**

**HUNAN UNIVERSITY**

《DM&BID》

Report 11

|  |  |
| --- | --- |
| **报告名称：** | AdaBoost和GBDT模型 |
| **学生姓名：** | 杨超然 |
| **学生学号：** | 202106060220 |
| **专业班级：** | 电商2102班 |
| **学 院：** | 工商管理学院 |
| **指导老师：** | 江资斌 |
| **日 期：** | 2023.4.10 |

目录

[一、AdaBoost模型代码实现 3](#_Toc1801552879)

[（一）AdaBoost分类模型 3](#_Toc1327298300)

[（二） AdaBoost回归模型 3](#_Toc1989886711)

[二、 AdaBoost案例：信用卡精准营销模型 4](#_Toc1263117046)

[（一） 案例背景 4](#_Toc1332341527)

[（二） 模型搭建 4](#_Toc852057020)

[1.1读取数据 4](#_Toc1101376944)

[1.2提取特征变量和目标变量 4](#_Toc1680744315)

[1.3划分训练集与测试集 4](#_Toc269809567)

[1.4模型训练及搭建 5](#_Toc1351413752)

[（三） 模型预测及评估 5](#_Toc1423879192)

[2.1数据预测 5](#_Toc1727301423)

[2.2数据对比 5](#_Toc1071076215)

[2.3查看预测准确度与分类概率 6](#_Toc1370016351)

[2.4绘制ROC曲线与查看AUC值 6](#_Toc545148123)

[2.5查看特征重要性 7](#_Toc1139265159)

[三、 GBDT算法代码实现 7](#_Toc665330661)

[（一） GBDT分类模型 7](#_Toc265069498)

[（二） GBDT回归模型 8](#_Toc1141969008)

[四、 GBDT案例：产品定价模型 8](#_Toc1011764217)

[（一） 模型背景 8](#_Toc945678173)

[（二）模型搭建 8](#_Toc486582164)

[1.1读取数据 8](#_Toc368702572)

[1.2分类型文本变量处理 9](#_Toc1293806009)

[（三）模型预测及评估 9](#_Toc1725667388)

[2.1预测测试集数据 9](#_Toc1525137381)

[2.2对比预测值与实际值 10](#_Toc619151875)

[2.3查看预测评分与特征重要性 10](#_Toc1527293410)

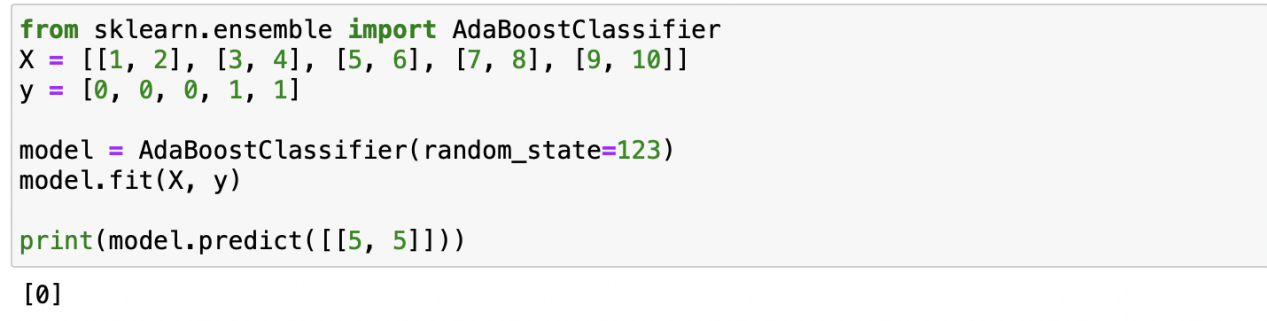
[五、 学习小记 11](#_Toc348309279)

[（一） AdaBoost与GBDT的算法原理 11](#_Toc2141114078)

