# 作业三 分类与聚类

**姓名：李晓波**

**学号：2620170054**

### 一、实验环境及语言

**语言及环境依赖**

语言： python

依赖的包：numpy,pandas,matplotlib,sklearn,math

### 二、实现要求

1. 对使用分类模型（至少2个）对数据集进行挖掘；
2. 对挖掘结果进行可视化，并解释其意义；
3. 使用聚类方法（至少2种）对数据集进行分析；
4. 对挖掘结果进行可视化，并解释其意义。

### 三、实现方法

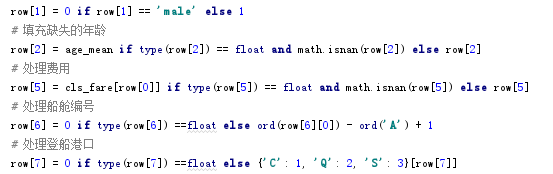
### 数据集：Titanic: Machine Learning from Disaster

#### 3.1 数据预处理

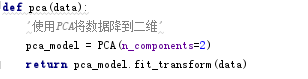
在此数据集中，根据属性分为：



然后进行缺失处理：



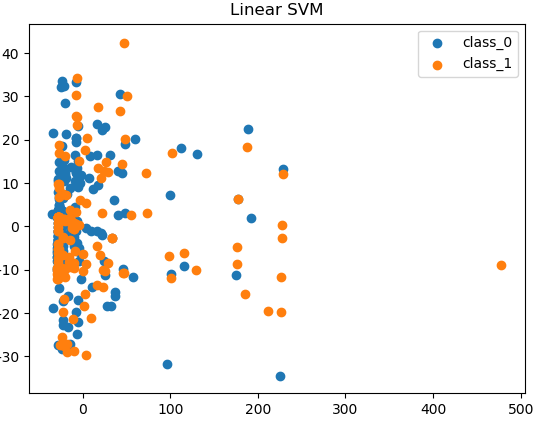
在利用sklearn库中的PCA进行降维操作



#### 3.2 分类

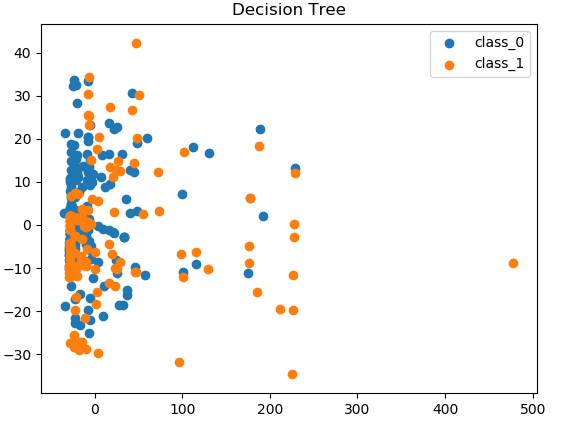
##### 3.2.1 SVM

将处理完的数据svm库中的svc()函数进行分类，得出的结果保存为svm.csv,可视化利用matplotlib进行可视化，其图像为：



##### 3.2.2决策树

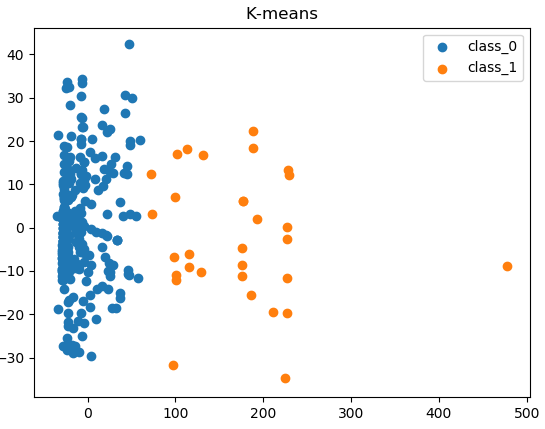
将处理完的数据tree库中的DecisionTreeClassifier()函数进行分类，得出的结果保存为decision\_tree.csv,可视化利用matplotlib进行可视化，其图像为：



#### 3.3 聚类

##### 3.3.1 K-means

将处理完的数据cluster库中的Kmeans()函数进行分类，得出的结果保存为k\_means.csv,可视化利用matplotlib进行可视化，其图像为：



##### 3.3.2 hierarchical

将处理完的数据cluster库中的AgglomerativeClustering()函数进行分类，得出的结果保存为hierarchical.csv,可视化利用matplotlib进行可视化，其图像为：

