算法	概括	优缺点
k-means	每次从类中求均值作为中心点 用到了EM的思想 目标是最小化sum of squared error	要求预设k值 易受噪音和离异点的影响 对不规则形状的类聚类效果不好 不保证全局最优
k-means++	目标是找到k个合理的初始种子点给k-means。 1. 随机挑个随机点当“种子点” 2. 对于每个点，计算其和最近的“种子点”的距离D(x)并保存，然后把这些距离加起来得到Sum(D(x))。 3. 再取一个随机值，用权重的方式来取计算下一个“种子点”。这个算法的实现是，先取一个能落在Sum(D(x))中的随机值Random，然后用Random -= D(x)，直到其<=0，此时的点就是下一个“种子点”。 4. 重复2和3直到k个中心被选出来 5. 利用这k个初始的聚类中心来运行标准的k-means算法	
k-modes	K-Means算法的扩展 对于分类型数据，用mode求中心点	
k-prototypes	结合了k-means和k-modes	
k-medoids	每次从类中找一个具体的点来做中心点。目标是最小化absolute error。 PAM是一种典型的k-medoids实现。	对噪音和离异点不那么敏感 然而计算量大很多
CLARA	先抽样，再用PAM	对于大数据比PAM好点 主要是看sample的效果
CLARANS	每次随机的抓一个medoid跟一般点，然后判断，这两者如果替换的话，能不能减小absolute-error	融合了PAM和CLARA两者的优点，是第一个用于空间数据库的聚类算法