

**实 验 报 告**

**（ 2024 / 2025 学年 第 一 学期）**

|  |  |
| --- | --- |
| 课程名称 | Python程序设计（混合式）（JS170101S） |
| 实验名称 | 实验二：线性回归与逻辑斯蒂分类实验 |
| 实验时间 |  2024 | 年 |  12  | 月 |  13  | 日 |
| 指导单位 |  计算机学院、软件学院、网络空间安全学院 |
| 指导教师 |  张伯雷  |

|  |  |  |  |
| --- | --- | --- | --- |
| 学生姓名 |   | 班级学号 |   |
| 学院(系) |   | 专 业 |   |

| **实 验 报 告** |
| --- |
| **实验名称** | **线性回归与逻辑斯蒂分类实验** | **指导教师** |  张伯雷  |
| **实验类型** | **验证** | **实验学时** | **2** | **实验时间** |  2学时  |
| 1. **实验目的和要求**
	1. 掌握线性回归的基本原理。
	2. 掌握逻辑斯蒂分类的基本原理。
	3. 学会使用Python语言实现线性回归算法。
	4. 学会使用Python语言实现逻辑斯蒂分类。
 |
| 二、**实验环境(实验设备)**硬件: 微型计算机软件: Windows /Linux/Mac操作系统、Python3.7+ |
| 1. **实验原理及内容**
2. 使用线性回归对房屋价格（数据集:题目1数据.txt）进行预测，需要使用前350行作为训练，351-414行作为测试数据，并对测试数据输出评价指标R方和残差平方和。
	1. 先用sklearn实现线性回归；
	2. 再用最小二乘或梯度下降法实现（参考书上代码）。

选用随机种子42：np.random.seed(42)提示：先把数据复制到文本文件，保存为data.txt，在用np.loadtxt()函数进行读取，示例代码如下import numpy as npa = np.loadtxt("data.txt") 注意data1.xlsx中数据的最后一列是y值，其它列是属性。1. 编写代码，使用Logistic Regression对训练数据(题目2训练数据.txt)进行模型训练，并针对测试数据(题目2测试数据.txt)输出二分类结果，结果用0-1之间的概率表示，用空格进行分割，分别计算在训练集和验证集（训练数据后20%）的AUC值。
* 数据集说明

该数据集需要通过特征数据分析，来预测二分类的结果。训练数据集中的二分类结果已经给出，为最后一列* 数据集信息

本数据集用于估计信用卡的违约概率。其中训练数据包括27997个样本，测试数据包括2001个样本。需要根据训练数据，建立Logistic Regression模型，对测试数据集中的数据预测结果。结果文件为一个txt文件，其中包含2000个0-1之间的实数值，保留两位小数，用空格分割。最后会在真实的结果上比较每个同学的预测结果。例如：0.01 0.24 0.55 0.67* 数据属性

本数据集供包括23个特征，具体说明如下：- 第1列: 信用卡的额度- 第2列: 性别 (1 = male; 2 = female).- 第3列: 教育 (1 = graduate school; 2 = university; 3 = high school; 4 = others).- 第4列: 婚姻状态 (1 = married; 2 = single; 3 = others).- 第5列: 年龄.- 第6 - 11列: 过去几个月的还款状态（4月到9月），-1，-2代表按时还款，其他代表延迟还款的月份数- 第12 - 17列: 过去几个月的账单记录- 第 18 – 23 列: 过去几个月的支付记录- 最后一列为类别y值，取值为0或1要求：（1）使用Sklearn实现（2）自定义编码使用梯度下降法实现（选做） |
| **四、实验小结**（包括问题和解决方法、心得体会、意见与建议、实验出错信息及解决方案等）(一)实验中遇到的主要问题及解决方法（二）实验心得（三）意见与建议（没有可省略）  |

|  |
| --- |
| 1. **支撑毕业要求指标点**
 |
| **六、指导教师评语** |
| **评 分 细 则** | **评分项** | **优秀** | **良好** | **中等** | **合格** | **不合格** |
| **遵守实验室规章制度** |  |  |  |  |  |
| **学习态度** |  |  |  |  |  |
| **算法思想准备情况** |  |  |  |  |  |
| **程序设计能力** |  |  |  |  |  |
| **解决问题能力** |  |  |  |  |  |
| **课题功能实现情况** |  |  |  |  |  |
| **算法设计合理性** |  |  |  |  |  |
| **算法效能评价** |  |  |  |  |  |
| **回答问题准确度** |  |  |  |  |  |
| **报告书写认真程度** |  |  |  |  |  |
| **内容详实程度** |  |  |  |  |  |
| **文字表达熟练程度** |  |  |  |  |  |
| **其它评价意见** |  |
| **本次实验能力达成评价（总成绩）** |  | **批阅人** |  | **日期** |  |