



# 5.1 机器学习助力 质量分类(下)

HUAWEI TECHNOLOGIES CO., LTD.



[www.huawei.com](http://www.huawei.com)

# 目录

## Contents

1

数据缺失处理

2

算法优化

3

建模

# 特征清洗

## 特征清洗

忽略：在特征上设定规则忽略满足规则的特征值

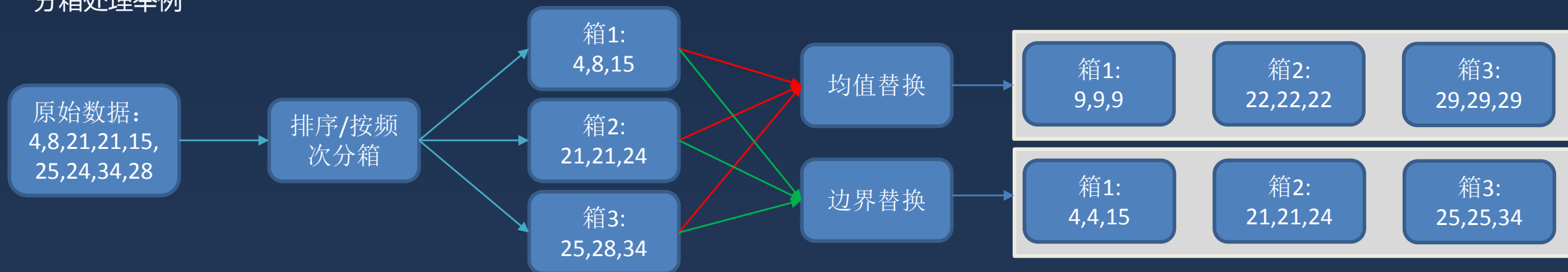
缺失值填充：使用特征上的全局统计量、K最近邻统计量进行缺失值填充

异常值处理：使用特征上的全局统计量、K最近邻统计量进行异常值的替换

分箱平滑：在特征上使用数据分箱，用每一个数据箱中的统计量代替特征本身的价值

回归平滑：使用可视化的技术、回归技术在特征上使用一个函数来进行特征拟合，特征上的离群点用函数值替换

### 分箱处理举例



本课程使用的数据倒数第二列“V”即是经过缺失值填充操作的。

# 目录

Contents

1

数据缺失处理

2