# 一、AdaBoost模型代码实现

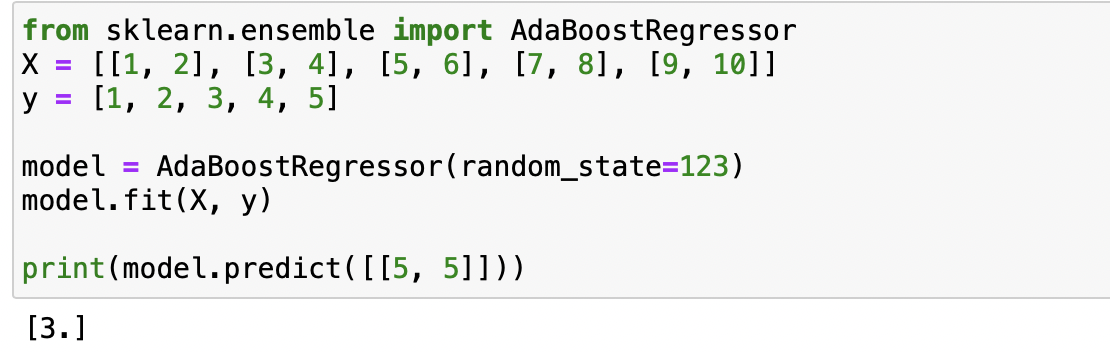
## （一）AdaBoost分类模型

导入sklearn.ensemble库，搭建分类模型并喂入相关数据。



## AdaBoost回归模型

定义AdaBoostRegressor模型，利用predict函数进行模型回归；其中random\_state=123表示在取样过程中喂给它随机参数种子。



# AdaBoost案例：信用卡精准营销模型

## 案例背景

当前经济增速下行，风控压力加大，各家商业银行纷纷投入更多资源拓展信

用卡业务，信用卡产业飞速发展。因为市场竞争激烈，信用卡产品同质化严

