

无监督学习

王昌栋

<https://cse.sysu.edu.cn/content/2465>

中山大学



机器智能与先进计算
教育部重点实验室

声明：该PPT只供非商业使用，也不可视为任何出版物。由于历史原因，许多图片尚没有标注出处，如果你知道图片的出处，欢迎告诉我们 at wszheng@ieee.org.



无监督学习

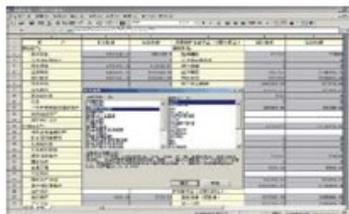
- 聚类算法
- 成分学习
- 主题模型



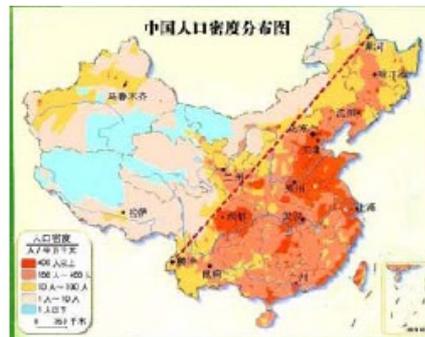
聚类算法



银行数据



企业报表



人口数据



信息大爆炸



如何有效地
分析和利用
这些数据?



交通数据



股票行情



多媒体数据

将所获得的数据分类是一种有效简化数据处理和分析的方法

聚类算法



给定一堆样本，我们希望根据某个相似性度量将这些样本**无监督**地分成若干类



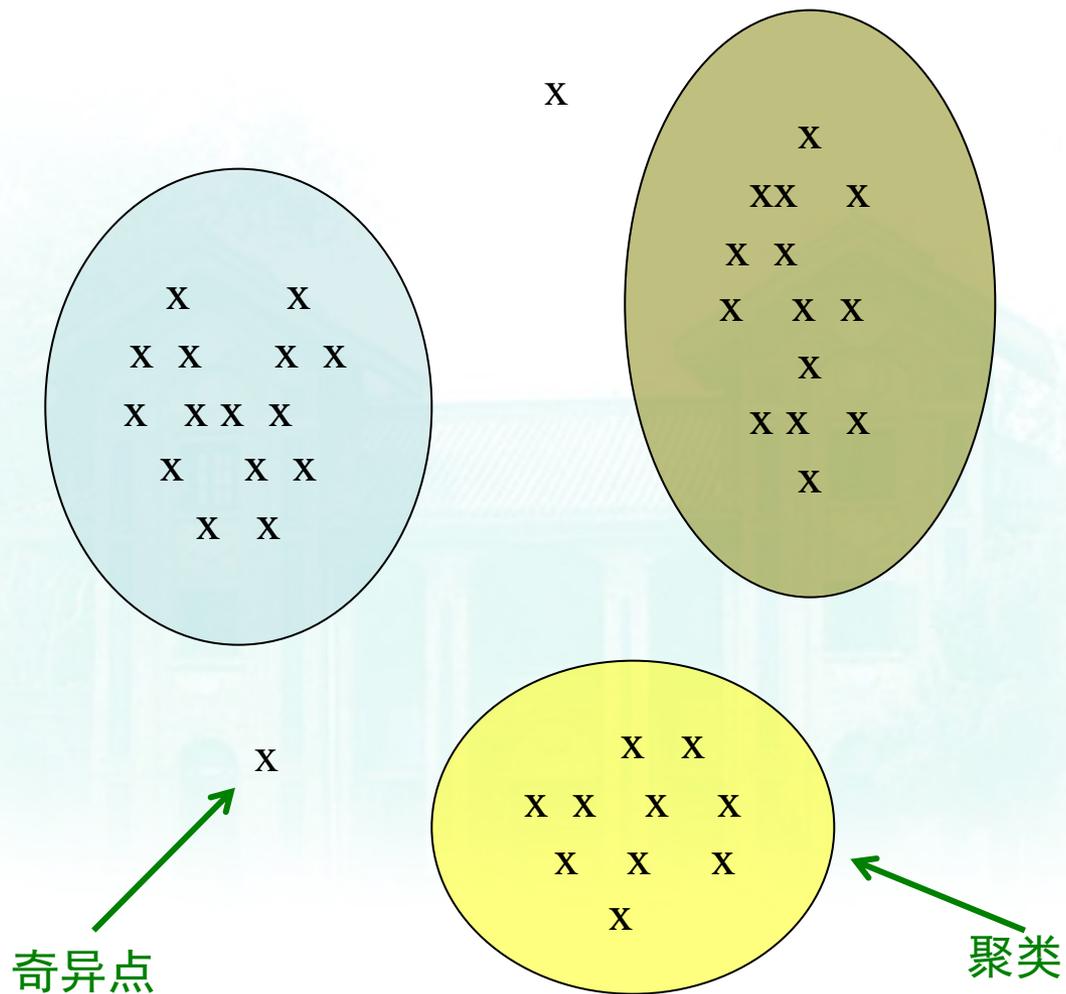
聚类问题定义：

- 给定一堆样本，我们希望根据某个相似性度量将这些样本**无监督**地分成若干类：
 - ✓ 属于同一个类的样本很相似
 - ✓ 属于不同类的样本不相似

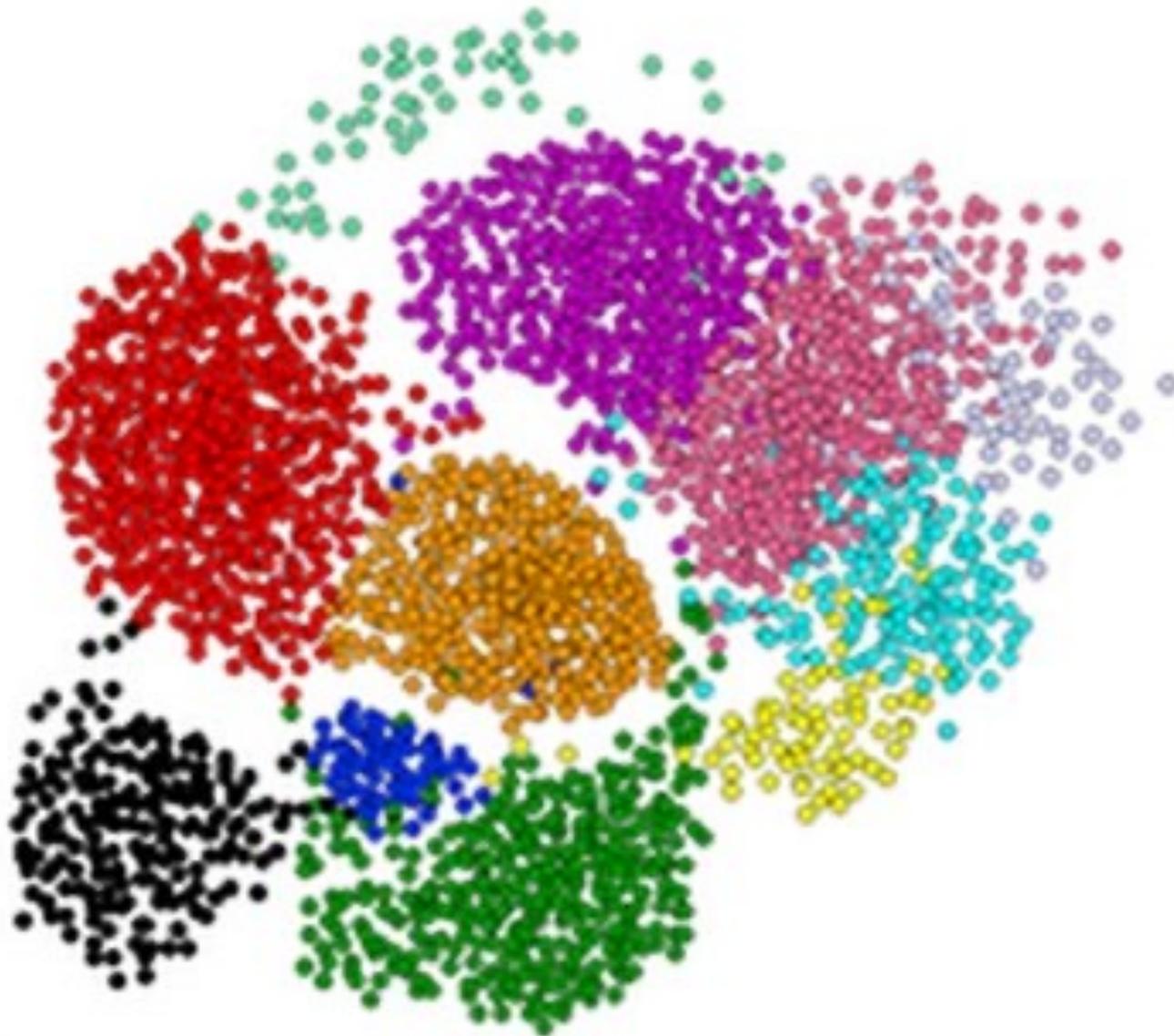
通常来说：

- 数据是位于高维空间的
- 相似性是通过一个距离度量来定义的：欧氏距离、Cos距离、Jaccard距离、编辑距离…

例：聚类&奇异点



聚类问题是一个难题！





聚类为何很难？

- ❑ 低维特别是2维样本聚类看起来很容易
- ❑ 小量样本的聚类看起来也很容易
- ❑ 在上述情况下，“所看到的”不存在欺诈性

但是

- ❑ 实际应用会涉及到成千上万维
- ❑ 高维空间与低维空间的“所见即所得”存在本质差异——在高维空间中，几乎所有的样本对的距离差别不大。



聚类问题：音乐唱片

- 将音乐唱片分成类，每个顾客只对其中很少几个类感兴趣
- 对任意一个唱片，通过已经购买该唱片的顾客集合来表示该唱片
- 相似的唱片有相似的顾客集合，反之亦然



聚类问题：音乐唱片

唱片表达空间

- 将每一个顾客看成是唱片表达空间的一个维度（二值）， k 个顾客：
- 一个唱片就是唱片表达空间中的一个样本点：
(x_1, x_2, \dots, x_k)：若顾客 i 购买了该唱片，则 $x_i=1$ ，否则 $x_i=0$

对于亚马逊来说，这个维数可能是上亿的！

目标：找到相似唱片的聚类



聚类问题：文档聚类

- 话题挖掘：
- 通过一个词向量表达一个文档：
 (x_1, x_2, \dots, x_k) ：若第 i 个词出现在该文档，
则 $x_i=1$ ，否则 $x_i=0$
- k 可能是无穷的。
- 具有相似词集合的文档很有可能是属于同一个话题的。



相似性度量的选择：Cosine, Jaccard, 欧氏？

- ❑ 在唱片或者文档聚类中，可以对“顾客集合”或者“词集合”采用如下相似性度量：
- ❑ Cosine：适用于向量
- ❑ Jaccard：适用于集合
- ❑ 欧氏：适用于样本点

