

1.kmeans初始聚类中心的选取

基本思想：初始的聚类中心之间的相互距离要尽可能的远

1. 从输入的数据点集合中随机选择一个点作为第一个聚类中心
2. 对于数据集中的每一个点 x ，计算它与最近聚类中心（指已选择的聚类中心）的距离 $D(x)$
3. 选择一个新的数据点作为新的聚类中心，选择的原则是： $D(x)$ 较大的点，被选取作为聚类中心的概率较大
4. 重复2和3直到 k 个聚类中心被选出来
5. 利用这 k 个聚类中心来运行标准的kmeans算法

2.如何确定k的大小

https://www.julvedu.com/question/big/kp_id/23/ques_id/2119