

《大数据应用》课程报告

选题名称：基于Orange3的数据挖掘与数据分析——以某kaggle竞赛为例

学 院： 计算机科学与通信工程学院

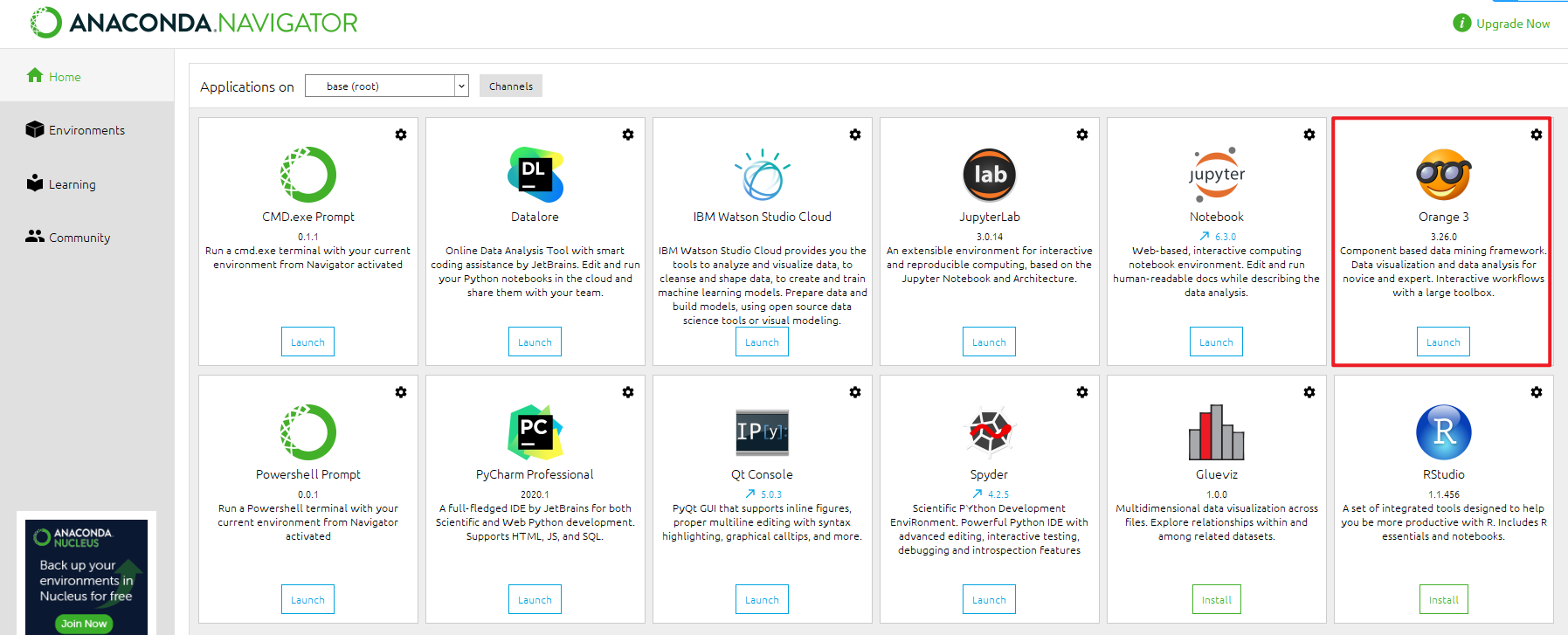
班 级： 智能科技1901

姓 名： 3190613024 周楷洋