Q：对于文档而言，我们是否可以采用编辑距离？

聚类算法：两大经典类型

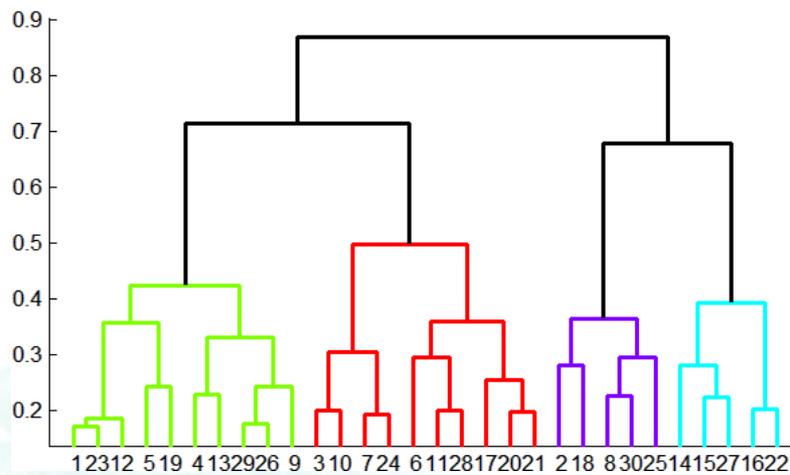
□ 层次聚类

✓ 聚合式（从下到上）：

1. 初始状态下：每个样本点都是一个类
2. 不断将最近的两个类合并成一个类

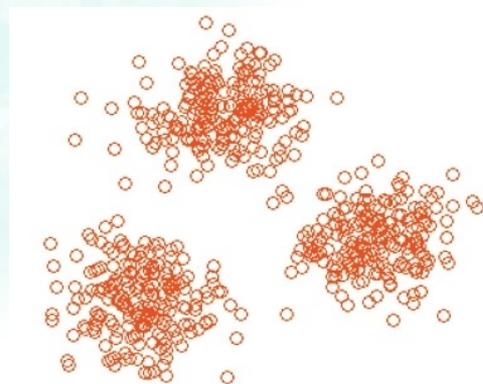
✓ 分裂式（从上到下）：

从一个类开始，不断地分裂



□ 划分聚类

1. 维护一个聚类的集合
2. 将每一个样本点划分给最近的类



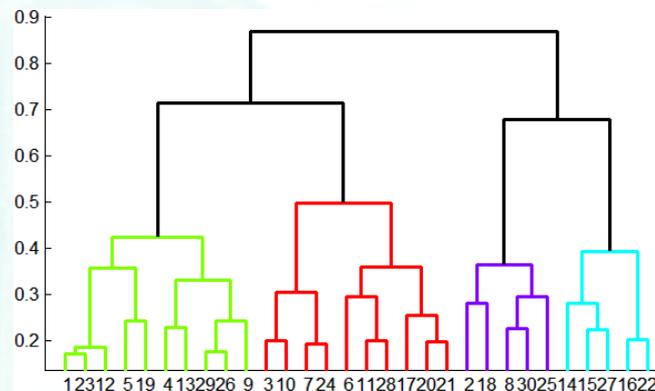
层次聚类---合并式

□ 关键操作：

不断地将两个最近的类合并成一个类

□ 三个重要问题：

1. 如何表示一个含有多个样本的类？
2. 如何判断“最近”的两个类？
3. 什么时候该终止合并？





层次聚类——合并式

□ **关键操作**：不断地将两个最近的类合并成一个类

1. 如何表示一个含有多个样本的类？

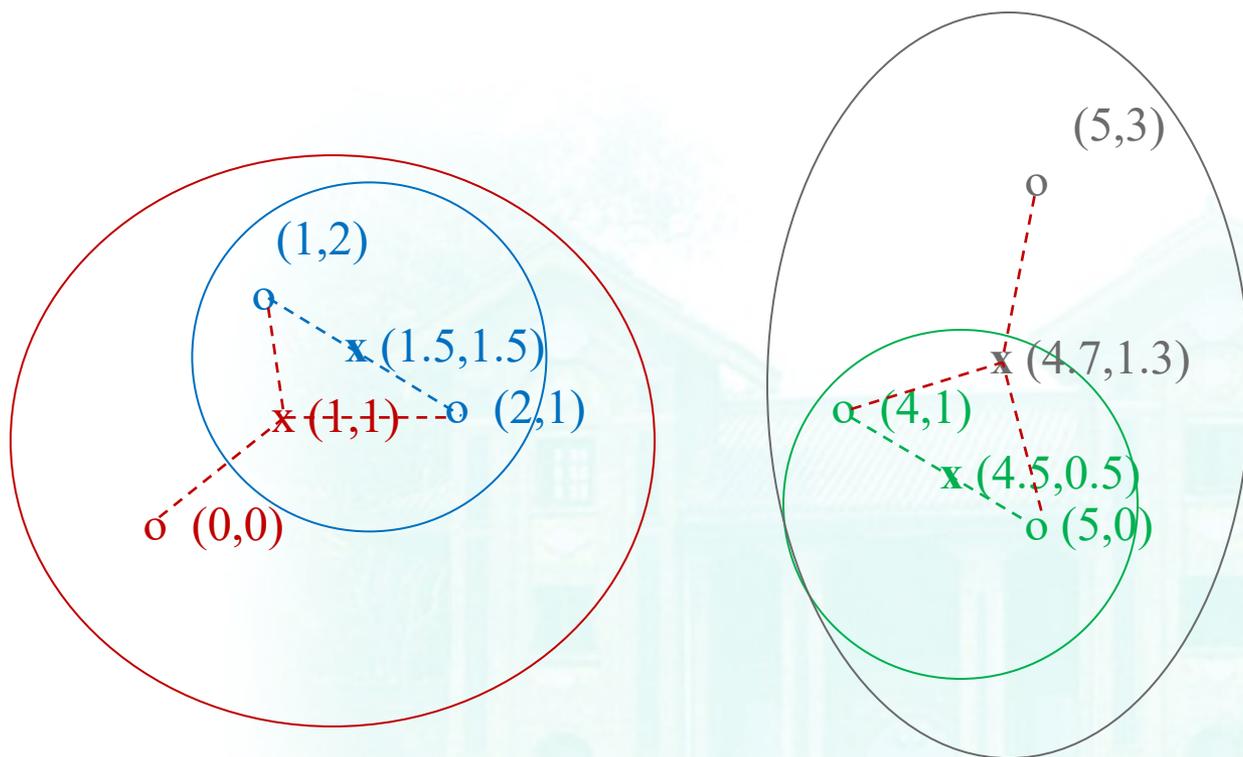
关键问题：当合并两个类的时候，如何表达每个类的“中心”，从而进一步判断两个类很近？

欧氏距离：每个类可以用类均值（centroid, 质心）来表示，即所有属于该类的样本的均值。

2. 如何判断两个类的距离？

用类均值的距离表示两个类的距离。

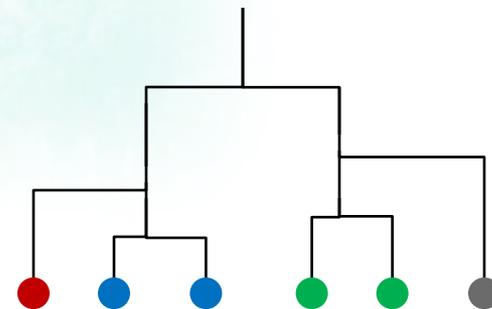
层次聚类——合并式聚类例子



数据:

o ... 样本点

x ... 类均值



树状图



非欧氏距离怎么办？

□ 非欧氏距离怎么办？

- 非欧氏距离不存在样本的均值
- 此时的均值位置只能是样本点本身

解决方案：

1) 如何表示一个含有多个样本的类？

类代表 (clustroid) = 最靠近其他样本点的样本点

2) 如何判断两个类的距离？

将类代表看成是类中心，用类代表的距离表示类的距离

类代表“最靠近”？

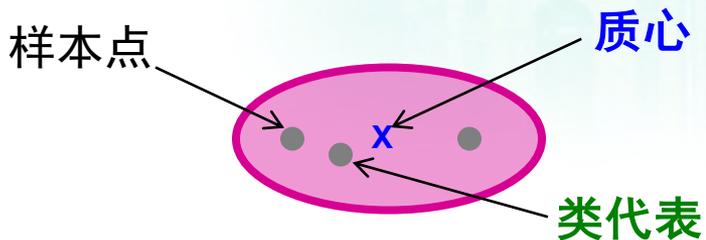
1) 如何表示一个含有多个样本的类？

类代表 (clustroid) = 最靠近其他样本点的样本点

类代表“最靠近”的多种可能的含义：

1. 与其他样本点的最大距离最小 (QA: 公式?)
2. 与其他样本点的平均距离最小 (QA: 公式?)
3. 与其他样本点的距离平方之和最小

$$\arg \min_c \sum_{x \in C} d(x, c)^2$$



含有3个样本的类

质心 vs. 类代表：
质心：虚拟的样本点
类代表：实际的样本点



两个类“最近”的定义

2) 如何判断两个类的距离?

- 类间距离=分别来自两个类的样本所形成的样本对的最小距离
- 内聚力度量：即将两个合并之后具有最强内聚力的类合并在一起
 1. 内聚力定义1: 类的最大直径即类内样本对的最大距离
 2. 内聚力定义2: 类的所有样本对的距离均值
 3. 内聚力定义3: 基于某种密度相关性

什么时候该终止合并?



K-means 聚类

- 假设欧氏距离/欧氏空间
- 从选择聚类个数 k 开始
- 随机初始化：为每个类选择一个初始中心点
 - ✓ 例如：随机选第一个样本作为第一个类的初始类中心，接下来随机选 $k-1$ 个样本作为相应类的初始类中心，每一个都尽量远离之前的样本



K-means 聚类更新过程

1. 类标重分配：对每一个样本，将它分配给距离最近的类中心所在的聚类
 2. 更新类中心：当所有样本的类标都更新之后，更新每一个类的类中心
- 不断重复上述两个步骤直至收敛为止
- 收敛：样本点的类标不再改变或者是类中心不再更新

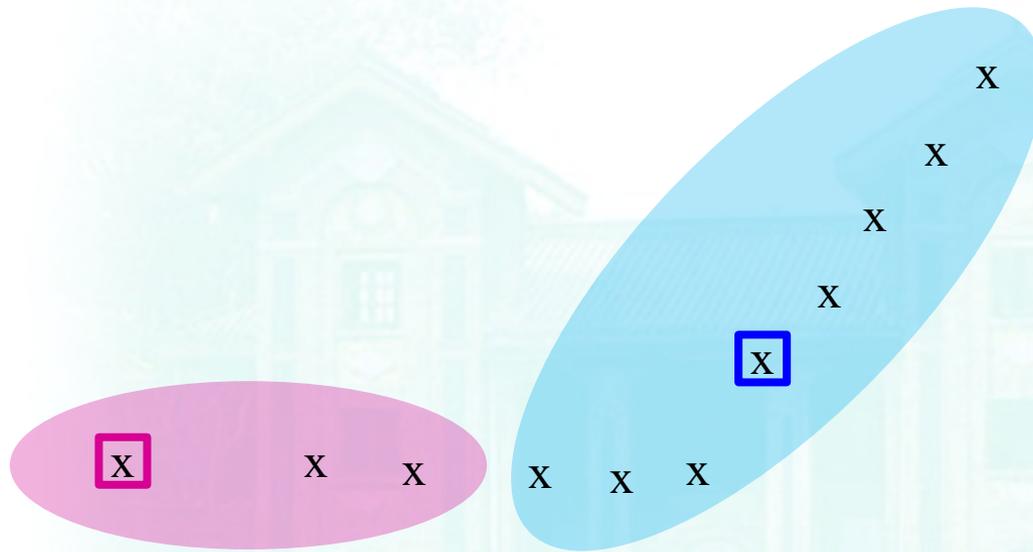


K-means 聚类

课堂Q&A:

- 1) k -means的目标函数?
- 2) 计算复杂度?

