

实验 5: 消费金融风控实验

一、实验目的

设计风控模型在金融数据集上进行实验并评估结果

二、实验步骤

1. Python 环境安装

建议安装 python3.7

1. 安装原生 Python 环境

直接在官网下载对应平台 (Linux、Windows、Mac OS) 最新版本的 python, 参考文档和教程安装。

Python 官网: <https://www.python.org>

Python 文档地址: <https://docs.python.org/3/>

2. 安装 Anaconda (推荐!)

Anaconda 提供了很多基础依赖库, 并支持自己创建多个不同虚拟环境, 且通过 `conda install` 命令能自动安装相关依赖库, 十分方便。

Anaconda 下载地址 <https://www.anaconda.com/distribution/> 下载后直接根据提示安装即可。

检验安装成功: 在命令行输入 `python --version` 能够看到安装的对版本 python。
或者输入 `conda` 可以看到 Anaconda 相应的提示即可。

2. 设计风控模型

1. 根据课程讲述内容设计风控模型

a) 安装 sklearn 学习使用 LR, GBDT 模型

Sklearn 资料:

<https://morvanzhou.github.io/tutorials/machine-learning/sklearn/1-2-install/>

b) 安装 lightgbm 学习使用集成树模型

Lightgbm 资料:

<https://zhuanlan.zhihu.com/p/99069186>

<https://www.cnblogs.com/jiangxinyang/p/9337094.html>

c) 根据数据集进行特征工程, 改进模型, 提升风控效果

2. 在数据集上运行结果

a) 给出自己划分的训练集和验证集的 auc 结果, 截图在实验报告中

b) 给出测试集的预测结果, csv文件保存

c) Bonus

- 手写 auc 实现, 对比sklearn 结果差异

- 手写逻辑回归, 对比sklearn 结果差异

- 结合是否归一化在 LR 和树模型的差异, 阅读相关资料分析什么情况下需要对特征做归一化并尝试给出原因

三、提交内容

1. 实验报告: 学号_姓名_lab5.pdf

- 1) 特征工程过程(如归一化、WOE 编码、均值填充、交叉特征衍生等)
- 2) 不同模型结果
- 3) 特征重要性分析 (matplotlib)
- 4) 个人体会

2. 实验代码: code 文件夹

3. 包含预测结果的 csv 文件: 学号_pre.csv

4. 如果是组队完成的: 分工说明.pdf

实验 5 的文件打包命名为: 学号_姓名_lab5.zip

实验 1 的结果打包命名为: 学号_姓名_lab1.zip

上传到[学在浙大](#)