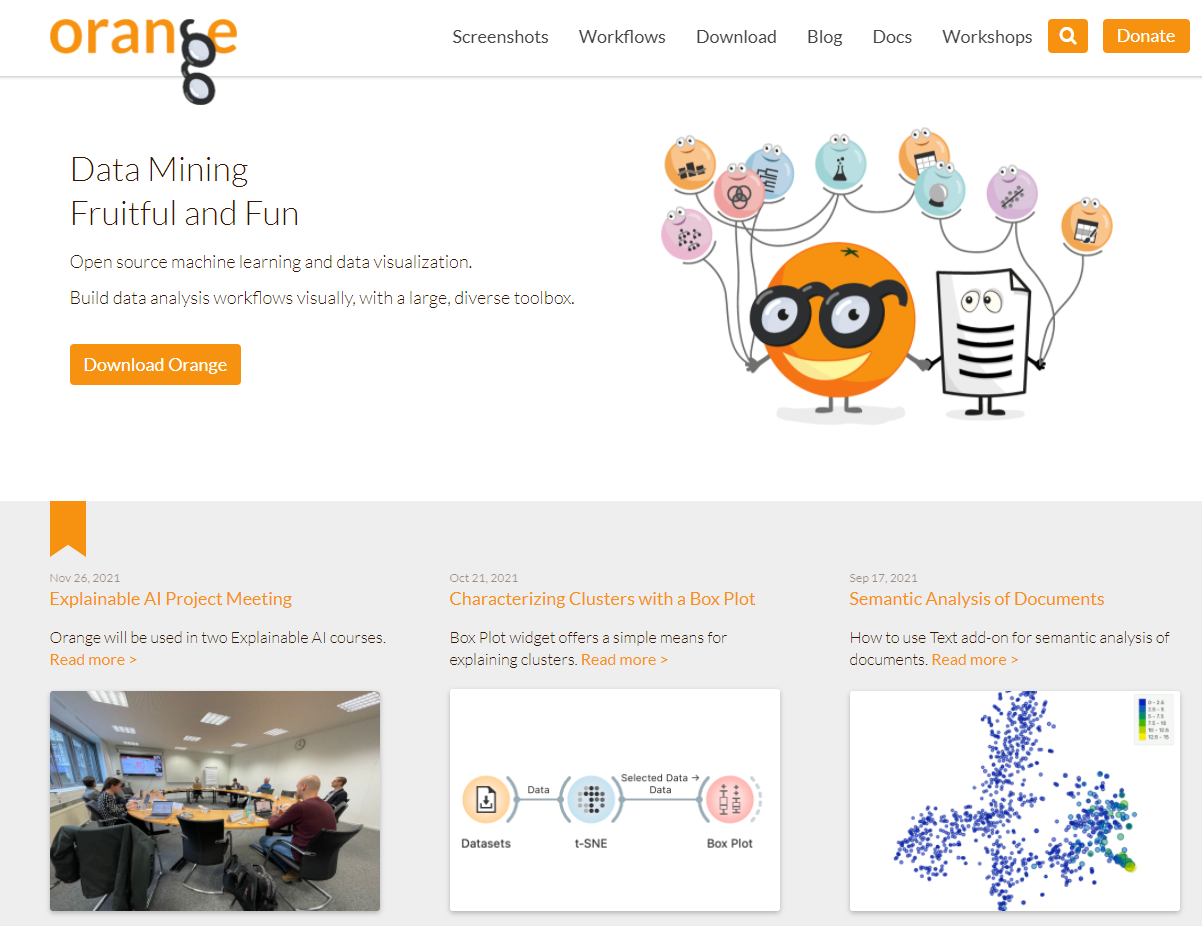
指导老师： 聂 文 惠

1. **orange3的基本用法**

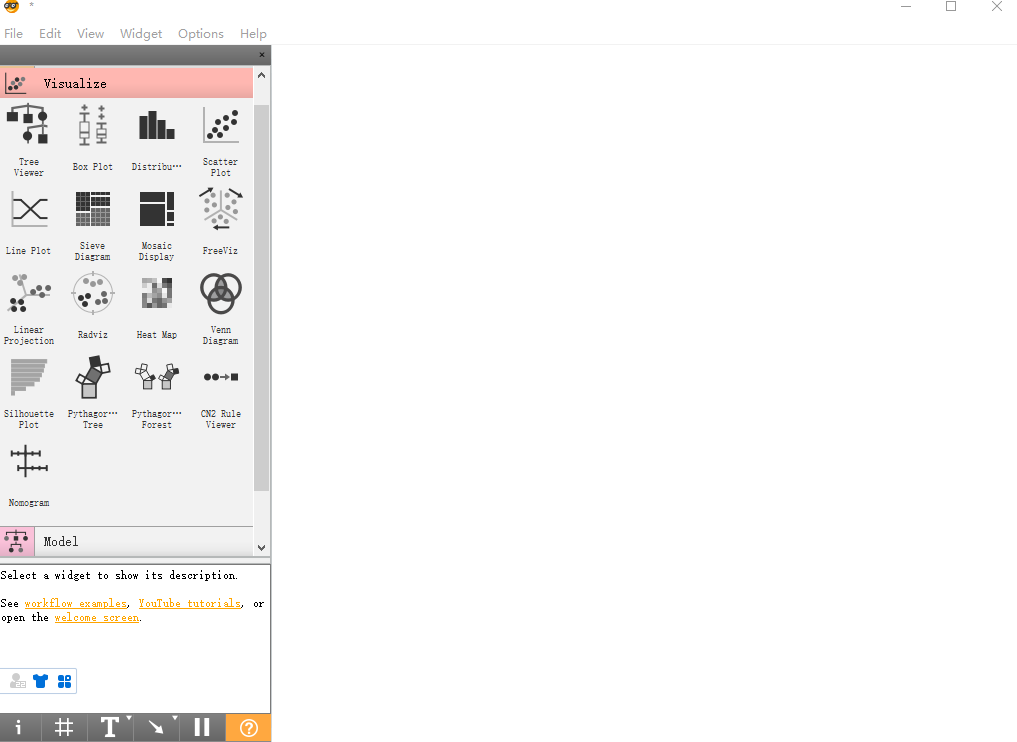
Anaconda 中集成了Orange3



