

3.2 机器学习助力客户分群（下）

1 任务介绍

本次任务将介绍如何使用MLS工作流中的k-means节点实现客户分群优化。

- 1) 本次课程使用的MLS实例是MLS标准版实例，区域是华北区-北京一。假设创建时提示没有标准版配额，则说明您已经拥有了标准版实例。
- 2) MLS标准版实例的创建过程请参看<https://support.huaweicloud.com/qs-mls/index.html>中左侧“快速入门”——“【标准版】创建MLS实例【01】”
- 3) 释放资源，MLS工作流是按需收费，最好在课程结束前不要删除实例。如果要释放资源请首先在MLS实例页左侧的“设置”中，将NoteBookServer进行关闭。然后在MLS实例管理控制台，将标准版实例进行删除。

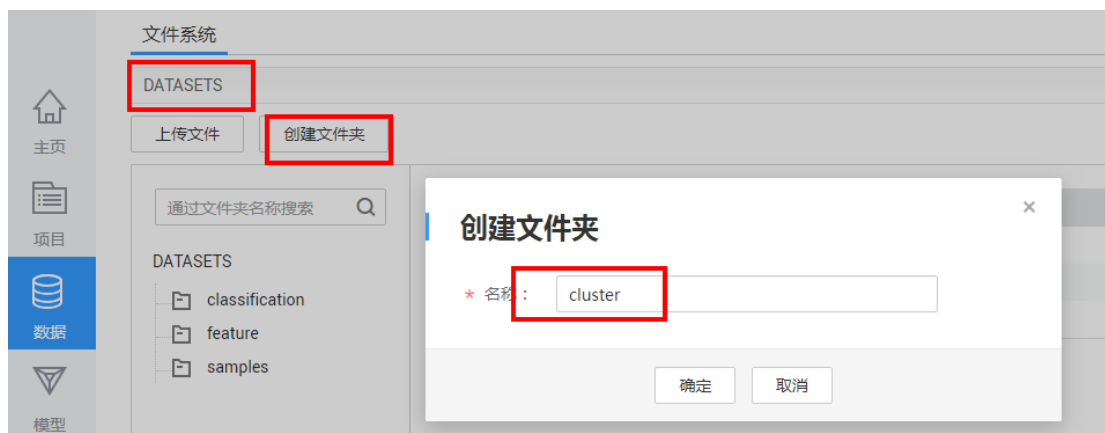
2 任务执行

2.1 数据上传

使用数据某生鲜渠道销售数据。

数据地址：<https://obs-mlsclass7.obs.cn-north-1.myhuaweicloud.com/feature-ex.csv>

在MLS实例主页上单击“数据”-----单击“DATASETS”-----单击“创建文件夹”，文件夹名称为“cluster”



