# 学生实验报告

|  |  |  |  |
| --- | --- | --- | --- |
| 学号 | 1120182525 | 学院 | 徐特立学院 |
| 姓名 | 梁瑛平 | 专业 | 计算机科学与技术 |

## 线性回归

## 实验简介

本实验采用线性回归，对波士顿房价数据集进行回归及预测，其目的是提升学生应用线性回归的实际解决问题能力。

## 实验目的

（1）帮助学生理解线性回归在数据科学中的应用。

（2）帮助学生掌握线性回归的原理。

（3）帮助学生掌握利用sklearn实现线性回归的方法。

## 相关理论与知识点

（1）线性回归的原理

（2）平均绝对误差，均分误差的理解

（3）sklearn中线性回归函数的应用

## 实验条件与环境

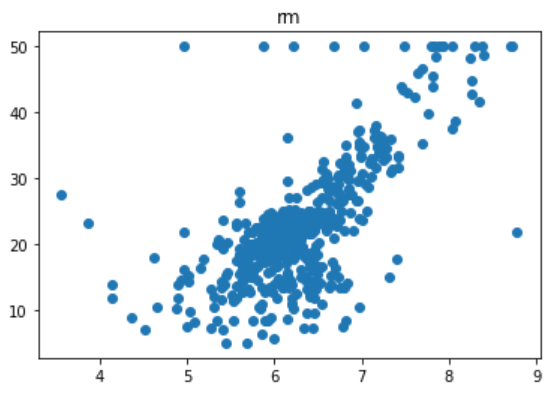
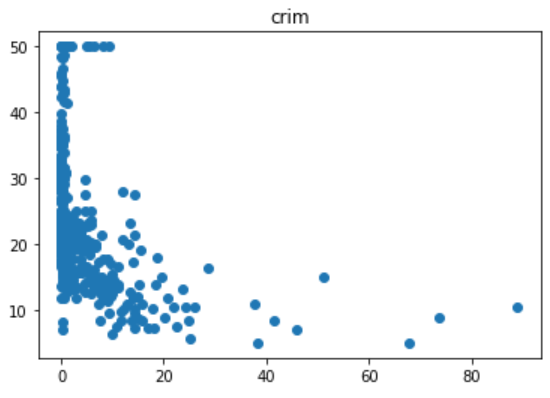
|  |  |  |  |
| --- | --- | --- | --- |
| 要求 | 名称 | 版本要求 | 备注 |
| **编程语言** | python | 3.6以上 |  |
| **开发环境** | Jupter | 无要求 |  |
| **第三方工具包/库/插件** | sklearn | 0.23.1 |  |
| **第三方工具包/库/插件** | Pandas | 1.0.5 |  |
| **其他工具** | 无 | 无要求 |  |
| **硬件环境** | 台式机、笔记本均可 | 无要求 |  |

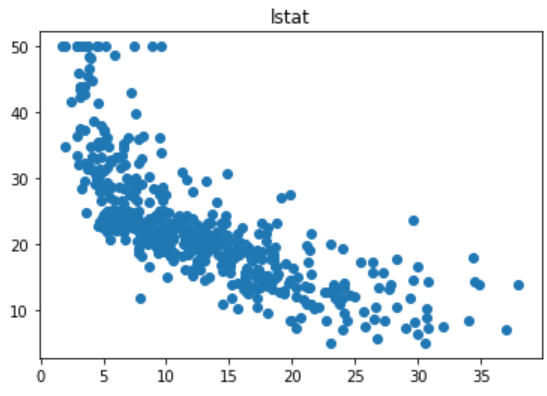
## 实验任务

|  |  |  |
| --- | --- | --- |
| **序号** | **任务名称** | **任务具体要求** |
| **1** | 数据理解 | 理解数据集背景以及数据含义。 |
| **2** | 数据读入 | 使用pandas 读入数据并输出读入的数据 |
| **3** | 定义特征值，目标值 | 使用'crim', 'rm', 'lstat'作为特征值，'medv'为目标值，输出特征值的描述性统计 |
| **4** | 区分训练集和测试集 | split\_num = int(len(features)\*0.7) |
| **5** | 线性回归 | 利用sklearn的LinearRegression()函数进行线性回归，输出模型的回归方程系数及方程的截距 |
| **6** | 预测 | 对数据集进行预测，输出预测值 |
| **7** | 获取平均绝对误差 | 求取预测值和真实值的mae |
| **8** | 获取均分误差 | 求取预测值和真实值的mse |

