

7.2 机器学习中的聚类问题

-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k-means进行聚类

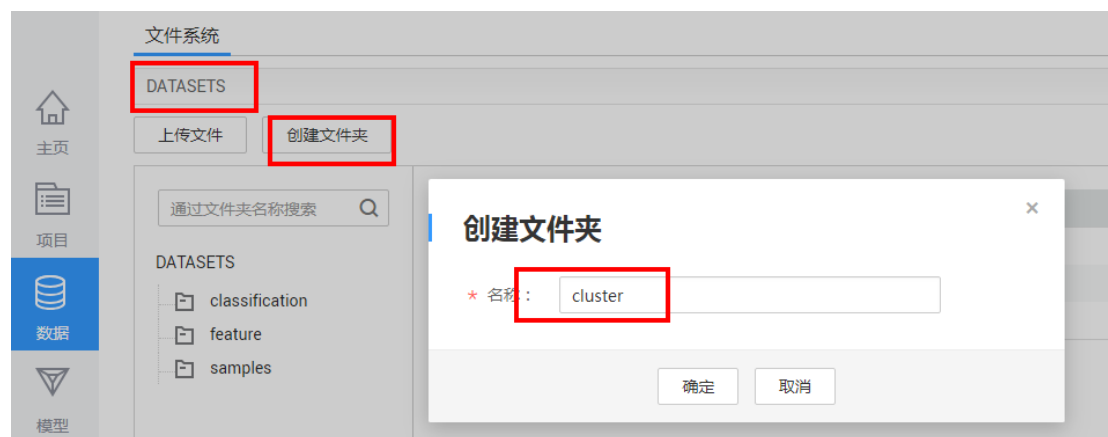
2 任务执行

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-mlsclass12.obs.cn-north-1.myhwclouds.com/customer.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”



点击“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”。



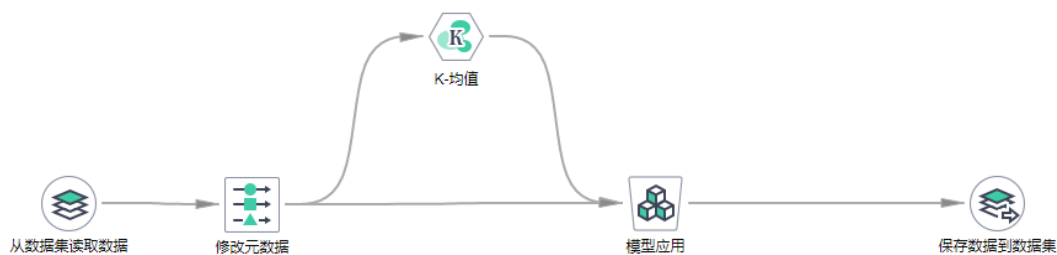
2.3 创建工作流

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”：数据文件地址：/cluster/customer.csv

从数据集读取数据

* 数据格式:

CSV

* 数据文件:

/cluster/customer.csv ...

导入元数据:

☐

是否包括表头:

☒

- 2) “修改元数据”：进行配置时，点击 ，然后将特征“id”的角色改为“None”

修改元数据

设置元数据



定义元数据

类型 格式

类型：显示全部

通过字段搜索

字段	类型	测量尺度	值	角色
id	Integer	Continuous	[1,440]	None
channel	Integer	Flag	1, 2	ID
region	Integer	Ordinal	1, 2, 3	Input
fresh	Integer	Continuous	[3,112151]	Target
milk	Integer	Continuous	[55,73498]	None
grocery	Integer	Continuous	[3,92780]	Frequency
frozen	Integer	Continuous	[25,60869]	Both
detergents_paper	Integer	Continuous	[3,40827]	Split

- 3) “模型应用”：预测类型：聚类
- 4) “k-均值”：

K-均值

* 聚类数:

* 最大迭代次数:

* 初始模式:

* 初始模式步数:

5) “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:



