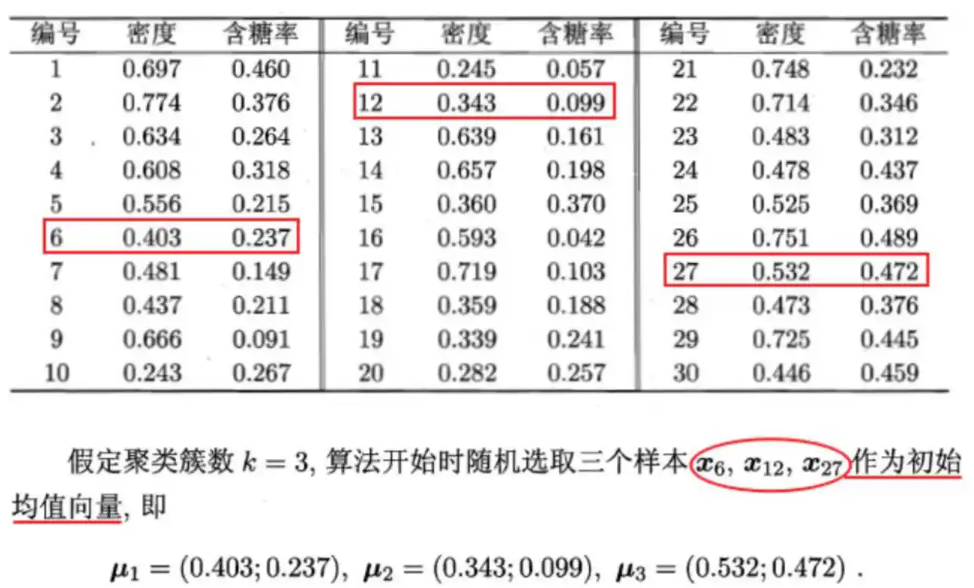
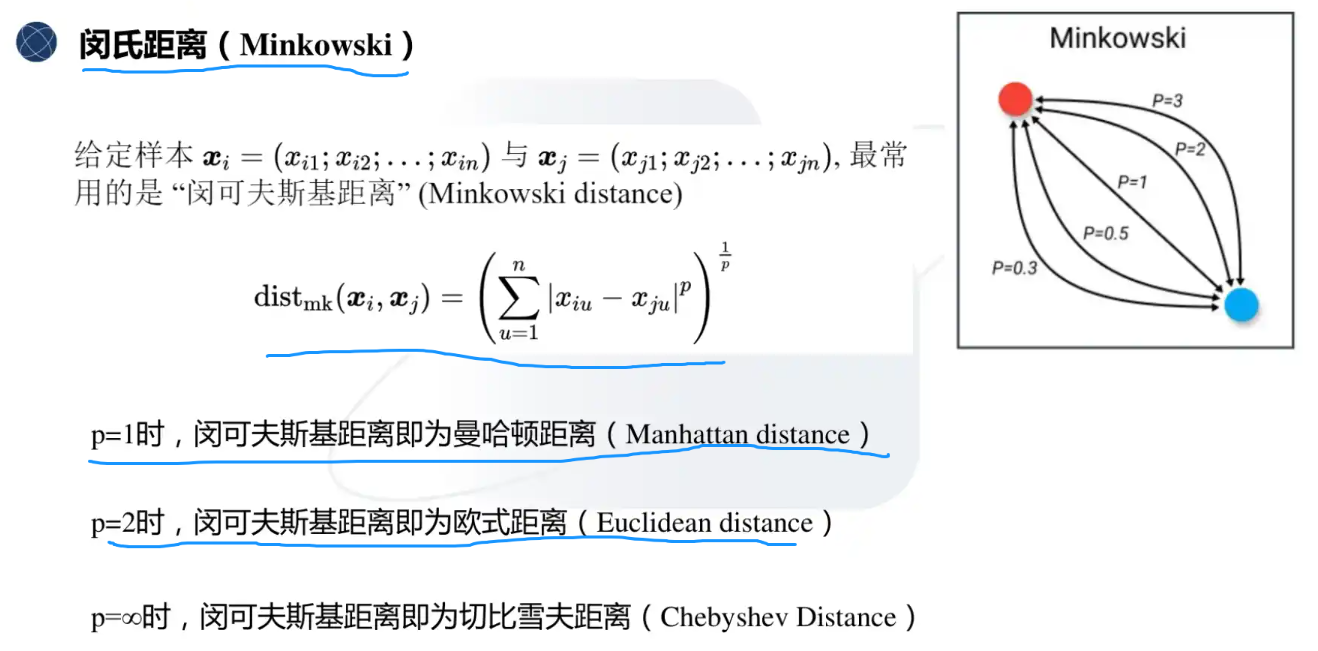
# 聚类算法

