

机器学习中的回归问题之随机森林和梯度提升树

机器学习服务操作指导

1 任务介绍

本次任务将介绍如何使用MLS的workflow训练一个随机森林回归模型和梯度提升树回归模型

2 任务执行

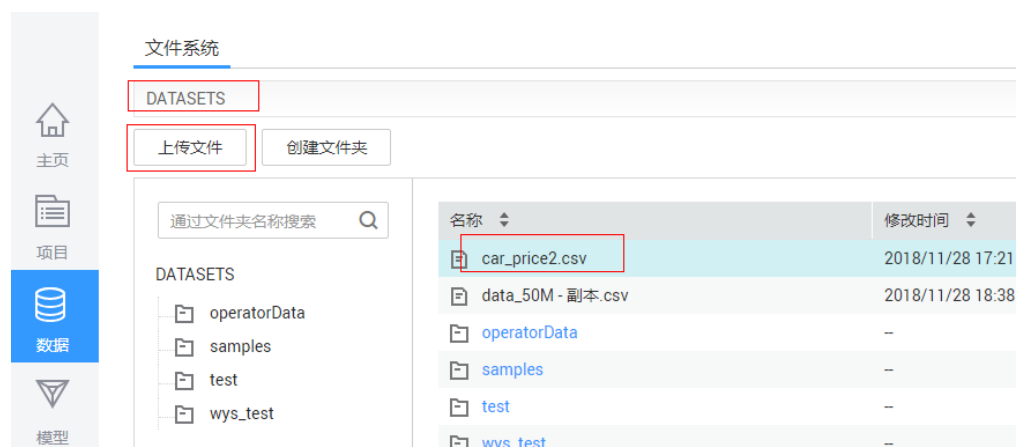
2.1 数据上传

使用数据为某车场的定价数据。

如果已经在第5天的课程进行过数据上传，则此步骤可以忽略。

数据地址：https://obs-mlsclass12.obs.cn-north-1.myhwclouds.com/car_price2.csv

在MLS实例主页上单击“数据”-----单击“DATASETS”-----单击“上传文件”，文件名称为“car_price2.csv”



2.2 创建项目

在MLS实例主页上单击“创建项目”，并写入项目名称，导入案例无需选择，完成后单击“确

定”。



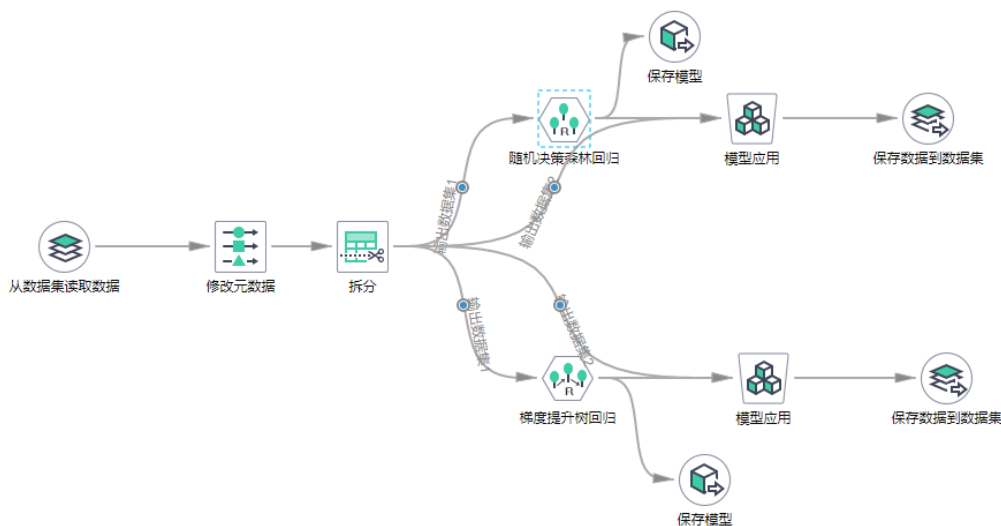
2.3 创建工作流

MLS实例主页单击“项目”—单击2.2中创建的项目名称----单击工作流-----单击“创建工作流”



2.4 编辑工作流

单击“工作流”—单击2.3中创建的工作流名称----打开一个空的工作流，然后按照下图的方式进行编辑，所有的算子在工作流页面的左侧“节点库”中都可以找到。



每个节点的配置如下：

- 1) “从数据集读取数据”：文件地址：/car_price2.csv

读取用户基本信息

* 数据格式:

CSV

* 数据文件:

/car_price2.csv

导入元数据:



是否包括表头:



2) “修改元数据”：选择price特征列

修改元数据

设置元数据



* 字段: price 角色: Target 测量尺度: Conti... * 值: [0,45400]

3) “拆分”：默认设置

4) “随机决策森林回归”：默认设置

5) 两个“模型应用”：预测类型：回归

6) 两个“保存模型”：模型格式：PMML；模型文件名

7) “梯度提升树回归”：默认设置

8) “梯度提升树回归”对应的“保存数据到数据集”：

保存数据到数据集

* 文件路径:

/samples/ ...