单击“cluster”进入文件夹，将数据文件上传至这个文件夹

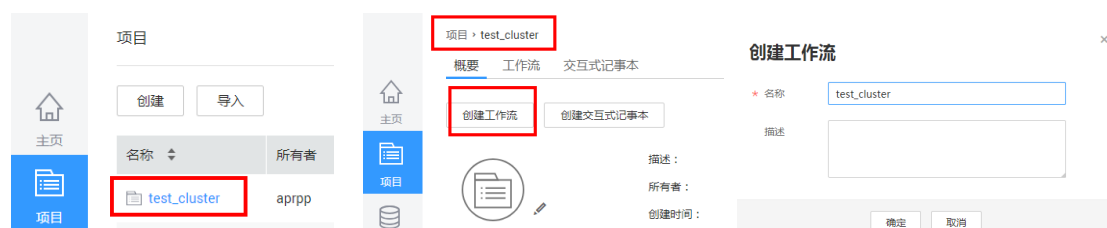
2.2 创建项目

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”。



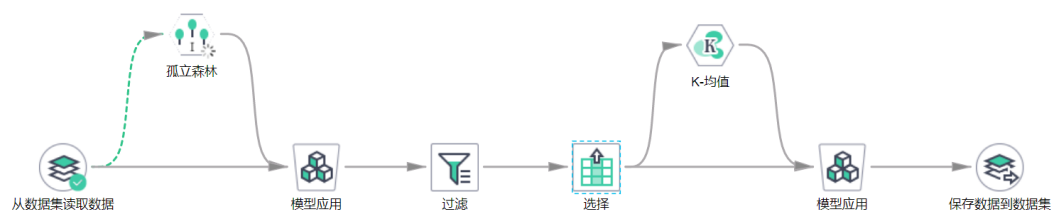
2.3 创建 workflow

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



2.4 编辑 workflow

单击“workflow”—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”：数据文件地址：请填写或者选择feature-ex.csv所在位置
- 2) “孤立森林”：默认配置
- 3) 与“孤立森林”相连的“模型应用”：预测类型：异常检测；异常检测阈值：0.5

4) “**过滤**”: 条件为: predictioncol=0, 作用是将异常检测结果为正常的的数据 (结果为0) 选择出来。

过滤

过滤条件  

* 字段:	predictioncol	...	* 操作类型:	=	▼	* 过滤条件值:	0
-------	---------------	-----	---------	---	---	----------	---

5) “**选择**”: 两个字段为predictioncol、score

选择

字段集  

* 字段类型:	predictioncol	...
---------	---------------	-----

* 字段类型:	score	...
---------	-------	-----

反选:



6) “**k-均值**”:

K-均值

* 聚类数:

* 最大迭代次数:

* 初始模式:

7) 与“**k-均值**”相连的“**模型应用**”：预测类型：聚类

8) “**保存数据到数据集**”：填写保存路径和文件名

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

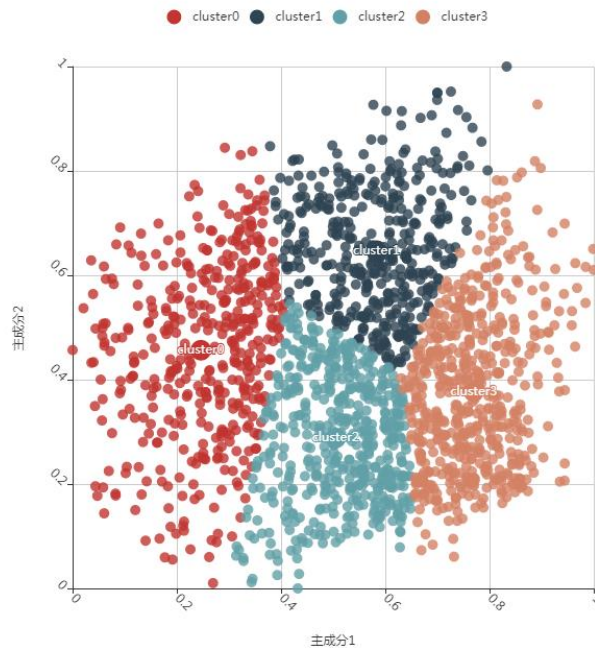
1) 单击  运行工作流。

在下方的运行日志查看运行结果。

运行日志

2) 在“k-均值”节点右键，单击“查看模型”，可以将聚类模型可视化

查看模型



3) 工作流运行完毕后，可以在“主页”——“数据”当中在自己设置的结果文件路径下找到结果文件，查看“kmeans”的聚类结果

先单击meta.desc查看每一列的意义，再单击kmeans_result.csv查看结果。最后一列是聚类结果列。

数据预览

```
115.79,138.0,265,61.7,547.1,3
124.74,140.6,285,67.5,527.2,0
121.45,137.6,278,64.3,548.3,3
110.22,139.1,252,59.1,541.0,2
106.63,136.5,245,56.0,543.4,2
106.5,138.4,243,56.9,554.6,2
106.73,138.4,245,56.9,553.3,2
112.57,138.4,258.60,1,530.1,3
```

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

