

第一部分

任务 1:

按照 requirement 文档配置虚拟环境

任务 2:

读入数据，查看数据异常值情况，删除缺失值所在行，对数据完成清洗工作

任务 3:

对数据中一些关键信息，可画图查看关键信息的分布。

任务 4:

开始进行本项目中需要使用的几类算法的系统学习

第二部分

任务 1:

导入数据集，对数据集进行训练集测试集分割

任务 2:

学习 sklearn 包相关内容，学习并实现决策数回归、线性回归、岭回归、拉索回归和 SVM 回归的算法原理和算法内容

任务 3:

利用以上模型对训练集进行训练。根据前一周所学内容，调整超参数，找到最优参数，对算法效果进行提升

任务 4:

对不同结果的算法进行对比，对每一种算法进行评估，找到效果最好的算法

任务 5:

分别建立 KNN 回归、随机森林回归、Adaboost 回归、GBRT、Bagging 和 ExtraTree 极端随机数回归算法

任务 6:

利用以上模型对训练集进行训练。根据前一周所学内容，调整超参数，找到最优参数，对算法效果进行提升

任务 7:

对不同结果的算法进行对比，对每一种算法进行评估，找到效果最好的算法

第三部分

任务 1:

对数据集运行结果，以及运行中的各种判别进行可视化

任务 2:

实现报告自动输出功能，输出的报告包括了训练中的文字信息以及上述可视化

任务 3:

项目总结