

$$H = \begin{pmatrix} \frac{1}{\sqrt{2}} & & & \\ & \frac{1}{\sqrt{2}} & & \\ & & \frac{1}{\sqrt{2}} & \\ & & & 1 \end{pmatrix}$$

对标签进行k均值聚类, 获得离散结果

NCut的近似谱求解: $K=2$

• **定义:** 指示矢量 \mathbf{f}

$$f_i = \begin{cases} \sqrt{\frac{\text{vol}(\bar{A})}{\text{vol}(A)}}, & v_i \in A \\ -\sqrt{\frac{\text{vol}(A)}{\text{vol}(\bar{A})}}, & v_i \in \bar{A} \end{cases}$$

1. $(D\mathbf{f})^t * \mathbf{1} = 0$

证明:

$$\begin{aligned} (D\mathbf{f})^t * \mathbf{1} &= \sum_{i=1}^n d_i f_i \\ &= \sum_{i \in A} d_i \sqrt{\frac{\text{vol}(\bar{A})}{\text{vol}(A)}} - \sum_{i \in \bar{A}} d_i \sqrt{\frac{\text{vol}(A)}{\text{vol}(\bar{A})}} \\ &= \text{vol}(A) * \sqrt{\frac{\text{vol}(\bar{A})}{\text{vol}(A)}} - \text{vol}(\bar{A}) * \sqrt{\frac{\text{vol}(A)}{\text{vol}(\bar{A})}} \\ &= 0 \end{aligned}$$

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

NCut的近似谱求解: $K=2$

2. $\mathbf{f}^t D \mathbf{f} = \text{vol}(V)$

证明:

$$\begin{aligned}\mathbf{f}^t D \mathbf{f} &= \sum_{i=1}^n d_i f_i^2 \\ &= \sum_{i \in A} d_i \frac{\text{vol}(\bar{A})}{\text{vol}(A)} + \sum_{i \in \bar{A}} d_i \frac{\text{vol}(A)}{\text{vol}(\bar{A})} \\ &= \text{vol}(\bar{A}) + \text{vol}(A) \\ &= \text{vol}(V)\end{aligned}$$

$$f_i = \begin{cases} \sqrt{\frac{\text{vol}(\bar{A})}{\text{vol}(A)}}, & v_i \in A \\ -\sqrt{\frac{\text{vol}(A)}{\text{vol}(\bar{A})}}, & v_i \in \bar{A} \end{cases}$$

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

NCut的近似谱求解: $K=2$

$$3. \mathbf{f}^t L \mathbf{f} = \text{vol}(V) \text{NCut}(A, \bar{A})$$

证明:

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

$$= \frac{1}{2} \sum_{\substack{i \in A \\ j \in \bar{A}}} w_{ij} \left(\sqrt{\frac{\text{vol}(\bar{A})}{\text{vol}(A)}} + \sqrt{\frac{\text{vol}(A)}{\text{vol}(\bar{A})}} \right)^2 + \frac{1}{2} \sum_{\substack{i \in \bar{A} \\ j \in A}} w_{ij} \left(-\sqrt{\frac{\text{vol}(\bar{A})}{\text{vol}(A)}} - \sqrt{\frac{\text{vol}(A)}{\text{vol}(\bar{A})}} \right)^2$$

$$= \text{Cut}(A, \bar{A}) * \left(\frac{\text{vol}(\bar{A})}{\text{vol}(A)} + \frac{\text{vol}(A)}{\text{vol}(\bar{A})} + 2 \right)$$

$$= \text{Cut}(A, \bar{A}) * \left(\frac{\text{vol}(V)}{\text{vol}(A)} + \frac{\text{vol}(V)}{\text{vol}(\bar{A})} \right)$$

$$= \text{vol}(V) \text{NCut}(A, \bar{A})$$

$$f_i = \begin{cases} \sqrt{\frac{\text{vol}(\bar{A})}{\text{vol}(A)}}, & v_i \in A \\ -\sqrt{\frac{\text{vol}(A)}{\text{vol}(\bar{A})}}, & v_i \in \bar{A} \end{cases}$$

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

NCut的优化问题

- 严格的优化问题:

$$\min_A \mathbf{f}^t L \mathbf{f}$$

约束:

$$(D\mathbf{f})^t \mathbf{1} = 0$$

$$\mathbf{f}^t D \mathbf{f} = \text{vol}(V)$$

$$f_i = \begin{cases} \sqrt{\frac{\text{vol}(\bar{A})}{\text{vol}(A)}}, & v_i \in A \\ -\sqrt{\frac{\text{vol}(A)}{\text{vol}(\bar{A})}}, & v_i \in \bar{A} \end{cases}$$

NCut的近似优化问题

- **近似的优化问题**：放松 \mathbf{f} 的离散性约束

$$\min_{\mathbf{f}} \mathbf{f}^t L \mathbf{f}$$

$$\text{约束: } (D\mathbf{f})^t \mathbf{1} = 0$$

$$\mathbf{f}^t D \mathbf{f} = \text{vol}(V)$$

- **问题的解**：对应矩阵 $(D^{-1}L)$ 第2小特征值的特征矢量。

证明：

1. 不考虑正交性约束，是一个广义的Rayleigh商问题，解是 $(D^{-1}L)$ 最小特征值对应特征矢量；
2. 最小特征矢量不满足正交性，第2小特征矢量满足。