K-means 聚类——演示

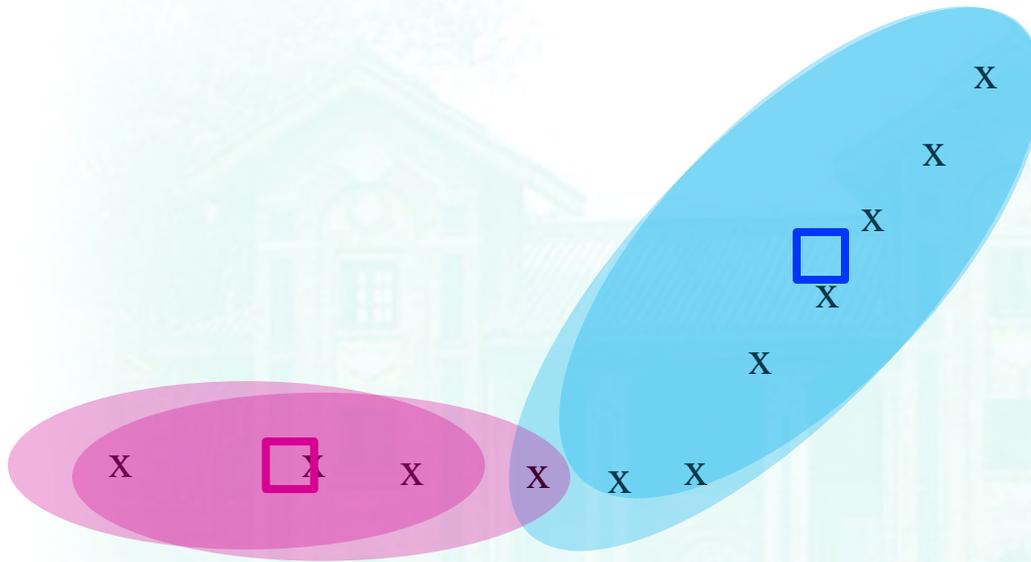


x ... data point

□ ... centroid

Clusters after round 1

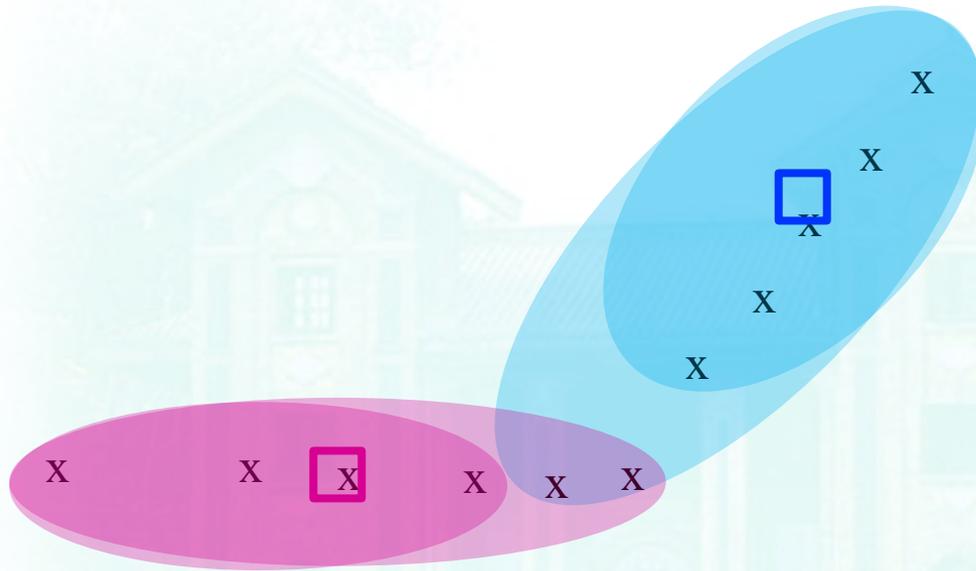
K-means 聚类——演示



x ... data point
□ ... centroid

Clusters after round 2

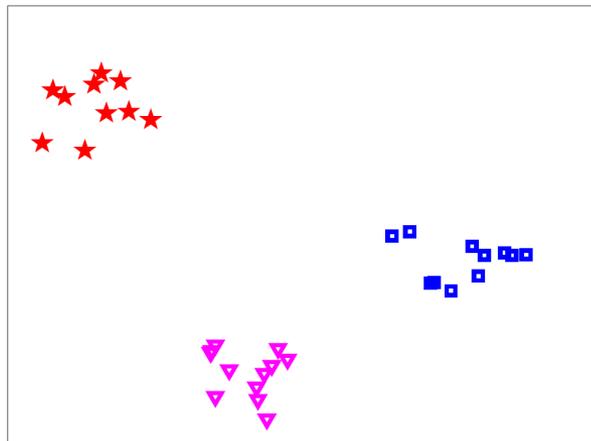
K-means 聚类——演示



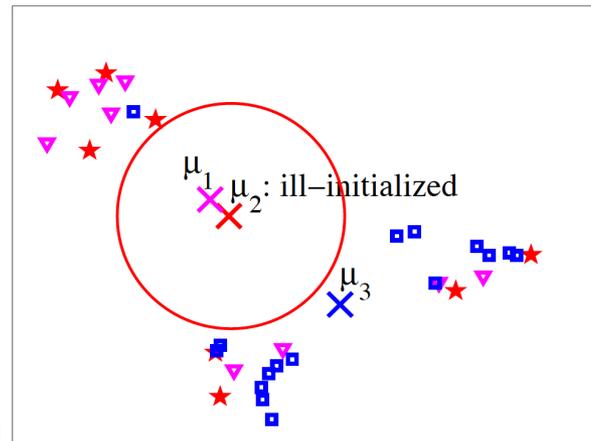
x ... data point
□ ... centroid

Clusters at the end

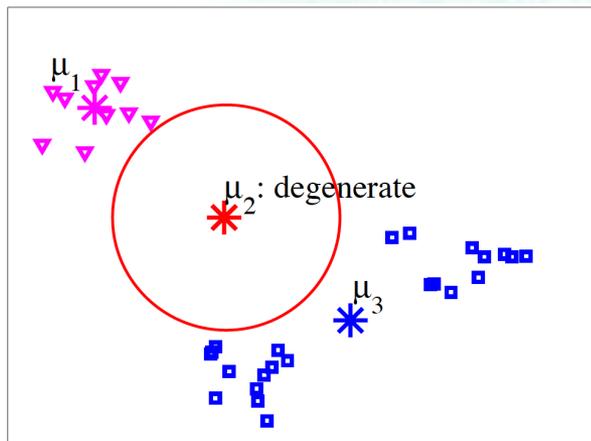
K-means 聚类——类中心病态初始化



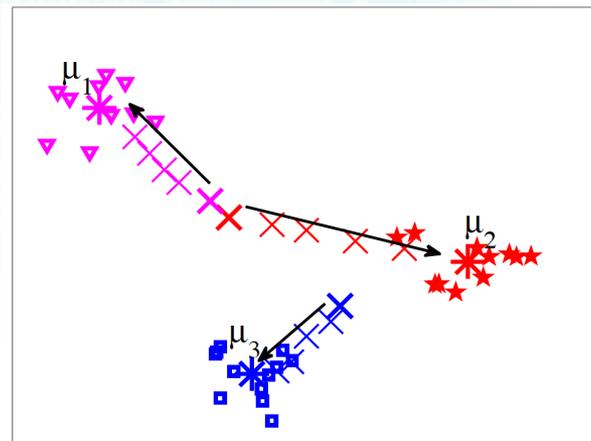
(a) 数据集



(b) 病态初始化



(c) k-means的聚类结果

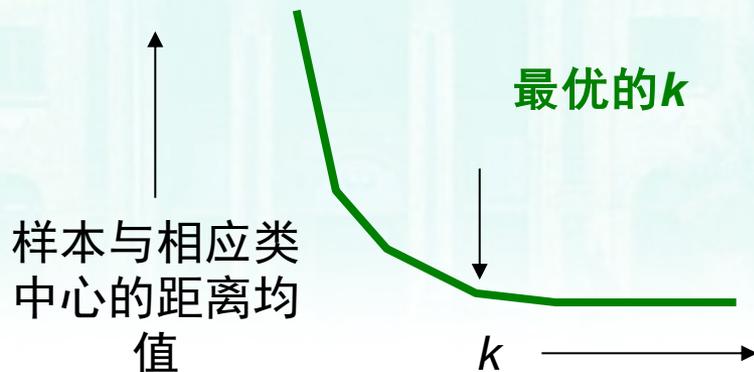


(d) COLL算法的聚类结果

选取合适的k

□ 如何选取合适的k?

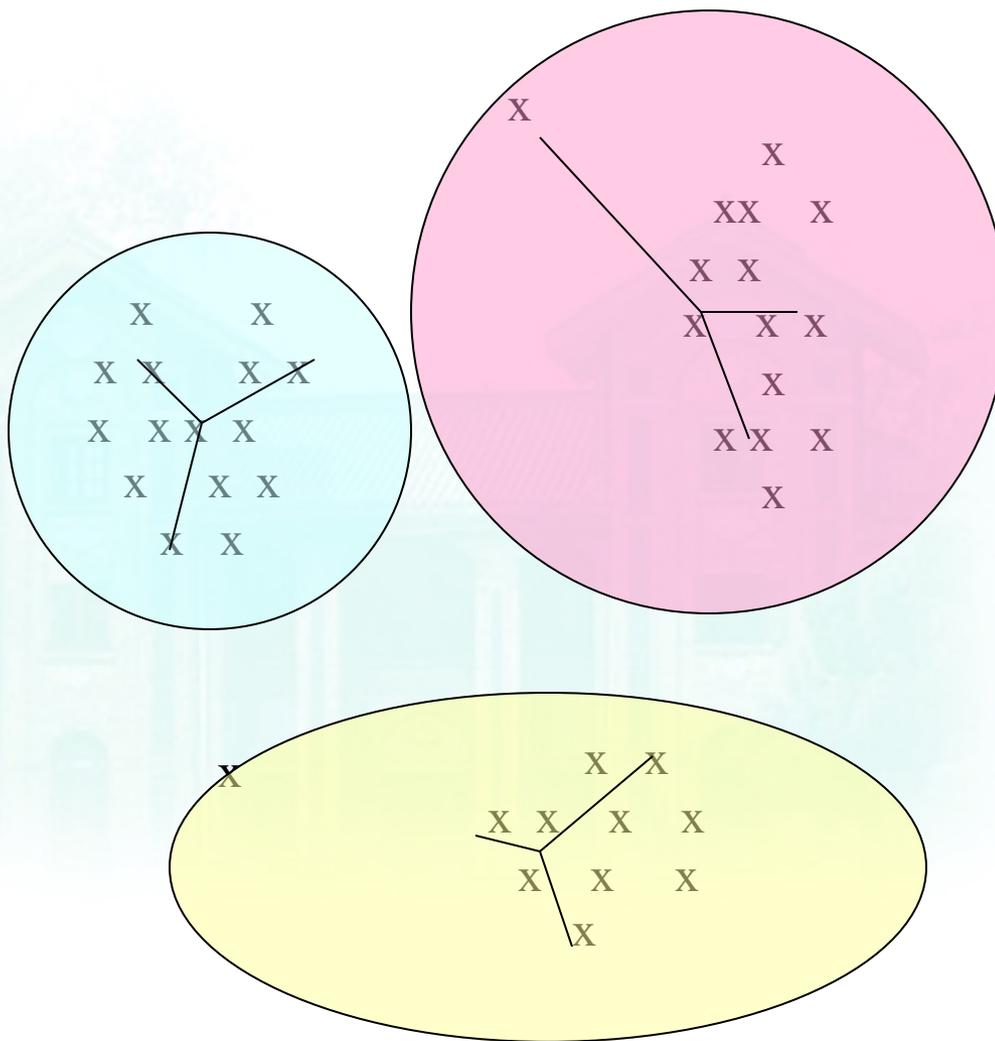
1. 在一定范围内尝试不同的k，运行k-means得到聚类结果，计算出每个样本与相应类中心的距离均值
2. 随着k增大，均值会快速下降，接着几乎保持不变，拐点即为最优的k所对应的点



k的选择是一个典型模型选择问题

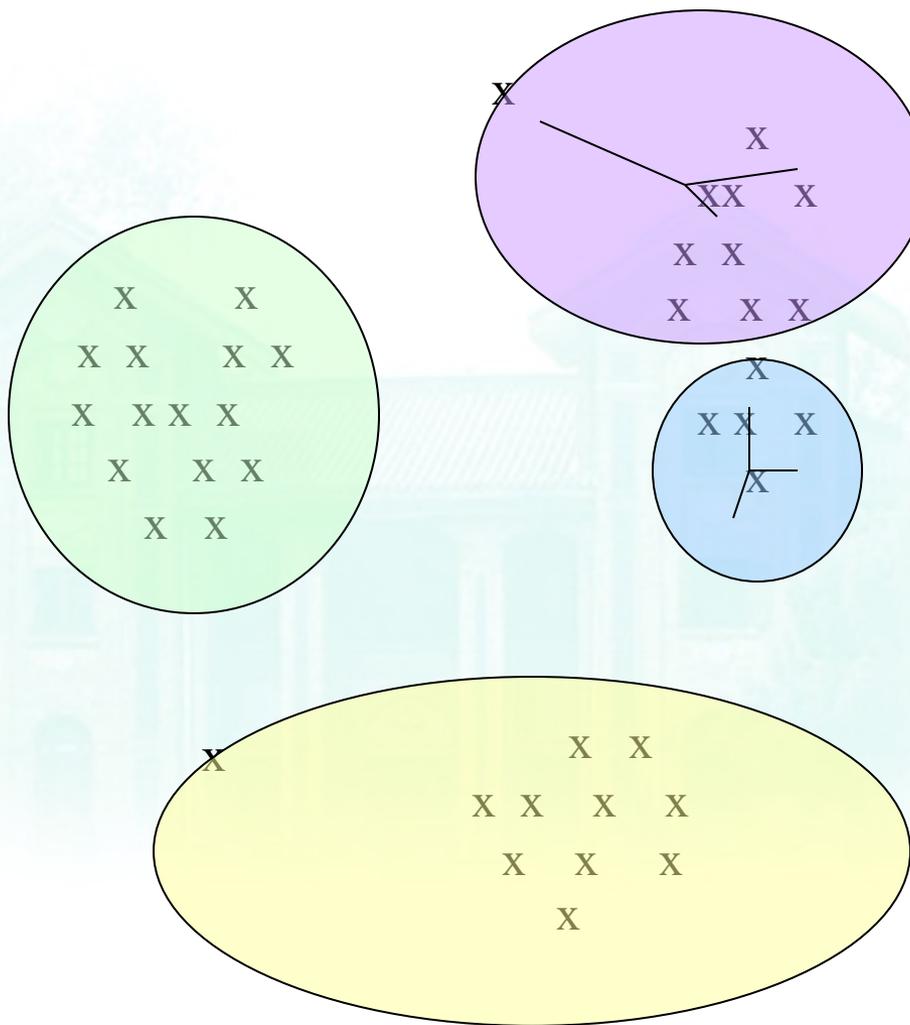
选取合适的k

刚好合适的k:
距离均值相对
较小

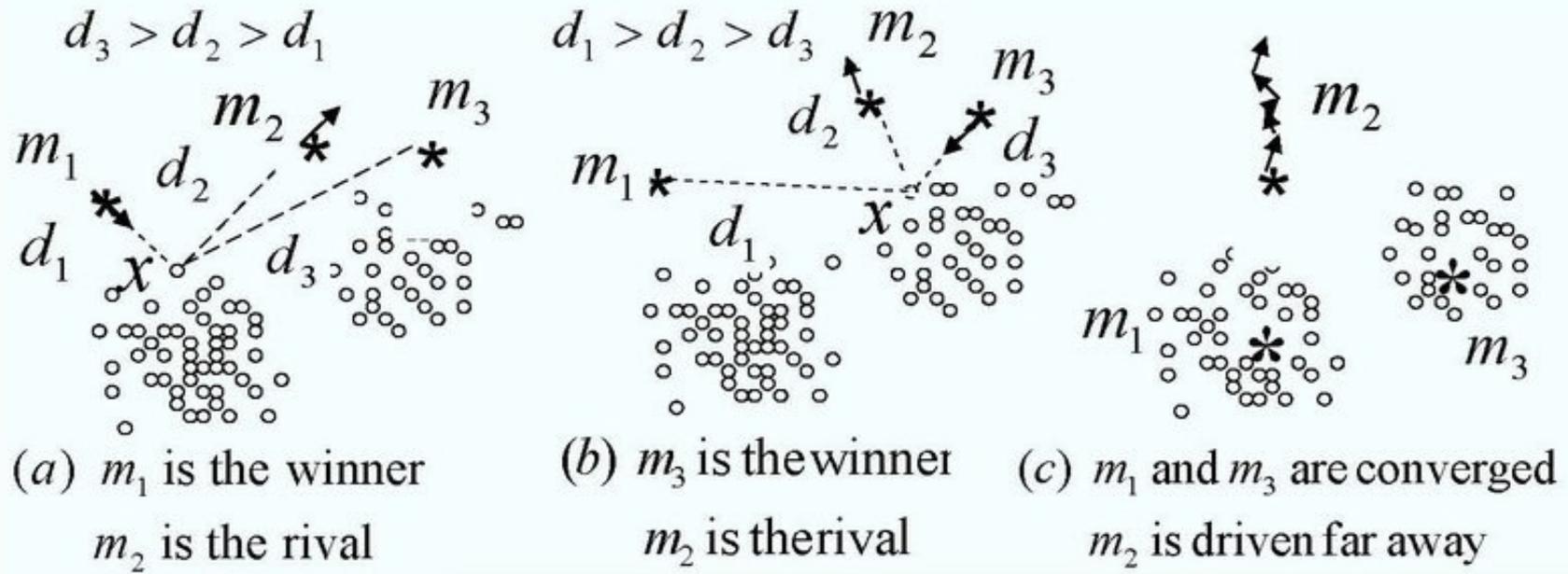


选取合适的k

k太大：
距离均值几
乎没有改变



选取合适的k: RPCL



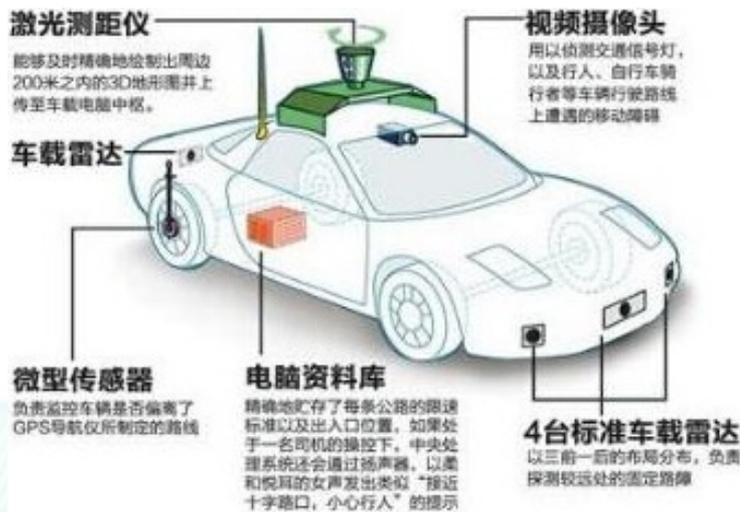
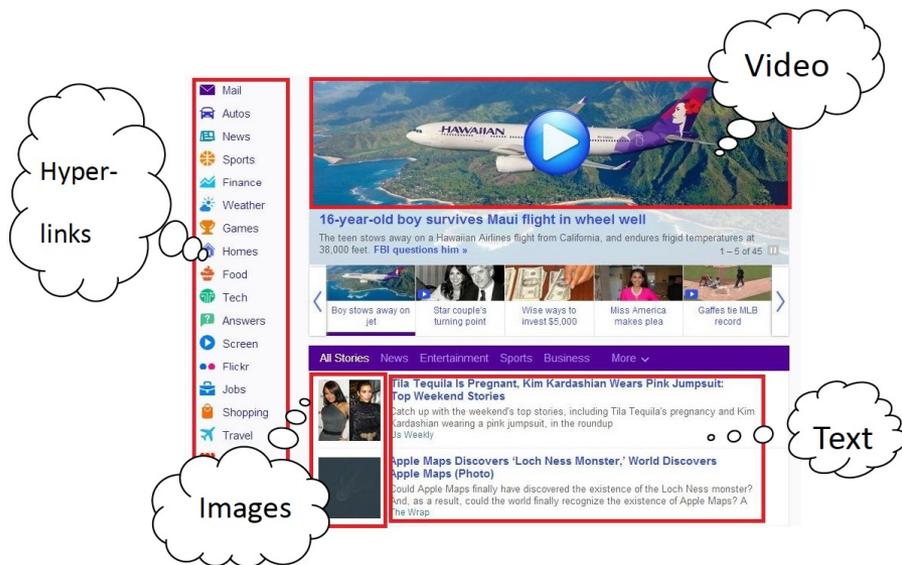


