****

**《鸢尾花数据可视化》上机报告**

# 实验名称：利用sklearn库实现Iris鸢尾花数据集可视化

**学号：** 21051907

**姓名：** 陈浩明

完成时间 2023 年 3 月

## 实验目的

将鸢尾花数据集可视化，直观了解其内在联系。

## 实验环境

谷歌在线python编程平台Colab（https://colab.research.google.com/）

## 项目内容（作业任务、算法原理）

利用sklearn库实现Iris鸢尾花数据集可视化：

直方图： 数据集在不同feature上的数量分布；

散点图： 数据集在坐标系中的直观分布显示；

箱图： 显示出一组数据的最大值、最小值、中位数、及上下四分位数；

平行坐标： 数据集中不同类别的特性和规律；

散点图矩阵： 组织成网格（矩阵）形式的散点图集合。每个散点图显示一对变量之间的关系。

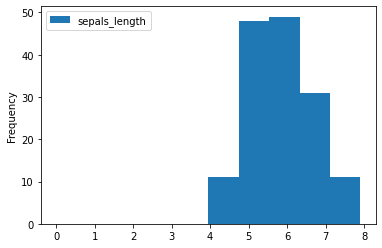
## 实验结果及结果分析

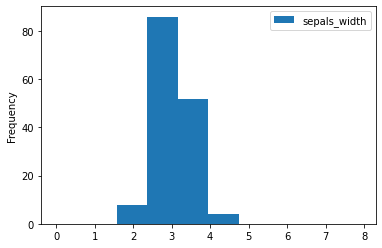
### 数据集及评估方法

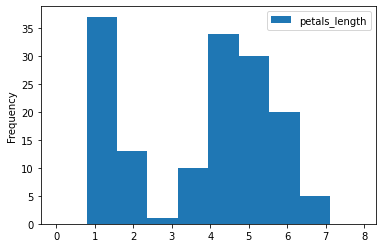
鸢尾花数据集是sklearn库自带的数据集，其包含三类鸢尾花（各50个样本），共150个样本。每个样本有4个features，分别是花萼宽度/长度，花瓣宽度/长度。我们需要对其可视化以发现其内在联系。

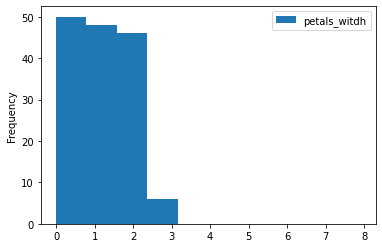
### 模型对比（表格、结果展示截图等）

直方图：

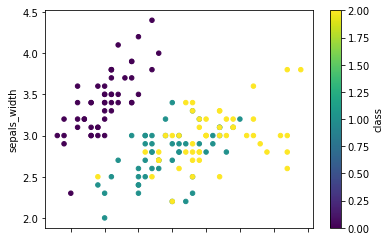




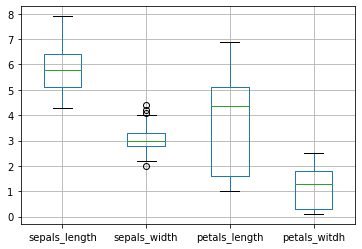




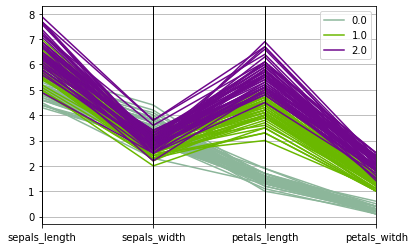
散点图：



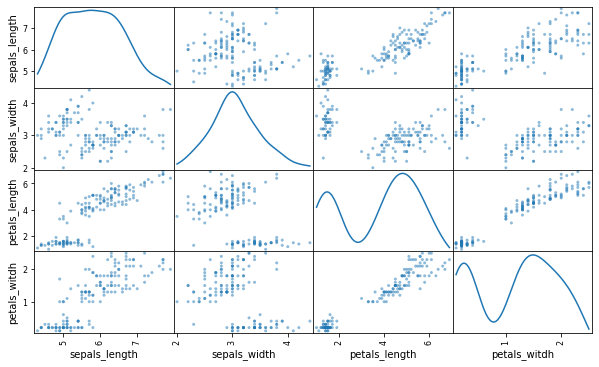
箱型图：



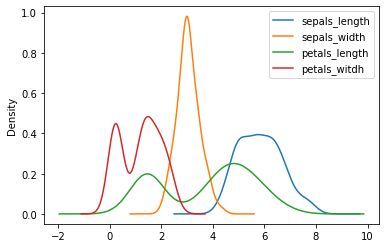
平行坐标：



散点图矩阵：



KDA图：



## 实验总结

鸢尾花数据集包含包含三类样本，每个样本有4个特征。在建模鸢尾花分类模型之前从不同角度对鸢尾花数据集可视化可以让我们大致了结其内部规律，如从直方图和箱型图中我发现不同种类的鸢尾花其花萼宽度相差较小，可能对分类模型的影响较小；从平行坐标图中我发现不同鸢尾花在4个features上的走势不同…这些可视化图让我门直观了解鸢尾花不同类别的内在联系，为后续分类建模提供了大大的帮助，所以数据可视化是数据科学中不可缺少的部分。