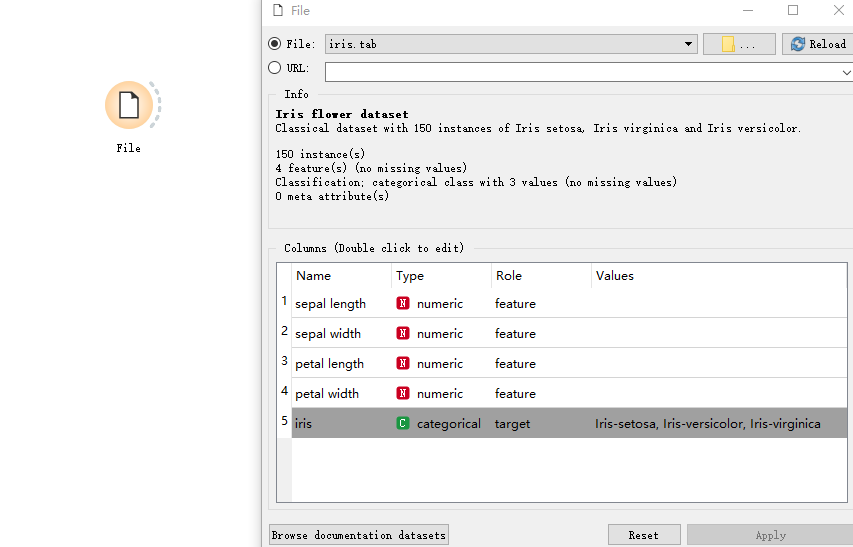
也可以通过官网下载安装



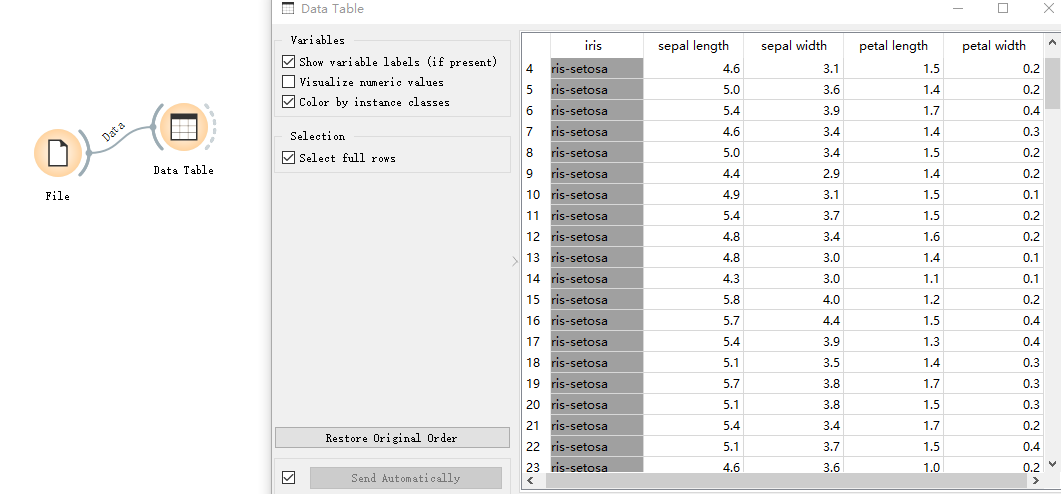
初始页面



导入数据



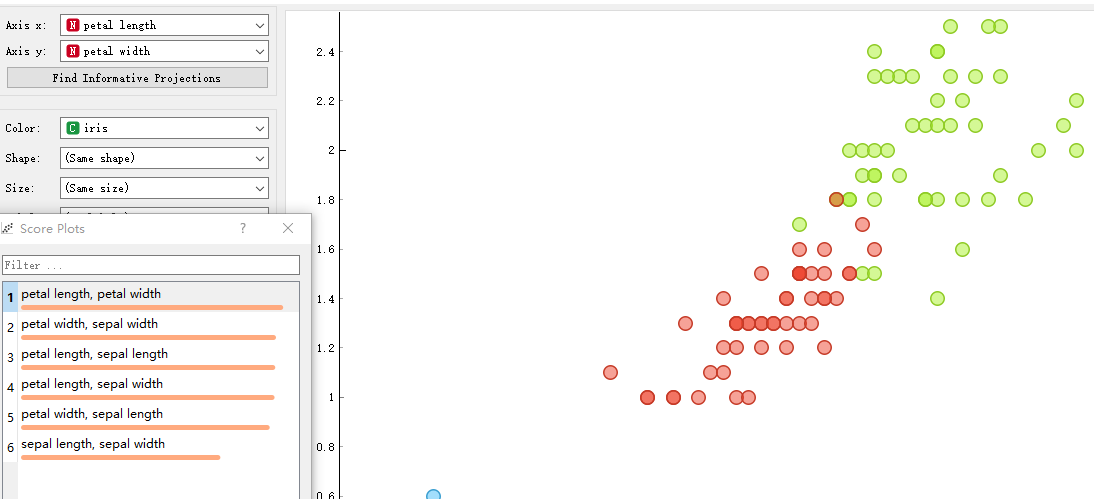
数据显示



