



2.1 机器学习助力 客户分群(上)

HUAWEI TECHNOLOGIES CO., LTD.



www.huawei.com

目录

Contents

1

客户分群场景定义

2

数据说明

3

算法与建模

客户分群场景定义

客户分群是指在营销、客户管理等场景下需要对客户进行分群，从而对不同的群体进行不同的营销策略制定、不同的管理方式等，一般来说进行客户分群时，客户本身的信息是不带标签的信息，需要使用聚类算法找到客户之间内在的联系，将相同的客户分在一起。

举例：在商品交易活动中，商品批发商会依据每个客户的年进货量大小，人工逐个判断为大客户还是小客户，再以此来确定第二年的销售策略和客户管理方针。随着买卖双方交易的时间变长，商品多种多样以及不同的客户之间差异较大，仅仅依靠商品批发商判断客户级别的大小，有可能出现遗漏或者分析结果偏差较大的情况。而且容易遗漏大客户，企业利润有下降的风险。需要使用基于机器学习的客户分群，更加细致和快捷区分出大、中、小客户，为提前指定适宜的商品销售策略和管理客户方针提供依据，最终提升客户满意度，带来更多的商业利润和深度合作伙伴关系。

目录

Contents

1

客户分群场景定义

2

数据说明

3

算法与建模

数据说明

字段说明

| 字段名 | 含义 | 类型 | 描述 |
|------------------|------|-----|------|
| Channel | 渠道 | Int | 渠道 |
| Region | 区域 | Int | 区域 |
| Fresh | 生鲜类 | Int | 生鲜类 |
| Milk | 奶制品 | Int | 奶制品 |
| Grocery | 杂货 | Int | 杂货 |
| Frozen | 冷冻类 | Int | 冷冻类 |
| Detergents_Paper | 洗涤类 | Int | 洗涤类 |
| Delicassen | 熟食类 | Int | 熟食类 |
| id | 客户ID | Int | 客户标识 |

数据说明

数据示例

| Channel | Region | Fresh | Milk | Grocery | Frozen | Detergents_Paper | Delicatessen | id |
|---------|--------|-------|------|---------|--------|------------------|--------------|----|
| 2 | 3 | 12669 | 9656 | 7561 | 214 | 2674 | 1338 | 1 |
| 2 | 3 | 7057 | 9810 | 9568 | 1762 | 3293 | 1776 | 2 |
| 2 | 3 | 6353 | 8808 | 7684 | 2405 | 3516 | 7844 | 3 |
| 1 | 3 | 13265 | 1196 | 4221 | 6404 | 507 | 1788 | 4 |
| 2 | 3 | 22615 | 5410 | 7198 | 3915 | 1777 | 5185 | 5 |
| 2 | 3 | 9413 | 8259 | 5126 | 666 | 1795 | 1451 | 6 |
| 2 | 3 | 12126 | 3199 | 6975 | 480 | 3140 | 545 | 7 |

