

数据挖掘-相关性分析

概述

相关性分析是用来反映变量之间的相关关系的密切程度。相关系数的取值一般介于-1和1之间。当相关系数为正的时候，意味着变量之间是正相关的；当相关系数为负的时候，意味着变量之间是负相关。

相关性分析常用在数据探索阶段，当我们并不了解原始数据各字段之间的关系时，通过相关性分析，可以看到各个字段之间的相关性，其后进行的数据分析工作可以围绕这些相关性展开。

- [概述](#)
- [输入/输出](#)
- [参数设置](#)
- [示例](#)

输入/输出

输入	一个输入端口，用于接收数据集。
输出	一个输出端口，用于输出相关系数的数据集。

参数设置

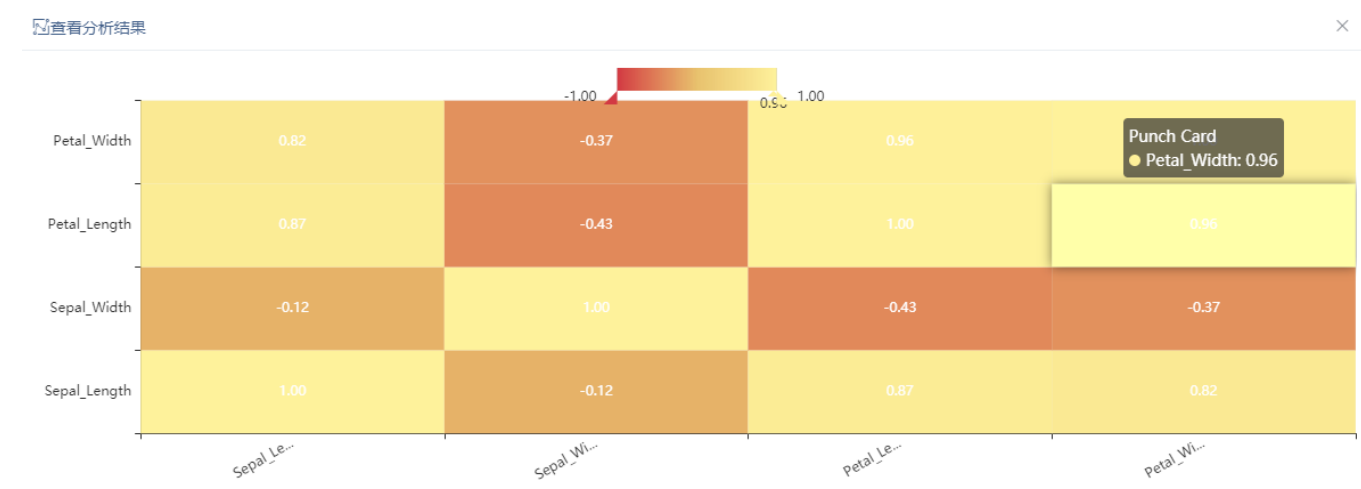
参数名称	说明	备注
选择列	用于选择进行相关性分析的字段列	必填

示例

使用“鸢尾花数据”，选择4个特征，分析这4个特征之间的相关性。输出结果如下：

Featrue	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width
Sepal_Length	1.0	-0.11756978413301303	0.8717537758865829	0.8179411262715678
Sepal_Width	-0.11756978413301303	1.0	-0.4284401043305347	-0.3661259325364441
Petal_Length	0.8717537758865829	-0.4284401043305347	1.0	0.9628654314027957
Petal_Width	0.8179411262715678	-0.3661259325364441	0.9628654314027957	1.0

鼠标点击右键查看分析结果“可视化”：



结合上图分析得出：特征“Petal_Length”和“Petal_Width”具有较高的相关性。