**市场分析03**

**锁定目标用户**

**实验报告**

****

**学号：201600301079**

**姓名：崔玉峰**

**班级：2016级**

## 问题描述

1. **问题题目**

本章基于4000多条用户对于变动后的产品响应情况的调查数据，运用定性与定量结合的研究方法，对用户的相应的数据变量与对产品订购情况进行分析。主要完成以下问题：

1. 筛选出与响应变量相关的某些变量，试确定其数学表达式；
2. 根据一个或几个变量的值，预测或控制另一个响应变量的取值，即建立模型，并且可以知道这种预测或控制能达到什么样的精确度；
3. 使用模型后对银行的营销的影响。
4. **解题思路**

**逻辑回归分类：**

**逻辑回归是一个通过线性回归模型的分类模型，线性回归结果通过**Sigmoid 函数激活之后可以处理二分类问题。

**虚拟变量**

对于自变量中的一些分型变量或者有序性变量的处理，设置虚拟变量。

对于本题目通过分析题目中的变量可以将，age jobtype education marital default balance housing loan 作为自变量，response作为响应变量。并且进行统计计算虚拟变量。喂给逻辑回归模型进行训练查看结果。

## 具体实现：

1. **加载数据**

根据所给的.csv数据，通过pandas库读进来

每个样本有8个属性，和一个响应变量。并且通过Counter类进行统计取值方便下面进行虚拟变量的设置。

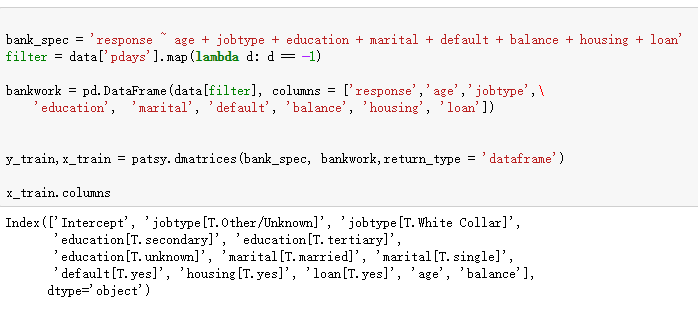


1. **分析样本数据，处理数据，设置虚拟变量**
2. Job将属性分为三类虚拟变量White Collar,Blue Collar, other
3. 将响应变量response映射成0 ，1取值方便逻辑回归训练。
4. 其他变量无需特别设置虚拟变量，只需将取值直接作为虚拟变量

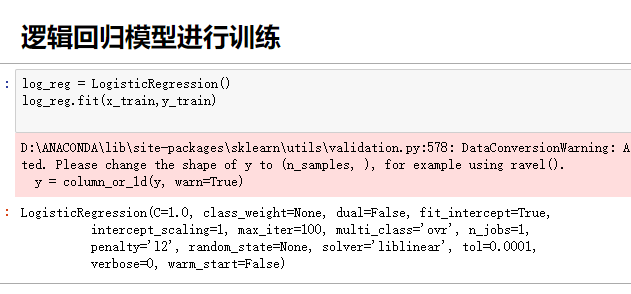


1. **将数据转换成训练数据集**

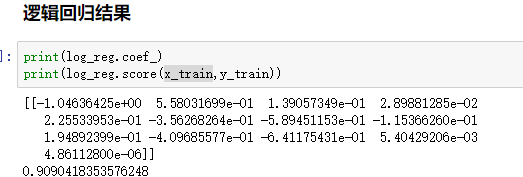
将每个变量的取值打开



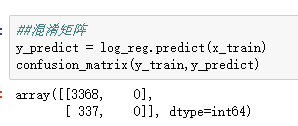
1. **建立逻辑回归模型，并得到回归系数**
2. 通过调用Scikit-learn 机器学习库的逻辑回归模块可以很容易的对样本进行回归分析：



1. 可以查看逻辑回归的系数，以及逻辑回归的准确度：



1. 调整阈值来调整分类结果：



## 结果分析：

通过逻辑回归拟合出的模型还是对数据有很好的预测作用的，可以根据这个模型预测银行用户的响应情况选择，并且这个模型很好储存，只有几个个参数可以很方便的进行使用。

## 实验结论和感悟：

本次实验的数据和代码会通过.ipynb的形式上传可以随时运行查看。

本次实验通过建立线性回归模型对样本数据进行了简单的训练，可以通过此模型对消费者的选择进行预测。线性回归模型是最简单的回归模型，但是功能很强大。通过实验感受到了数据科学中机器学习算法的应用。通过实验也学习到了虚拟变量在逻辑回归中的使用，对于取值不为数字的属性可以转换成虚拟变量进行处理。