**实验题目**：特征处理

**实验目的**：掌握特征处理的基本实现方法

**实验题目**：我们提供了在双11促销期间一组商人他们相应增长的买家。你的任务是预测未来这些新买家是否会成为这些商家的忠诚客户。换句话说，您需要预测这些新购买者在6个月内再次从同一商家购买商品的可能性。

**实验步骤**：

1. 数据清洗

检测异常样本：根据对具体业务的理解和认识去除一些异常极端的数据

缺省值处理：采用均值，众数，中位数等填充办法。

1. 数据采样
2. 将各种类型的值进行特征分类
3. 特征的处理与分析：主要采用离散化的方式
4. 特征选择：特征选择的目标是寻找最优特征子集。特征选择能剔除不相关或冗余的特征，从而达到减少特征个数，提高模型精确度，减少运行时间的目的。另一方面，选取出真正相关的特征简化模型，协助理解数据产生的过程。
5. 模型评估

**个人心得：**

1.添加xlrd和xlwt模块。Pip install xlrd;pip install xlwt

2.应先将**csv**格式的数据文件转换为**xls**格式的数据文件，xlrd和xlwt打不开csv文件

3.学会如何使用xlrd和xlwt读写xls文件及其基本用法，xlrd是读xls文件，xlwt是写xls文件。

4.添加sklearn模块。Pip install sklearn.

5.学会sklearn的基本使用方法

6.将年龄转化为特征值，并将特征值进行离散化

7.将性别转化为特征值，并将特征值进行离散化

8.分割数据文件中的Log数据，并进行处理

9. 采用均值，众数，中位数等填充办法。进行缺省值处理

9.调试python代码，使其能够成功在机器上运行

**实验总结：**通过本次实验，我掌握了python中xlrd，xlwt，sklearn模块的基本用法，能够使用python语言实现机器学习的第一步数据的预处理，能够将实验数据进行初步特征处理，

为下一步的实现打下了坚实的基础。通过调试python代码，已经能够是python代码在自己的计算机上运行，并能输出正确的结果。为下一次实验打下了坚实的基础。

**源代码：  
import** numpy **as** np  
**import** xlrd *#引入xlrd模块***import** xlwt *#引入xlwd模块  
  
#data = xlrd.open\_workbook('*