重，商业银行需要采用更快捷有效的方式扩大客户规模，实现精准营销，从

而降低成本提高效益，增强自身竞争力。该精准营销模型也可以应用其他领

域的精准营销，例如信托公司信托产品的精准营销等。

## 模型搭建

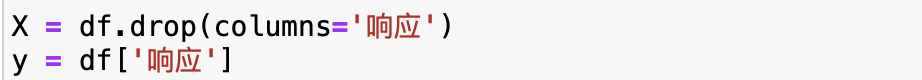
1.1读取数据

通过pandas库读取信用卡营销模型表格，通过head()函数展示前五行数据，可以看到年龄、月收入、月消费、响应等数据。



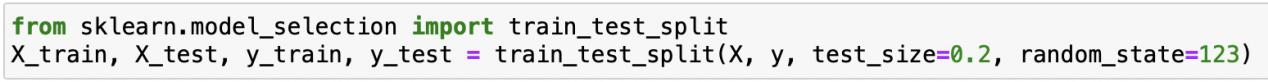
1.2提取特征变量和目标变量

将“响应”作为特征变量和目标变量，代码如下：

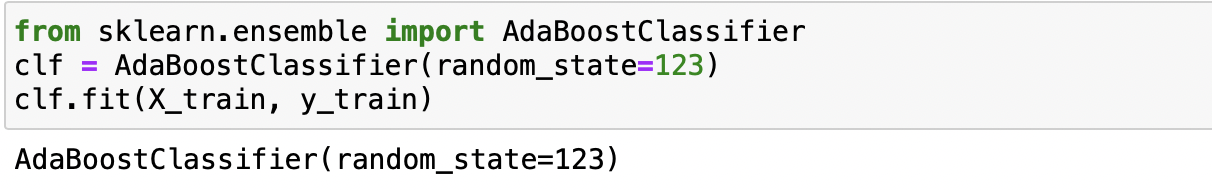