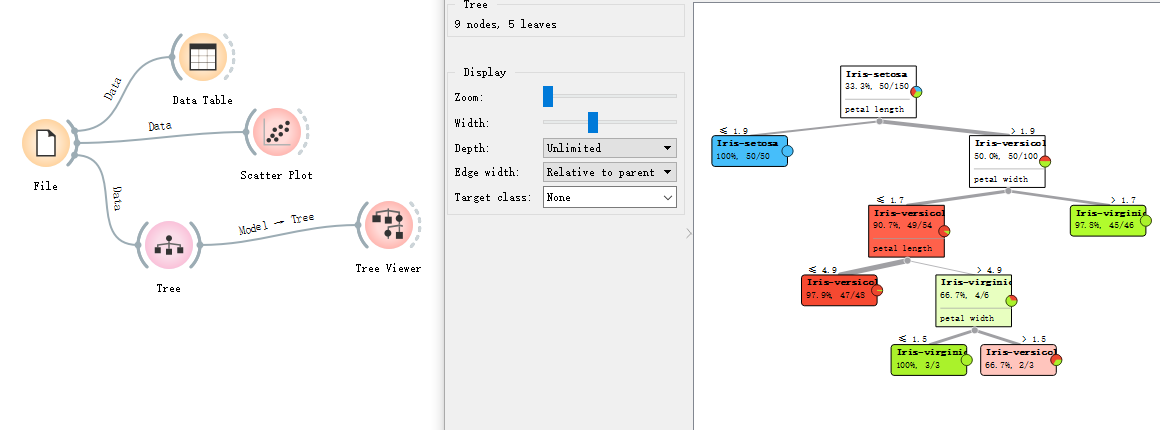
绘制散点图



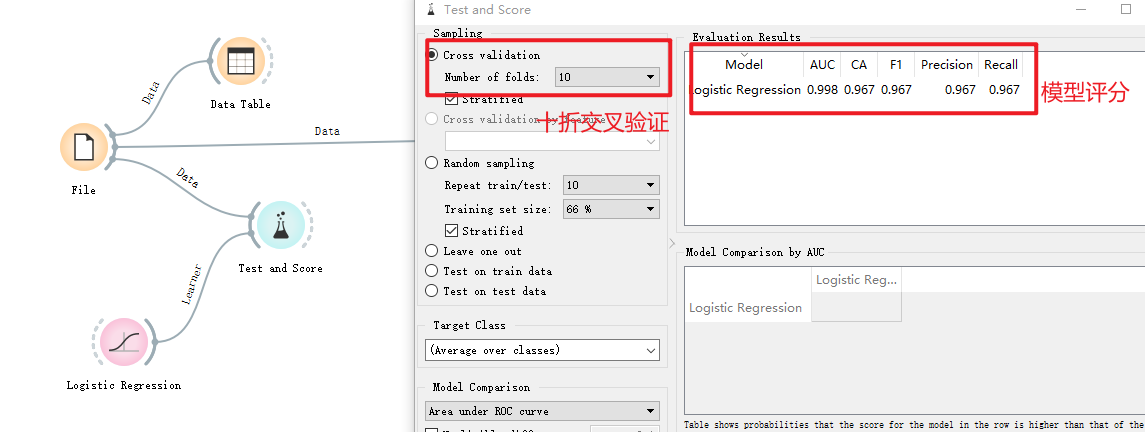
特征评分，并选择最高评分绘制散点图



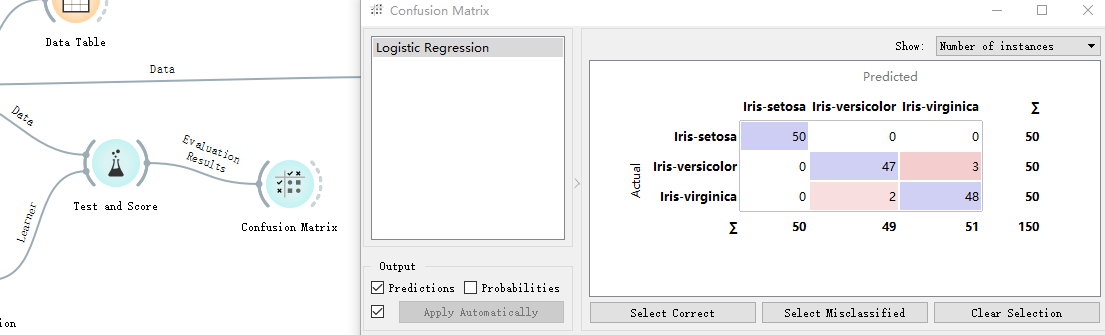
