

K-means 是一种常用的聚类算法，用于将数据集分成 K 个互不重叠的子集，每个子集被称为一个簇。这个算法的目标是最小化簇内数据点的方差，使得同一簇内的数据点相似度较高，而不同簇之间的相似度较低。

以下是 **K-means** 聚类的基本过程：

1.初始化： 选择 K 个初始聚类中心点，通常是从数据集中随机选择。这些聚类中心点将成为每个簇的代表。

2.分配： 对于每个数据点，计算其与每个聚类中心的距离，并将其分配到离其最近的聚类中心所属的簇。

3.更新： 重新计算每个簇的中心，即将该簇中所有数据点的均值作为新的聚类中心。

4.重复： 重复步骤 2 和 3，直到簇分配不再变化，或者达到预定的迭代次数。

K-means 的结果包括分配给每个数据点的簇标签和最终的聚类中心。这些簇标签可用于将数据点分组，而聚类中心代表每个簇的中心位置。聚类结果具有以下特点：

1.簇内相似度高： 同一簇内的数据点相似度较高，即它们彼此之间的距离较小。

2.簇间相似度低： 不同簇之间的数据点相似度较低，即它们彼此之间的距离较大。

K-means 是一种迭代的优化算法，但它可能受到初始聚类中心的选择和局部最小值的影响。在实际应用中，可以多次运行 **K-means** 算法，并选择具有最小总簇内方差的结果，以提高聚类的稳定性。