

预览数据			
id	m	r	f
848	23.24	4	1
773	28.43	20	1
915	33.58	17	1
445	45.23	7	1
813	49.61	16	1
776	49.7	4	1
294	50.14	3	2
129	53	28	1

共 940 条 25 条/页 < 1 2 3 4 5 6 ... 38 > 前往 1 页

图 95

3.4.2.11 数据拆分

图标: 

描述: 数据拆分对全量数据进行简单随机抽样, 将数据拆分为训练数据和测试数据。

字段属性

特征列: 必选。选择进行拆分的数据, 勾选的列将传入下个组件, 如图 96 所示。

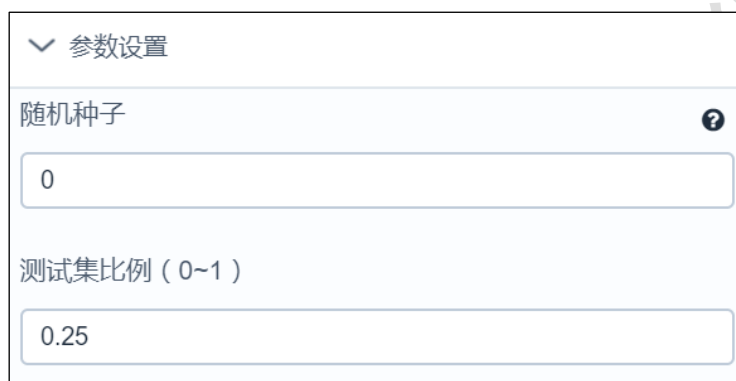
特征列			
<input checked="" type="checkbox"/>	字段	类型	取值范围
<input checked="" type="checkbox"/>	sepal_width	数值	2.2-3.9
<input checked="" type="checkbox"/>	petal_length	数值	1.2-6.7
<input checked="" type="checkbox"/>	species	字符	setosa,vers...
<input checked="" type="checkbox"/>	sepal_length	数值	4.4-7.7
<input checked="" type="checkbox"/>	petal_width	数值	0.4-0.5

图 96

参数设置

随机种子：可以理解为一个序号，这个序号交给一个数列管理器，通过这个序号，从管理器中取出一个数列，这个数列就是通过那个序号得到的随机数。

训练集比例：设置训练数据的比例，范围在 0-1 之间，默认 0.25，如图 97 所示。



参数设置

随机种子 ?

0

测试集比例 (0~1)

0.25

图 97

输出

表结果：训练集与测试集。

报告：无。

示例

下面将某数据拆分为训练集、测试集。

- 勾选需要进行数据拆分的数据，如图 98 所示。
- 保留默认的随机种子，设置测试集比例为 0.25，如图 99 所示。
- 运行该组件，选择查看数据，查看对应的数据集，如图 100 所示。



图 98

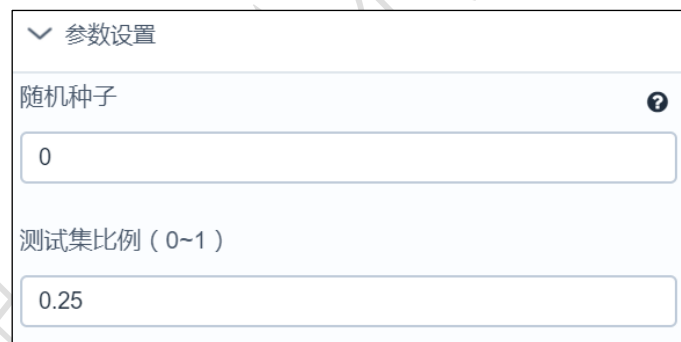


图 99



图 100