

第一次课后作业

0-1 提交要求

- 提交内容：
 1. 实验源代码（ipynb文件）：命名方式为 学号+姓名+班级+课后作业1.ipynb，例如 202312345+张三+班级+课后作业1.ipynb；
 2. 实验报告：可以是源代码导出的PDF文件，但需要有markdown的文字描述分析，命名方式同上，PDF文件中需要留存运行痕迹；
 3. 以上两个文件压缩打包，压缩包命名方式同上。
- 提交方式：
 - 提交至1350650232@qq.com；
 - 邮件主题：【25春机器学习】202312345+张三+班级+课后作业1。
- 截止时间：
 - 北京时间：2025年4月10日 23:59。

0-2 环境配置

- Conda: Conda可以帮助用户安装、运行、更新和管理软件包及其依赖关系，建议通过Conda管理计算机中的Python环境，Windows/Linux/macOS用户可在[Anaconda官网](#)下载相应安装包进行安装，虚拟环境配置可参考[conda官网教程](#)或自行搜索。
- Jupyter Notebook / JupyterLab: Jupyter Notebook是一个开放的、实时交互的Web应用程序，允许创建和分享包含实时代码、方程、可视化输出和说明文本的文档，应用范围包括数据清洗与转换、数值模拟、统计建模、机器学习等领域。建议使用Jupyter Notebook进行实验代码编写及结果展示。可参考[官方指导](#)或Conda进行notebook的安装。
- 一些会用到的Python库：numpy、pandas、matplotlib、seaborn、scikit-learn。

1 特征工程

- 作业目的：
 - 理解数据预处理的重要性；
 - 掌握特征工程的方法；
 - 掌握使用pandas、numpy等科学计算工具以及matplotlib、seaborn等可视化工具进行数据探索和可视化。
- 作业内容：
 - 下载数据集[红酒质量评估](#)（只需要用到winequality-red.csv数据集），使用pandas库加载和探索数据集；
 - 将任务定义为二分类任务，相应的需要转换样本标签以适配任务；
 - 使用可视化方法进行特征工程分析，可视化结果需要留存（散点图、heatmap等）；

2 逻辑回归模型的实现与评估

- 作业目的：
 - 理解并实现逻辑回归、梯度下降以及交叉验证；
 - 掌握使用机器学习工具库。
- 作业内容：
 - 手动实现逻辑回归、梯度下降，并在红酒质量数据集上进行交叉验证；
 - 使用scikit-learn库快速实现逻辑回归、梯度下降、交叉验证等方法。