* 文件名:

GT

* 文件格式:

CSV ▼

* 字段分隔符:

, ▼

允许覆盖:



9) “随机森林决策回归”对应的“保存数据到数据集”:

保存数据到数据集

* 文件路径:

* 文件名:

* 文件格式:

* 字段分隔符:

允许覆盖:

☒

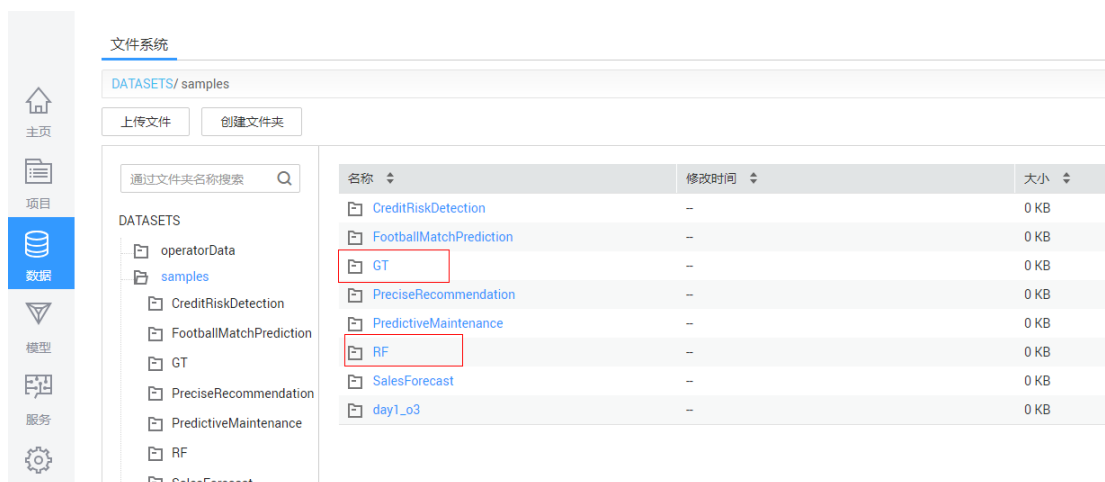
2.5 运行工作流

1) 单击  运行工作流。

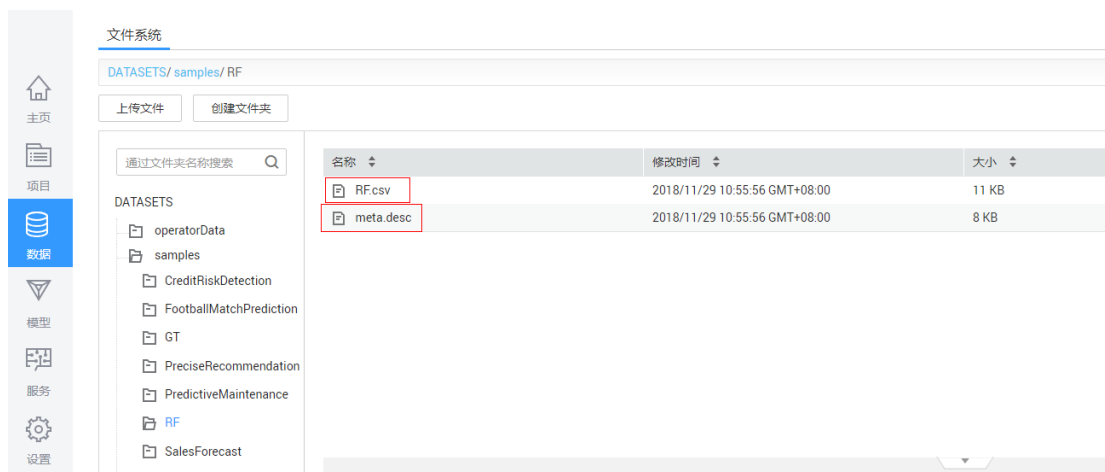
在下方的运行日志查看运行结果。

运行日志

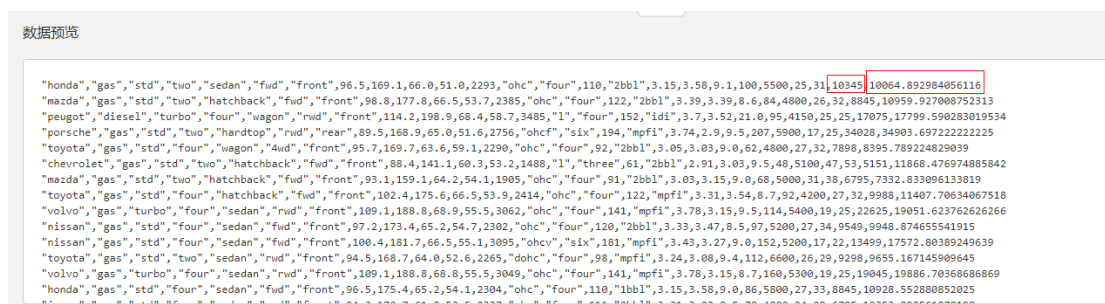
2) 工作流运行完毕后，可以在“主页”—“数据”当中找到两个结果文件，进行查看



3) 比如查看“随机森林决策回归”的预测结果



先单击meta.desc查看每一列的意义，再单击RF.csv查看结果。梯度提升树回归的结果同理。



倒数第二列为真实的price值，最后一列为预测值。

3 打卡任务

3.1 完成单元测试

3.2 任务截图

1、在2.4 workflow界面进行截图：

1) 右上角为用户名、下方为“工作流运行成功”

2) 工作流与图示相同

