

# XXXX 项目总结

## 1 引言

### 1.1 项目背景

### 1.2 参考文档

## 2 项目完成情况总结

### 2.1 最开始获取过程

根据课堂上老师的讲解进行知识的学习

### 2.2 复习过程

把上课老师敲过的代码自己先照抄一遍，然后再自己盲打一遍，以求自己能够敲出能运行的代码的目的

### 2.3 学过什么

1. 一个项目的基本运行流程是什么样的：  
制作训练集 → 进行数据的训练 → 对结果进行评测
2. 对数据集的处理
  - a. 缺失值的处理方式
  - b. 交叉验证
  - c. 数据 EDA
  - d. 数据相关性分析
  - e. 数据可视化
3. 回归算法
  - a. 基础线性回归
  - b. Ridge 回归
  - c. Lasso 回归
  - d. Xgboost 训练

- e. Stacking 集成算法
- 4. 算法的手动实现
  - a. 如何制作训练集（加噪声）
  - b. 线性回归的基本模型框架实现
  - c. 决策树的手动实现
  - d. 评估模型的实现
  - e. 带 L1、L2 正则化的线性回归
  - f. 通过自定义的算法进行基本的回归
- 5. 使用 sklearn 进行房价预测模型的实现
  - a. sklearn 对机器学习模型各个阶段的接口实现
  - b.

## 2.4 怎么把项目进行下来

1. 了解整个项目的目的
2. 根据目的进行数据集结构的分析，从中选择出有意义的数据集和特征
3. 进行数据的清洗，对缺失值的数据进行填充，对无用数据进行过滤
4. 根据分析出的结果选择合适的训练模型与算法
5. 对数据进行归一化处理，使数据更加容易学习出结果
6. 通过处理后的数据划分训练集与测试集，传入不同的超参，选择出最优的超参作为模型的超参，完成训练
7. 把训练出的模型运行真实数据，看训练的结果如何

# 3 开发工作总结

## 3.1 在开发过程中学到什么

1. 一个机器学习项目整个流程是什么样的
2. 数据的处理与清洗非常重要
3. 数据可视化的方法以及视图的各种表示法
4. 基本常用回归算法的手动实现以及基本的原理
5. 超参是什么以及怎么调参

## 3.2 希望在未来的开发过程中学到什么

1. 如何通过房价预测项目来举一反三，把房价预测项目的经验切实转换成实际开发中的经验
2. 当面临真实数据时，如何选择处理方式，才能使筛选出的数据质量更高
3. 数据的量级如何选择，如何根据不同的业务需求选择数据的数量和方式
4. 不同的项目如何选择训练模型
5. 调参最好能够总结出一个相对比较能用的方式，有一个准则可以遵守

### 3.3 开发学习过程中优点和缺点

### 3.4 还希望开发过程中在哪方便做更多尝试

1. 能够手动实现更多的机器学习算法模型，了解各种算法的运行原理
2. 能够学习更加实际的项目，用的数据最好是比较真实的数据，能够在实际应用中进行检测训练模型的优劣

## 4 经验和教训

### 4.1 项目中习得经验

1. 了解原理后再敲代码感觉非常清晰
2. 看再多的原理不如自己敲代码来的实际，来的深刻

### 4.2 项目中习得教训

## 5 建议和展望

### 5.1 对于今后项目开发工作的建议

### 5.2 其他建议

### 5.3 展望