

# 数据挖掘-K均值

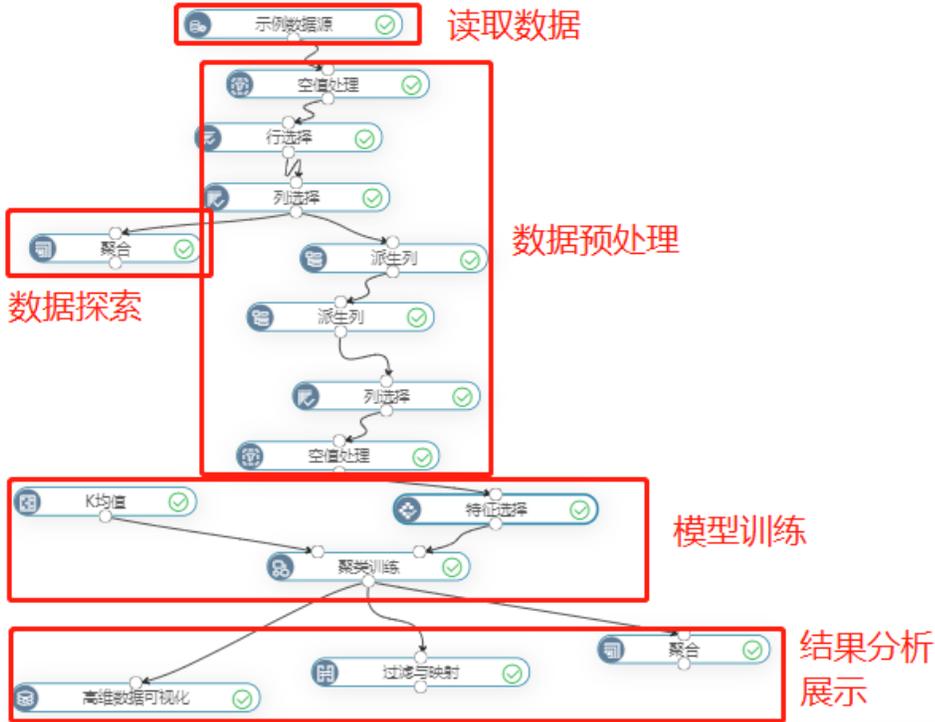
## 概述

K均值算法是一种简单的迭代型聚类算法，采用距离作为相似性指标，从而发现给定数据集中的K个类，且每个类的中心是根据类中所有值的均值得到，每个类用聚类中心来描述。它的聚类目标是以欧式距离作为相似度指标，使得各类的聚类平方和最小。

- 概述
- 示例

## 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。



聚类算法参数如下：

参数名称	值	说明
归一化	正则化	详情请参考 <a href="#">归一化</a> 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；
随机种子	参数范围：任意整数，默认值为：2。	初始化时随机选择类中心点的随机种子。seed设置为固定值，每次聚类结果是稳定的。
收敛阈值	参数范围：>=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。 <b>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。</b>