使用决策树算法



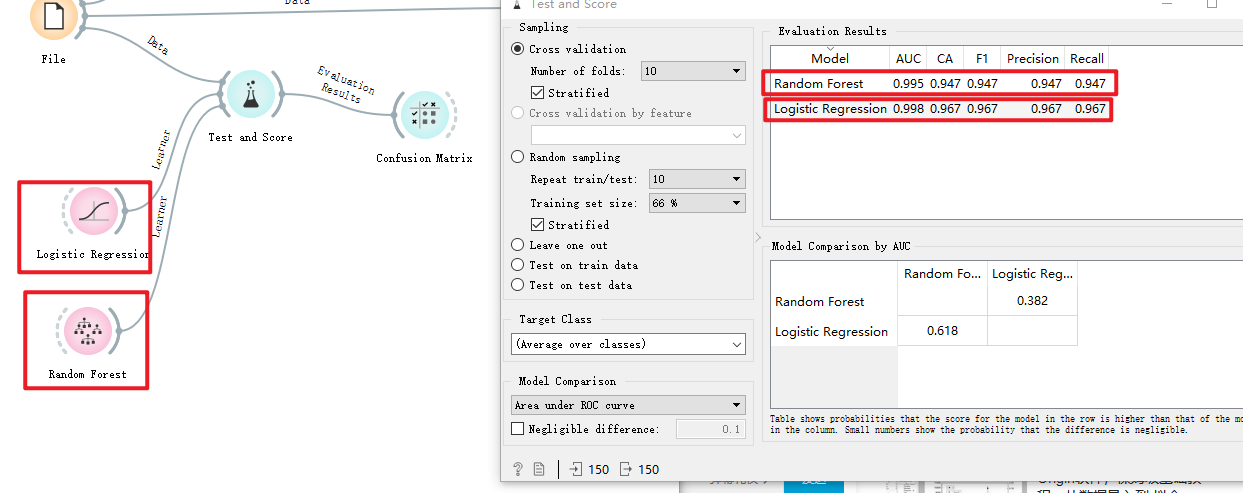
利用逻辑回归算法进行分类并评分（十折交叉验证）



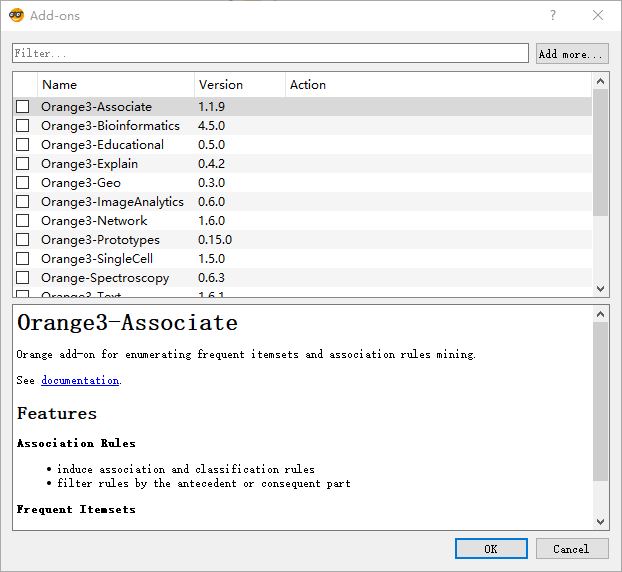
查看模型混淆矩阵

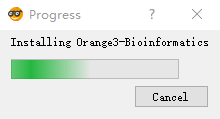