聚类算法：前沿

- 多视图聚类
- 多视图流数据聚类



多视图聚类



多视图数据:

例1: 网页数据通常同时包含**超链接**、**视频**、**图像**、**文本**等元素, 每一种元素便是该网页的一个**视图**。

例2: 无人驾驶汽车会采集来自多种视图的数据, 比如: **多个角度的摄像头**、**红外线传感器**、**超声波传感器**等等设备返回的数据。

不同视图之间存在着很强的**互补性**、**一致性**。

多视图聚类主要方法

- ❑ 相似性矩阵串联
- ❑ 视图约束策略
- ❑ 张量分解
- ❑ 多视图谱聚类
- ❑ 相似性矩阵学习
- ❑

refine by year

2021 (1)

2020 (110)

2019 (102)

2018 (63)

2017 (53)

2016 (45)

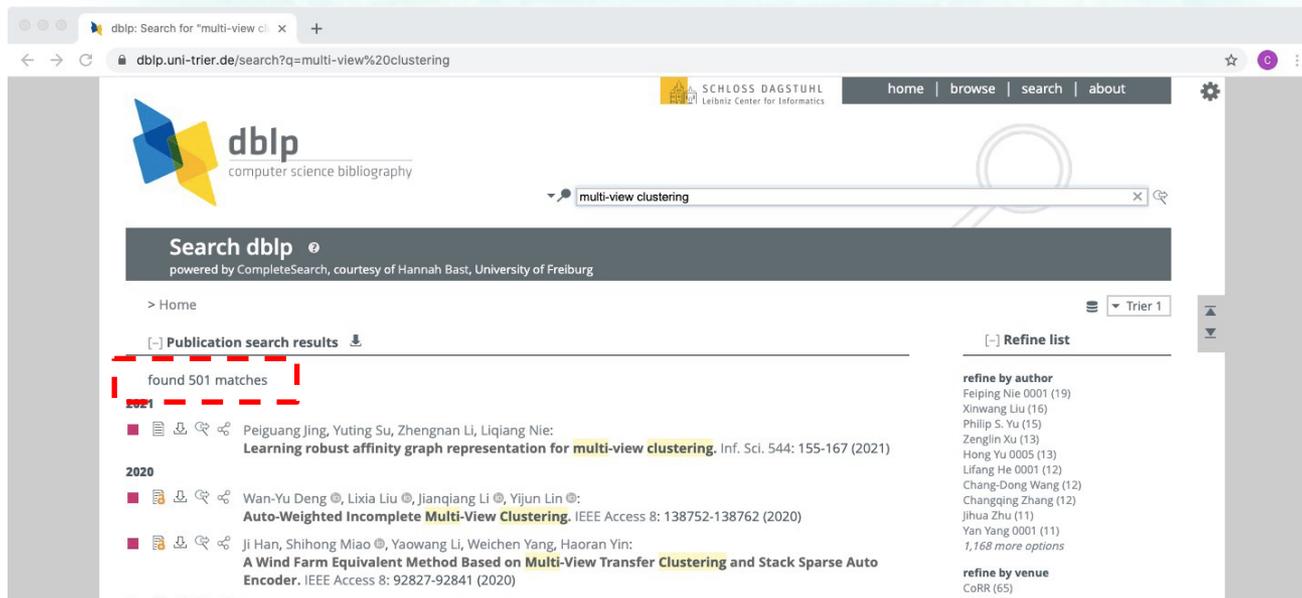
2015 (32)

2014 (19)

2013 (16)

2012 (17)

11 more options



The screenshot shows the dblp search results for 'multi-view clustering'. The search bar contains 'multi-view clustering' and the results show 501 matches. The first result is from 2021, titled 'Learning robust affinity graph representation for multi-view clustering' by Peiguang Jing, Yuting Su, Zhengnan Li, and Liqiang Nie. The second result is from 2020, titled 'Auto-Weighted Incomplete Multi-View Clustering' by Wan-Yu Deng, Lixia Liu, Jianqiang Li, and Yijun Lin. The third result is also from 2020, titled 'A Wind Farm Equivalent Method Based on Multi-View Transfer Clustering and Stack Sparse Auto Encoder' by Ji Han, Shihong Miao, Yaowang Li, Weichen Yang, and Haoran Yin. On the right side, there is a 'Refine list' section with 'refine by author' and 'refine by venue' options.

多视图相似性传播聚类 (MVAP)

模型的标目标函数

视图内聚类质量

视图间聚类一致性

$$J(\{\mathbf{c}^1, \dots, \mathbf{c}^m\}) = \sigma \sum_{p=1}^m J_w(\mathbf{c}^p) + (1 - \sigma) \sum_{p=1, q=1, p \neq q}^m J_b(\mathbf{c}^p, \mathbf{c}^q)$$

$$J_w(\mathbf{c}^p) = \sum_{i=1}^n \tilde{S}^p(i, \mathbf{c}_i^p) + \sum_{k=1}^n \tilde{\delta}_k^p(\mathbf{c}^p)$$

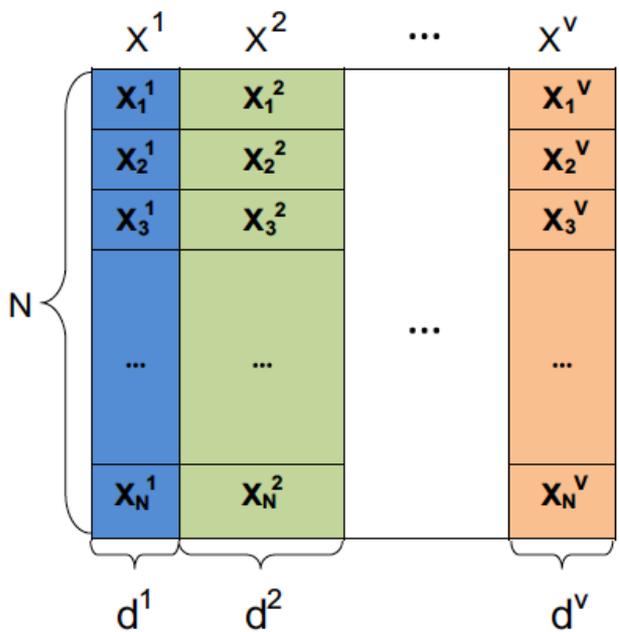
$$J_b(\mathbf{c}^p, \mathbf{c}^q) = \sum_{i=1}^n \tilde{L}^{pq}(\mathbf{c}_i^p, \mathbf{c}_i^q)$$

$$\tilde{\delta}_k^p(\mathbf{c}^p) = \begin{cases} -\infty & \text{If } \mathbf{c}_k^p \neq k \text{ but } \exists i : \mathbf{c}_i^p = k \\ 0 & \text{Otherwise} \end{cases}$$

$$\begin{aligned} \tilde{L}^{pq}(i, j) &= \text{JaccardSim}(\mathcal{N}_k^p(\mathbf{x}_i), \mathcal{N}_k^q(\mathbf{x}_j)) \\ &= \frac{|\mathcal{N}_k^p(\mathbf{x}_i) \cap \mathcal{N}_k^q(\mathbf{x}_j)|}{|\mathcal{N}_k^p(\mathbf{x}_i) \cup \mathcal{N}_k^q(\mathbf{x}_j)|} \end{aligned}$$

$$J(\{\mathbf{c}^1, \dots, \mathbf{c}^m\}) = \sum_{p=1}^m \left(\sum_{i=1}^n \sigma \cdot \tilde{S}^p(i, \mathbf{c}_i^p) + \sum_{k=1}^n \sigma \cdot \tilde{\delta}_k^p(\mathbf{c}^p) \right) + \sum_{p=1, q=1, p \neq q}^m \left(\sum_{i=1}^n (1 - \sigma) \cdot \tilde{L}^{pq}(\mathbf{c}_i^p, \mathbf{c}_i^q) \right)$$

带特征提取的视图加权多视图聚类 (WMCFS)



$$\min_{\{\delta_{ik}\}_{k=1}^M, \{\omega_v\}_{v=1}^V, \{\tau_l^v\}_{l=1}^{d^v}} \mathcal{E}_H, \quad \mathcal{E}_H = \sum_{v=1}^V (\omega_v)^p \sum_{i=1}^N \sum_{k=1}^M \delta_{ik} \|\phi^v(\mathbf{x}_i^v) - \mathbf{m}_k^v\|^2$$

subject to $\sum_{k=1}^M \delta_{ik} = 1, \quad \forall i, \delta_{ik} \in \{0, 1\},$

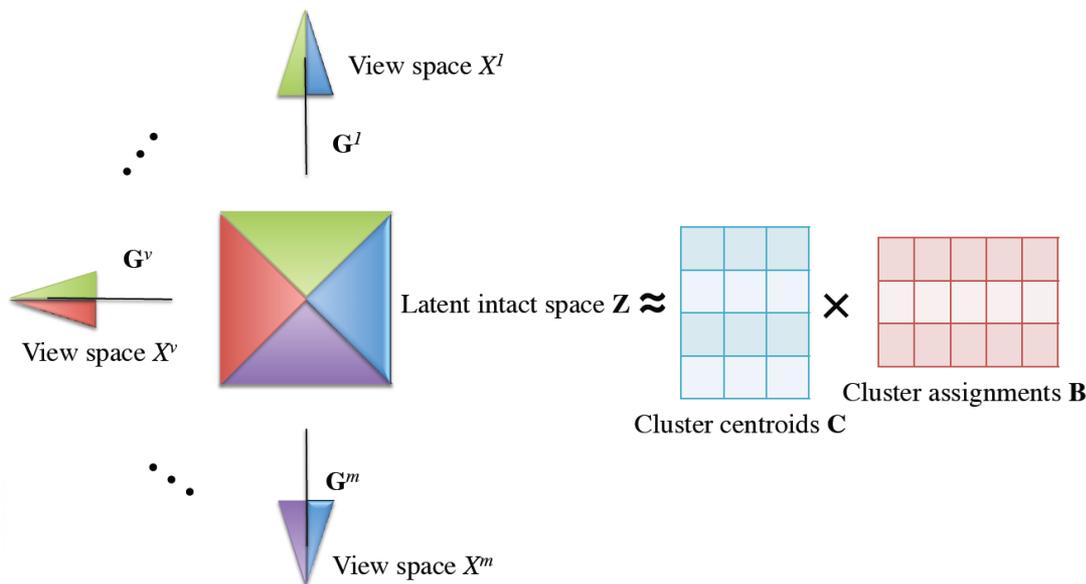
$$\sum_{v=1}^V \omega_v = 1, \omega_v \geq 0,$$

$$\sum_{l=1}^{d^v} \tau_l^v = 1, \tau_l^v \geq 0, \quad \forall v.$$

Datasets		Average of single view k-means	Average of single view EM	Average of single view LLC-fs	MVKKM	Co-regspec	WMCFS
Mfeat	mfeat-fac&mfeat-fou	0.639 ± 0.011	0.655 ± 0.010	0.776 ± 0.009	0.825 ± 0.012	0.792 ± 0.013	0.835 ± 0.010
	mfeat-fac&mfeat-zer	0.610 ± 0.013	0.595 ± 0.009	0.708 ± 0.010	0.718 ± 0.015	0.652 ± 0.007	0.794 ± 0.009
	All 5 views	0.678 ± 0.012	0.636 ± 0.010	0.815 ± 0.006	0.646 ± 0.014	0.735 ± 0.010	0.836 ± 0.010
Reuters	EN&FR	0.698 ± 0.007	0.770 ± 0.008	0.682 ± 0.010	0.626 ± 0.009	0.915 ± 0.012	0.925 ± 0.008
	EN&GR	0.762 ± 0.009	0.770 ± 0.011	0.782 ± 0.013	0.638 ± 0.008	0.907 ± 0.010	0.926 ± 0.009
	All 5 views	0.748 ± 0.011	0.683 ± 0.010	0.825 ± 0.012	0.690 ± 0.007	0.925 ± 0.007	0.927 ± 0.007
Corel	Col-h&Col-m	0.423 ± 0.007	0.553 ± 0.010	0.480 ± 0.011	0.502 ± 0.010	0.621 ± 0.011	0.657 ± 0.007
	Coo-t&Col-hl	0.344 ± 0.013	0.495 ± 0.015	0.423 ± 0.012	0.513 ± 0.007	0.656 ± 0.010	0.690 ± 0.012
	All 4 views	0.384 ± 0.013	0.505 ± 0.010	0.459 ± 0.013	0.508 ± 0.014	0.698 ± 0.008	0.712 ± 0.008

Yu-Meng Xu, Chang-Dong Wang* and Jian-Huang Lai. Weighted Multi-view Clustering with Feature Selection. Pattern Recognition 2016.

