

市场调研与数据分析

“正大杯”第十二届全国市场调查与分析大赛公益培训



聚类分析

第十二届市赛公益培训课件
(Credamo 见数 版权所有)

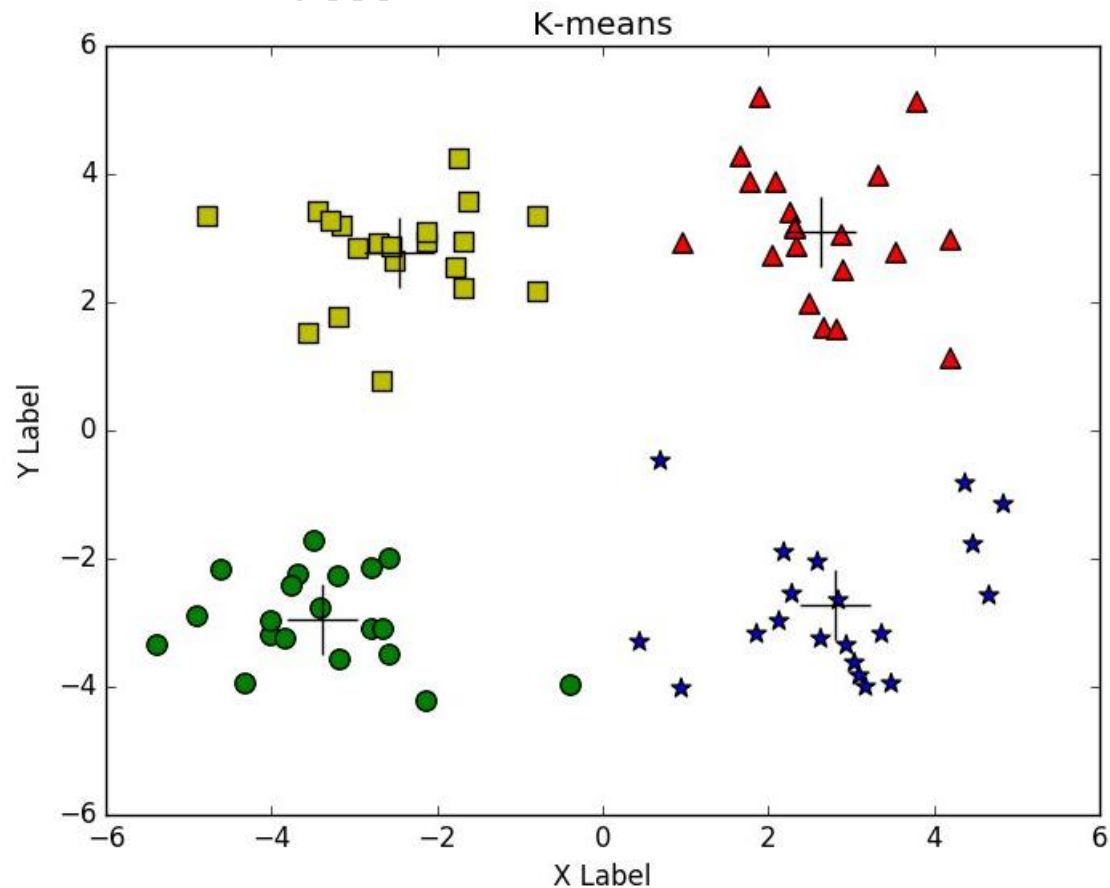
聚类分析

- **聚类分析** (cluster analysis) 是一种将研究对象分为相对同质的群组 (clusters) 的统计分析技术。
- **特征:** 聚类分析是一种探索性的分析, 在分类的过程中, 人们不必事先给出一个分类的标准, 聚类分析能够从样本数据出发, 自动进行分类。聚类分析所使用方法的不同, 常常会得到不同的结论。不同研究者对于同一组数据进行聚类分析, 所得到的聚类数未必一致。
- **目的:** 可用来发现不同的客户群, 并且通过购买模式刻画不同的客户群的特征; 可用于研究消费者行为, 寻找新的潜在市场; 通过分组聚类出具有相似浏览行为的客户, 并分析客户的共同特征, 可以更好的帮助电子商务的用户了解自己的客户, 向客户提供更合适的服务。

