**KNN节点使用说明**

**作者：王龙强**

[**1.    节点设置参数页面 1**](#_Toc425927970)

[**2.   节点使用过程 2**](#_Toc425927971)

[**3.   注意事项 4**](#_Toc425927972)

KNN节点是 Data Studio 建模节点分类中的一个建模节点，模型训练过程中的参与变量只能是数值类型的变量，K最近邻(k-Nearest Neighbor，KNN)分类算法，是一个理论上比较成熟的方法，也是最简单的机器学习算法之一。该方法的思路是：如果一个样本在特征空间中的K个最相似(即特征空间中最邻近)的样本中的大多数属于某一个类别，则该样本也属于这个类别。KNN算法中，所选择的邻居都是已经正确分类的对象。该方法在定类决策上只依据最邻近的一个或者几个样本的类别来决定待分样本所属的类别。

# 1.     节点设置参数页面

如图 1 所示的KNN节点设置界面设置项意义如下：

1. 目标变量：即可以理解为样本所属类别；
2. 数值变量：参与变量，依据参与变量对样本进行分类；



 图1KNN设置界面

# 2.    节点使用过程

在以前讲过的建模节点使用说明中可以知道，建模节点配置完并执行后，会在左下角模型区域生成指定的训练模型，双击训练模型可以展示预测数据的前50条记录。如图2



 图2,KNN训练模型

  将训练模型拖入工作区，配置训练模型，训练模型的配置界面如下图3，

 邻居个数：分类数据的邻居个数

 输出分类概率：输出待分类数据隶属于各分类的个数

 考虑距离权重：默认选中，不以邻居同分类个数的多少，判定待分类数据的隶属分类



 图3,训练模型配置界面

保存配置，连接预测数据即可执行并且预测，如图4



 图4,训练模型进行预测

 可以右键预测模型“预览结果集”也可以文件输出查看最终分类结果，如下图4、图5



 图4,预测模型结果预览



图5,预测模型结果预览

# 3.    注意事项

1. 注意不要将训练数据与预测数据搞混淆了，即训练数据不应该含有预测数据，如果训练数据包含预测数据的话，分类将变得没有意义，因为预测数据本来就已经分类。