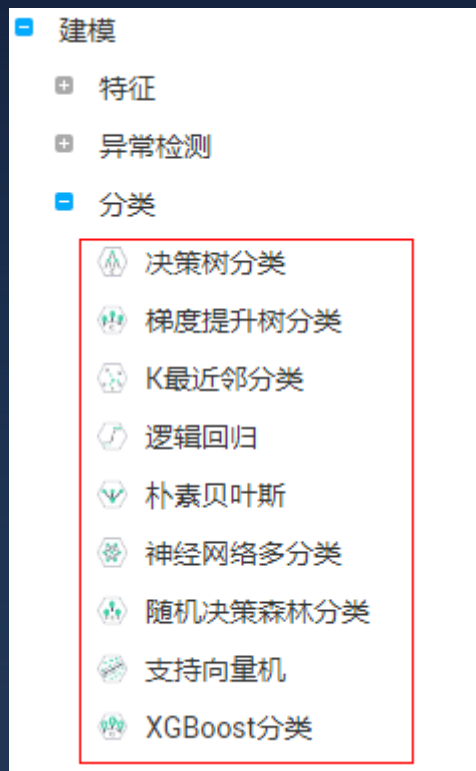
算法优化

3

建模

# 算法优化

- 1、课程4.2中采用的是随机决策森林分类算法，可通过改变算法的各项参数，观察分类结果。
- 2、更换算法，华为云MLS提供多种分类算法进行选择、尝试。



# 目录

## Contents

1

数据缺失处理

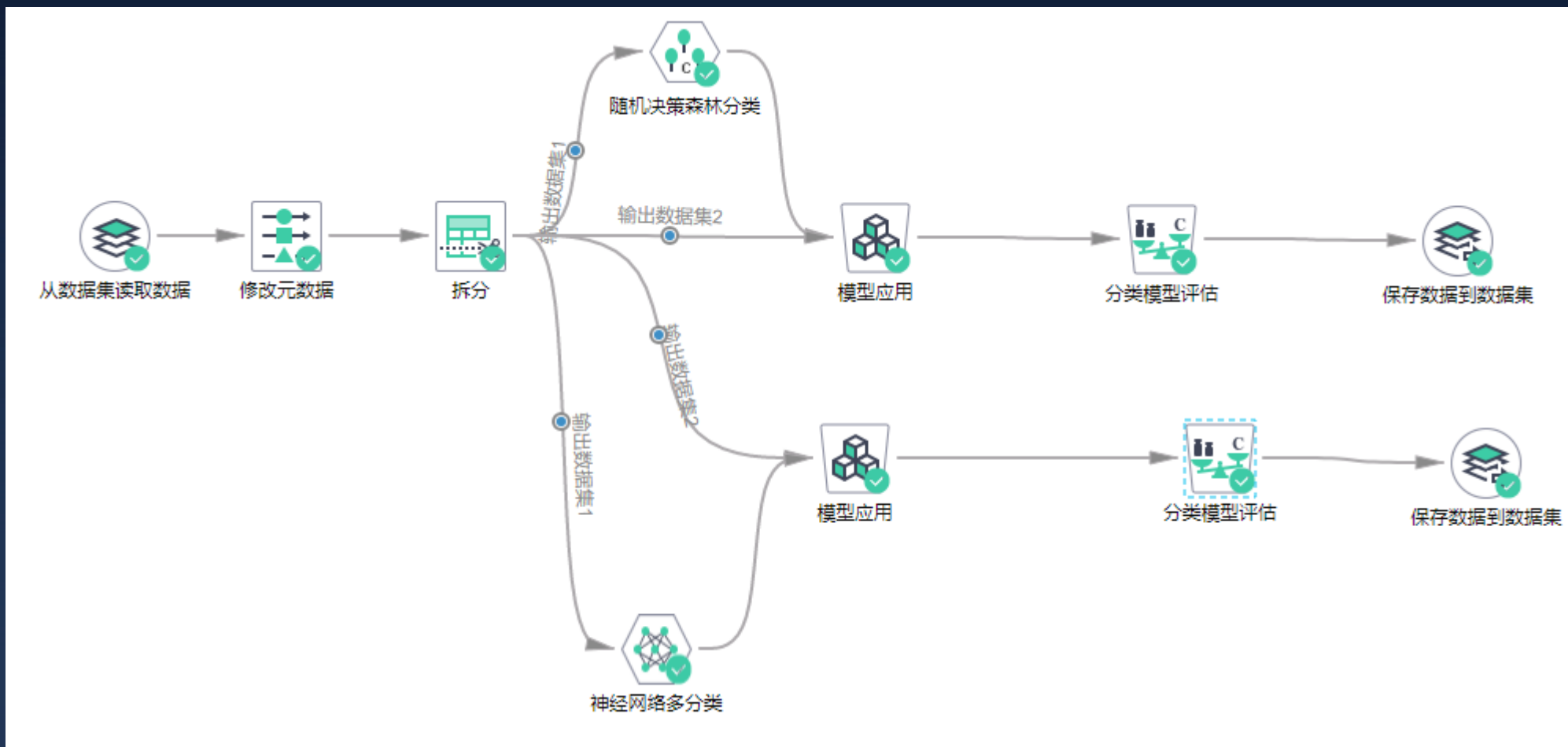
2

算法优化

3

建模

# 建模



详见操作指导。



# Thank You.

**Copyright©2016 Huawei Technologies Co., Ltd. All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

华为云机器学习服务MLS  
[www.huaweicloud.com/product/mls.html](http://www.huaweicloud.com/product/mls.html)