

机器学习中的回归问题之线性回归和KNN

-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训练一个线性回归模型和KNN模型

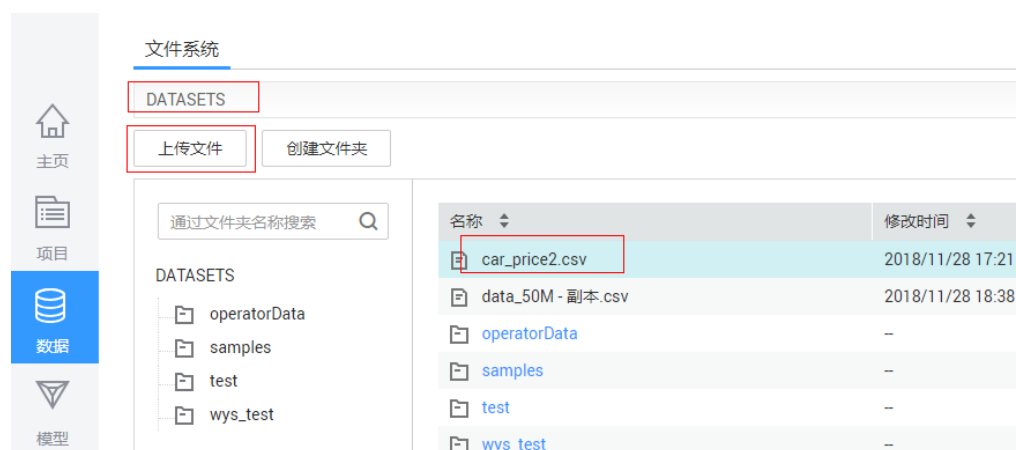
2 任务执行

2.1 数据上传

使用数据为某车场的定价数据。

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”。



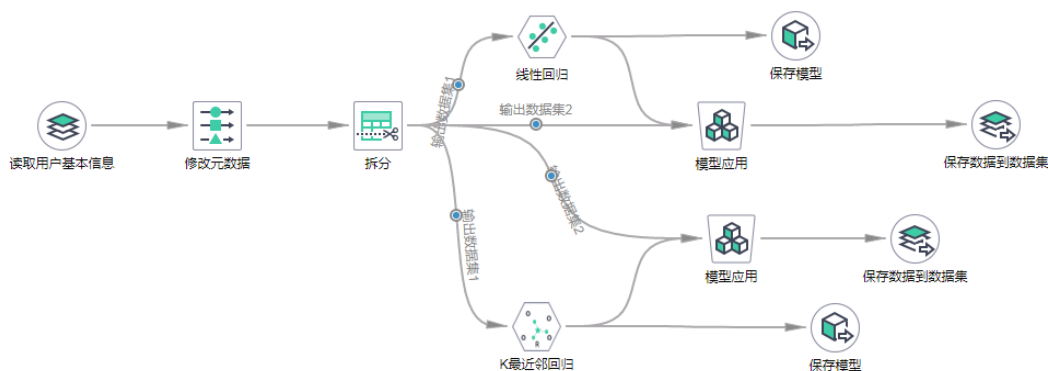
2.3 创建工作流

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) “线性回归”：默认设置

5) 两个“模型应用”：预测类型：回归

6) 与“线性回归”相连的“保存模型”：模型格式：PMML；模型文件名

7) 与“K最近邻回归”相连的“保存模型”：模型格式：INNER；模型文件名

8) “K最近邻”：默认设置

9) “K最近邻”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:



10) “线性回归”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 workflow

1) 单击  运行 workflow。

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

文件系统

DATASETS/ test

上传文件 创建文件夹

通过文件夹名称搜索

DATASETS

- operatorData
- samples
- test
 - KNN
 - linR
 - result
 - spatiotemporal
 - text
 - wys_test