多模型对比



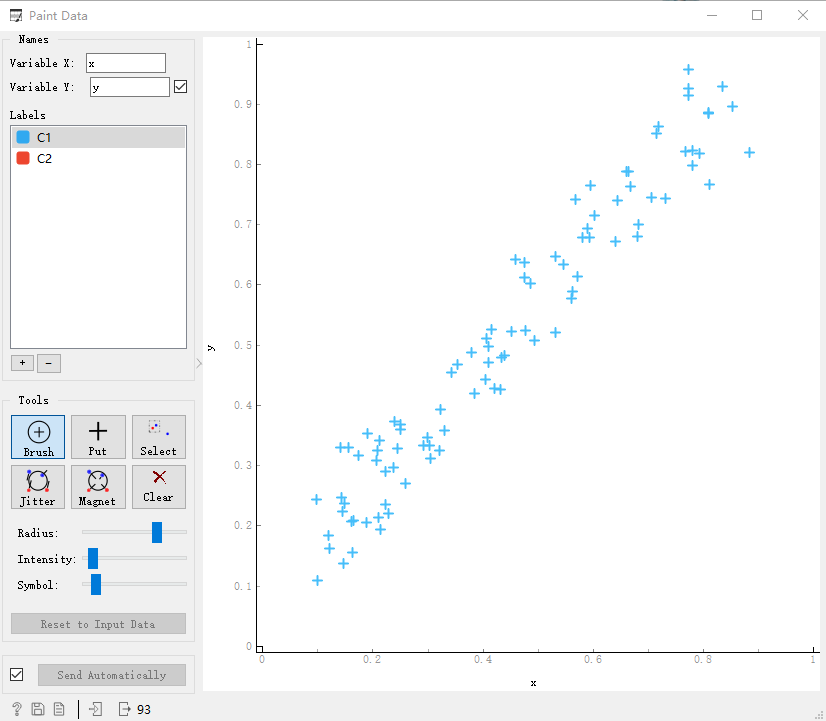
安装其他工具包 Options->Add-one



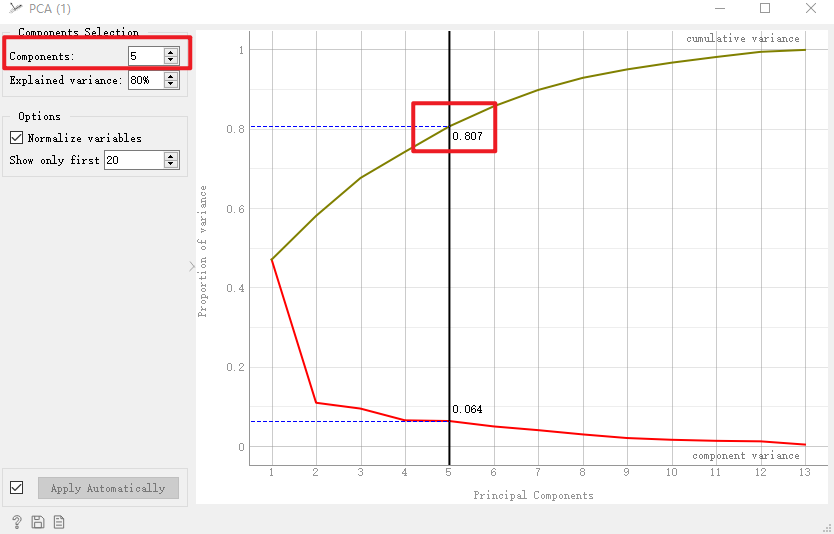


主成分分析 PCA

