

数据挖掘-高斯混合模型

概述

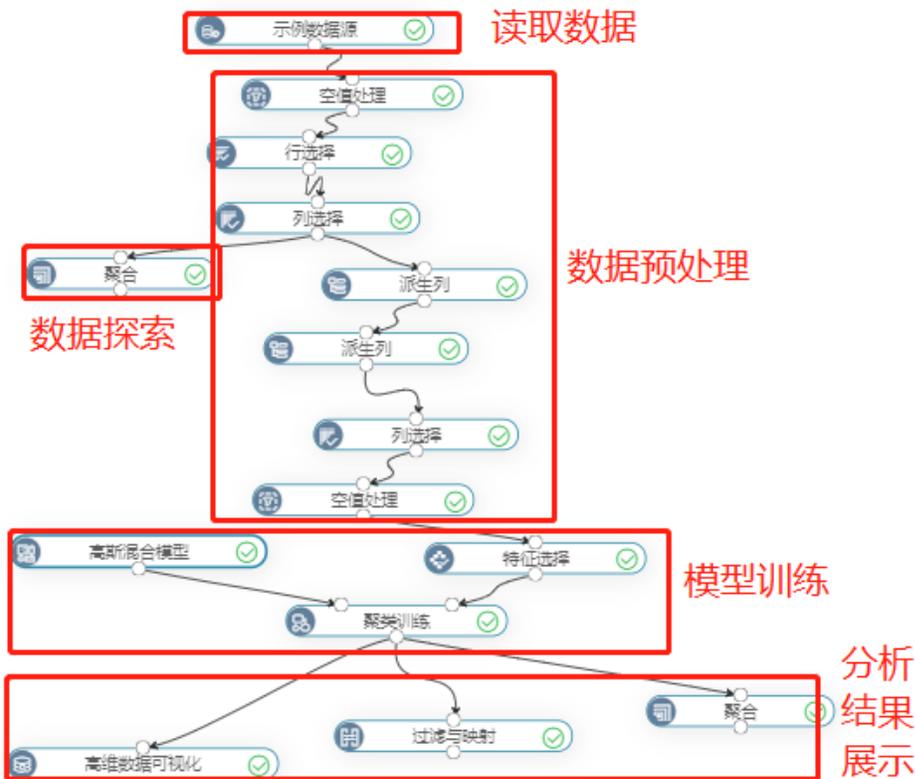
高斯混合模型就是用高斯概率密度函数（正态分布曲线）精确地量化事物，它是一个将事物分解为若干的基于高斯概率密度函数（正态分布曲线）形成的模型。高斯混合模型与K均值聚类不同，K均值是考虑每个数据点到某个类簇中心点的距离，而高斯混合模型是考虑数据点被分配到每个类簇的概率。

高斯混合模型适用于聚类问题中各个类别的尺寸不同，聚类间有相关关系的情况。

示例

使用“航空公司客户价值分析”案例数据，分析客户为高价值客户、一般客户、低价值客户。

- 概述
- 示例



高斯混合模型参数如下：

参数名称	值	说明
归一化	正则化	详情请参考 归一化 介绍说明。
	标准化	
	最小最大值归一化	
	最大绝对值归一化	
K值	取值范围是：>=2的整数，默认值为2	期待将数据聚类的数目；
收敛阈值	参数范围为：>=0的数，默认值为：0.000001。	收敛误差值；当各类的聚类平方和小于收敛阈值时则停止迭代；
迭代次数	参数范围：>=0的正整数，默认值为20。	算法的最大迭代次数，达到最大迭代次数即退出。 最大迭代次数的值越大，模型训练更充分，但会耗费更多时间。