1. means算法选择说明

K-means是机器学习中的一种无监督聚类算法，由于其简单和容易理解，所以是人们最常用的聚类算法之一。它是一种迭代聚类算法，它试图将数据点准确分配给我们预定义的K个聚类中的一个聚类。

与其他任何聚类算法一样，它试图使一个聚类中的项目尽可能相似，同时也使聚类之间彼此尽可能不同。通过确保群集中的数据点与该群集的质心之间的平方距离之和最小。群集的质心是群集中所有值的平均值。

1. Means算法的实现步骤，主要分为四个步骤：

1、从样本集合中随机抽取k个样本点作为初始族的中心。

2、将每个样本点划分到距离它最近的中心点所代表的族中。

3、用各个族中所有样本点的中心点代表族的中心点。

4、重复2和3，直到族的中心点不变或达到设定的选代次数或达到设定的容错范围。

技术指标说明

样本轮廓系数

轮廓系数（Silhouette Coefficient），是聚类效果好坏的一种评价方式。

最佳值为1，最差值为-1。接近0的值表示重叠的群集。负值通常表示样本已分配给错误的聚类，因为不同的聚类更为相似。

计算步骤：

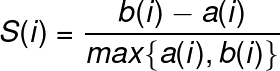
1）对于簇中的每个向量，分别计算它们的轮廓系数。

      对于其中的一个点 i 来说：

      计算 簇内不相似度a(i) ：i向量到同簇内其他点不相似程度的平均值，体现凝聚度

      计算 簇间不相似度b(i) ：i向量到其他簇的平均不相似程度的最小值，体现分离度

      那么第i个对象的轮廓系数就为：



所有样本的si的均值称为聚类结果的轮廓系数，定义为S，是该聚类是否合理、有效的度量。聚类结果的轮廓系数的取值在【-1,1】之间，值越大，说明同类样本相距约近，不同样本相距越远，则聚类效果越好。