NCut的近似谱求解: $K > 2$

- **定义**: k 个指示矢量 $\mathbf{h}_1, \dots, \mathbf{h}_k$

$$h_{i,j} = \begin{cases} 1/\sqrt{\text{vol}(A_j)}, & v_i \in A_j \\ 0, & v_i \in \bar{A}_j \end{cases}$$

1. $\mathbf{h}_i^t D \mathbf{h}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$

证明: $i \neq j$ 显然;

$$\mathbf{h}_i^t D \mathbf{h}_i = \sum_{t=1}^n d_t h_{t,i}^2 = \frac{1}{\text{vol}(A_i)} \sum_{t \in A_i} d_t = 1$$

NCut的近似谱求解: $K > 2$

$$2. \text{NCut}(A_1, \dots, A_k) = \sum_{j=1}^k \mathbf{h}_j^t L \mathbf{h}_j$$

证明:

$$\begin{aligned} \mathbf{h}_i^t L \mathbf{h}_i &= \frac{1}{2} \sum_{j,t=1}^n w_{jt} (h_{j,i} - h_{t,i})^2 \\ &= \frac{1}{2} \sum_{\substack{j \in A_i \\ t \in \bar{A}_i}} \frac{w_{jt}}{\text{vol}(A_i)} + \frac{1}{2} \sum_{\substack{j \in \bar{A}_i \\ t \in A_i}} \frac{w_{jt}}{\text{vol}(A_i)} \\ &= \text{Cut}(A_i, \bar{A}_i) / \text{vol}(A_i) \end{aligned}$$

$$\text{NCut}(A_1, \dots, A_k) = \sum_{i=1}^k \frac{\text{Cut}(A_i, \bar{A}_i)}{\text{vol}(A_i)} = \sum_{i=1}^k \mathbf{h}_i^t L \mathbf{h}_i$$

NCut的优化问题: $K > 2$

- 严格的优化问题:

$$\min_{\mathbf{h}_1, \dots, \mathbf{h}_k} \sum_{i=1}^k \mathbf{h}_i^t L \mathbf{h}_i$$

约束:

$$\mathbf{h}_i^t D \mathbf{h}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

$$h_{i,j} = \begin{cases} 1 / \sqrt{\text{vol}(A_j)}, & v_i \in A_j \\ 0, & v_i \in \bar{A}_j \end{cases}$$

NCut的近似优化问题: $K > 2$

- **近似的优化问题**: 放松 \mathbf{h} 的离散性约束

$$\min_{\mathbf{h}_1, \dots, \mathbf{h}_k} \sum_{i=1}^k \mathbf{h}_i^t L \mathbf{h}_i$$

$$\text{约束: } \mathbf{h}_i^t D \mathbf{h}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

- **问题的解**: 广义特征值

$L\mathbf{u} = \lambda D\mathbf{u}$ 的最小 k 个特征值对应的特征矢量。

谱聚类算法

https://blog.csdn.net/SL_World/article/details/104423536

- **正则化算法**

1. 计算相似图的邻接矩阵 W ;
2. 计算Laplacian矩阵 L ;
3. 求解广义特征值问题 $L\mathbf{u} = \lambda D\mathbf{u}$ 的前 k 个(最小)特征矢量 $\mathbf{u}_1, \dots, \mathbf{u}_k$;
4. 用 $\mathbf{u}_1, \dots, \mathbf{u}_k$ 作为列矢量构造矩阵 U ;
5. $\mathbf{y}_1, \dots, \mathbf{y}_n$ 为 U 的行矢量, 用K均值算法将其聚成 k 个类别。

- **对角矩阵**: $D = \text{diag}(d_i), \quad d_i = \sum_{j=1}^n w_{ij}$

- **Laplacian矩阵**: $L = D - W$

算法的实现

1. 相似图的构造

- 全连接:

$$w_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right)$$

- ε -近邻:

$$w_{ij} = \begin{cases} 1, & \|\mathbf{x}_i - \mathbf{x}_j\| < \varepsilon \\ 0, & otherwise \end{cases}$$

- 互K-近邻:

$$w_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right), & \mathbf{x}_i \in KNN(\mathbf{x}_j) \wedge \mathbf{x}_j \in KNN(\mathbf{x}_i) \\ 0, & otherwise \end{cases}$$

算法的实现

2. **特征矢量的计算**：矩阵具有稀疏性，存在快速算法 Lanczos method，最小特征值0对应的特征矢量为**1**，其它矢量与其正交；
3. **Laplacian矩阵的选择**：
 - 非正规化： $L = D - W$
 - 随机游走正规化： $L_{rw} = D^{-1}L = I - D^{-1}W$
 - 对称正规化： $L_{sym} = D^{-1/2}LD^{-1/2} = I - D^{-1/2}WD^{-1/2}$
 - 使用 L_{sym} 时，特征矢量需要乘 $D^{-1/2}$ 。推荐使用 L_{rw} 。

算法的实现

4. 聚类数的选择:

- 理想情况：样本按照类别的顺序排列，类别之间相似度为0，Laplacian矩阵前K个特征值为0

$$L = \begin{bmatrix} L_1 & & & \\ & L_2 & & \\ & & \ddots & \\ & & & L_k \end{bmatrix}$$

- 一般情况：根据特征值的分布确定聚类数

