

Case VI: How to Measure the Difference

主讲教师：詹德川

The bottom of the slide features three horizontal bars of equal width. The top bar is light pink, the middle bar is a slightly darker shade of pink, and the bottom bar is light green.

k 近邻学习器

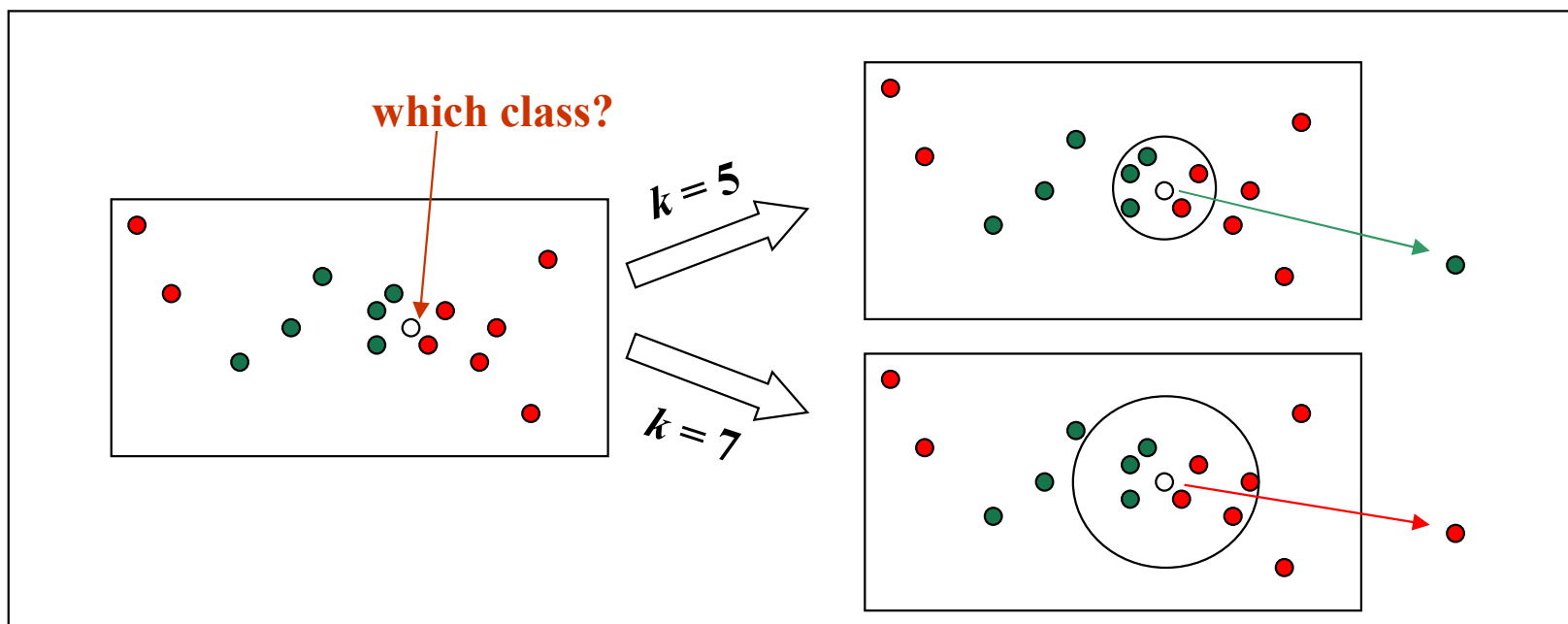
k 近邻 (k -Nearest Neighbor, k NN)

懒惰学习 (lazy learning) 的代表

基本思路：

近朱者赤，近墨者黑

(投票法；平均法)



关键： k 值选取； 距离计算

Information Retrieval

Baidu 百度

information retrieval

百度一下

网页 图片 资讯 贴吧 文库 视频 知道 采购 地图 更多

百度为您找到相关结果约39,500,000个

搜索工具

information retrieval - 百度翻译

information retrieval

英[ˌnfəˈmeɪʃən rɪˈtriːvl] 美[ˌnfərˈmeɪʃən rɪˈtriːvl]

[词典] 信息检索; 情报检索。

[例句] But now, in the day of the "information retrieval system," such a reverence is not being placed on the reading, and then saving, of books. 但是现在, 在这个"信息检索系统"的时代, 读书和藏书已不能获得这种尊重。

进行更多翻译

fanyi.baidu.com

information retrieval - Bing 词典 必应

4. 信息提取 5. 信息恢复 您要找的是不是 information retrieval information retrieve information server information service information system 下载手机版必应词典 iOS...

Bing 百度快照

Information Retrieval - 百度百科

《Information Retrieval》是The MIT Press出版的图书, 作者是Stefan Böttcher, Charles L. A. Clarke, Gordon V. Cormack

baike.baidu.com

信息检索(Information Retrieval)相关概念_土豆同学的...

2021年3月4日 信息检索(Information Retrieval,简称IR)从大规模非结构化数据(通常是文本)的集合(通常保存在计算机上)中找出满足用户信息需求的资料(通常是文档)的过程。非结构化数据(unstructure...

CSDN技术社区 百度快照

information retrieval是什么意思_information retrieval...

英[ˌnfəˈmeɪʃən rɪˈtriːvl] 美[ˌnfərˈmeɪʃən rɪˈtriːvl] 信息检索 information retrieval的用法和样例: 例句 Dr. Li is a specialist in information r...

dict.cn/information retrieval 百度快照

其他人还在搜

Information AND retrieval privacy policy 什么意思 generalizations manufactures fundamental infotainment administration 是什么意思啊 information cocoons