目录

Contents

1

客户分群场景定义

2

数据说明

3

算法与建模

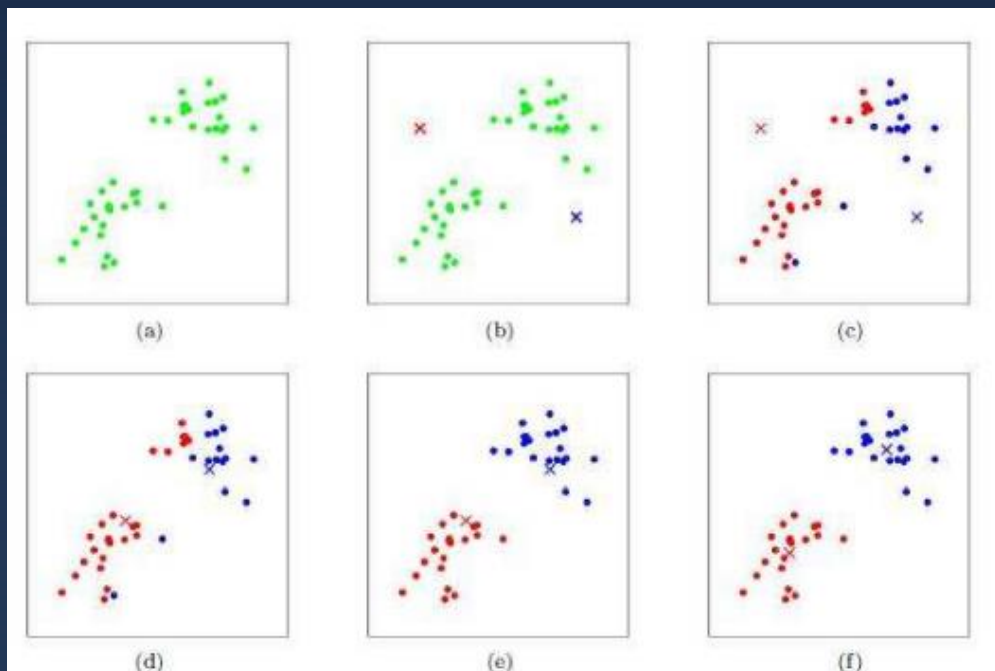
划分聚类算法：k-means 原理

输入：k（簇的数目）、要进行分类的数据集D

输出：k个簇的集合

过程：

- 1) 从D中以某种规则选择k个样本或者k个在值域范围内的点作为初始簇的中心
- 2) 计算簇中心之外的每个样本和每个簇中心的距离，将样本归属于最近的簇。
- 3) 计算簇内均值，将均值作为簇的中心。
- 4) 训练2和3，直到簇中心的变化小于一定的阈值或者即便簇中心变动，但是簇内样本不变动。



