

3.4.7 聚类

3.4.7.1 K-Means

图标: 

描述: K-Means 是 Mac Queen 提出的一种非监督实时聚类算法, 在最小化误差函数的基础上将数据划分为预定的类数 K。

字段属性

特征列: 需要进行聚类的列, 请选择数值型数据, 如果勾选了非数值类型数据, 则会自动过滤, 下个组件可能无法获取所有列。如图 351 所示。



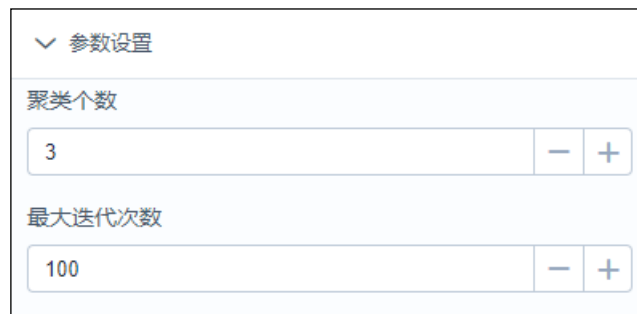
图 351

参数设置

聚类个数: 聚类的个数, 默认 3。

最大迭代次数: 迭代的次数。

如图 352 所示。



参数设置

聚类个数

3

最大迭代次数

100

图 352

输出

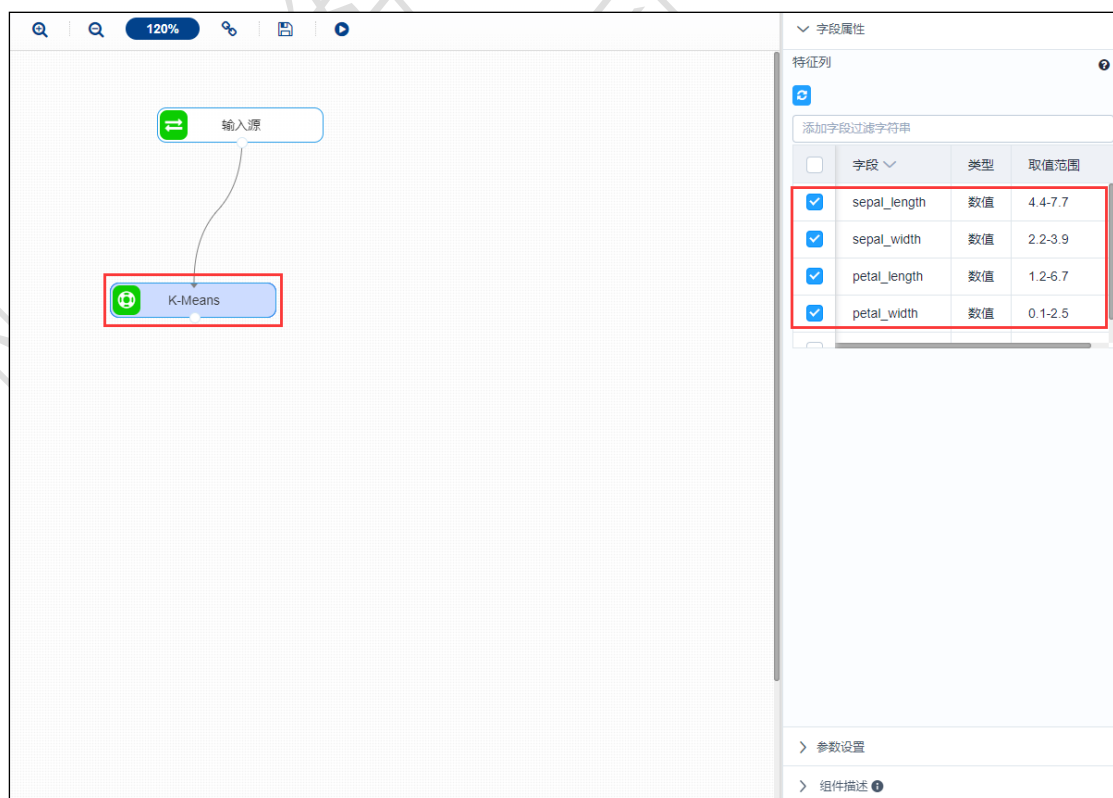
表结果：包含聚类结果的数据表。

报告：聚类中心、饼图。

示例

下面对某数据进行 K-Means 聚类。

- 选择待聚类的序列，数据必须为数值型。如图 353 所示。
- 点击参数设置，聚类个数设置为 3，最大迭代次数设置为 100。如图 354 所示。
- 运行该组件，对组件右击，选择查看数据与报告，结果如图 355 与图 356 所示。



字段属性

特征列

添加字段过滤字符串

字段	类型	取值范围
<input checked="" type="checkbox"/> sepal_length	数值	4.4-7.7
<input checked="" type="checkbox"/> sepal_width	数值	2.2-3.9
<input checked="" type="checkbox"/> petal_length	数值	1.2-6.7
<input checked="" type="checkbox"/> petal_width	数值	0.1-2.5

> 参数设置

> 组件描述

图 353

∨ 参数设置

聚类个数

- +

最大迭代次数

- +

图 354

sepal_length	sepal_width	petal_length	petal_width	cluster_id
5.1	3.5	1.4	0.2	2
4.9	3	1.4	0.2	2
4.7	3.2	1.3	0.2	2
4.6	3.1	1.5	0.2	2
5	3.6	1.4	0.2	2
5.4	3.9	1.7	0.4	2
4.6	3.4	1.4	0.3	2
5	3.4	1.5	0.2	2

共 150 条 25 条/页 < 1 2 3 4 5 6 > 前往 1 页

图 355


聚类中心

cluster_id	sepal_length	sepal_width	petal_length	petal_width
1	5.901612903225806	2.7483870967741937	4.393548387096774	1.4338709677419355
2	5.006	3.428	1.4619999999999997	0.246000000000000033
3	6.85	3.0736842105263156	5.742105263157894	2.0710526315789473

饼图

图 356

3.4.7.2 GMM(高斯混合模型)

图标: 

描述: GMM (高斯混合模型) 是用高斯概率密度函数精确地量事物, 将一个事物分解为