聚类分析

聚类分析算法

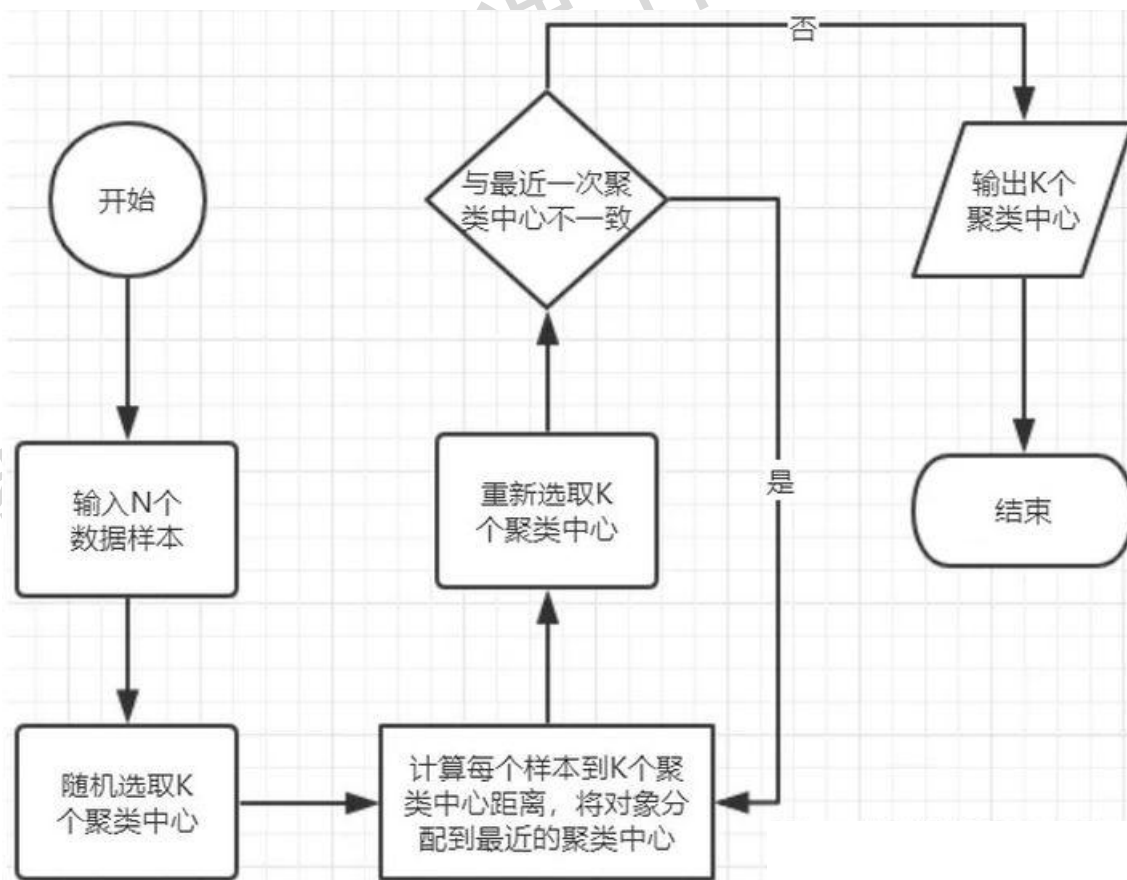
- 聚类分析是数据挖掘中的一个很活跃的研究领域，并提出了许多聚类算法。传统的聚类算法可以分为五类：划分方法、层次方法、基于密度方法、基于网格方法和基于模型方法。
- 最常用且高效的分析算法为划分方法中的k-means（K均值）算法，首先创建k个划分，k为要创建的划分个数；然后利用一个循环定位技术通过将对象从一个划分移到另一个划分来帮助改善划分质量。