k-means的优缺点：

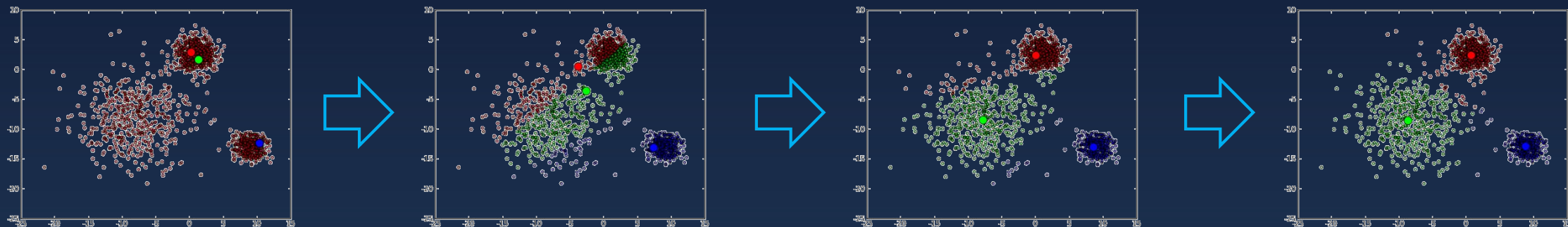
优点：

- 1、能够比较快速的收敛。
- 2、在大数据集上是相对可伸缩和有效的。
- 3、当无法计算均值时，可以通过定义一些簇中心（比如众数簇心）来改写算法。

缺点：

- 1、严重依赖于k值的确定。
- 2、不适合发现非球形簇或者大小差别非常大的簇。
- 3、对噪声和离群点非常的敏感。

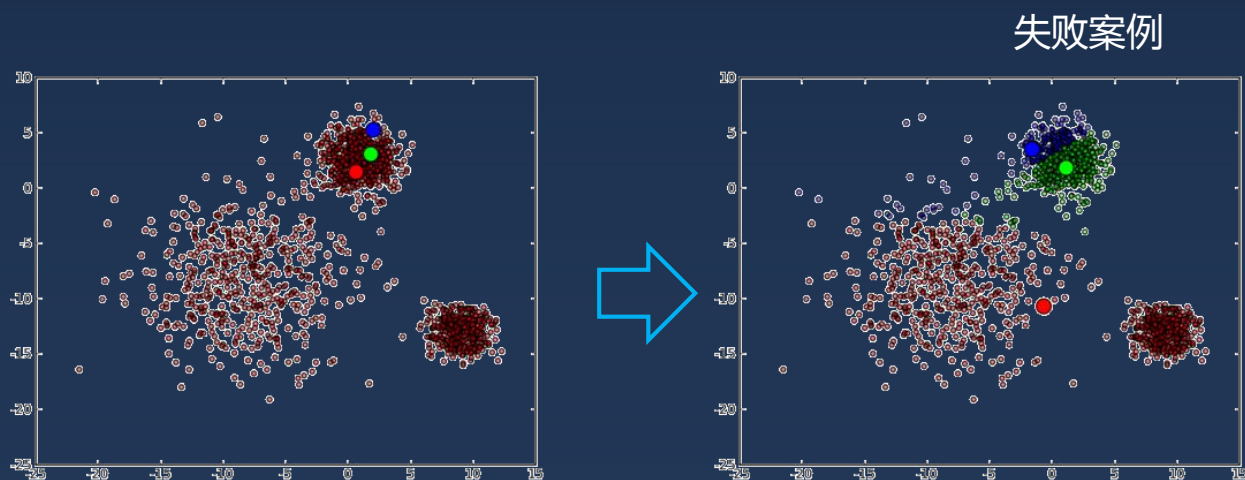
划分聚类算法：k-means 图示



最小化目标函数

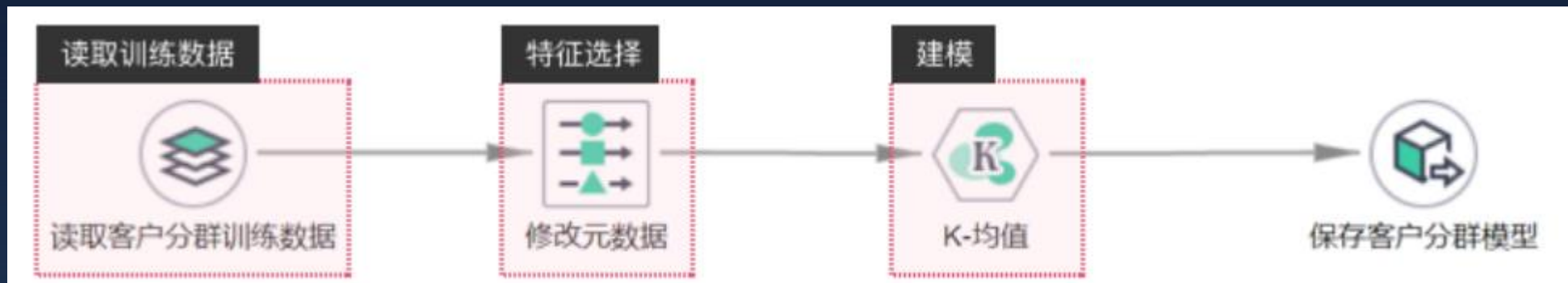
$$J = \sum_{n=1}^N \sum_{k=1}^K \tau_{nk} \|x_n - \mu_k\|^2$$

$$\mu_k = \frac{\sum_n \tau_{nk} x_n}{\sum_n \tau_{nk}}$$

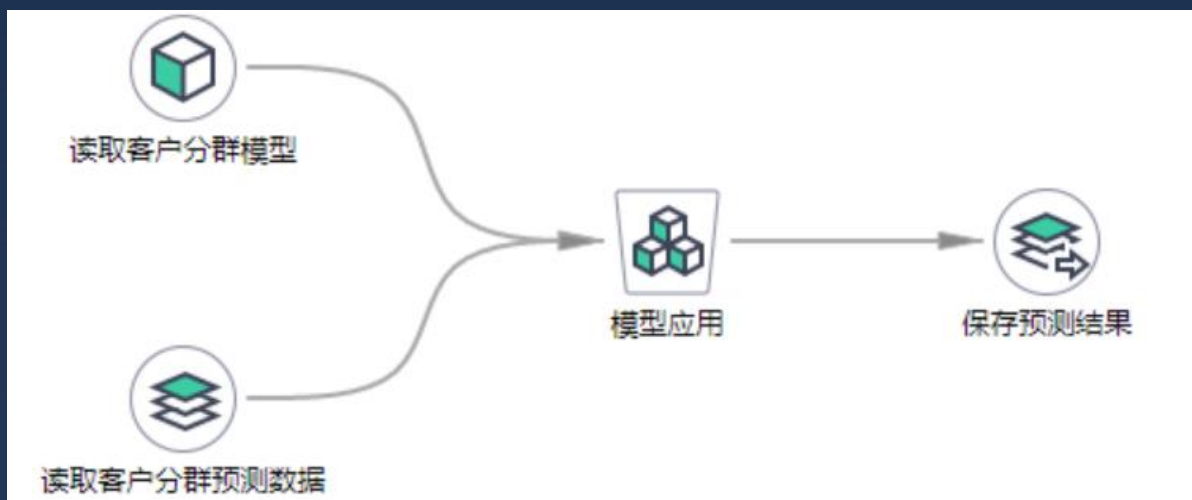


使用MLS进行客户分群的建模与预测

建模过程



预测过程



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