



浙江工业大学

人工智能原理及应用 实验报告

实验名称: Kmeans 鸢尾花聚类-完整版

学 号: 202205240220

姓 名: 潘家航

专业班级: 自动化 2304

学 院: 信息工程学院

指导教师: 付明磊

目录

一、实验目的.....	3
二、实验设备.....	3
三、实验内容.....	3
四、实验过程.....	3
五、实验结果分析.....	3
六、实验小结.....	4
七、其它.....	5

一、实验目的

1. 掌握 K-Means 聚类算法的基本原理与实现过程；
2. 理解无监督学习的概念以及聚类任务的应用场景；
3. 利用 scikit-learn 库中的 KMeans 模型对鸢尾花 (Iris) 数据集进行聚类；
4. 对比聚类结果与数据分布特征，分析模型性能与不足。

二、实验设备

1. 硬件设备： 游侠 G15。
2. 软件环境： 主流浏览器（如 Chrome、Edge）、代码编辑器 pycharm、Python 编程环境。
3. 开发平台： 百度 AI 开放平台 (ai.baidu.com) 的 EasyDL 产品。

三、实验内容

使用 K-Means 聚类算法对鸢尾花数据集进行无监督聚类实验。

鸢尾花数据集包含 150 个样本，3 种类型（山鸢尾、变色鸢尾、维吉尼亚鸢尾），

每个样本具有 4 个特征：萼片长度、萼片宽度、花瓣长度、花瓣宽度。

通过 K-Means 算法将样本聚成 3 类，并进行可视化展示与结果分析。

四、实验过程

1. 导入相关包
2. 加载数据集
3. 绘制二维数据分布图
4. 定义模型与训练函数
5. 模型训练
6. 获取聚类结果并可视化

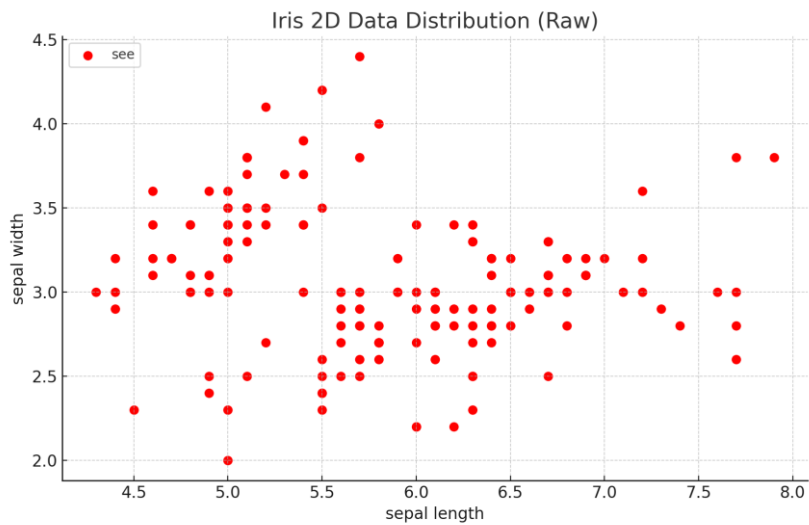


图 1 鸢尾花二维数据分布图（未聚类）

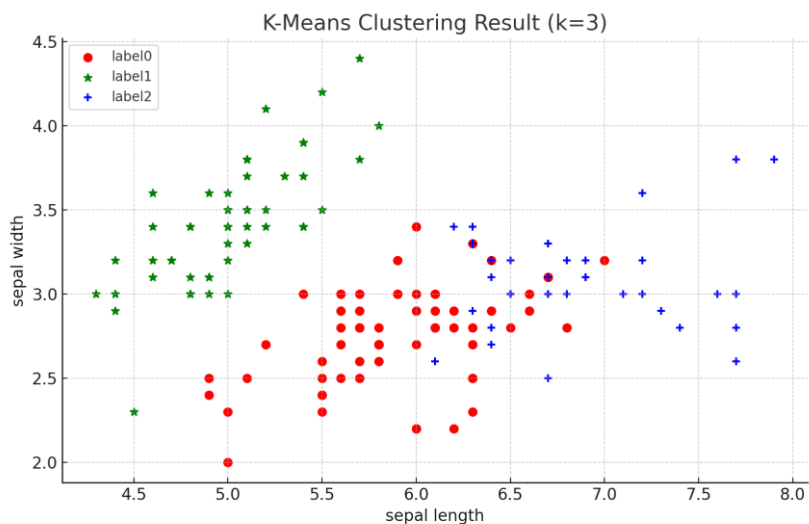


图 2 K-Means 聚类结果可视化图

五、实验结果分析

K-Means 成功将鸢尾花样本划分为 3 个簇，结果与实际类别较为接近。

聚类中心反映出不同花种在特征空间的分布差异。虽然在边界区域存在一定重叠，但整体聚类效果良好，说明该算法能够有效识别数据内在结构。

六、实验小结

1. 本实验实现了基于 K-Means 的鸢尾花聚类分析，掌握了无监督学习的核心思想；

2. 通过可视化结果验证了聚类算法的自动分组能力;
3. 发现不同簇的特征差异与实际花种分布基本一致;
4. K-Means 对初始质心敏感,需提前设定 K 值,后续可考虑使用层次聚类或 DBSCAN 改进。

七、其它

本实验的所有代码均在 Python 环境中实现,运行于 PyCharm 编辑器。实验过程中采用了 matplotlib 库进行数据可视化,scikit-learn 库实现 K-Means 算法,实验结果稳定可复现。此外,实验还生成了两幅结果图:图 1 展示了鸢尾花数据集中前两个特征(萼片长度与萼片宽度)的二维分布,可直观看出样本在特征空间的整体结构;图 2 为基于 K-Means 的聚类结果,不同颜色代表聚类后的三类簇分布情况,可观察到模型在样本区分方面的良好表现。为进一步验证算法性能,可引入以下改进与拓展:增加评价指标——使用轮廓系数(Silhouette Coefficient)或 Calinski-Harabasz 指数评估聚类效果;比较不同 K 值的聚类表现——通过肘部法(Elbow Method)确定最优簇数;结合降维算法——使用 PCA(主成分分析)将高维特征降到二维或三维,以更直观地展示聚类结果;与其他聚类算法对比——如层次聚类(Hierarchical Clustering)或密度聚类(DBSCAN),分析各自优缺点;结合实际应用场景——探索聚类算法在药品分类、图像分组、用户画像分析等任务中的扩展潜力。通过本次实验,掌握了从数据加载、算法训练到结果可视化的完整流程,初步具备了将无监督学习算法应用到实际数据分析任务中的能力。