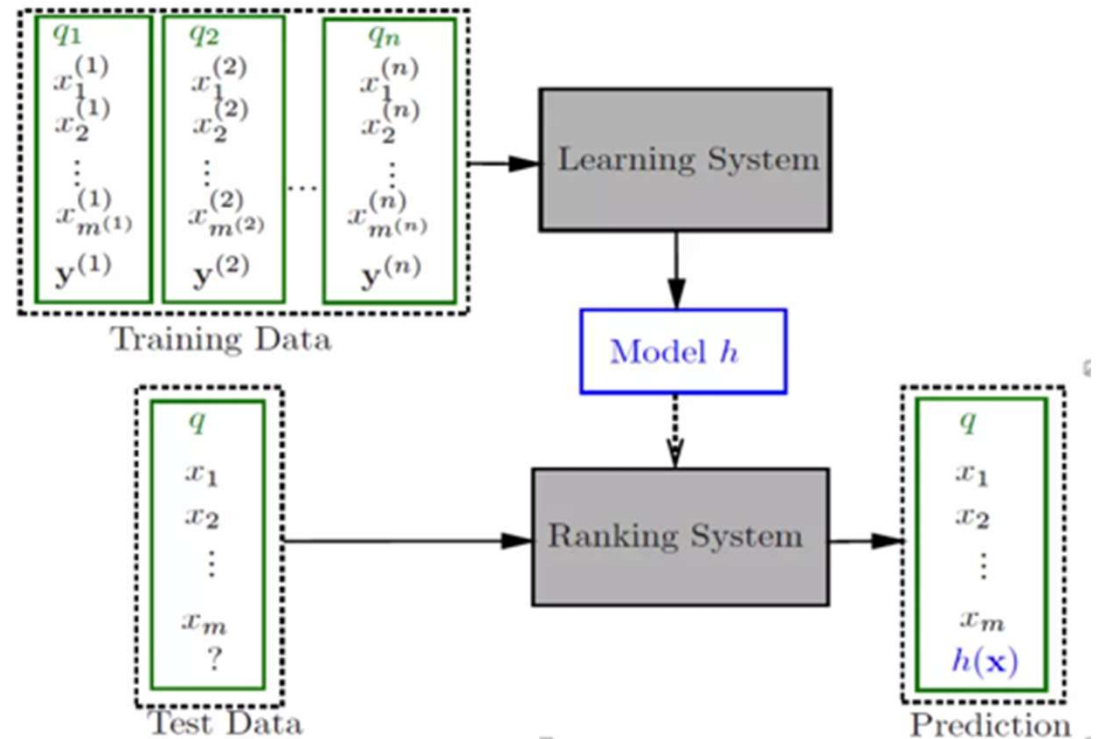
information retrieval是什么意思_翻译information retrie...

沪江网校精选information retrieval是什么意思、英语单词推荐 信息恢复, 信息检索, 情报检索 相似短语 information retrieval 信息恢复, 信息检索, 情报检索 information retrieval language 情报...

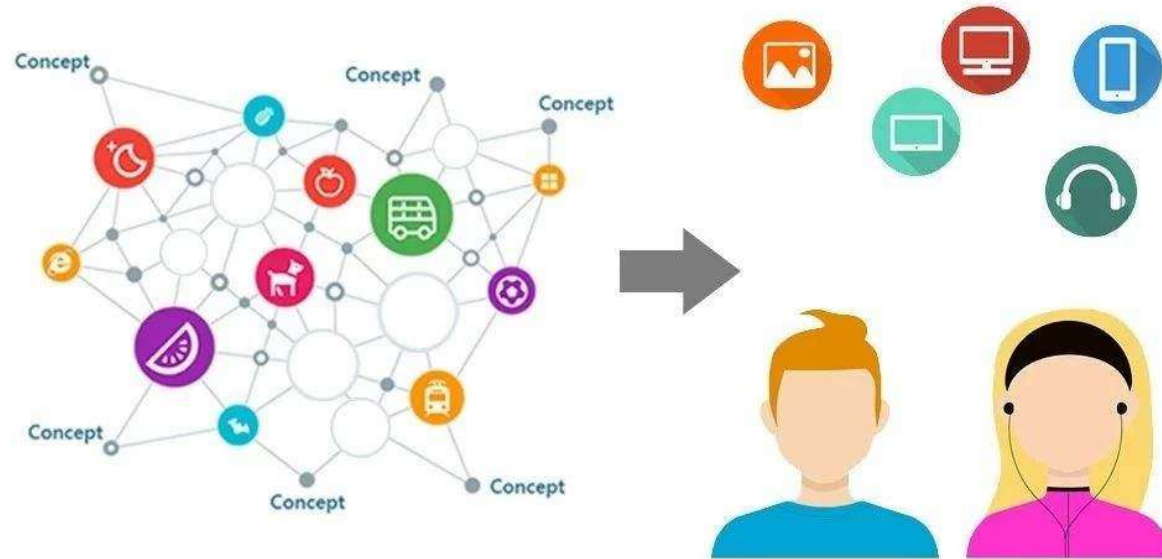
沪江网 百度快照

information retrieval - 相关论文(共760篇) - 百度学术

Ranking based Similarity Measure



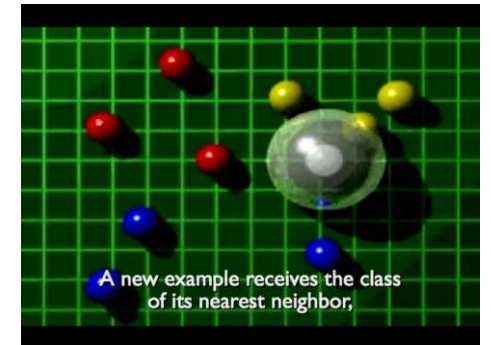
Recommendation Systems



Similarity between
Concepts
Users
Items
Hybrids

距离度量的重要性

- 人类很早就意识到距离的重要性，并且在文明发展之初就以统一**度量衡**作为进步的标识
- 在信息检索中如何比较提交的查询和检索结果之间是否相似？
 - 使用**距离**来表示样本之间的不相似度
- 使用最近邻分类器如何判别一个样本属于哪个类别？
 - 近**朱**者，**赤**；近**墨**者，**黑**
- Kernel Machine核的生成也和**距离度量**有着密切的关系



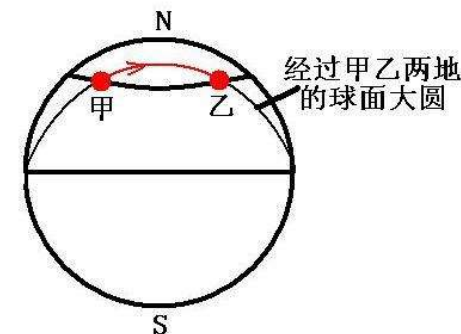
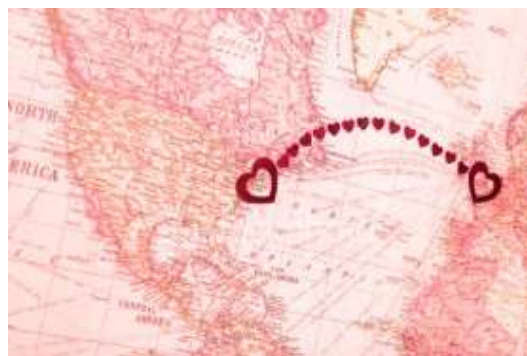
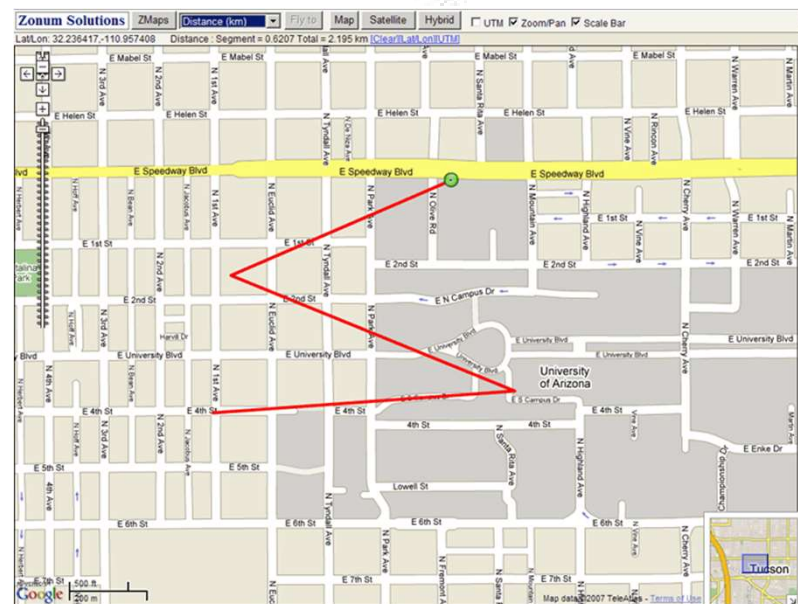
距离度量的种类

