**贝叶斯节点使用说明**

作者：张青松

目录

[1. 贝叶斯节点 2](#_Toc426114948)

[1.1. 朴素贝叶斯分类基本原理 2](#_Toc426114949)

[1.2. 贝叶斯节点 2](#_Toc426114950)

[2. 贝叶斯设置 3](#_Toc426114951)

[2.1. 建立贝叶斯节点的工作流 3](#_Toc426114952)

[2.1.1. 设置 3](#_Toc426114953)

[3. 贝叶斯分类结果 4](#_Toc426114954)

# 贝叶斯节点

贝叶斯节点使用了经典的朴实贝叶斯（NaiveBayes）算法对数据进行分类，其核心思想是贝叶斯公式：对于给出的待分类项，求解在此项出现的条件下各类别出现的概率，取概率最大的类别作为对该项的分类结果。

## 朴素贝叶斯分类基本原理

朴素贝叶斯正式定义如下：

1. 设$x=\{a\_{1},a\_{2},…,a\_{m}\}$为一个待分类项，而每个a为x的一个特征属性.
2. 有类别集合$C=\{y\_{1},y\_{2},…y\_{n},\}$。
3. 计算$P\left(x\right)，P\left(x\right)，…，P\left(x\right)$。
4. 如果$P\left(x\right)=max⁡\{P\left(x\right)，P\left(x\right)，…，P\left(x\right)\}$，则$x\in y\_{k}$。

针对训练数据中，某个类别下的某个特征划分没有出现时，会令分类器的质量大大降低。为了解决这个问题，引入了Laplace校准。其思想就是对每类别下所有划分的计数加1，或者概率加上一个很小的值，这样如果训练样本集数据充分大时，并不会对结果产生影响，并且解决了概率为0的尴尬局面。

## 贝叶斯节点

在DataStudio中，通过设置面板在输入的训练数据表中，选择某个标称量字段作为分类目标字段以及另一部分列作为参与分类的特征属性，训练朴素贝叶斯模型，然后使用该模型对位置的数据进行分类。

# 贝叶斯设置

## 建立贝叶斯节点的工作流



图1 贝叶斯节点工作流

首先为贝叶斯节点连接输入数据。输入数据中必须包含类型为标称量的字段。以iris.csv数据集为例。

### 设置



图2 贝叶斯节点数据选择设置

选择数据集中的标称量字段作为分类的目标列，并且在下方表格中选中要作为特征属性参与分类的列。

切换到模型页签，如图。



图3 贝叶斯算法参数设置

注意：模型页签中的默认概率表示上文中提到的Laplace校准参数，最大分类个数不能小于分类目标列标称量的个数。

# 贝叶斯分类结果

使用训练结果对数据进行分类得到结果如图



图5 分类结果

在上图中，前几列是分类数据中的列，最后一列是朴素贝叶斯算法对该项的分类，中间的几列是该行在各个分类下的概率，取最大概率的分类做为该行的分类。