1.3划分训练集与测试集

从sklearn中引入train\_test\_split，进行训练集与测试集的划分。



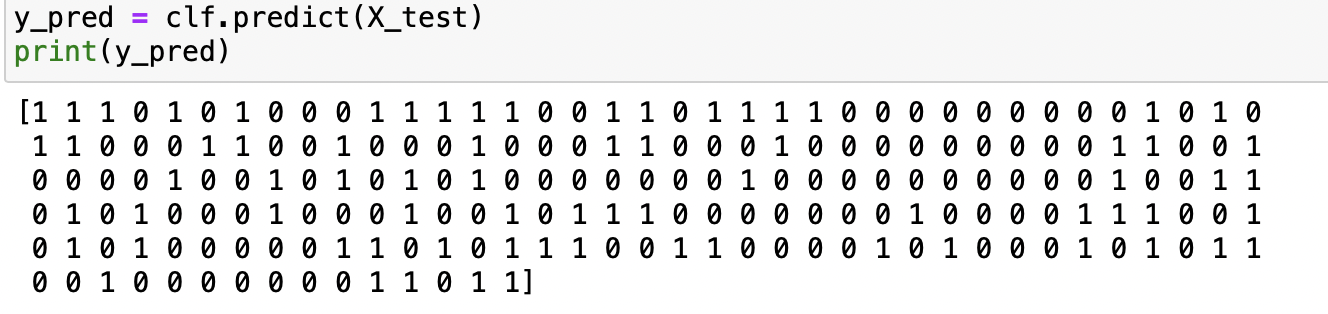
1.4模型训练及搭建



## 模型预测及评估

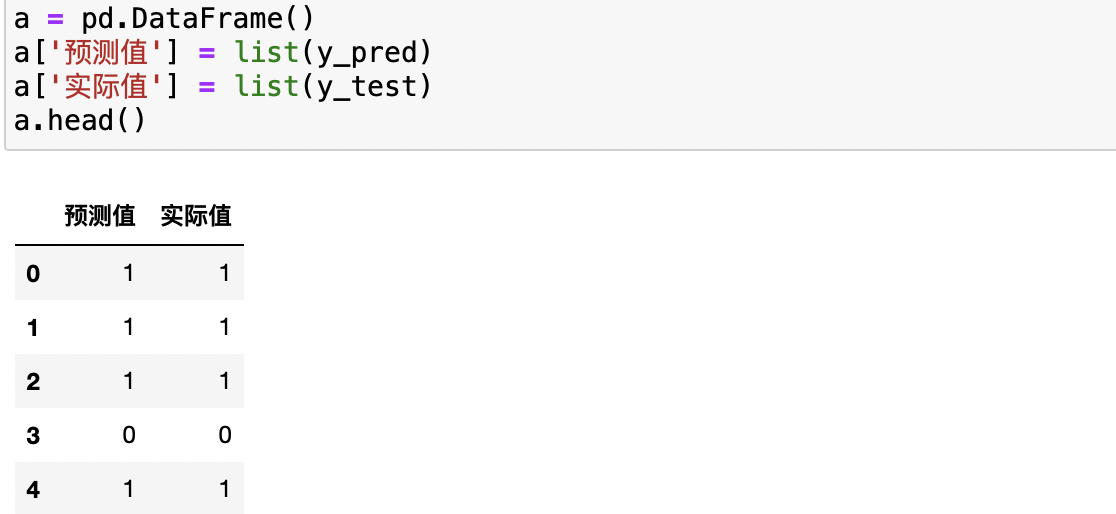
2.1数据预测

模型搭建完毕后，通过如下代码进行数据预测。

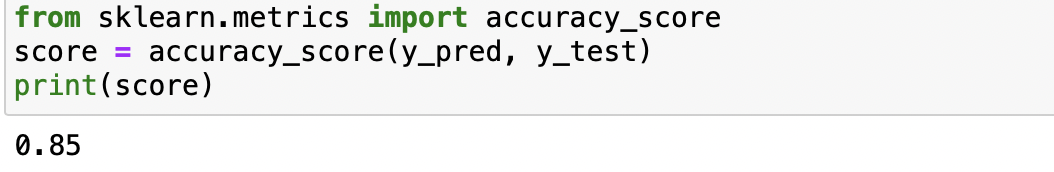


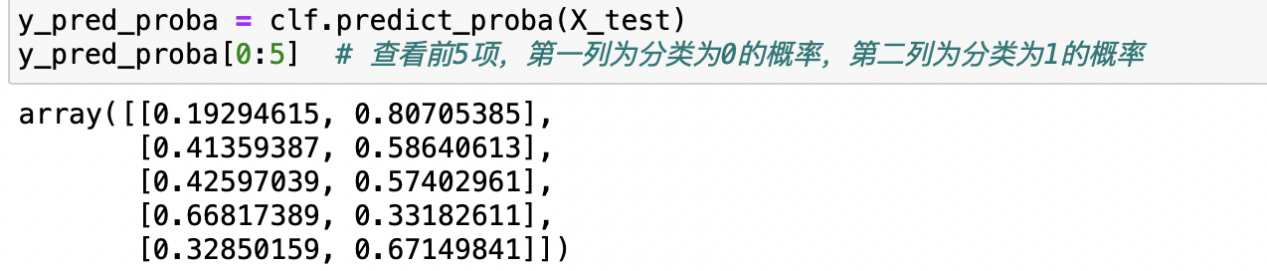
2.2数据对比

