1. 实验介绍

1.1关于本实验

使用K-means模型对鸢尾花数据集（iris.csv）进行聚类。首先，随机将数据划分为训练集和测试集，划分比例为8:2；其次，至少尝试 3 个不同的K值，并可视化出不同K下的聚类结果，以及训练集和测试集上的损失；最后，对结果进行讨论，并给出最佳K值。

1.2 数据集介绍

Iris数据集是常用的分类实验数据集，由Fisher于1936年收集整理。Iris也称鸢尾花卉数据集，是一类[多重变量分析](https://baike.baidu.com/item/%E5%A4%9A%E9%87%8D%E5%8F%98%E9%87%8F%E5%88%86%E6%9E%90/1617367?fromModule=lemma_inlink" \t "/Users/yanrong/Documents\\x/_blank)的数据集。数据集包含150个数据样本，分为3类，每类50个数据，每个数据包含4个属性。可通过花萼长度（Sepal.Length），花萼宽度（Sepal.Width），花瓣长度（Petal.Length），花瓣宽度（Petal.Width）4个属性预测鸢尾花卉属于Setosa（山鸢尾），Versicolour（杂色鸢尾），Virginica（维吉尼亚鸢尾）三个种类中的哪一类。更多数据集介绍可参考：*<https://archive.ics.uci.edu/ml/datasets/Iris/>*

1. 实验步骤
   1. 数据集划分

随机将数据划分为训练集和测试集，划分比例为8:2。

* 1. 手写K-means算法

手写K-means算法，并在服务器上运行。注意，至少尝试 3 个不同的 K值，得出不同K下的聚类结果、训练集和测试集上的损失。

* 1. 实验结果可视化

可视化不同K值下的聚类结果（需用颜色区分不同类别），并将对应训练集和测试集上的损失值绘制成折线图（横轴为训练轮次，纵轴为损失值）。

* 1. 分析实验结果

对结果进行讨论，并给出最佳K值。