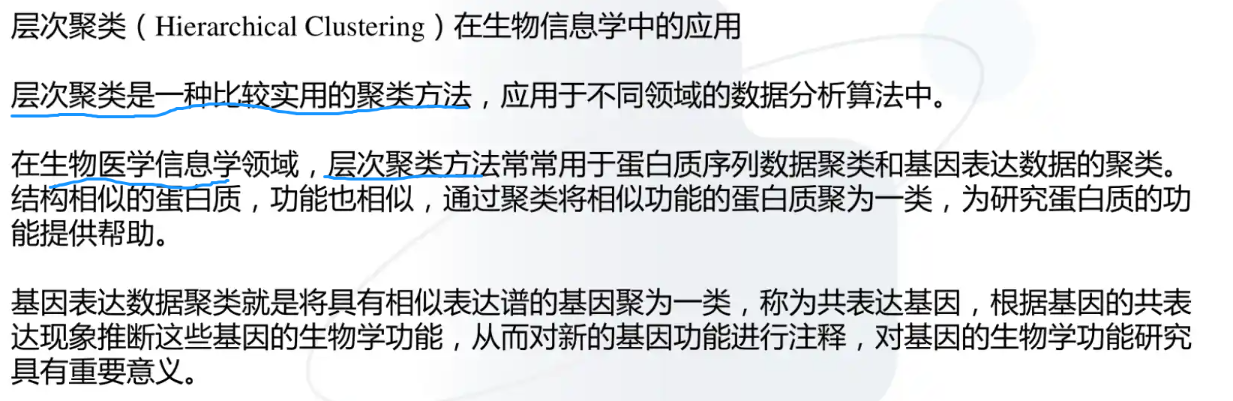
1. 类别：无监督学习
2. K均值聚类算法
   1. 步骤
      1. 选定k值
         1. 
      2. 随机选取样本作为初始的均值向量
      3. 不断计算距离，找到新的中心点，直到每一类的数据数量不再发生变化
         1. 求距离的算法
            1. 曼哈顿距离
            2. 切比雪夫距离
            3. 欧氏距离
            4. 闵式距离



* + - * 1. 余弦相似度
        2. 汉明距离

1. 密度距离DBSCAN
   1. 优势在于能够通过样本之间的连接性来确定样本空间的
      1. 文本

         低可信度描述已自动生成
   2. 刻画点与点之间的连接性
   3. 相关概念
      1. 图片包含 文本

         描述已自动生成
2. 层次聚类
   1. 自底向上的聚类思想
   2. 每个样本当做一个簇
   3. 计算每一簇之间的距离
   4. 把距离最近的簇归为一类
   5. 重复步骤
   6. 直到达到终止条件（到达目标分类簇数或者样本变成一簇）
      1. 
      2. 多应用于生物科学领域
3. 高斯混合模型
   1. 算法涉及很多数学计算，不想看了
4. 总结图形用户界面, 文本, 应用程序

   描述已自动生成
5. 性能度量
   1. 簇内的样本相似度尽可能的高
      1. 平均距离（离散程度）
         1. 图示

            描述已自动生成
      2. 簇间样本距离
         1. 图示

            描述已自动生成
      3. DB指数（越小越好）
         1. 图示

            描述已自动生成
         2. 图示

            描述已自动生成