创建一个空的DataFrame，将预测值与实际值进行对比。



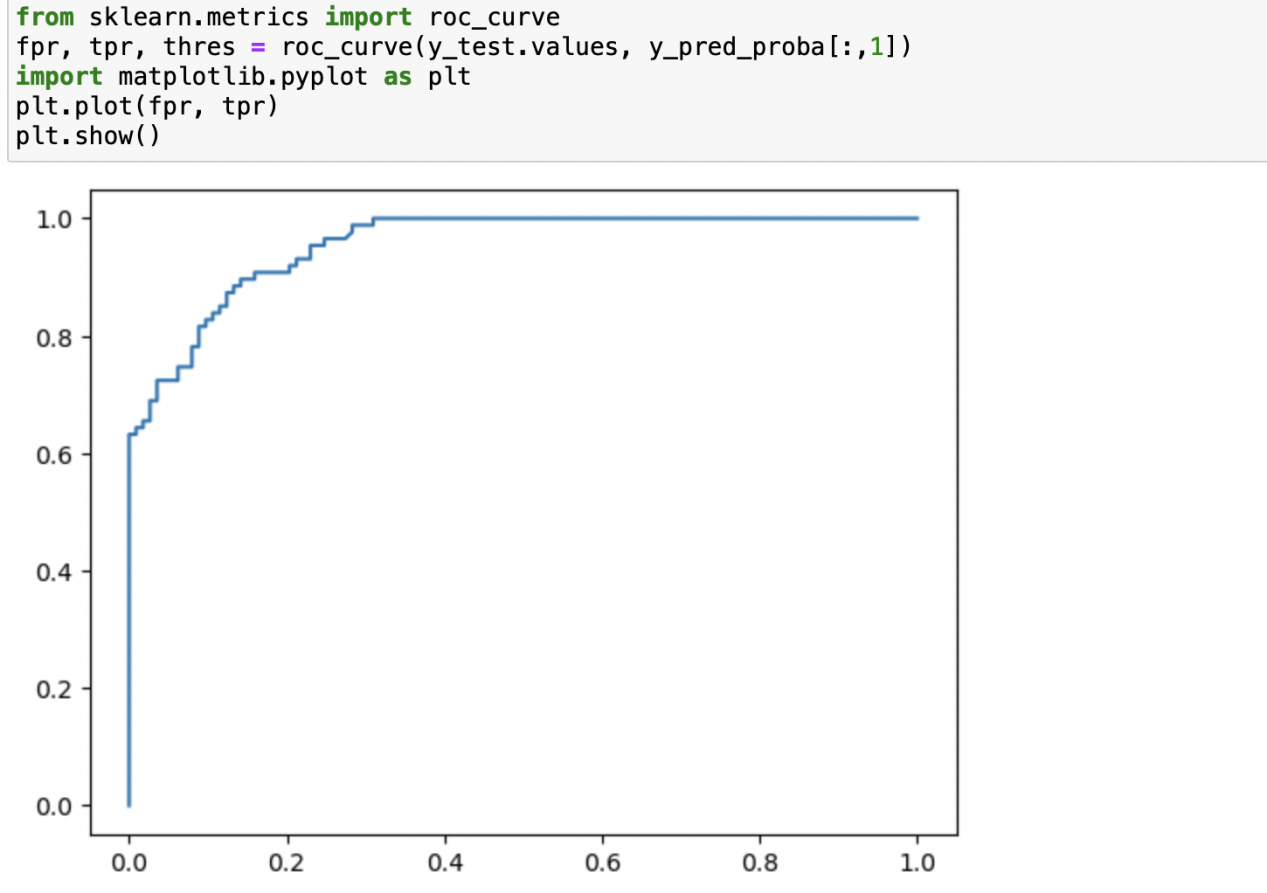
2.3查看预测准确度与分类概率





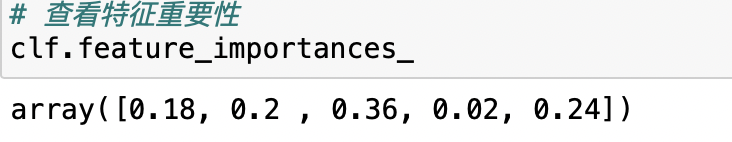
2.4绘制ROC曲线与查看AUC值

截屏2023-04-10 10.48.40

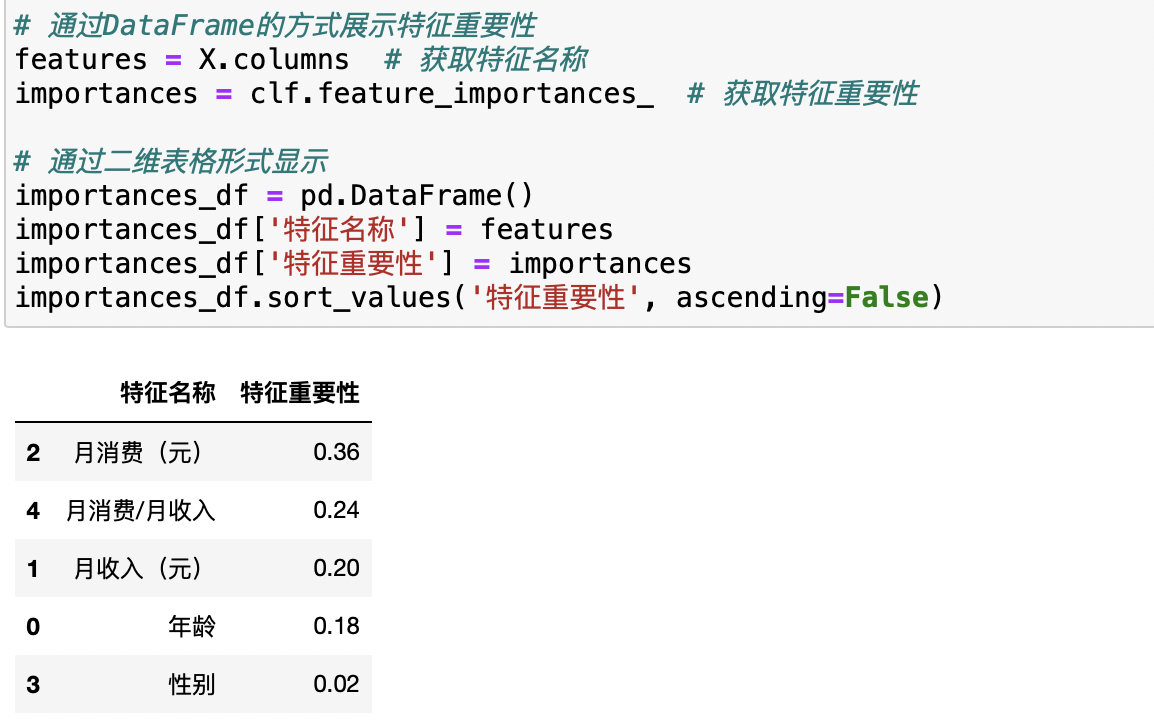




2.5查看特征重要性

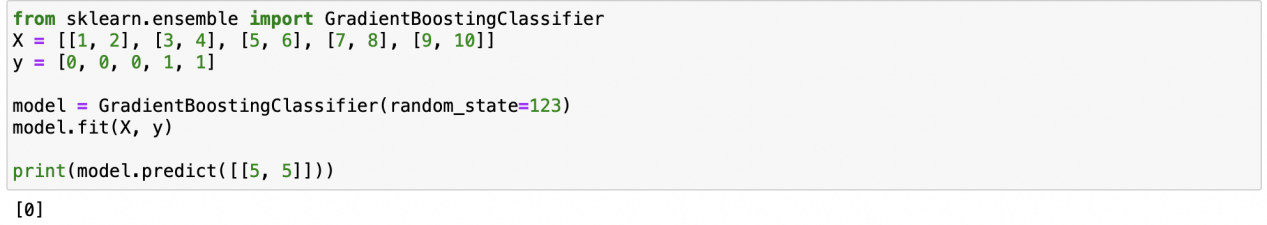


通过如下代码，以DataFrame方式显示特征重要性，更加直接美观。



