

K-means 算法是一种基于距离度量的迭代聚类算法，其原理基于最小化簇内数据点的方差，通过不断迭代来优化簇的中心位置。以下是 K-means 算法的详细描述：

1.初始化：

选择聚类数 K 。

随机初始化 K 个聚类中心，每个中心代表一个簇。

2.迭代优化过程：

分配数据点到簇：对每个数据点，计算其与各个聚类中心的距离，并将其分配到距离最近的簇中。

更新簇中心：对每个簇，计算其所有数据点的平均值，将该平均值作为新的聚类中心。

3.重复迭代：

重复步骤 2，直到簇中心不再发生明显变化，或达到预定的迭代次数。

4.输出结果：

算法收敛后，每个数据点都被分配到一个簇中，形成最终的聚类结果。

K-means 算法的关键步骤是通过迭代过程不断优化簇中心，使得每个簇内的数据点更加相似。这一过程是通过最小化簇内方差的目标函数来实现的。然而，K-means 对初始聚类中心敏感，因此可以多次运行算法，选择效果最好的聚类结果。另外，用户需要在应用前自行确定聚类数 K ，这一选择直接影响了最终聚类效果。