名称	修改时间
frequency.txt	2018/11/26 15:12:10 GMT+08:00
num_string.csv	2018/11/27 16:59:07 GMT+08:00
prefixspan.txt	2018/11/26 15:26:28 GMT+08:00
txt_discrete_data.desc	2018/11/28 12:44:00 GMT+08:00
txt_discrete_data.txt	2018/11/27 09:18:48 GMT+08:00
KNN	-
linR	-
result	-
spatiotemporal	-
text	-

数据预览

3) 比如查看“线性回归”的预测结果

文件系统

DATASETS/ test/ linR

上传文件 创建文件夹

通过文件夹名称搜索

DATASETS

- operatorData
- samples
- test
 - KNN
 - linR
 - result
 - spatiotemporal
 - text
 - wys_test

名称	修改时间
linR.csv	2018/11/29 10:12:12 GMT+08:00
meta.desc	2018/11/29 10:12:12 GMT+08:00

先单击meta.desc查看每一列的意义，再单击lineR.csv查看结果。KNN的结果同理。

数据预览

```
"honda","gas","std","two","sedan","fwd","front",96.5,169.1,66.0,51.0,2293,"ohc","four",110,"2bbl",3.15,3.58,9.1,100,5500,25,31,10345,18225.58959062324
"mazda","gas","std","two","hatchback","fwd","front",98.8,177.8,66.5,53.7,2385,"ohc","four",122,"2bbl",3.39,3.39,8.6,84,4800,26,32,8845,10333.362360636012
"peugot","diesel","turbo","four","wagon","rwd","front",114.2,198.9,68.4,58.7,3485,"1","four",152,"1di",3.7,3.52,21.0,95,4150,25,25,17075,16029.300101784029
"porsche","gas","std","two","hardtop","rwd","rear",89.5,168.9,65.0,51.6,2756,"ohcf","six",194,"mpfi",3.74,2.9,9.5,207,5900,17,25,34028,33973.19964581157
"toyota","gas","std","four","wagon","4wd","front",95.7,169.7,63.6,59.1,2290,"ohc","four",92,"2bbl",3.05,3.03,9.0,62,4800,27,32,7898,-4273.615301265316
"chevrolet","gas","std","two","hatchback","fwd","front",88.4,141.1,60.3,53.2,1488,"1","three",61,"2bbl",2.91,3.03,9.5,48,5100,47,53,5151,10240.918337251234
"mazda","gas","std","two","hatchback","fwd","front",93.1,159.1,64.2,54.1,1905,"ohc","four",91,"2bbl",3.03,3.15,9.0,68,5000,31,38,6795,7435.603578384187
"toyota","gas","std","four","hatchback","fwd","front",102.4,175.6,66.5,53.9,2414,"ohc","four",122,"mpfi",3.31,3.54,8.7,92,4200,27,32,9988,10598.604580898536
"volvo","gas","turbo","four","sedan","rwd","front",109.1,188.8,68.9,55.5,3062,"ohc","four",141,"mpfi",3.78,3.15,9.5,114,5400,19,25,22625,21123.017757727466
"nissan","gas","std","four","sedan","fwd","front",97.2,173.4,65.2,54.7,2302,"ohc","four",120,"2bbl",3.33,3.47,8.5,97,5200,27,34,9549,9203.507234715886
"nissan","gas","std","four","sedan","fwd","front",100.4,181.7,66.5,55.1,3095,"ohcv","six",181,"mpfi",3.43,3.27,9.0,152,5200,17,22,13499,15778.799612009687
"toyota","gas","std","two","sedan","rwd","front",94.5,168.7,64.0,52.6,2265,"ohc","four",98,"mpfi",3.24,3.08,9.4,112,6600,26,29,9298,9828.03434824488
"volvo","gas","turbo","four","sedan","rwd","front",109.1,188.8,68.8,55.5,3049,"ohc","four",141,"mpfi",3.78,3.15,8.7,160,5300,19,25,19045,16720.98943160991
"honda","gas","std","four","sedan","fwd","front",96.5,175.4,65.2,54.1,2304,"ohc","four",110,"1bbl",3.15,3.58,9.0,86,5800,27,33,8845,8408.206187927433
```

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

