

k-means

K-means 算法步骤

1. **初始化**: 选择 K (簇的数量) 个点作为初始中心点。在我们的例子中, $K=2$, 我们随机选择两个点作为初始中心。比如说, 我们选择 $(1, 2)$ 和 $(10, 4)$ 。
2. **分配**: 将每个数据点分配到最近的中心点所代表的簇。这里, 我们计算每个数据点到两个中心点的距离, 并将它们分配到最近的簇。
3. **更新**: 对于每个簇, 重新计算中心点, 这是通过取簇中所有点的平均值来完成的。
4. **重复**: 重复步骤 2 和 3, 直到中心点不再显著变化或达到预定的迭代次数。

首先随机选择样本点作为中心, 这里有几个簇就分几个样本点, 之后对于每一个样本点来讨论, 计算单个样本点到这些簇的距离, 并把这个样本点分配到距离最短的簇, 重复以上步骤对每个样本点都这样操作, 得到了第一次分类结果。

之后取这些簇的中心点, 也就是平均中心, 然后更新簇的中心。

重复再根据这些新的簇中心来划分所有样本点, 直到最后中心点不变时找到最佳簇分类结果。