**市场分析02**

**预测消费者选择**

**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300多条人们对交通工具选择的调查数据，运用定性与定量结合的研究方法，对交通工具的被选择现状及其影响因素进行了分析。主要完成以下问题：

1. 确定几个特定的变量之间是否存在相关关系，如果存在的话，找出它们之间合适的数学表达式；
2. 根据一个或几个变量的值，预测或控制另一个变量的取值，并且可以知道这种预测或控制能达到什么样的精确度；
3. 进行因素分析。例如在对于共同影响一个变量的许多变量（因素）之间，找出哪些是重要因素，哪些是次要因素，这些因素之间又有什么关系。
4. **解题思路**

建立线性回归模型，用cartime carcost traintime traincost 四个变量作为输入，choice作为输出建立线性预测模型。

**线性模型方程求解：**







可以通过数据拟合出线性模型。

**逻辑回归分类：**

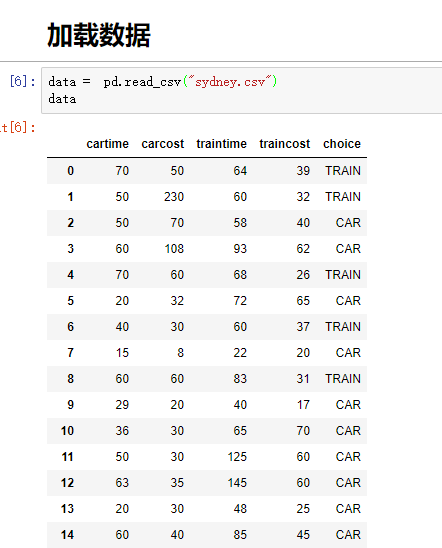
**逻辑回归是一个通过线性回归模型的分类模型，线性回归结果通过**Sigmoid 函数激活之后可以处理二分类问题。最终预测目标choice只有train和car两种选择，可以转化成0和1，因此响应变量是二值变量，所以我们可以用logit逻辑回归。

## 具体实现：

1. **加载数据**

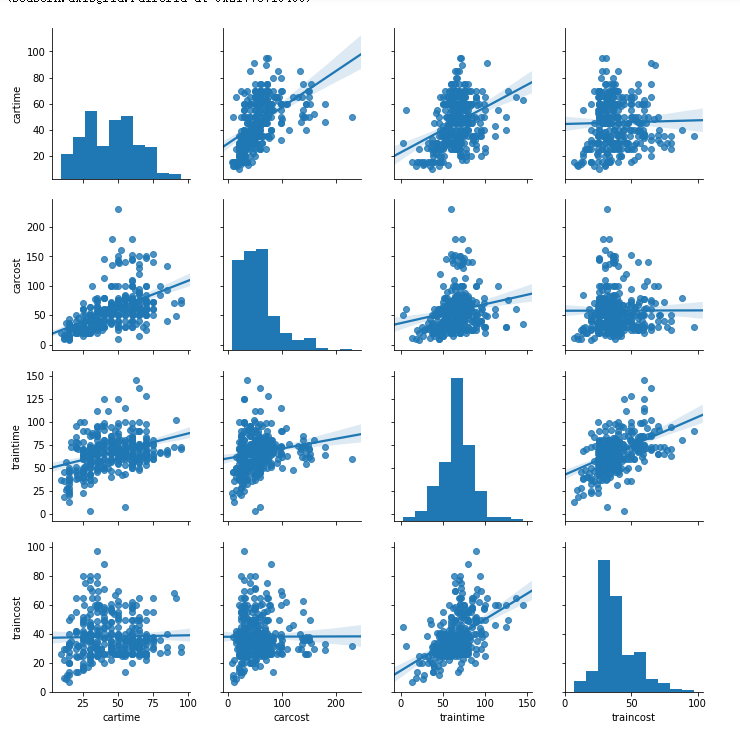
根据所给的.csv数据，通过pandas库读进来

每个样本有4个属性

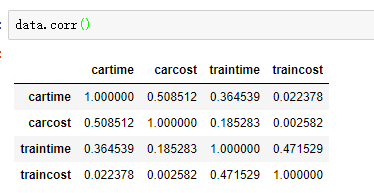


1. **分析样本数据，找到样本属性间关系**
2. **通过python的 seaborn库可以画出解释变量的散点图：**





1. **可以进而求得样本变量之间的相关系数：**



相关系数的的绝对值越接近1，表明两变量相关性越强；越接近于0，表明相关性越差。相关系数大于零代表正相关，相关系数小于零代表负相关

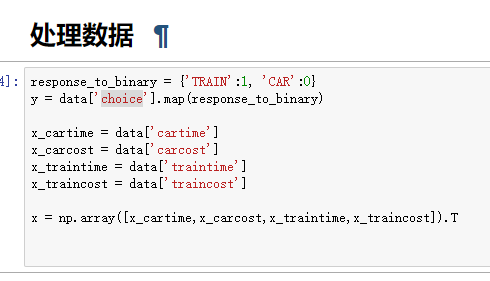
1. **建立逻辑回归模型，并得到回归系数**

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：

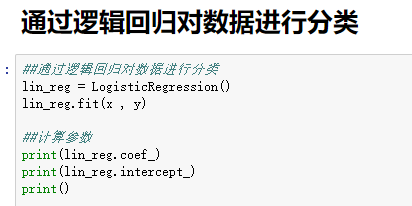
1. **将dateframe 的数据转换成numpy的矩阵数据，方便模型训练**

X\_train: cartime carcost traintime traincost

y\_train: choice



1. **将建立逻辑回归模型并训练**

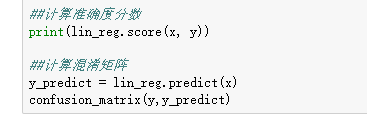


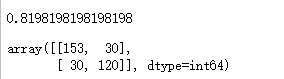
**求得回归系数，和截距：**



1. **将样本数据带回模型查看模型训练成果：**

计算预测准确度和分类混淆矩阵：





通过计算可以看到模型的分类准确度达到了0.82；可以确定建立线性模型还是可以很好的拟合样本数据的。

## 结果分析：

通过线性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消费者对于交通工具的选择，并且这个模型很好储存，只有五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。