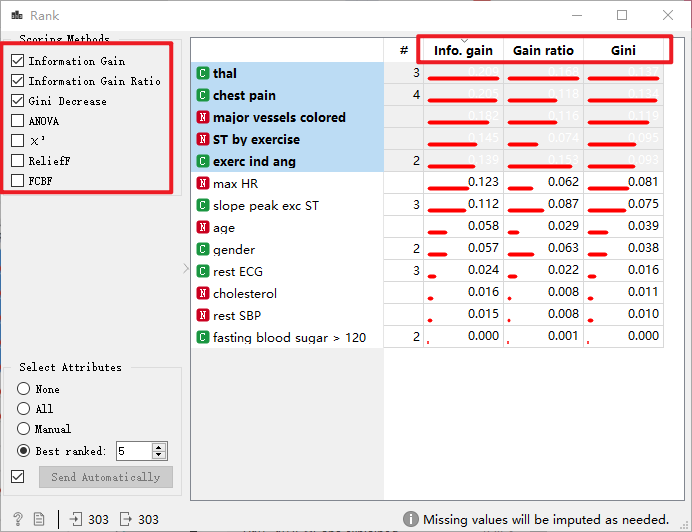
Paint data



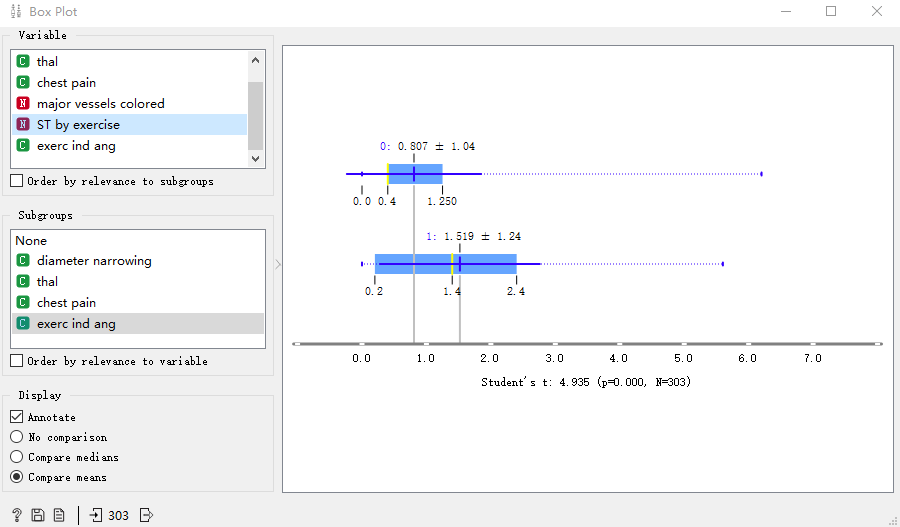
5个主成分可以解释80.7%的数据



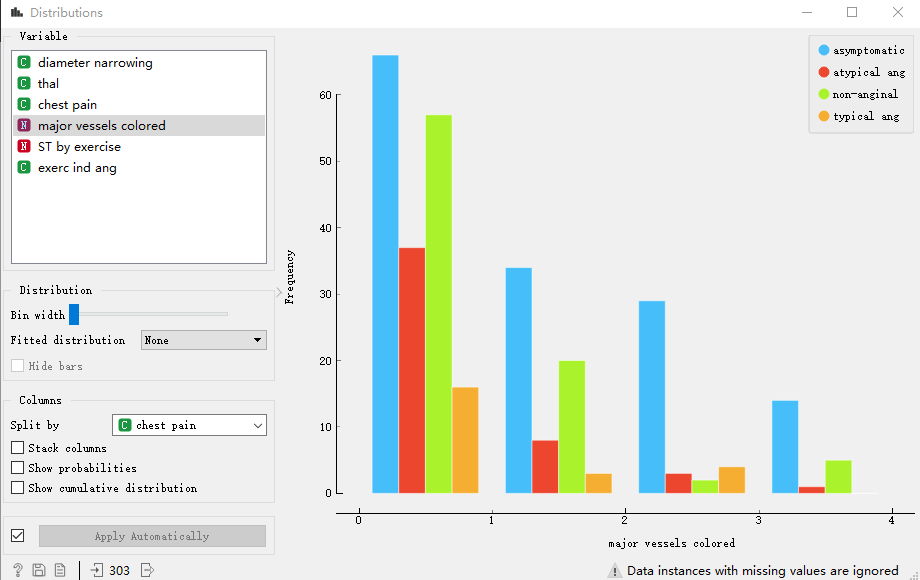