多视图完整空间聚类 (MVIC)

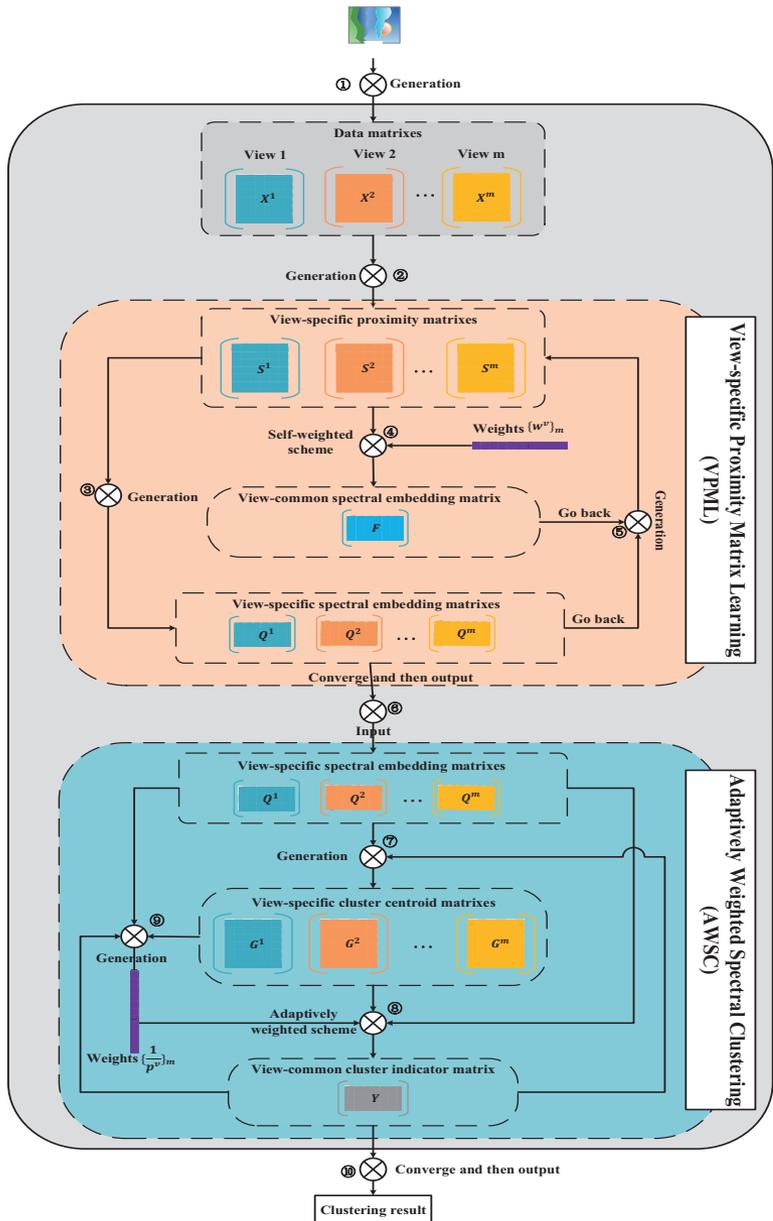


$$\min_{\mathbf{G}, \mathbf{Z}, \mathbf{C}, \mathbf{B}} \mathcal{J}(\mathbf{G}, \mathbf{Z}, \mathbf{C}, \mathbf{B}) = \frac{1}{mn} \sum_{v=1}^m \sum_{i=1}^n \|\mathbf{x}_i^v - \mathbf{G}^v \mathbf{z}_i\|^2 + \frac{1}{n} \|\mathbf{Z} - \mathbf{CB}\|_F^2 \quad (1)$$

$$+ \lambda_1 \frac{1}{m} \sum_{v=1}^m \|\mathbf{G}^v\|_F^2 + \lambda_2 \frac{1}{n} \sum_{i=1}^n \|\mathbf{z}_i\|^2$$

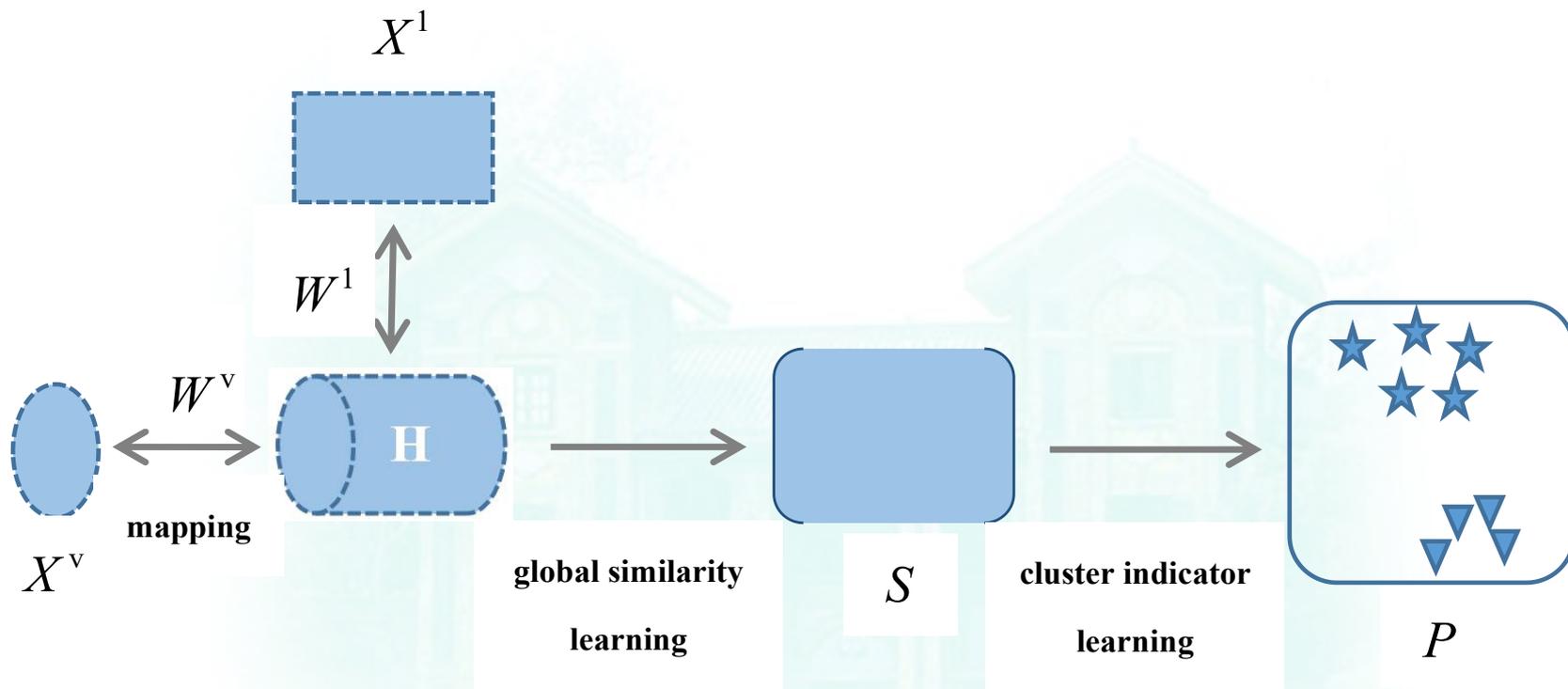
$$\text{s.t. } \mathbf{B}_{ij} \in \{0, 1\}, \sum_i \mathbf{B}_{ij} = 1, \forall j = 1, 2, \dots, n.$$

自适应加权多视图相似性学习聚类 (AWMVPL)

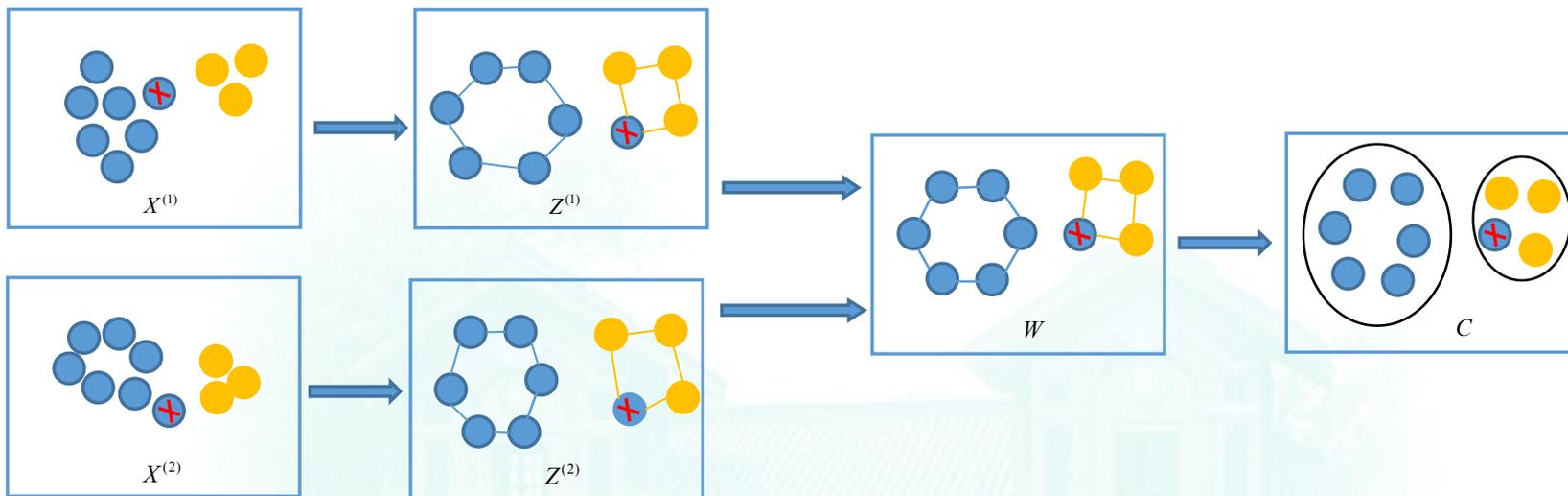


Bao-Yu Liu, Ling Huang, Chang-Dong Wang*, Suohai Fan and Philip S. Yu. Adaptively Weighted Multiview Proximity Learning for Clustering. IEEE TCYB, 2019

基于潜在嵌入空间的多视图聚类 (MCLES)



群效应保持的多视图子空间聚类 (MvSCGE)



0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.6	0.6	0.0	0.0	0.0	0.0	0.0
0.0	0.6	0.6	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.5	0.5	0.5	0.0	0.0
0.0	0.0	0.0	0.5	0.5	0.5	0.0	0.0
0.0	0.0	0.0	0.5	0.5	0.5	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4

$Z^{(1)}$

0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.3	0.3	0.0	0.3	0.3	0.0
0.0	0.0	0.3	0.3	0.0	0.3	0.3	0.0
0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0
0.0	0.0	0.3	0.3	0.0	0.3	0.3	0.0
0.0	0.0	0.3	0.3	0.0	0.3	0.3	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6

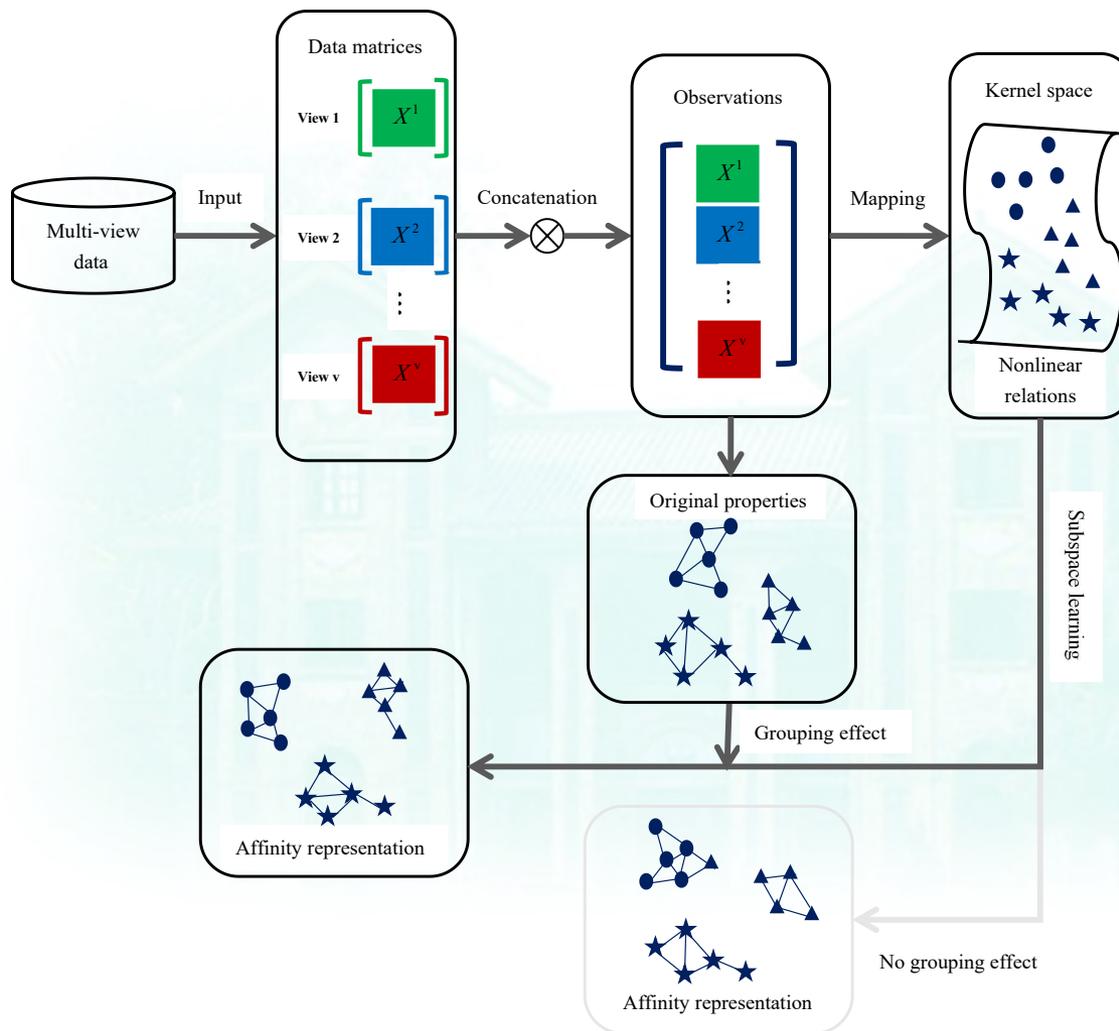
$Z^{(2)}$

0.3	0.3	0.3	0.0	0.0	0.0	0.0	0.0
0.3	0.3	0.3	0.0	0.0	0.0	0.0	0.0
0.3	0.3	0.3	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.0
0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2
0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2