□ Euclidean Distance

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

□ Block Distance

□ Geodesic Distance



研究者对距离度量的研究

□ 为特定应用设计距离度量

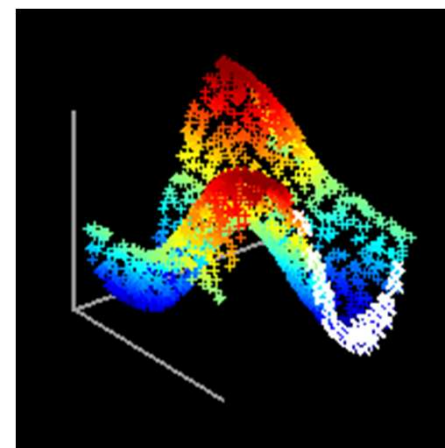
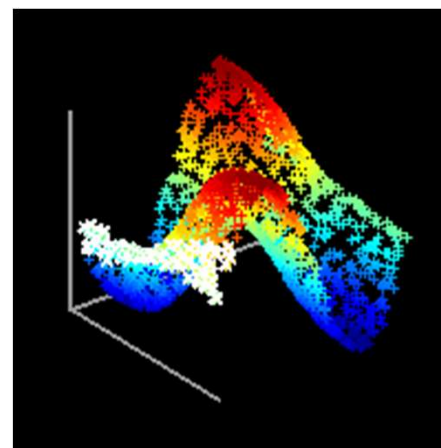
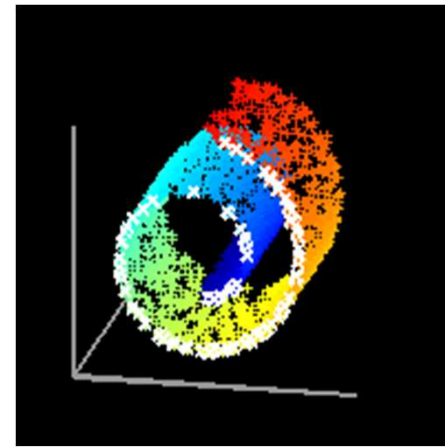
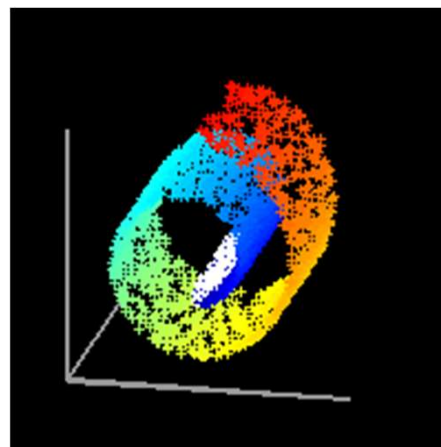
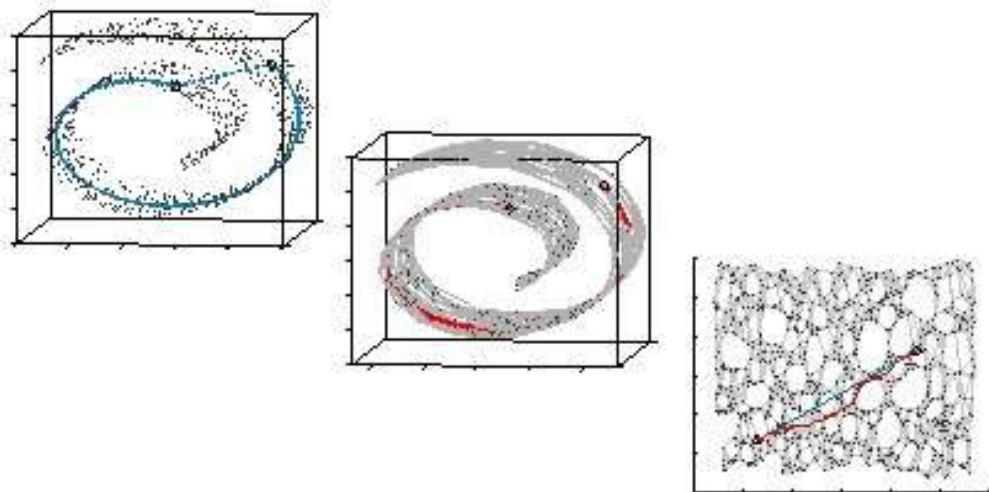
□ 距离度量学习