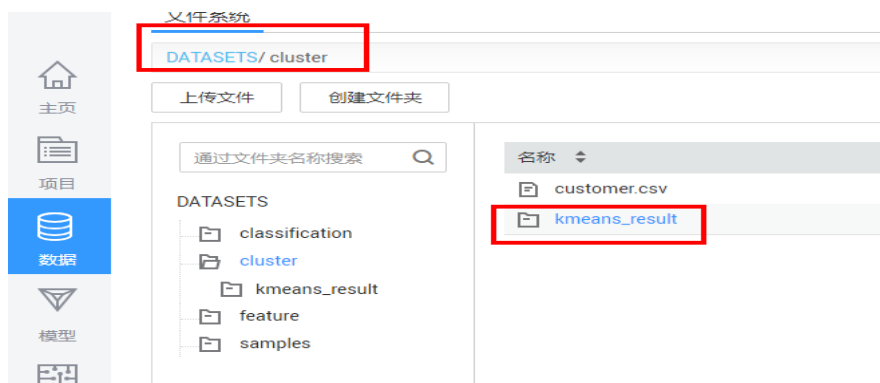
2.5 运行工作流

1) 单击  运行工作流。

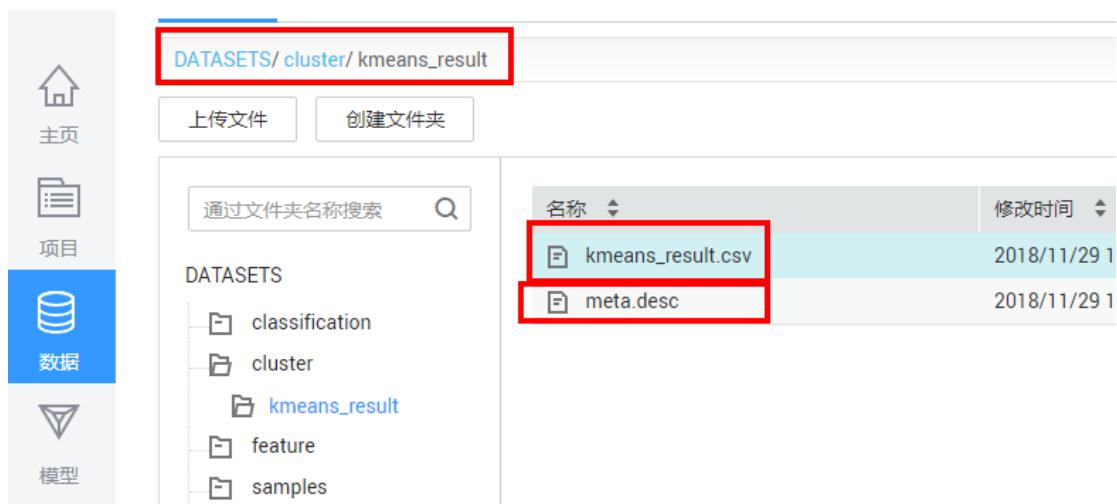
在下方的运行日志查看运行结果。

运行日志

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



3) 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

```
16,1,3,10253,1114,3821,397,964,412,2
231,2,1,11072,5989,5615,8321,955,2137,2
396,1,3,9602,1316,1263,2921,841,290,2
226,1,1,12680,3243,4157,660,761,786,2
883,1,3,34454,7435,8469,2540,1711,2893,1
272,1,1,2083,5007,1563,1120,147,1550,2
372,1,3,20105,1887,1939,8164,716,790,2
63,2,3,7823,6245,6544,4154,4074,964,2
371,2,3,39679,3944,4955,1364,523,2235,1
184,1,3,36847,43950,20170,36534,239,47943,1
275,1,3,894,1703,1841,744,759,1153,2
98,1,3,403,254,610,774,54,63,2
429,1,3,3047,5970,4910,2198,850,317,2
347,2,3,2771,6939,15541,2693,6600,1115,2
```

这两个样本被聚集到了不同的簇当中。

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“workflow 运行成功”
- 2) workflow 与图示相同