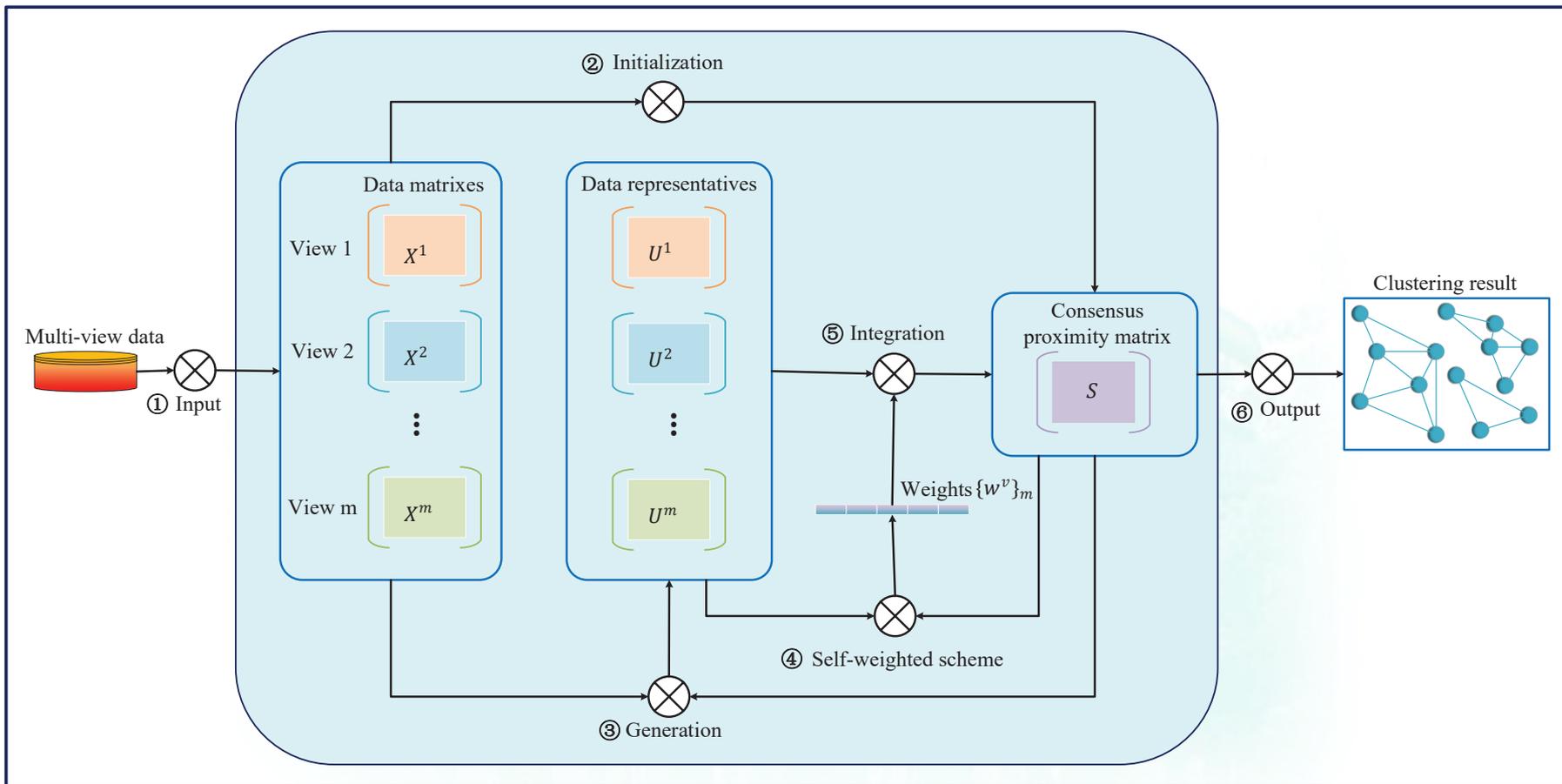
$Z^{(3)}$



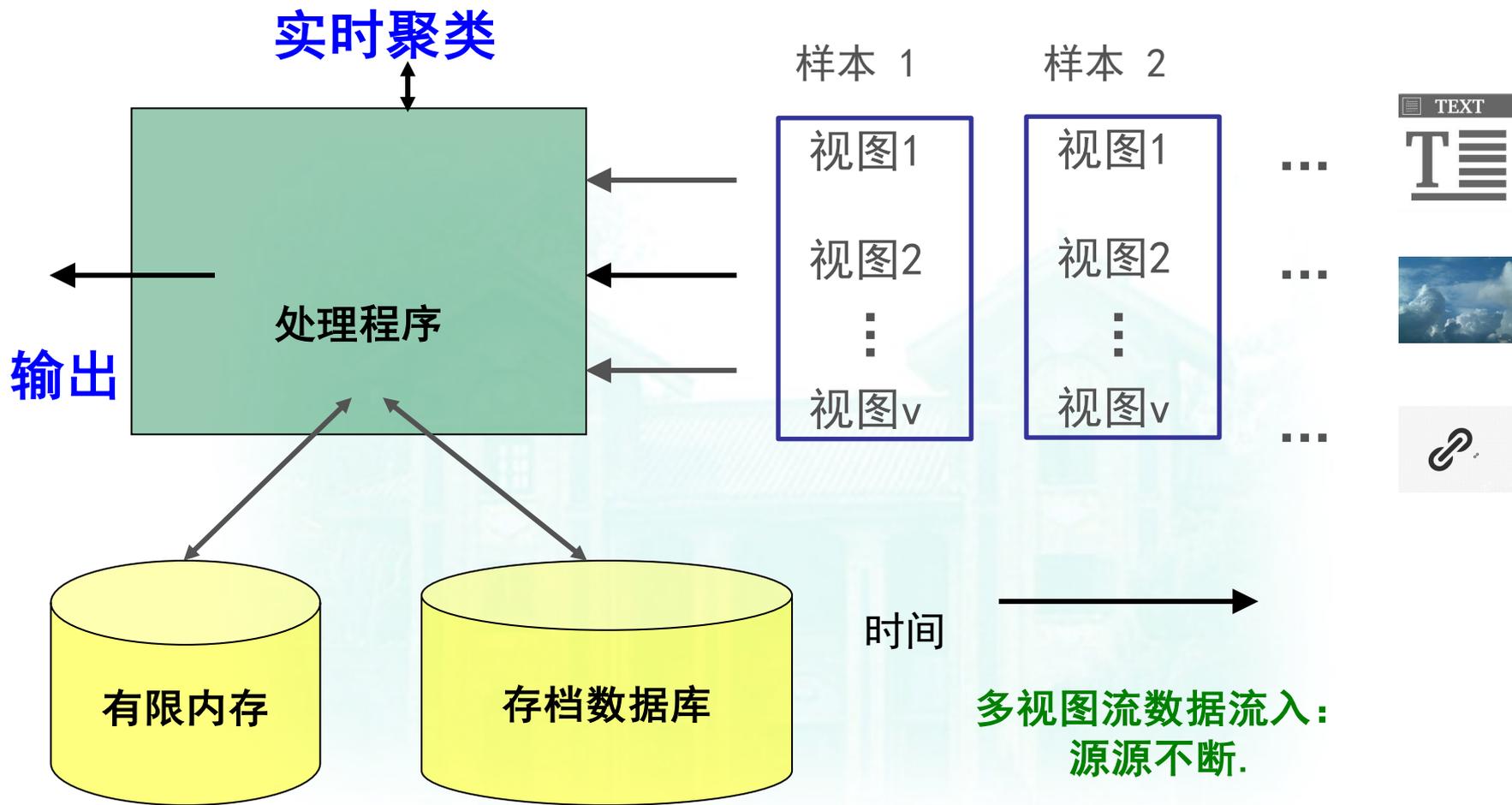
平滑正则化核学习多视图子空间聚类 (SMSOK)



多视图一致相似性学习聚类 (MCPL)



多视图流数据聚类问题定义



多视图流数据流入：
源源不断。

多视图流数据：

例1：多视图网页流数据

例2：无人驾驶汽车多视图实时流数据

$$\underbrace{x_{1,1}^p, \dots, x_{m,1}^p, \dots, x_{1,t-1}^p, \dots, x_{m,t-1}^p}_{\mathcal{X}_1^p}, \underbrace{x_{1,t}^p, \dots, x_{m,t}^p, \dots}_{\mathcal{X}_{t-1}^p}, \underbrace{\dots}_{\mathcal{X}_t^p}$$

数据块流入方式



多视图流数据聚类：挑战

多视图聚类

定义：融合多个视图的数据信息进行聚类

挑战：如何融合多个视图？

流数据聚类

定义：从历史流数据中提炼摘要信息进行聚类

挑战：如何在有限计算资源中快速提炼摘要信息？

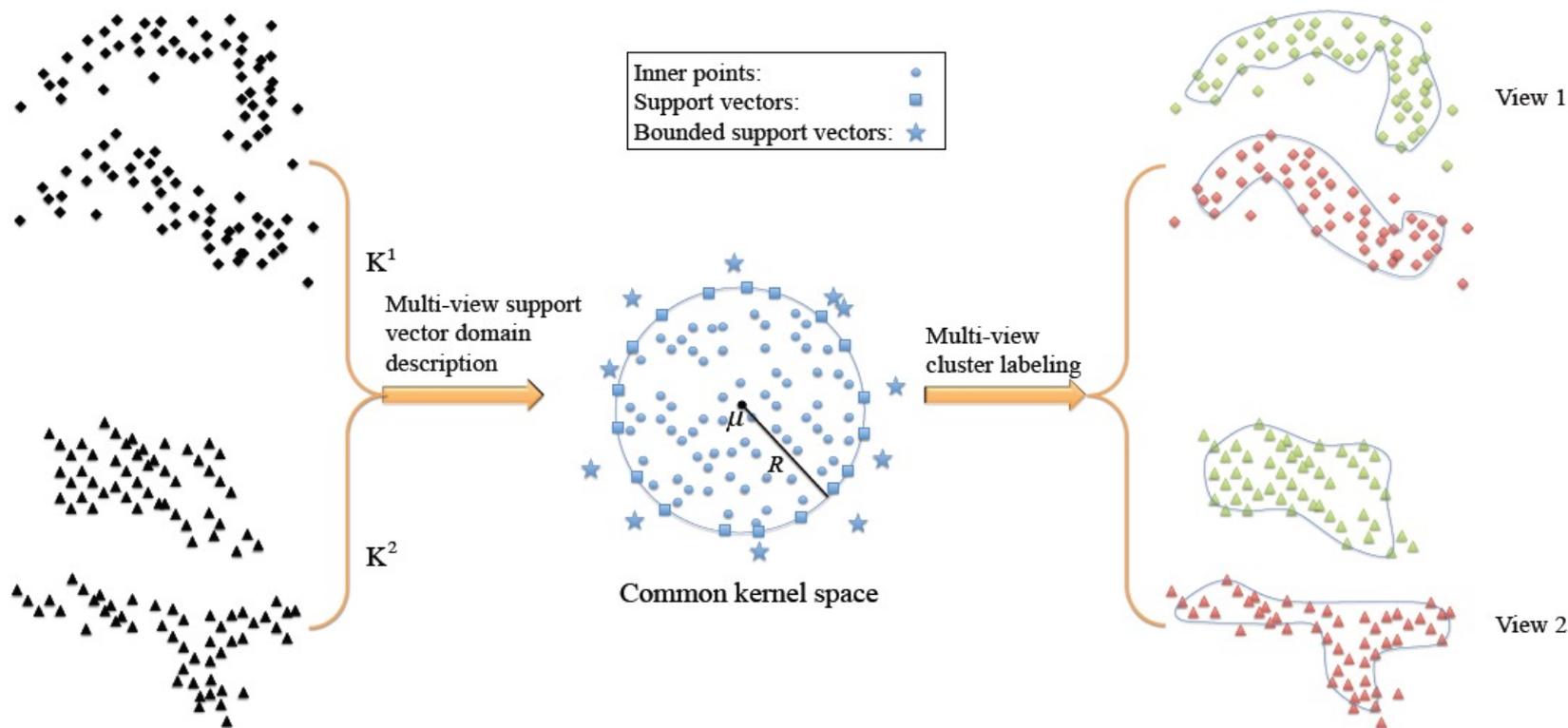
多视图流数据聚类

定义：融合多个视图的流数据信息进行聚类

挑战：1+1>>2!

1. 如何以流的方式快速融合多个视图的信息并同时提炼得到历史数据的摘要信息？
2. 如何刻画每个视图的由于概念漂移导致的类演变？
3. 如何挖掘任意形状类结构？
4. 如何解决有限的计算资源？

多视图流数据聚类算法 (MVStream)



算法的两个关键:

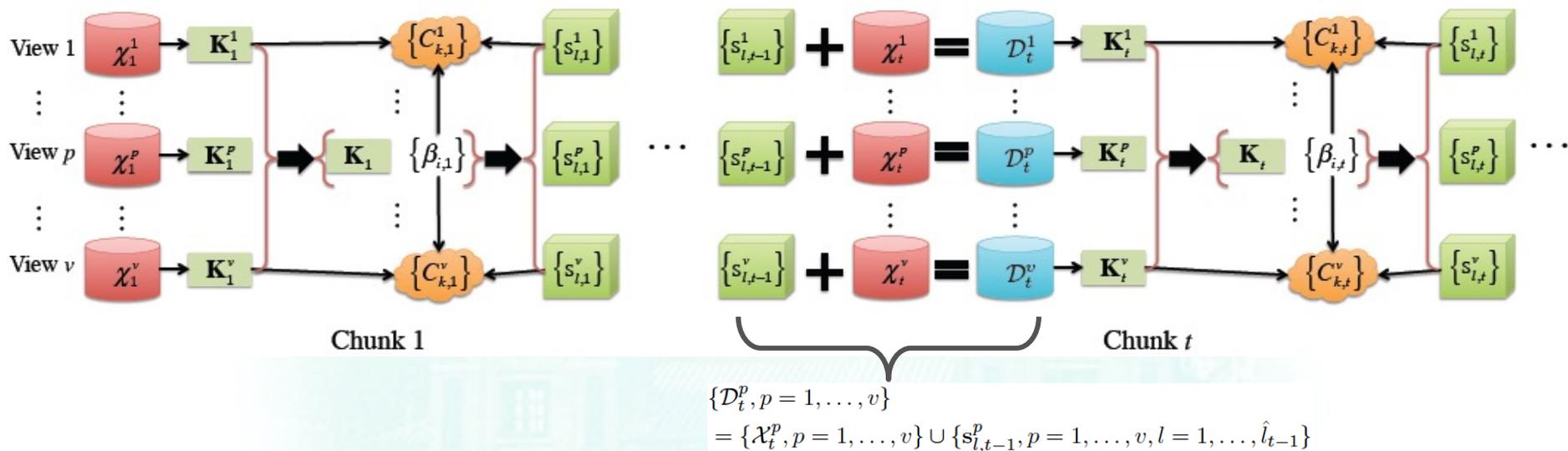
- **多视图支持向量域描述 (MVSVD)**

1. 解决挑战1: 以流的方式快速融合多个视图的信息并同时提炼得到历史数据的摘要信息 (支持向量)
2. 解决挑战4: 轻量级的历史数据的摘要信息用于解决计算资源瓶颈