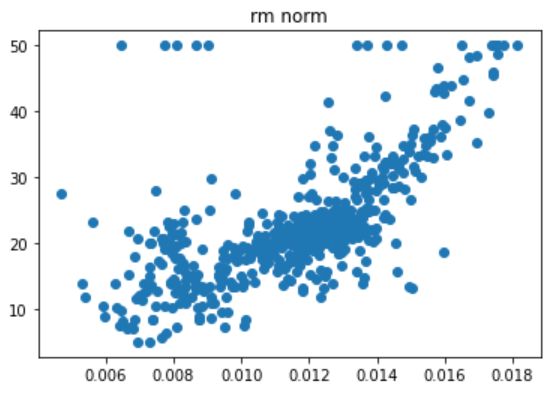
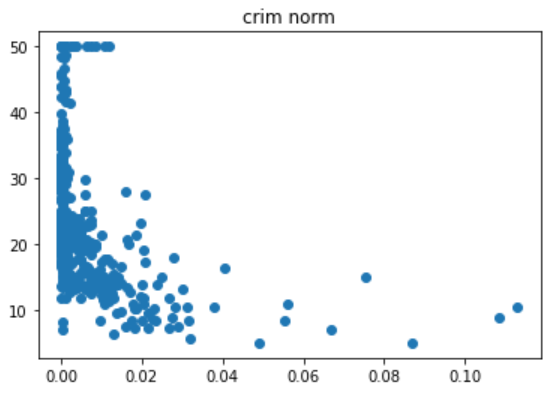
## 实验结果及分析

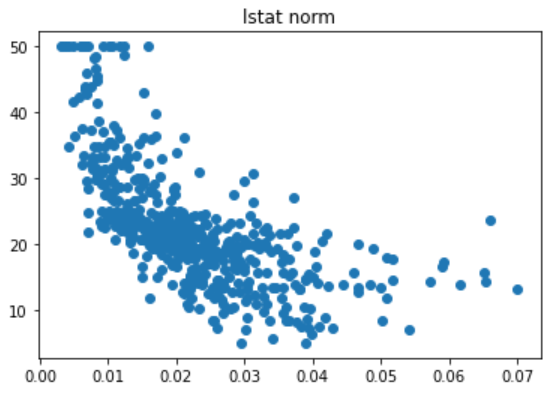
1. 相关性分析：



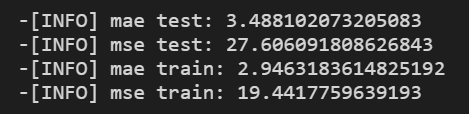


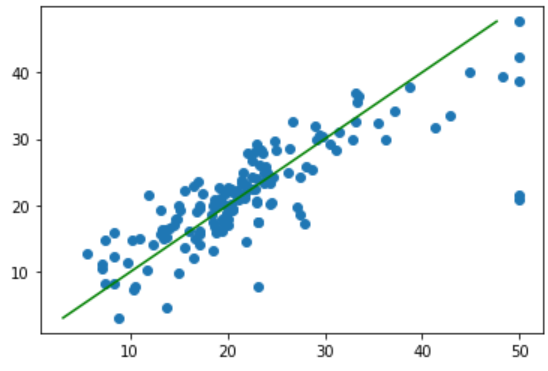
1. 归一化后相关性分析:





1. 实验结果：





## 收获与体会

通过这学期的学习，对机器学习有了更进一步的了解和认识。我觉得机器学习是一门具有挑战性的科学，想要学好这门课程必须要懂得计算机知识以及基本的算法认识。机器学习包括了十分广泛的科学，它由不同的领域组成，例如机器学习，计算机视觉等，总的来说，机器学习研究的主要目标是使机器能够胜任一些通常需要人类智能才能完成的复杂工作。虽然在此门课程中对算法的实现不能独立完成，但在一些简单的基本的算法上还是有一定的理解和认识。我也在此次课程设计的过程中不断的学习，反复的调式和思考问题，终于在我的坚持下能够很好地理解线性回归算法转换为实际代码的过程，也对算法有了更加清晰的思路。因此，我更加确信在自己的不断努力下总是会有收获的，只有坚持下去，才有成功的希望。在之后的学习中我会积极主动的查漏补缺，尽快完善自己欠缺的部分

## 备注及其他

无