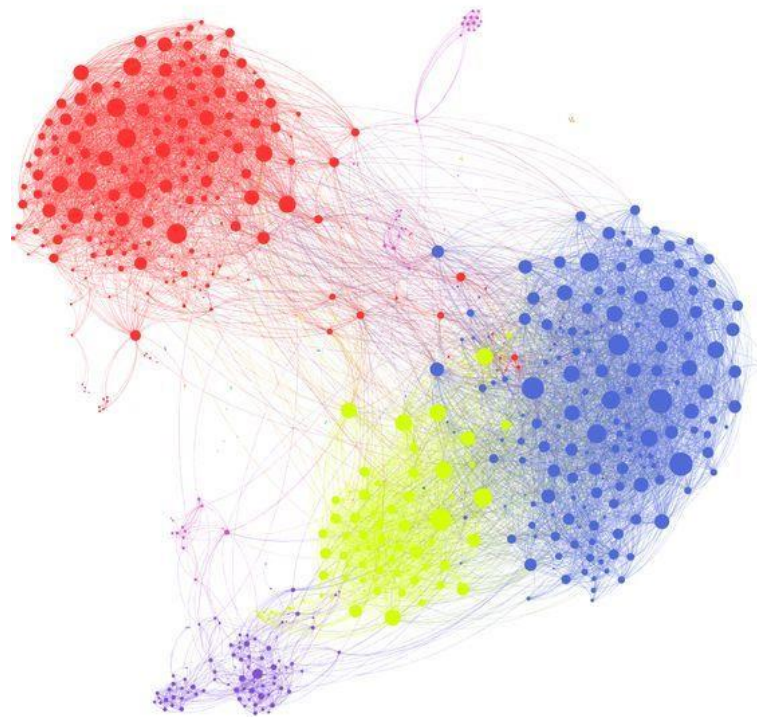


法律声明

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。



关注 **小象学院**



聚类分析

--Robin

K-Means聚类

- 聚类（clustering）属于无监督学习（unsupervised learning）
- 无类别标记
- 在线demo <http://syskall.com/kmeans.js/>
- 数据挖掘经典算法之一
- 算法接收参数k；然后将样本点划分为k个聚类；同一聚类中的样本相似度较高；不同聚类中的样本相似度较小
- 算法思想：
以空间中k个样本点为中心进行聚类，对最靠近它们的样本点归类。通过迭代的方法，逐步更新各聚类中心，直至达到最好的聚类效果

K-Means聚类

- 算法描述：
 1. 选择k个聚类的初始中心
 2. 在第n次迭代中，对任意一个样本点，求其到k个聚类中心的距离，将该样本点归类到距离最小的中心所在的聚类
 3. 利用均值等方法更新各类的中心值
 4. 对所有的k个聚类中心，如果利用2,3步的迭代更新后，达到稳定，则迭代结束。

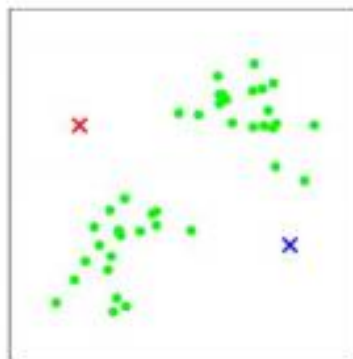
优点	缺点
<ul style="list-style-type: none">• 速度快，简单	<ul style="list-style-type: none">• 最终结果和初始点的选择相关，容易陷入局部最优• 需要给定k值

K-Means聚类

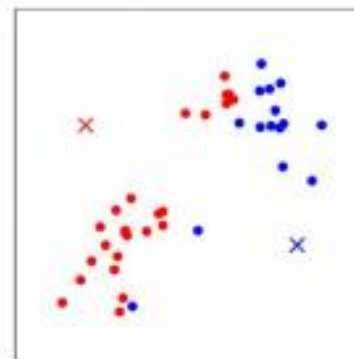
- 算法演示：



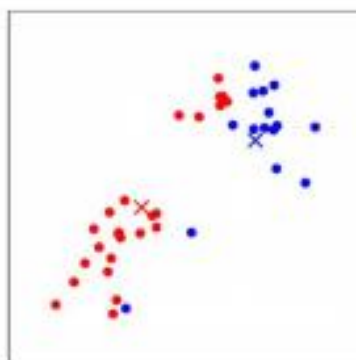
(a)



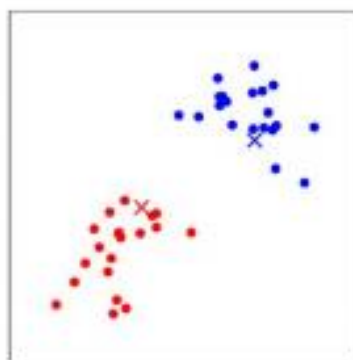
(b)



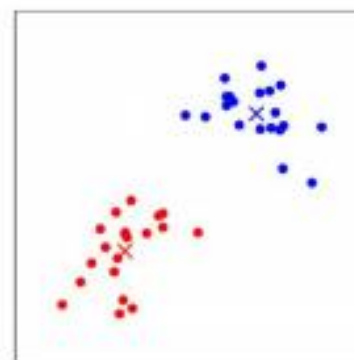
(c)



(d)



(e)



(f)

联系我们

小象学院：互联网新技术在线教育领航者

— 微信公众号：**小象学院**