# GBDT算法代码实现

## GBDT分类模型



## GBDT回归模型



# GBDT案例：产品定价模型

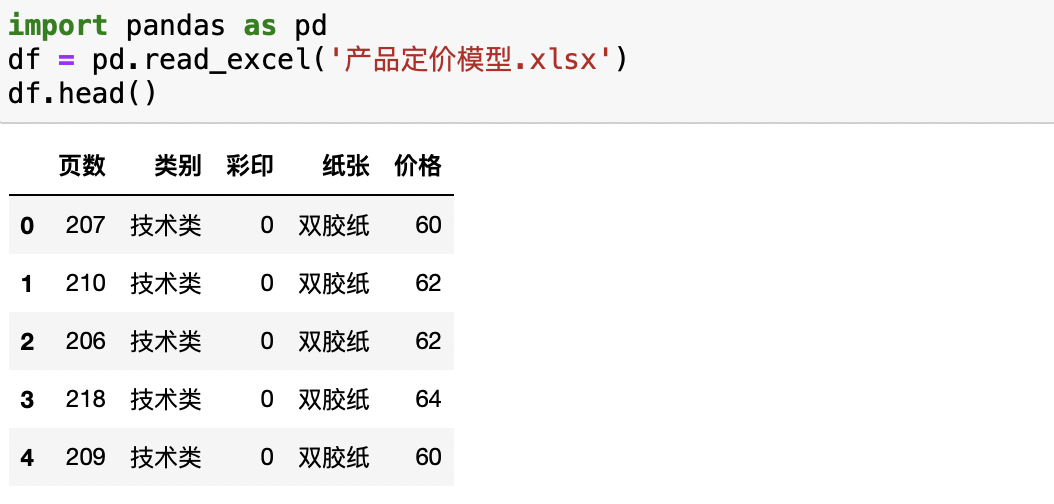
## 模型背景

出版社在对图书进行定价的时候会考虑图书的页数，纸张，类别，作者，内

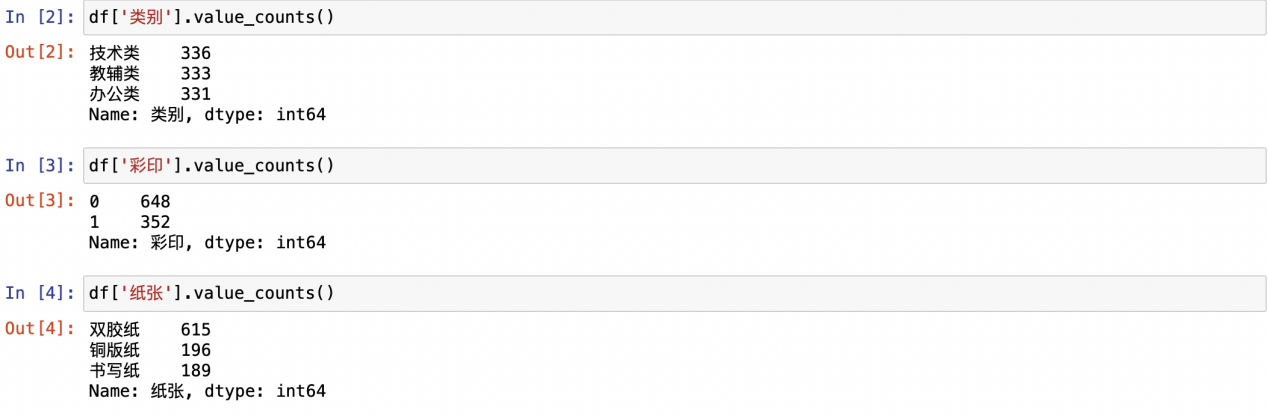
容，读者等各个方面。因为定价所需要考虑的因素较多，人工考虑起来较为繁琐，并且容易遗漏。如果我们能够建立一个模型综合考虑各方面因素并对图书进行定价，那么就能节约成本，提升效率，并契合买者需求，提高销售 量挖掘潜在利润。该产品定价模型也可以用于其他领域的产品定价模型，例 如金融产品的定价模型。

