

图 359

sepal_length	sepal_width	petal_length	petal_width	cluster_id
5.1	3.5	1.4	0.2	1
4.9	3	1.4	0.2	1
4.7	3.2	1.3	0.2	1
4.6	3.1	1.5	0.2	1
5	3.6	1.4	0.2	1
5.4	3.9	1.7	0.4	1
4.6	3.4	1.4	0.3	1
5	3.4	1.5	0.2	1

共 150 条 25 条/页 < 1 2 3 4 5 6 > 前往 1 页

图 360

3.4.7.3 密度聚类



描述: 密度聚类的核心思想是从某个核心点出发, 不断向密度可达的区域扩张, 从而得

到一个包含核心点和边界点的最大化区域，区域中任意两点密度相连。对于噪声样本，其簇标记为-1。

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 361 所示。



图 361

参数设置

邻域半径：设置某个半径，默认 0.5。

邻域内最小数目：设置邻域内最小点的个数，默认 5。

如图 362 所示。

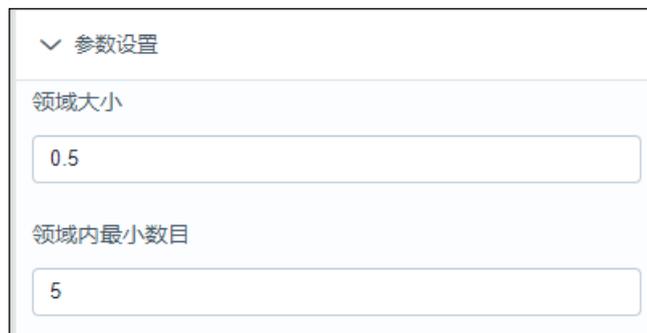


图 362

输出

表结果：包含聚类结果的数据表。

报告：无。

示例

下面对某数据进行密度聚类。

- 选择待聚类的序列，数据必须为数值型，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 363 所示。
- 运行该组件，对组件右击，选择查看数据，结果如图 364 所示。

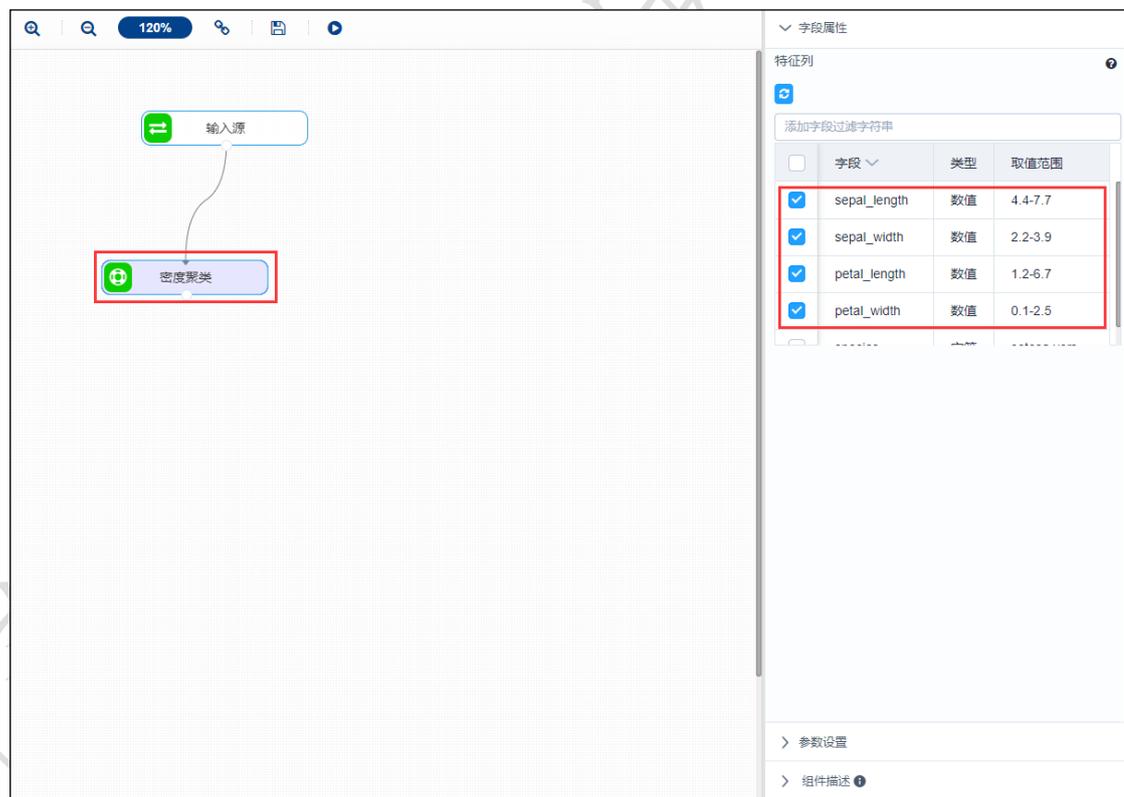


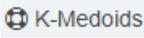
图 363

sepal_length	sepal_width	petal_length	petal_width	cluster_id
5.1	3.5	1.4	0.2	0
4.9	3	1.4	0.2	0
4.7	3.2	1.3	0.2	0
4.6	3.1	1.5	0.2	0
5	3.6	1.4	0.2	0
5.4	3.9	1.7	0.4	0
4.6	3.4	1.4	0.3	0
5	3.4	1.5	0.2	0

共 150 条 25 条/页 < 1 2 3 4 5 6 > 前往 1 页

图 364

3.4.7.4 K-Medoids

图标: 

描述: K-Medoids 是 Kmeans 算法的改进, 减轻了 Kmeans 算法对孤立点的敏感性, 选用簇中离平均值最近的对象作为簇中心。

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 365 所示。

字段属性

特征列 ?



添加字段过滤字符串

<input type="checkbox"/>	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

图 365