

作业题目：实现并测试 K-means 聚类算法

作业要求：

1. 实现 K-means 算法，采用三种不同初始化方法（见 Lecture11-Clustering-K-Means 第 19 页）得到三个不同版本。

2. 在 UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets.php>) 找到自己认为合适的 5 个数据集，进行如下分析：

1) 聚类个数设定为正确类个数情况下，三种不同初始化方法版本的聚类效果（用 Normalized Mutual Information 度量效果，见 ANMI\_analytical\_11.m，非 Matlab 版本的度量方法同学们可以网上找到）。

2) 设置不同聚类个数  $K$  的情况下（如  $K=2, 3, 4, \dots, n/2$ ），三种不同初始化方法版本的“目标函数  $J$  随着聚类个数变化的曲线”（见 Lecture11-Clustering-K-Means 第 17 页）。

3. 提交代码+数据集+详细实验报告及分析（编程语言不限、报告字数不限，需要透彻分析），压缩包提交：学号+姓名。

4. 提交日期：5 月 1 日。提交邮箱：sysumldm2022@163.com