## （二）模型搭建

1.1读取数据



可通过如下代码查看各个分类（类别、彩印、纸张）的数据量：



1.2分类型文本变量处理

即将文本内容转换为数值，涉及LabelEncoder函数的知识，处将类别一列处理后，我们可以使用value\_counts()方法查看转化效果：



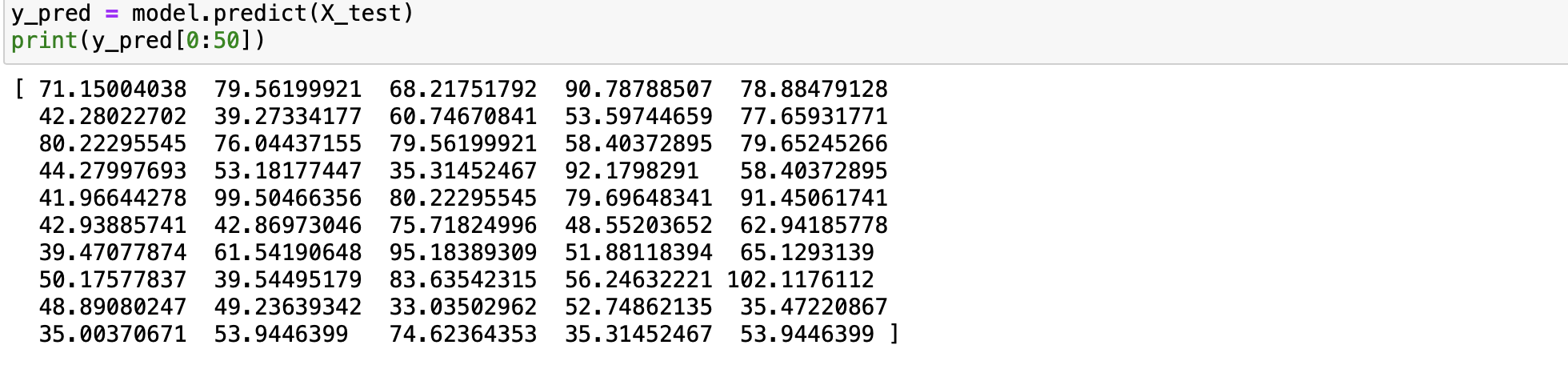
1.3变量提取、数据集划分与模型训练

由于这三步的代码本质相同，只有引用的库存在差异，故直接作为一个三级步骤放在报告中：



## （三）模型预测及评估

2.1预测测试集数据

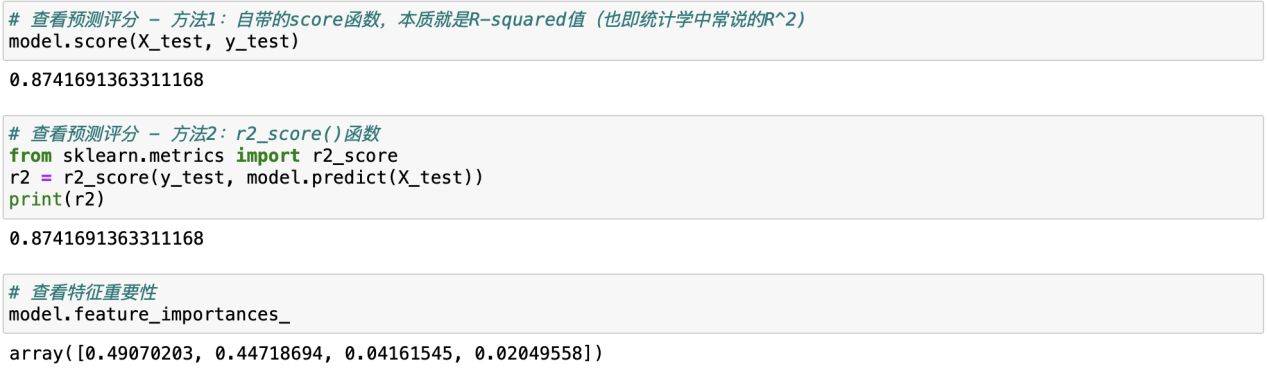


2.2对比预测值与实际值