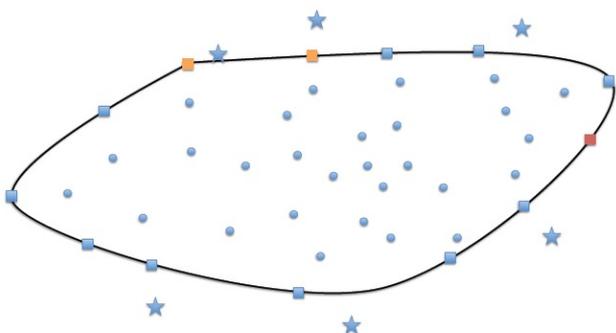
- **多视图聚类类标学习 (MVCL)**

1. 解决挑战2: 刻画每个视图的由于概念漂移导致的类演变
2. 解决挑战3: 挖掘任意形状的一类结构

多视图流数据聚类算法 (MVStream)

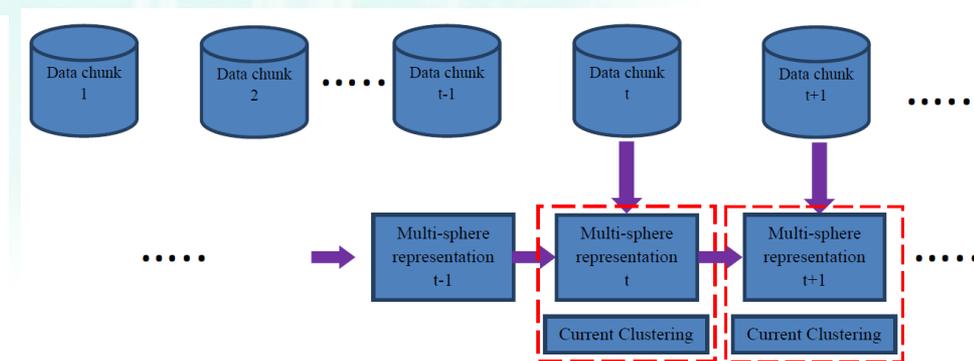


聚类类标跟踪:
支持向量的“多数表决”法



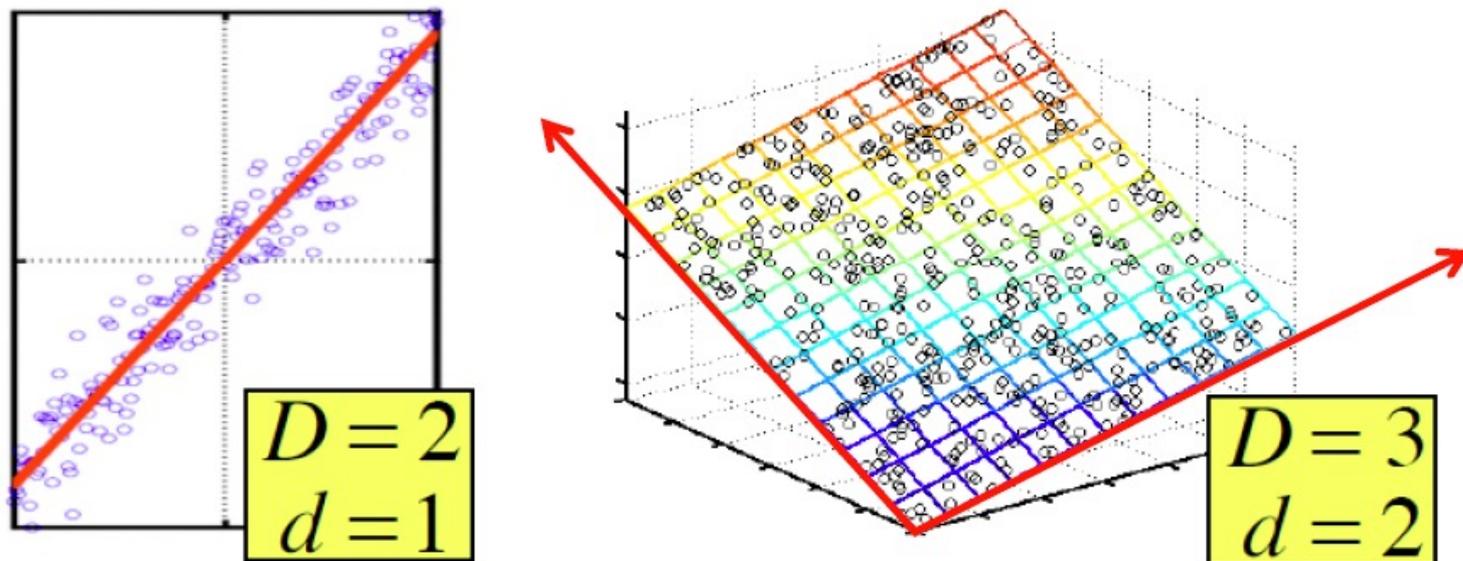
- 1
- 2
- 10 ✓

VS.



Wang, et. al. SVStream: A Support Vector Based Algorithm for Clustering Data Streams. TKDE 2013

成分学习 (SVD)



- ❑ **假设：** 数据分布或者几乎分布在一个低的 d 维子空间
- ❑ **该子空间的数轴有效地刻画了数据的特性分布**

降维

□ 压缩数据降维：

- 静态数据： 10^6 行； 10^3 列；
- 允许在有微小误差的情况下做数据压缩

customer	day	We 7/10/96	Th 7/11/96	Fr 7/12/96	Sa 7/13/96	Su 7/14/96
ABC Inc.		1	1	1	0	0
DEF Ltd.		2	2	2	0	0
GHI Inc.		1	1	1	0	0
KLM Co.		5	5	5	0	0
Smith		0	0	0	2	2
Johnson		0	0	0	3	3
Thompson		0	0	0	1	1

上述的矩阵确实是“2-维”数据：即所有的行都可以通过从行向量 $[1\ 1\ 1\ 0\ 0]$ 或 $[0\ 0\ 0\ 1\ 1]$ 做伸缩得到。



矩阵的秩

□ Q: 矩阵A的秩是什么?

□ A: 矩阵A的线性独立行的个数

□ 例如:

○ 矩阵 $A = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ 的秩是 $r=2$

❖ **为何?** 前两行是线性独立的, 从而该矩阵的秩至少为2, 但是三行是线性相关的 (比如: 第一行等于第二第三行之和), 则秩必然小于3.

□ 为何需要得到低秩矩阵?

○ 我们可以将矩阵A写成两个“基”向量: $\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \end{bmatrix}$

○ 从而得到新的坐标: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}$

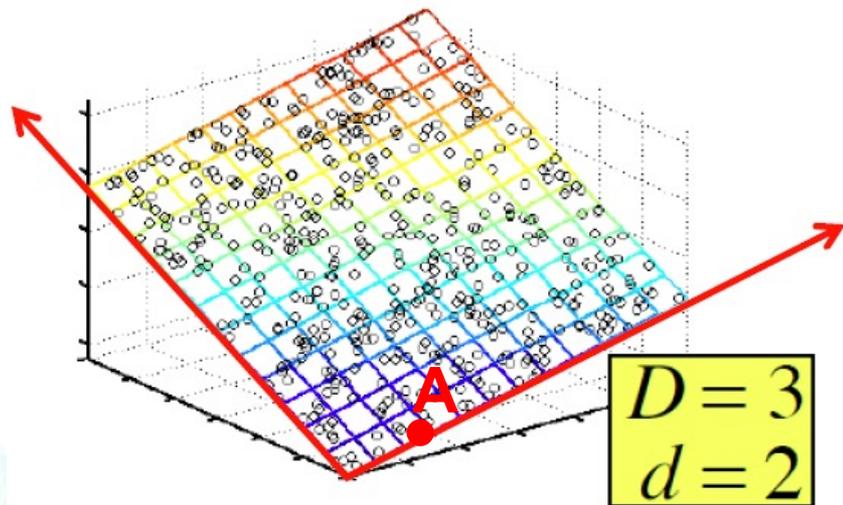
秩即维数

3维空间点云:

- 把样本点看成矩阵:

每个样本一行:

$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \begin{matrix} A \\ B \\ C \end{matrix}$$



我们可以更有效地表达样本坐标!

- 旧的基向量: $[1 \ 0 \ 0]$ $[0 \ 1 \ 0]$ $[0 \ 0 \ 1]$

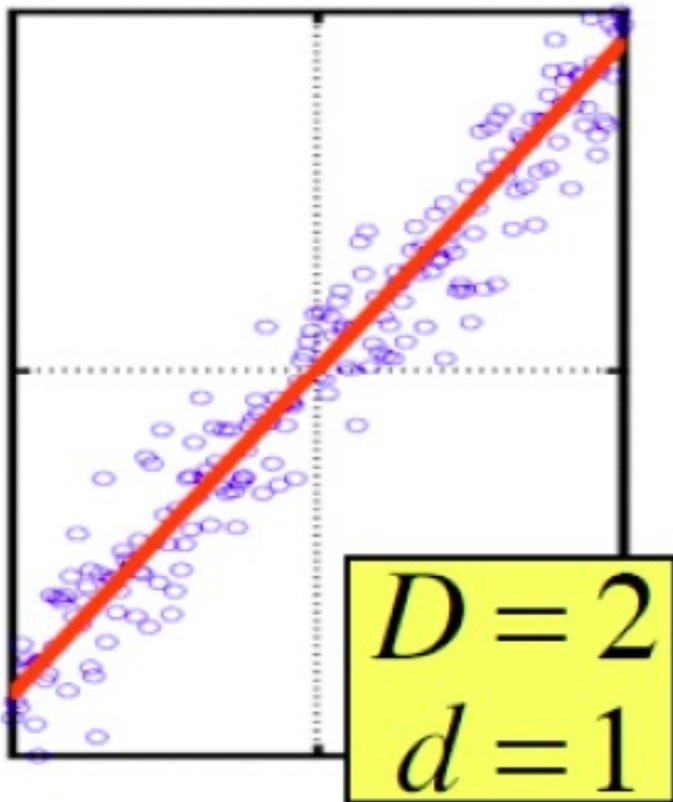
- 新的基向量: $[1 \ 2 \ 1]$ $[-2 \ -3 \ 1]$

- 则矩阵中的样本具有新的坐标: A: $[1 \ 0]$,
B: $[0 \ 1]$, C: $[1 \ -1]$

❖ 注意: 我们将坐标的维数降低了:

降维

- 降维的目标是找到数据的坐标!



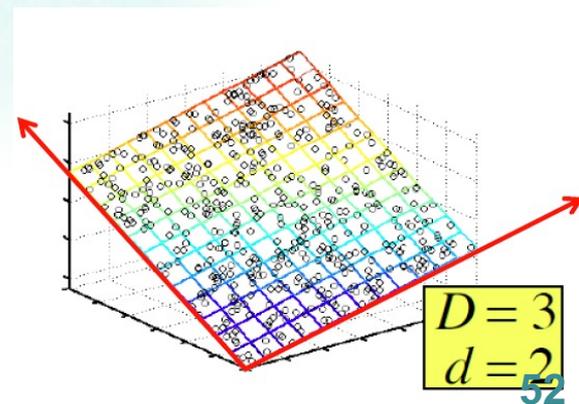
如图：我们可以用一个坐标（对应于样本在红线的位置）而不是两个坐标来表示任意一个样本。

由于样本点不是严格位于直线上，则这个过程可能会导致一些误差。

为何要降维（找主成分）？

为何要降维（找主成分）？

- 挖掘潜在的相关性/主题
 - 频繁共同出现的单词
- 去除冗余以及噪声特征
 - 并非所有单词都是有用的
- 可视化目的
- 更容易存储和处理



SVD: 定义

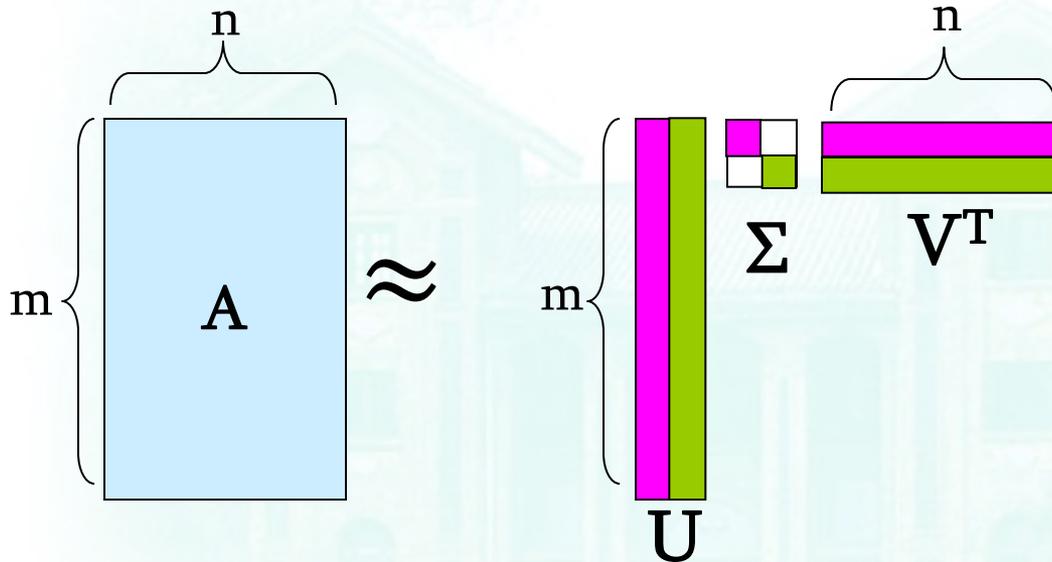
$$\mathbf{A}_{[m \times n]} = \mathbf{U}_{[m \times r]} \mathbf{\Sigma}_{[r \times r]} (\mathbf{V}_{[n \times r]})^T$$