Isomap

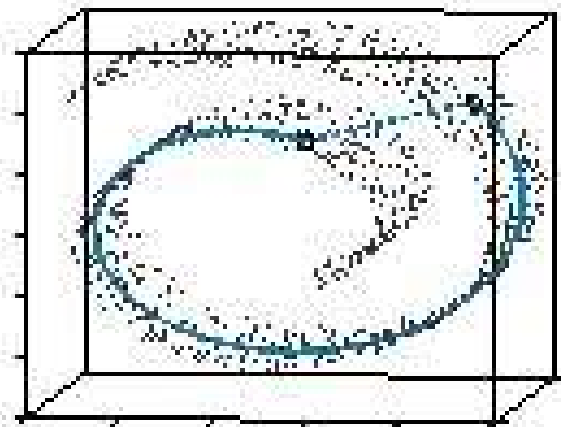
□ 背景

- 高维数据具有更加自然的低维结构
- 使用低维结构上面的直线距离更加能够反映样本之间不相似的程度

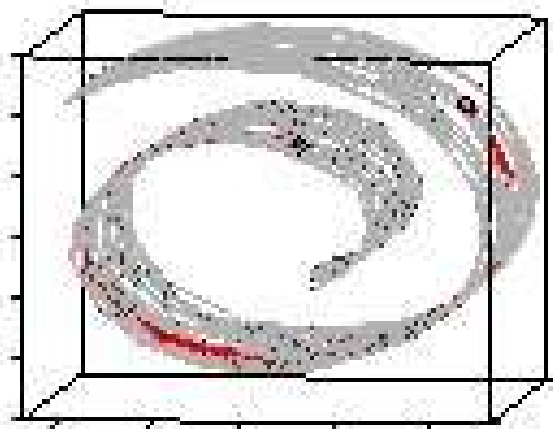


Isomap

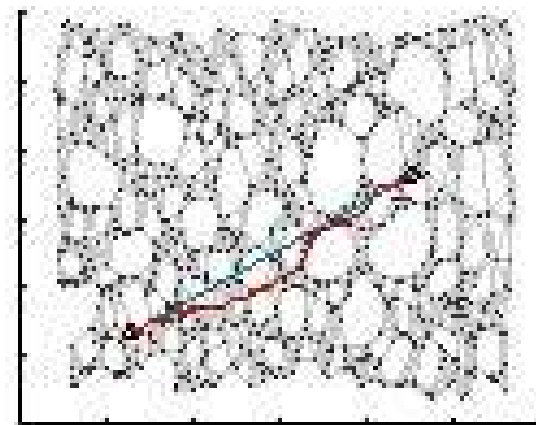
要求的是两点之间的测地线距离



无法获得测地线！怎么办！



近似获得测地线距离：



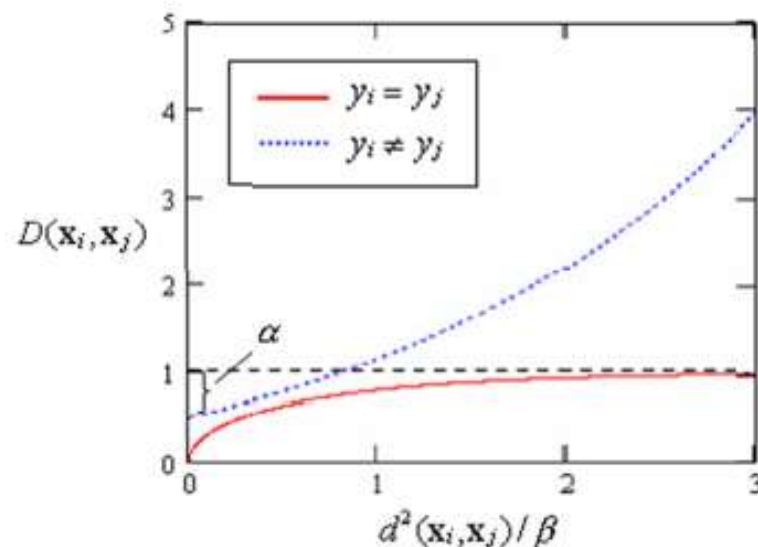
Isomap的缺陷及对其的改进

- Isomap能够度量样本之间在低维本真空间上的距离
- 却容易受到噪音的影响，并且不利于分类

□ 改进方法：

- 引入类别信息，对距离度量进行改进 Supervised-Isomap [Geng, Zhan and Zhou 05].

$$D(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} \sqrt{1 - e^{\frac{-d^2(\mathbf{x}_i, \mathbf{x}_j)}{\beta}}} & y_i = y_j, \\ \sqrt{e^{\frac{d^2(\mathbf{x}_i, \mathbf{x}_j)}{\beta}} - \alpha} & y_i \neq y_j \end{cases}$$



研究者对距离度量的研究

□ 为特定应用设计距离度量

□ 距离度量学习

- 顾名思义，即利用学习的方法获得更好的度量距离的方式
- 大多数研究者针对马氏距离的度量矩阵 **A** 进行学习



$$d_{ij} = \sqrt{\|\mathbf{x}_i - \mathbf{x}_j\|_A^2} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{A} (\mathbf{x}_i - \mathbf{x}_j)}$$

对马氏距离的进一步解释

□ 为什么要马氏距离 $d_{ij} = \sqrt{\|\mathbf{x}_i - \mathbf{x}_j\|_A^2} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^\top A (\mathbf{x}_i - \mathbf{x}_j)}$?

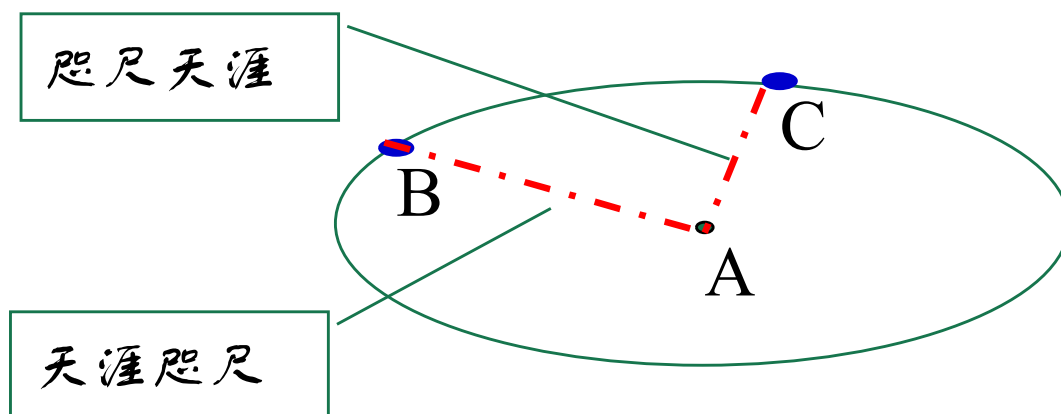
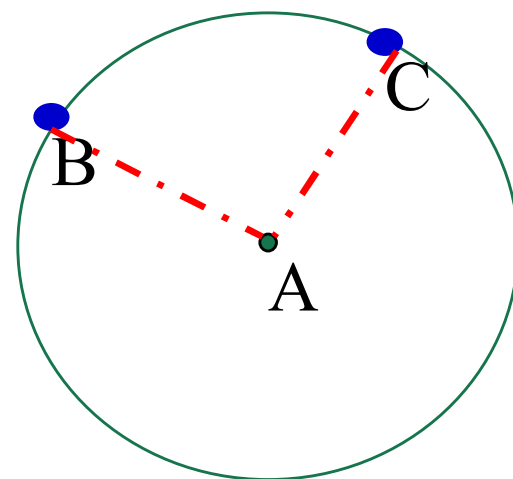
□ 我们回顾一下“什么是距离？”再思考一下“距离度量”
度量的是什么？

- It's a long distance to **walk**....
- 旅行的开销！

□ 欧氏距离的缺陷 —— 各向同性

□ 但是：

- 有缘千里来相会
(欧氏距离大但开销少)
- 无缘对面手难牵
(优势距离小但开销大)
- 马氏距离应运而生



利用边信息的距离度量学习方法

□ 基本思想:

- 一个好的距离度量方法，应该能够使得同类之间的距离小于1，异类之间的距离大于1

□ DML实现方法

$$\begin{array}{ll} \min & \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}} \|\mathbf{x}_i - \mathbf{x}_j\|_A^2 \\ \text{s.t.} & A \in S_+, \quad \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}} \|\mathbf{x}_i - \mathbf{x}_j\|_A^2 \geq 1. \end{array}$$

最小化同类之间的距离

保持异类样本之间的距离大于某个阈值

限制学习到的距离度量矩阵是半正定的；保证样本之间的距离为正(或者0)，也即： $d_{ij} \geq 0$

是否马氏距离就一定符合人们的认知？

- 在某些应用中，距离的定义偏重于一些特定的属性，并且对于不同的样本，这些属性是各不相同的
- 距离的定义应该是样本自适应的

- 例如：

- 当我们将描述天空的图片和其他图片进行比较的时候
 - 关注的是图片的颜色（蓝色）、纹理（有着特殊的光线）等
- 当我们将菲尔普斯II和其他游泳运动员比较的时候
 - 关注的是他脚的形状，游泳的速度



是否马氏距离就一定符合人们的认知？



图像1



图像2



图像3



查询1: 森林



查询2: 豹子

距离计算方法和
特定的样本个体相关

- 当用户提交“查询1”时，图像2与图像1应该比与图像3更接近，因为前两者都包含了描述“森林”这个查询概念的特征；而在用户提交“查询2”时，图像2与图像3应该比与图像1更接近，因为图像2和图像3都包含了描述“豹子”这个概念的图像特征。

样本自适应方面已有的工作

- 问题：不同样本有着各自不同的
视角、语义着重
- 或者说，将样本表示在高维空间中，样本的距离度量和其本身的局部特性相关
- 解决方案：对每个样本赋以不同的距离度量



- QSim [Zhou and Dai, ICDM'06] [Athitsos et al., TDS'07]
- Local distance functions [Frome et al., NIPS'06, ICCV'07]

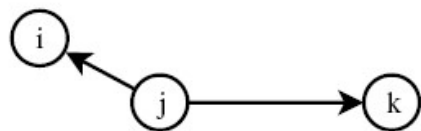
以往样本自适应方法的缺陷

□ Qsim:

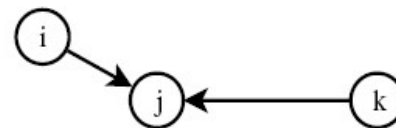
- 用于基于内容的图像检索，对象之间的距离会受到提交的检索的影响
- 主要问题：该方法完全基于启发式想法
 - 启发式想法 —— 拍脑袋想办法：
 - 人在解决问题时所采取的一种根据经验规则进行发现的方法
 - 利用过去的经验,选择已经行之有效的方法，而不是系统地、以确定的步骤去寻求答案

□ Local Distance Functions:

- [Frome et al. NIPS'06]



- [Frome et al. ICCV'07]