2.3查看预测评分与特征重要性

查看预测评分有两种方法，一是用自带的score函数，本质是R-square值；二是使用r2\_score()函数



同样，也可以利用DataFrame方式查看特征的重要性

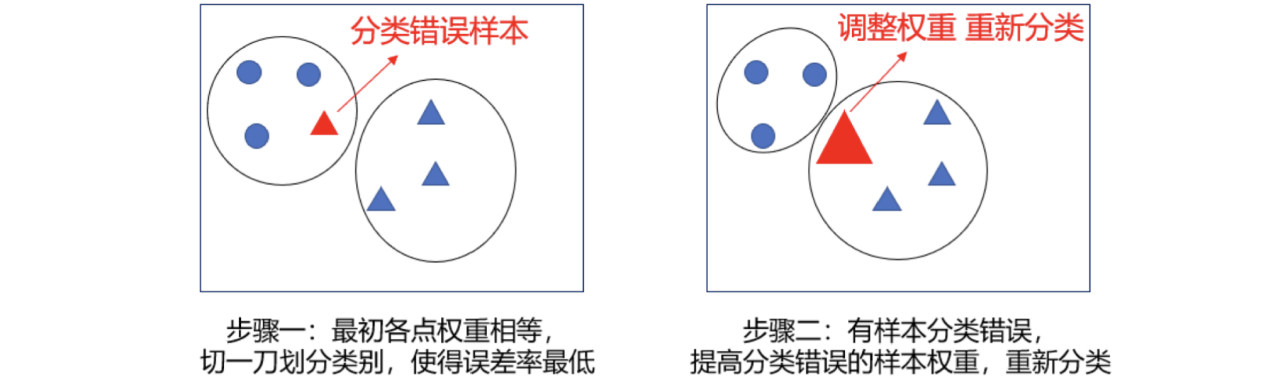


# 学习小记

## AdaBoost与GBDT的算法原理

两者一个为**调整权重**，一个为**拟合残差**

AdaBoost模型：即通过调整权重进行迭代，（分类错误则重点关注，增大权重）从而将弱学习器训练为强学习器。



GBDT算法：将损失函数的负梯度作为残差近似值，不断使用残差迭代与拟合回归树生成强学习器（最终的模型是多个模型集成在一起的结果，充分体现了算法集成思想）



