

考核一

史莱姆王

对于附件1, 进行数据清洗后, 使用线性回归算法, 进行回归预测, 并通过模型评估判断预测效果

内容

- 学习方向
  - 数据清洗
    - 缺失值、异常值处理
    - 字符数字化 (独热编码)
    - 特征构造、特征选择、数据转换等
    - 降维、分箱、聚类分析
  - 线性回归
  - 模型评估
    - rmse评估
    - 残差图
- 资料
  - 线性回归
    - b站吴恩达机器学习系列视频
  - 数据清洗
    - 链接发群里

骷髅老大

对于附件2, 进行数据清洗后, 使用逻辑回归算法, 进行分类预测, 并通过模型评估判断预测效果

内容

- 学习方向
  - 数据清洗
  - 逻辑回归
  - 模型评估
    - 精确率
    - 召回率
    - roc曲线
    - auc
- 资料
  - 逻辑回归
    - b站吴恩达机器学习系列视频
    - csdn博主

注意事项

1. 算法请使用python语法和科学计算库实现
2. 爬虫请使用基础爬虫库和正则表达式实现
3. 多推导、多记笔记、多动手实现代码
4. 心得和行程安排文档记得写
5. 对于刚入坑的小朋友, 难度较大, 如果做不完, 考核二可以继续做

地穴蠕虫

基础: 通过爬虫, 爬取京东任意3个笔记本的信息

进阶: 通过爬虫, 爬取作文吧的任意10篇作文 (标题+正文)

内容

- 学习方向
  - urllib库 (或者request库)
  - 正则表达式
- 资料
  - 崔庆才爬虫系列