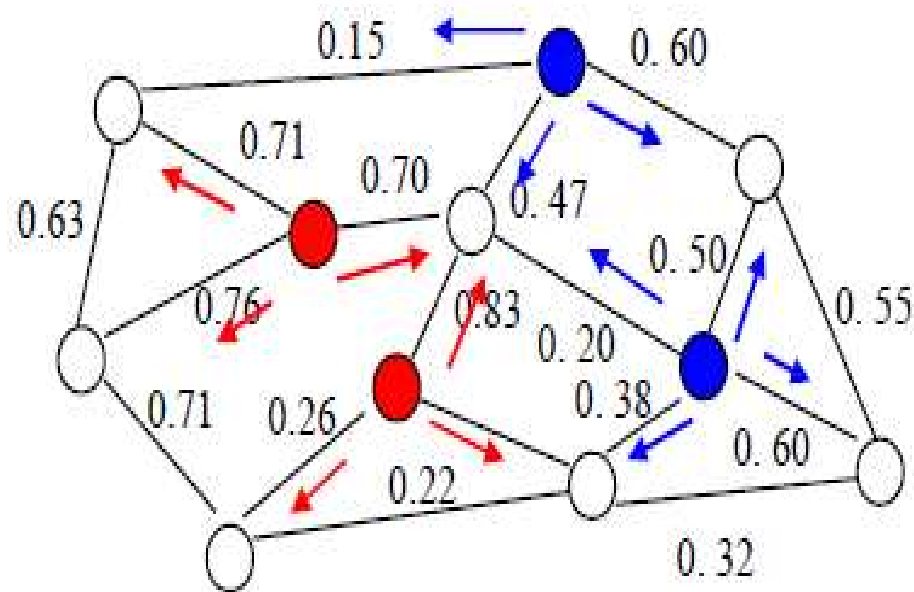


是否能够同时为标记样本和未标记样本学习
得到样本自适应距离度量呢？

答案是肯定的，我们可以尝试使用类似于标
记传播的思想来进行距离度量传播！

什么是标记传播？

- 一种基于图的半监督学习算法
 - 边的权重往往和点之间的欧氏距离相关

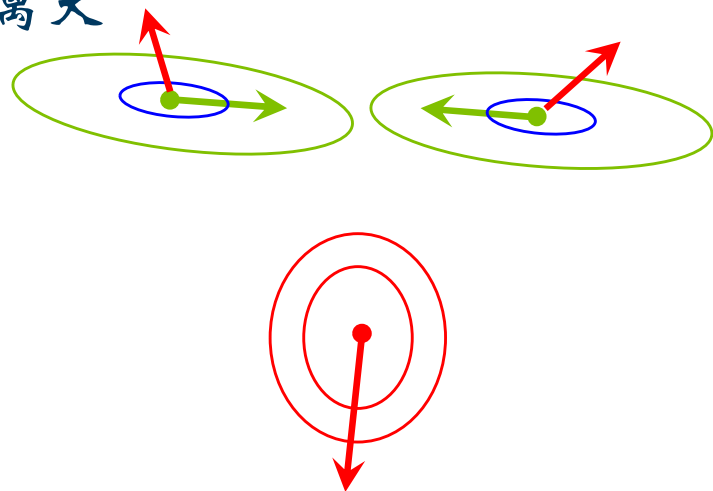


如何生成和传递样本自适应距离度量？

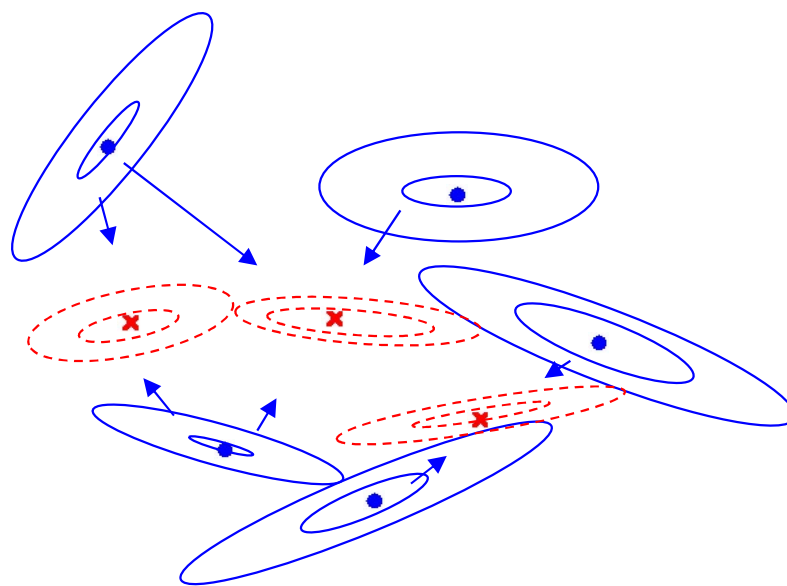
□ 从标记样本中生成样本自适应距离度量，并且通过邻域关系（图）将这种样本自适应度量传播出去

- 标记样本上的样本自适应距离度量生成方式：

生成的度量应该使得同类样本之间的距离小，异类样本之间的距离大



- 未标记样本的自适应距离度量应该遵从相近相似的原则（Metric Propagation）从近邻样本处获得



形式化模型

- 进行度量（标记）传播有两种选择：1. 采用随机游走的策略进行迭代求解；2. 将度量（标记）传播形式化成一个优化问题，这里我们选择的是后者

$$\begin{aligned} \min_{\mathbf{W}} \quad & \lambda \sum_{i=1}^n \sum_{j \in \mathcal{C}_i} \ell(\hat{y}_{ij}, D_i(\mathbf{x}_j)) + \Omega(\mathbf{W}, \mathbf{G}) \\ \text{s.t.} \quad & \mathbf{w}_i \geq 0, i = 1, \dots, n+u, \end{aligned}$$

学习的度量在标记样本上犯下的错误

并不需要考虑所有的样本，仅仅只该正则化项负责隐式地进行度量传播，要考虑特定样本邻域内的关系即可

受[Zhu 2003]的启发，该正则化项可以定义为：

对于同类样本来说， E_{ij} 表示样本 i 和 j 是否属于同一类， $E_{ij} = 1$ 表示属于同一类， $E_{ij} = 0$ 表示不属于同一类。

但是对于和特定 i 相关的样本 j ， E_{ij} 表示样本 i 和 j 是否属于同一类， $E_{ij} = 1$ 表示属于同一类， $E_{ij} = 0$ 表示不属于同一类。

$$\Omega(\mathbf{W}, \mathbf{G}) = \sum_{i,j=1}^{n+u} E_{ij} \|\mathbf{w}_i - \mathbf{w}_j\|^2 = 2\text{tr}(\mathbf{W}^\top \mathbf{L} \mathbf{W})$$

不需要考虑

形式化模型的特例和泛化

虽然在我们的工作中仅仅考虑到了样本对之间的信息，
但是整个ISD框架可以用于更普遍的情况

$$\begin{aligned} \min_{\mathbf{W}} \quad & \lambda \sum_{i=1}^n \sum_{j \in \mathcal{C}_i} \ell(\hat{y}_{ij}, D_i(\mathbf{x}_j)) + \Omega(\mathbf{W}, \mathbf{G}) \\ \text{s.t.} \quad & \mathbf{w}_i \geq 0, i = 1, \dots, n+u, \end{aligned}$$

可以在此使用其他不同的监督信息，
例如triplets information

FSM [Frome et al.
NIPS'06] is a
special case of
ISD

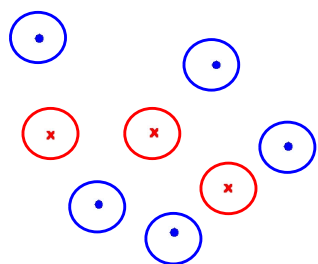
$$\Omega(\mathbf{W}, \mathbf{G}) = \sum_{i,j=1}^{n+u} E_{ij} \|\mathbf{w}_i - \mathbf{w}_j\|^2 = 2\text{tr}(\mathbf{W}^\top \mathbf{L} \mathbf{W}).$$

\mathbf{L} is set to identity matrix

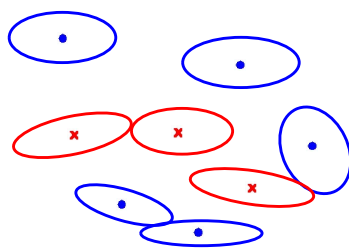
图的构建和精化

$$\begin{aligned} \min_{\mathbf{W}} \quad & \lambda \sum_{i=1}^n \sum_{j \in \mathcal{C}_i} \ell(\hat{y}_{ij}, D_i(\mathbf{x}_j)) + \Omega(\mathbf{W}, \mathbf{G}) \\ \text{s.t.} \quad & \mathbf{w}_i \geq 0, i = 1, \dots, n + u, \end{aligned}$$

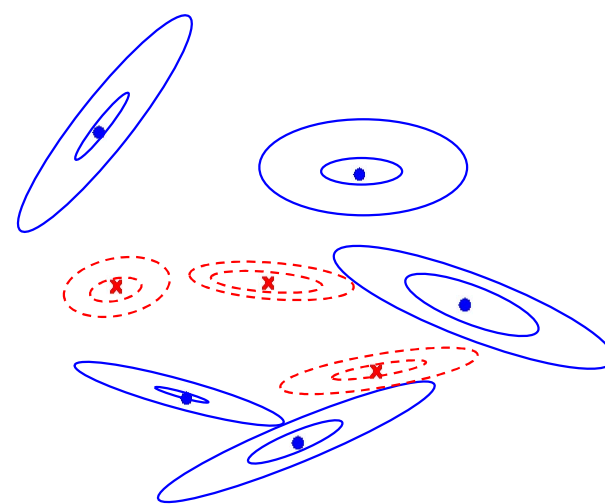
我们假设 \mathbf{G} 是预先给定的，但是如果没给定.....



