K-means聚类算法

K-means 是一个反复迭代的过程, 算法分为四个步骤:

1. 随机选取数据空间中的K个对象作为初始中心, 每个对象代表一个聚类中心;
2. 对于样本中的数据对象, 根据它们与这些聚类中心的欧氏距离, 按距离最近的准则将它们分到距离它们最近的聚类中心所对应的类;
3. 更新聚类中心: 将每个类别中所有对象所对应的均值作为该类别的聚类中心
4. 如果聚类中心和上一次聚类中心的距离小于某一个阈值。（表示重新计算的质心的位置变化不大，趋于稳定， 或者说收敛），可以认为我们进行的聚类已经达到期望的结果，算法终止。
5. 如果不满足阈值则迭代计算聚类中心。