

第一周任务

任务提交方式：github仓库

可以用的py库：numpy, pandas, matplotlib, gym, time

禁止使用pytorch, tensorflow等深度学习库，禁用sklearn等直接调用模型的库（题目特殊要求除外）

前置

1. 安装好 python 环境
2. 掌握 python 的基本语法 + numpy + pandas + matplotlib
3. 学会基本的git操作

一、数据处理

数据集：泰坦尼克号

https://uniquestudio.feishu.cn/drive/folder/fldcnV0PzAB5J8ZaoMp8WXho8if?from=from_copylink

1: 缺失值处理

包括但不限于，knn填补，众数，均值填补，补零

2: 数据标准化，归一化

3: 对于离散型变量的 OneHotEncoder

二、逻辑回归

实现逻辑回归

防止过拟合的算法

尝试多种优化方式，包括但不限于，sgd, adam

三、机器学习（两个小题二选一即可）

第1题：SVM

- 确保自己明白数学的推导过程,丢了一本书在下面 svm讲的蛮好
- smo实现

- 合页损失函数实现
- 尝试多种核函数，比较不同的效果
- 尝试可视化



机器学习.pdf

2.70MB



第2题：决策树&adaboost

决策树

- 实现ID3，C4.5，CART
- 实现预剪枝和后剪枝
- *决策树可视化

随机森林

- 实现随机森林

集成学习

- 实现 Bagging
- 实现adaboost

数据集还是泰坦尼克号，决策树是对离散+连续的数据集进行构建的！！