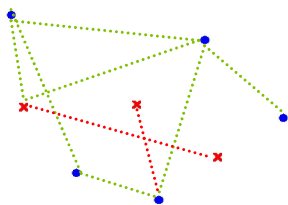
Initialize



In new ISD space



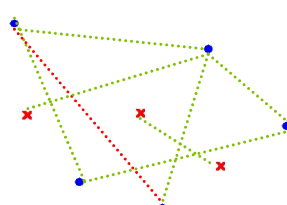
Final ISD



Graph

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

Weights



Updated Graph

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

Weights

损失函数的选择 ISD-L1

$$\ell(\hat{y}_{ij}, D_i(\mathbf{x}_j)) = \max(0, \hat{y}_{ij}(D_i(\mathbf{x}_j) - \eta))$$

通过引入裕量

$$D_i(\mathbf{x}) = \mathbf{w}_i^\top \delta_{\mathbf{x}_i, \mathbf{x}}$$

$$\delta_{\mathbf{x}_i, \mathbf{x}_j} = (\mathbf{x}_i - \mathbf{x}_j) \odot (\mathbf{x}_i - \mathbf{x}_j)$$

$$\begin{aligned} \min_{\mathbf{W}, \xi_{i,j}} \quad & \lambda \sum_{i=1}^n \sum_{j \in \mathcal{C}_i} \xi_{i,j} + 2\text{tr}(\mathbf{W}^\top \mathbf{L} \mathbf{W}) \\ \text{s.t} \quad & \hat{y}_{ij}(\mathbf{w}_i \delta_{\mathbf{x}_i, \mathbf{x}_j} - 1) \leq \xi_{i,j}, i = 1, \dots, n \\ & \xi \geq 0, \mathbf{w}_i \geq 0, i = 1, \dots, n + u \end{aligned}$$

考虑到 \mathbf{W} 的数量很多，同时求解有困难：

计算量太大！通过alternating descent的方法进行优化，也即：

每次固定其他的 \mathbf{W} s，优化其中的一个 \mathbf{W} ，然后再反复迭代，直到收敛



求解空间的选择

Primal:

$$\begin{aligned} \min_{\mathbf{w}_i, \xi_j} \quad & \sum_{i,j=1}^{n+u} E_{ij} (\mathbf{w}_i^\top \mathbf{w}_i - 2\mathbf{w}_i^\top \mathbf{w}_j) + \lambda \sum_{j \in \mathcal{C}_i} \xi_j \\ \text{s.t.} \quad & \hat{y}_{ij} (\mathbf{w}_i^\top \delta_{\mathbf{x}_i, \mathbf{x}_j} - 1) \leq \xi_j, \quad j \in \mathcal{C}_i \\ & \xi \geq 0, \mathbf{w}_i \geq 0 \end{aligned}$$

Dual:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & (\hat{\mathbf{D}}_i \boldsymbol{\alpha} - \boldsymbol{\gamma})^\top (\hat{\mathbf{D}}_i \boldsymbol{\alpha} - \boldsymbol{\gamma}) \\ & - 4(\hat{\mathbf{D}}_i \boldsymbol{\alpha} - \boldsymbol{\gamma})^\top \mathbf{C}_i + 4\theta_i \boldsymbol{\alpha}^\top \mathbf{y}_i. \\ \text{s.t.} \quad & 0 \leq \boldsymbol{\alpha} \leq \lambda, \boldsymbol{\gamma} \geq 0 \end{aligned}$$

$$\hat{\mathbf{D}}_i = \mathbf{D}_{i.} \mathbf{Y}_i \quad \mathbf{D}_{i.} = [\delta_{\mathbf{x}_i, \mathbf{x}_1}, \delta_{\mathbf{x}_i, \mathbf{x}_2}, \dots, \delta_{\mathbf{x}_i, \mathbf{x}_p}]$$

$$\mathbf{C}_i = \sum_j E_{ij} \mathbf{w}_j \quad \theta_i = \sum_j E_{ij}$$

加速方法

在加速算法方面，我们已经做出的努力是：

- 使用alternating descend方法进行优化
- 对同类之间产生的约束的数量进行了消减

但是异类之间产生的约束的数量仍然可能十分巨大，从而导致算法异常耗时



从nu-SVM中获得灵感，我们是否可以利用类似的方法得到一种更有效率的方法呢？ 答案是肯定的

$$\ell(\hat{y}_{i,j}, D_i(\mathbf{x}_j)) = \max(0, \hat{y}_{i,j}(D_i(\mathbf{x}_j) - \eta))^2$$

加速方法 ISD-L2

$$\begin{aligned} \min_{\mathbf{W}, \xi_{i,j}, \rho} \quad & \lambda \sum_{i=1}^n \sum_{j \in \mathcal{C}_i} \xi_{i,j}^2 + 2\text{tr}(\mathbf{W}^\top \mathbf{L} \mathbf{W}) - \rho \\ \text{s.t.} \quad & \hat{y}_{ij}(\mathbf{w}_i^\top \mathbf{x}_j - 1) \leq \xi_{i,j} - \rho \end{aligned}$$

Dual:

$$\begin{aligned} \min_{\alpha} \quad & \alpha^\top (\hat{\mathbf{D}}_i^\top \hat{\mathbf{D}}_i + \frac{\theta_i}{\lambda} \mathbf{I}) \alpha + 4(\theta_i \mathbf{y}_i^\top - \mathbf{C}_i^\top \hat{\mathbf{D}}_i) \alpha \\ \text{s.t.} \quad & \alpha^\top \mathbf{1} = 1, \alpha \geq 0 \end{aligned}$$

注意：这里存在一个等式约束

为了简化问题，我们可以首先把这个约束条件给去掉，然后因为等式约束的存在，这个Dual问题可以快速地使用SMO (Sequential Minimal Optimization) 求解。在求解出W之后再下式让W满足最终

$$\hat{\mathbf{w}}_i = (\mathbf{C}_i - \hat{\mathbf{D}}_i \alpha / 2) / \theta_i$$

$$\mathbf{w}_i = \max(0, \hat{\mathbf{w}}_i)$$

如果标记样本特别少，怎么办？

$$\begin{array}{ll} \min & \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}} \|\mathbf{x}_i - \mathbf{x}_j\|_A^2 \\ \text{s.t.} & A \in S_+, \quad \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}} \|\mathbf{x}_i - \mathbf{x}_j\|_A^2 \geq 1. \end{array}$$