- **A: 输入的数据矩阵**
 - $m \times n$ 矩阵 (e. g., m 个文档, n 个单词)
- **U: r 个左奇异值向量组成的矩阵**
 - $m \times r$ 矩阵 (m 个文档, r 个概念)
- **Σ : 奇异值**
 - $r \times r$ 对角线矩阵 (每个概念的强度)
(r : 矩阵A的秩)
- **V: r 个右奇异值向量组成的矩阵**
 - $n \times r$ 矩阵 (n 个单词, r 个概念)

SVD



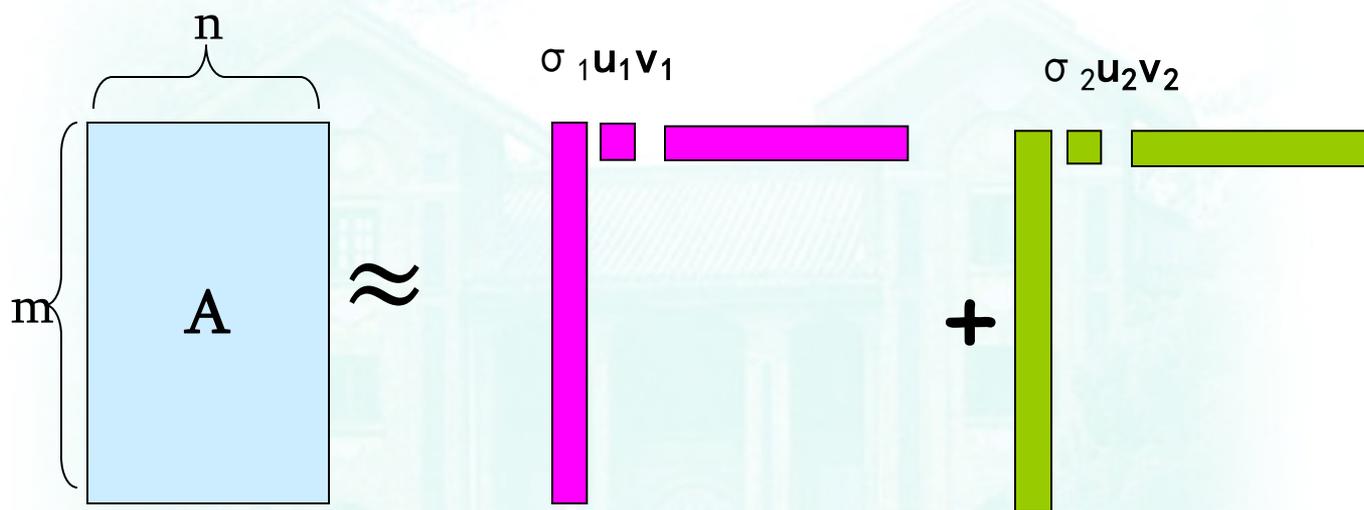
$$\mathbf{A} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i^T$$



SVD



$$\mathbf{A} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i^T$$



$\sigma_i \dots$ 标量
 $\mathbf{u}_i \dots$ 向量
 $\mathbf{v}_i \dots$ 向量



SVD: 特性

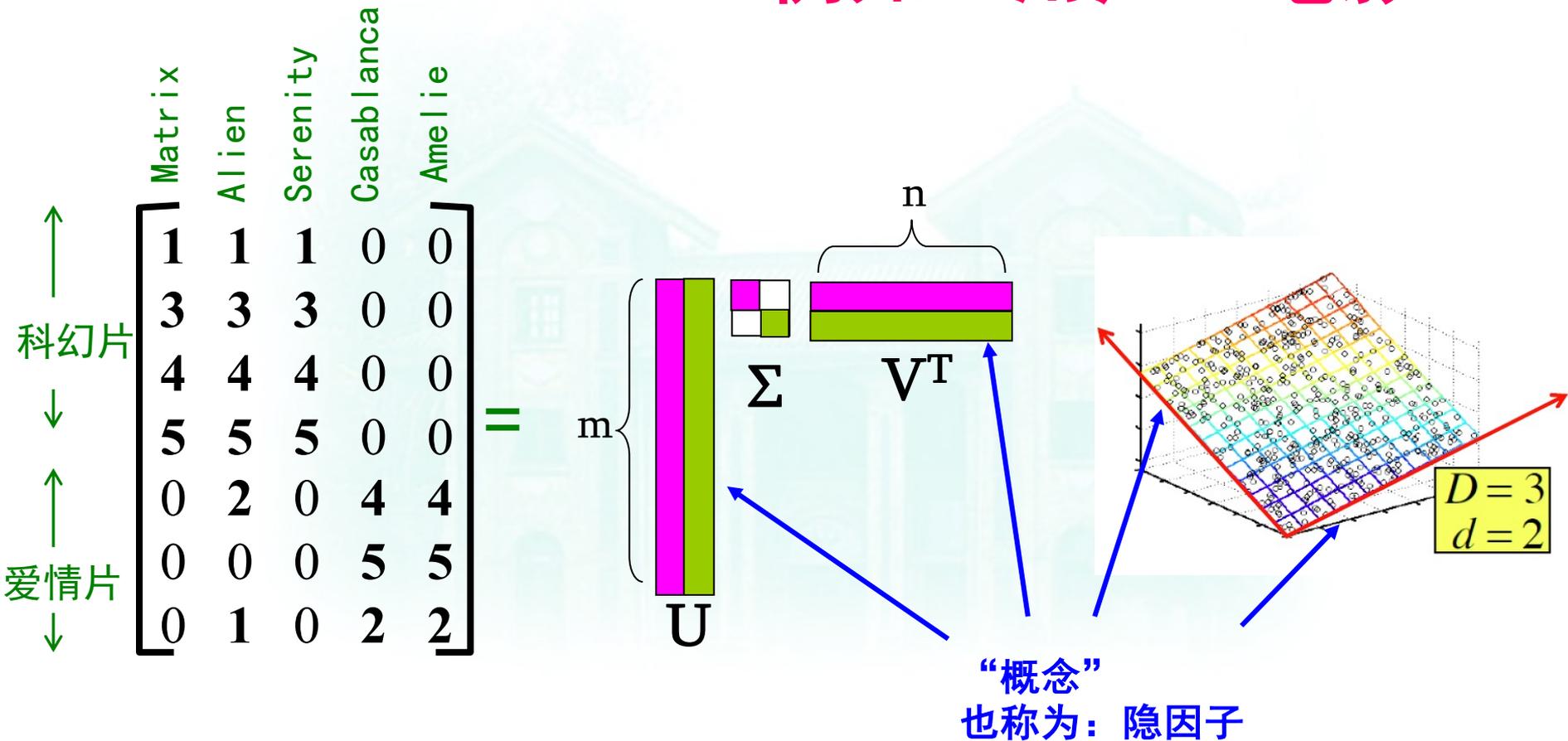
对于一个实数域矩阵 A , 总是能将 A 分解为

$A = U \Sigma V^T$, 其中

- U, Σ, V : 唯一的
- U, V : 列正交
 - $U^T U = I; V^T V = I$ (I : 单位矩阵)
 - (列向量是正交单位向量)
- Σ : 对角线矩阵
 - 对角线元素(奇异值)是正的, 并且以降序排序, 即($\sigma_1 \geq \sigma_2 \geq \dots \geq 0$)

SVD例子：用户~电影

$A = U \Sigma V^T$ - 例如：用户~电影



SVD例子：用户~电影

$A = U \Sigma V^T$ - 例如：用户~电影

	Matrix	Alien	Serenity	Casablanca	Amelie						
↑ 科幻片	1	1	1	0	0	=	0.13	0.02	-0.01		
↓	3	3	3	0	0		0.41	0.07	-0.03		
↑	4	4	4	0	0		0.55	0.09	-0.04		
↓	5	5	5	0	0		0.68	0.11	-0.05		
↑	0	2	0	4	4		0.15	-0.59	0.65		
↓	0	0	0	5	5		0.07	-0.73	-0.67		
↑	0	1	0	2	2	0.07	-0.29	0.32			
↓							12.4	0	0		
						x	0	9.5	0		
						x	0	0	1.3		
							0.56	0.59	0.56	0.09	0.09
							0.12	-0.02	0.12	-0.69	-0.69
							0.40	-0.80	0.40	0.09	0.09

科幻片概念

↓

科幻片概念的强度

↓



SVD例子：用户~电影

$A = U \Sigma V^T$ - 例如：用户~电影

V 是 “电影-概念” 相似性矩阵

Matrix Alien Serenity Casablanca Amelie

↑ 科幻片
↓
↑ 爱情片

1	1	1	0	0
3	3	3	0	0
4	4	4	0	0
5	5	5	0	0
0	2	0	4	4
0	0	0	5	5
0	1	0	2	2

科幻片概念

0.13	0.02	-0.01
0.41	0.07	-0.03
0.55	0.09	-0.04
0.68	0.11	-0.05
0.15	-0.59	0.65
0.07	-0.73	-0.67
0.07	-0.29	0.32

\times

12.4	0	0
0	9.5	0
0	0	1.3

\times

0.56	0.59	0.56	0.09	0.09
0.12	-0.02	0.12	-0.69	-0.69
0.40	-0.80	0.40	0.09	0.09

科幻片概念



SVD例子：用户~电影

更多的细节：

- Q: SVD降维是如何实现?
- A: 将最小奇异值所对应的项为0

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} \approx \begin{bmatrix} 0.13 & 0.02 & -0.01 \\ 0.41 & 0.07 & -0.03 \\ 0.55 & 0.09 & -0.04 \\ 0.68 & 0.11 & -0.05 \\ 0.15 & -0.59 & 0.65 \\ 0.07 & -0.73 & -0.67 \\ 0.07 & -0.29 & 0.32 \end{bmatrix} \times \begin{bmatrix} 12.4 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.3 \end{bmatrix} \times \begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & -0.69 & -0.69 \\ 0.40 & -0.80 & 0.40 & 0.09 & 0.09 \end{bmatrix}$$



SVD例子：用户~电影

更多的细节：

- Q: SVD降维是如何实现?
- A: 将最小奇异值所对应的项为0

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} \approx \begin{bmatrix} 0.13 & 0.02 \\ 0.41 & 0.07 \\ 0.55 & 0.09 \\ 0.68 & 0.11 \\ 0.15 & -0.59 \\ 0.07 & -0.73 \\ 0.07 & -0.29 \end{bmatrix} \times \begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix} \times \begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & -0.69 & -0.69 \end{bmatrix}$$

SVD例子：用户~电影

更多的细节：

□ Q: SVD降维是如何实现？

□ A: 将最小奇异值所对应的项为0

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} \approx \begin{bmatrix} 0.92 & 0.95 & 0.92 & 0.01 & 0.01 \\ 2.91 & 3.01 & 2.91 & -0.01 & -0.01 \\ 3.90 & 4.04 & 3.90 & 0.01 & 0.01 \\ 4.82 & 5.00 & 4.82 & 0.03 & 0.03 \\ 0.70 & 0.53 & 0.70 & 4.11 & 4.11 \\ -0.69 & 1.34 & -0.69 & 4.78 & 4.78 \\ 0.32 & 0.23 & 0.32 & 2.01 & 2.01 \end{bmatrix}$$

Frobenius norm:

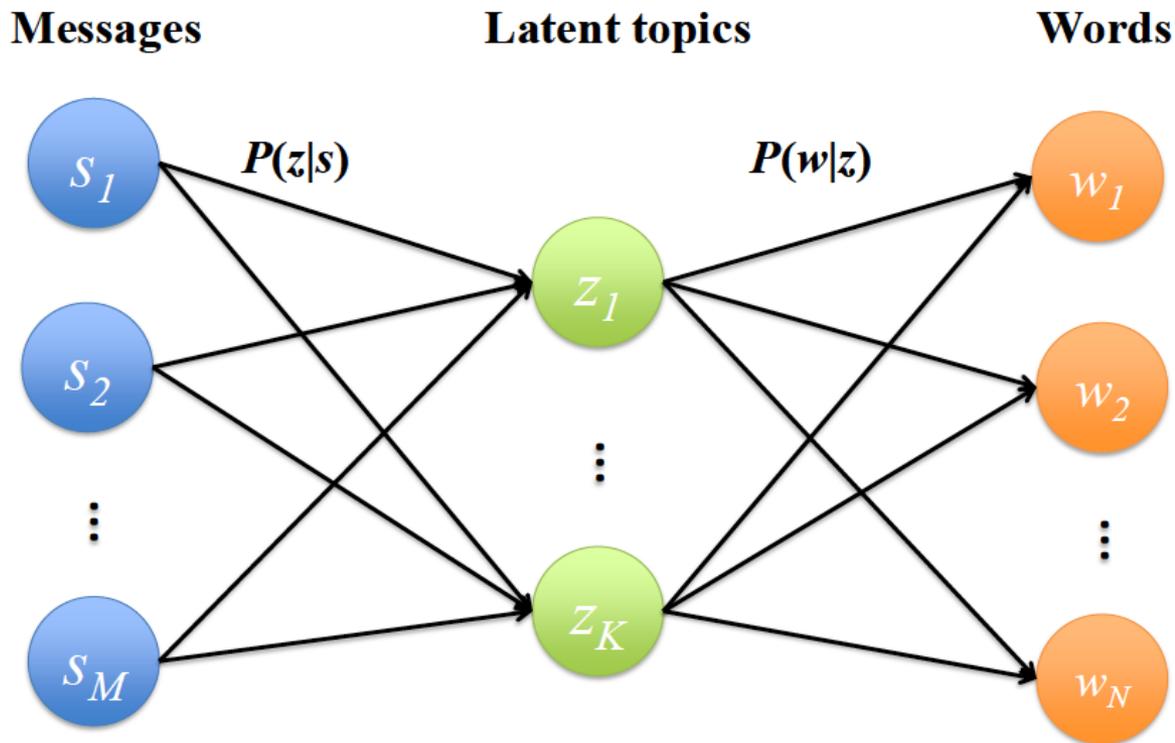
$$\|M\|_F = \sqrt{\sum_{ij} M_{ij}^2}$$

$$\|A-B\|_F = \sqrt{\sum_{ij} (A_{ij}-B_{ij})^2}$$

is "small"

主题模型

PLSA主题模型



$$\ell = \sum_{s \in S} \sum_{w \in W} n(s, w) \log P(s, w)$$

其中

$$\begin{aligned} P(s, w) &= P(s) \sum_{z \in Z} P(w|z)P(z|s) \\ &= \sum_{z \in Z} P(z)P(s|z)P(w|z) \end{aligned}$$