Rank 成分评分



属性与类的关联性



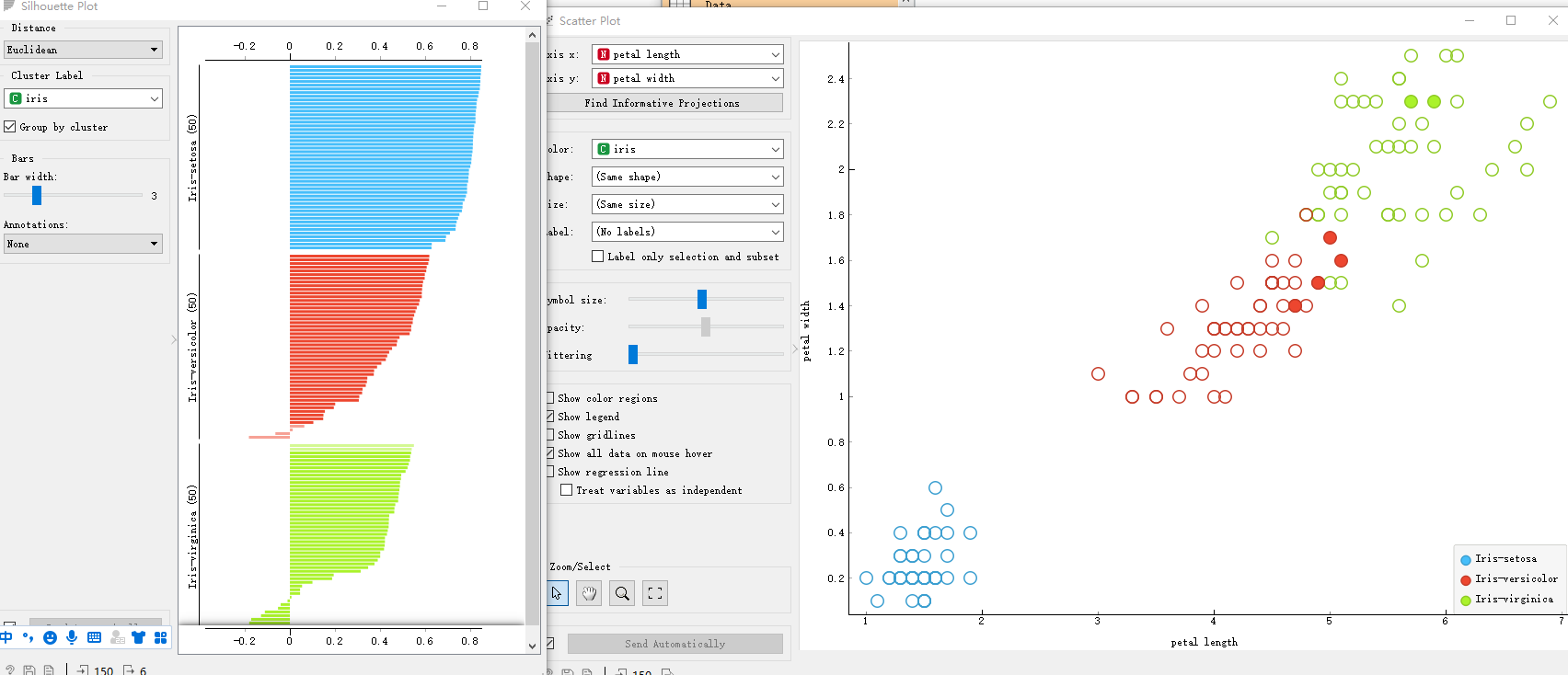
属性密度分布图



K-mean

Educational插件

Silhouette Plot轮廓图



图像转数字

Add-one ImageAnalytics