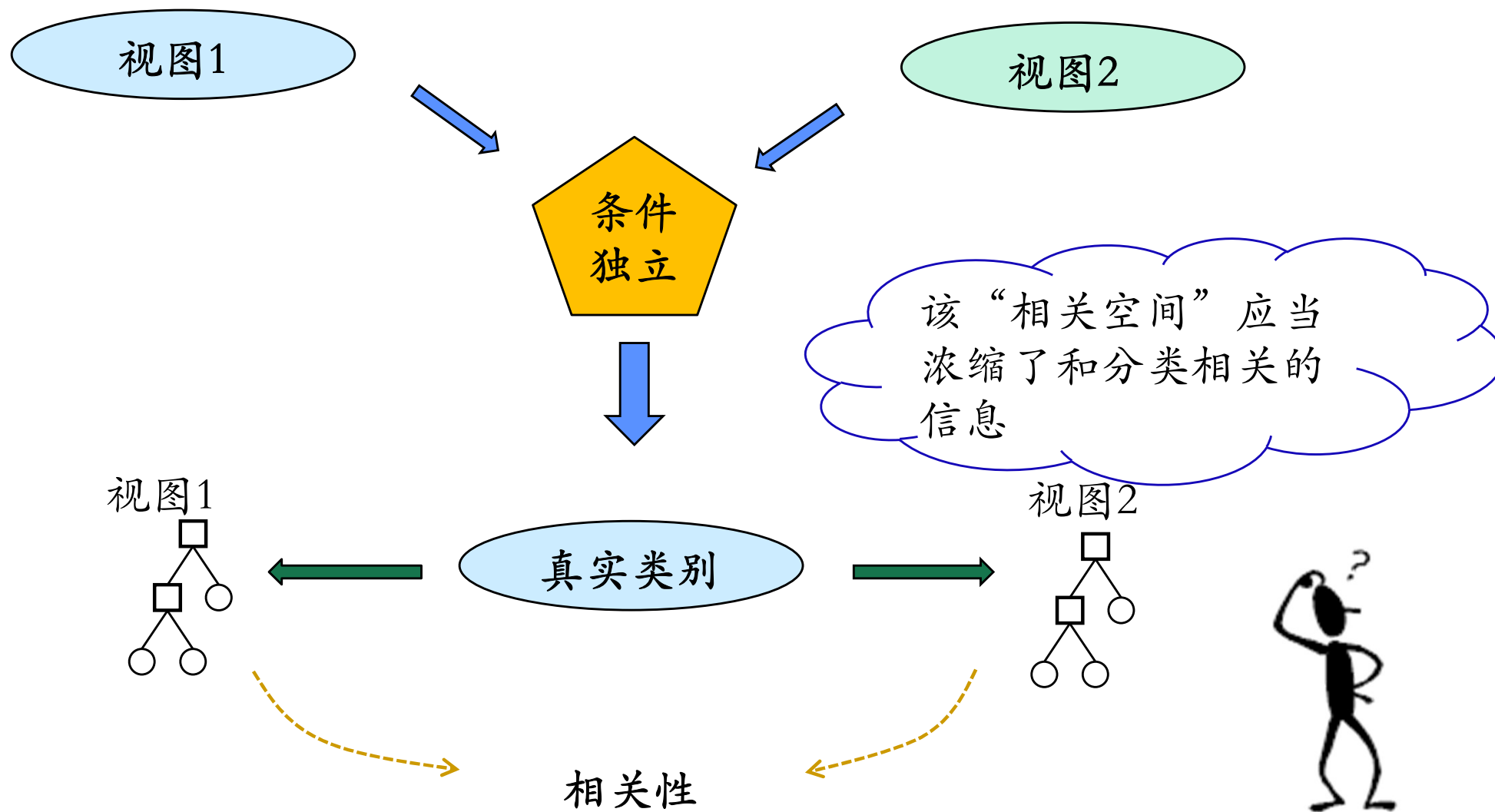
最小化同类之间的距离

保持异类样本之间的距离大于某个阈值

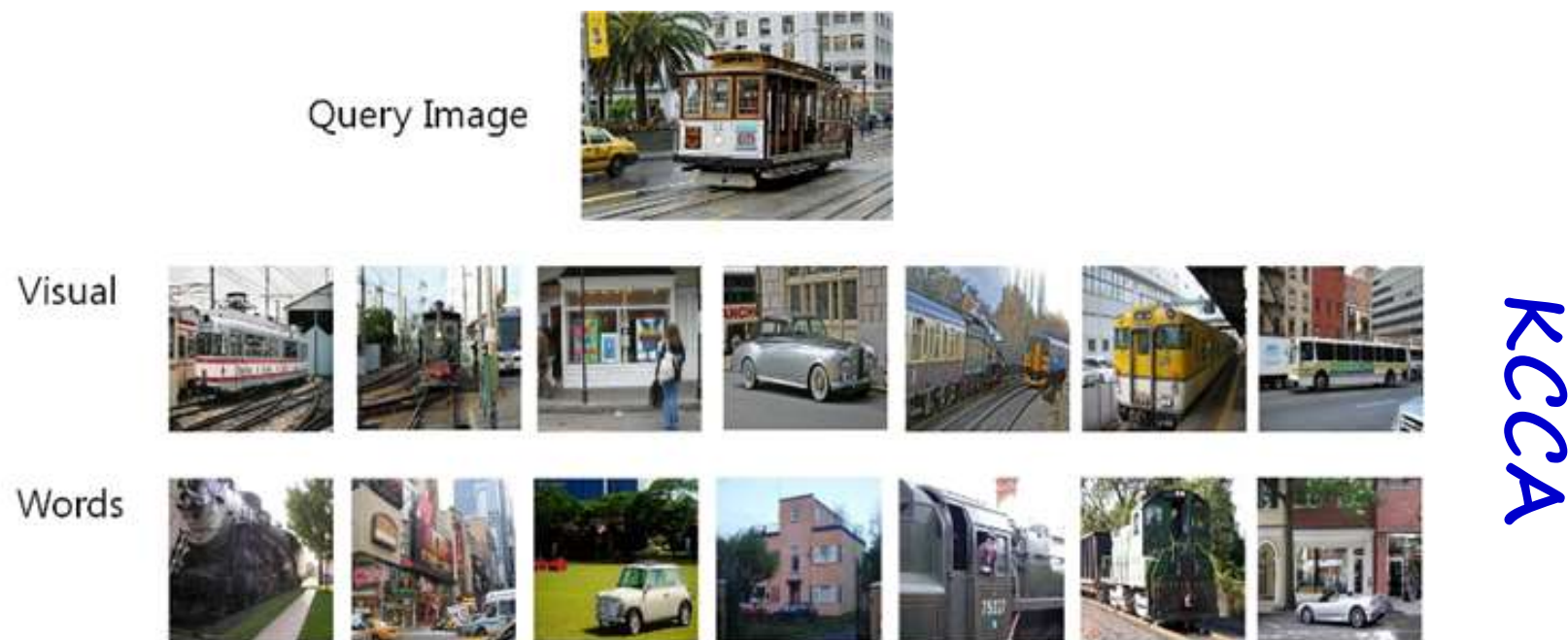
当标记样本极少的时候，无法得到这些约束

但是，如果你在web上面搜索的时候会提交多个查询样本么？
(将查询样本看成标记正例)

基于双视图关联距离度量的半监督学习



基于双视图关联距离度量的半监督学习



1. KCCA寻找“相关性”大的空间
2. 在此空间中定义距离度量
3. 使用该距离度量扩充标记样本集
4. 进行半监督学习Co-training

THANKS