聚类分析

K均值算法原理

- K-Means 算法又称为快速聚类法，是基于最小误差法的原则将数据划分为预定的类别K。这种算法原理简单且计算高效。
- Scikit-learn 中的 K-Means算法使用欧式距离去度量样本到聚类中心的距离，并把误差平方和SSE作为度量聚类效果的目标函数，选取误差平方和最小的分类结果作为最终的聚类结果。

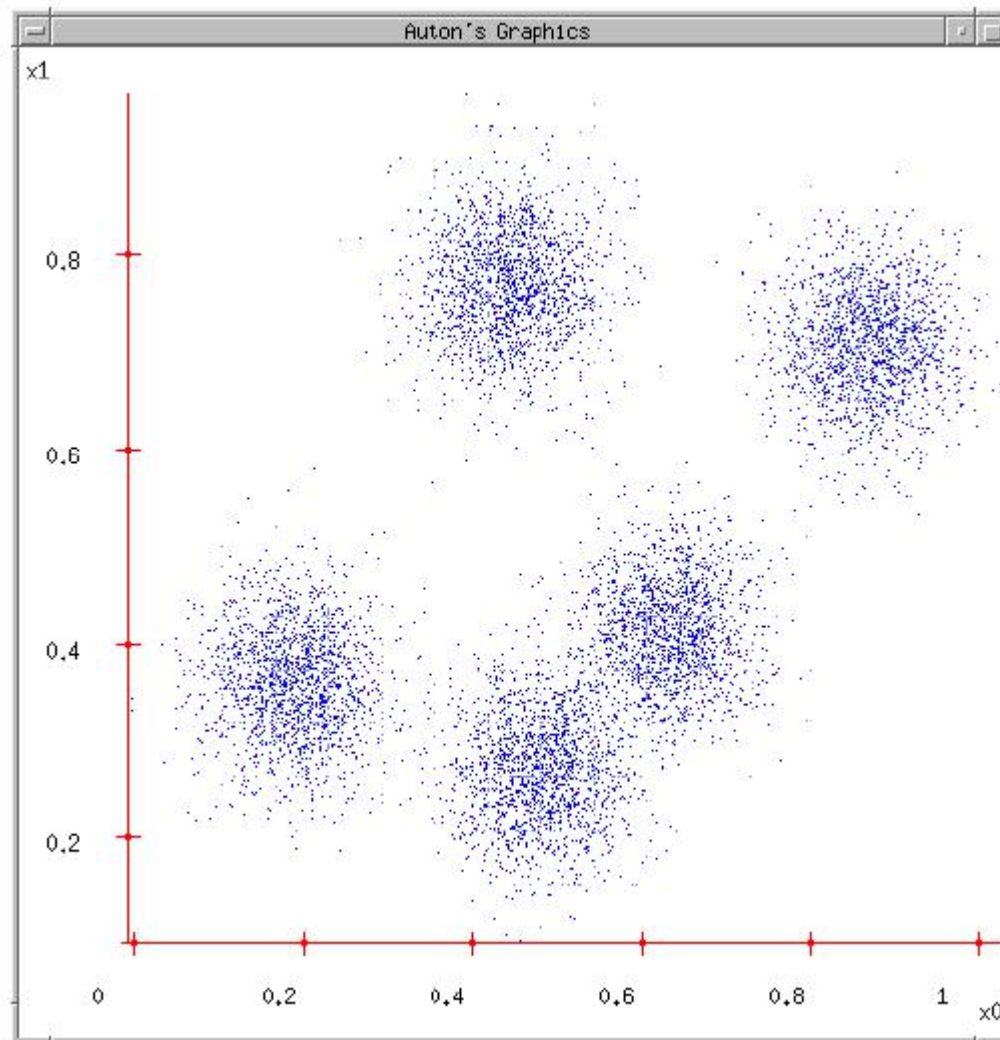


K均值算法过程图

聚类分析

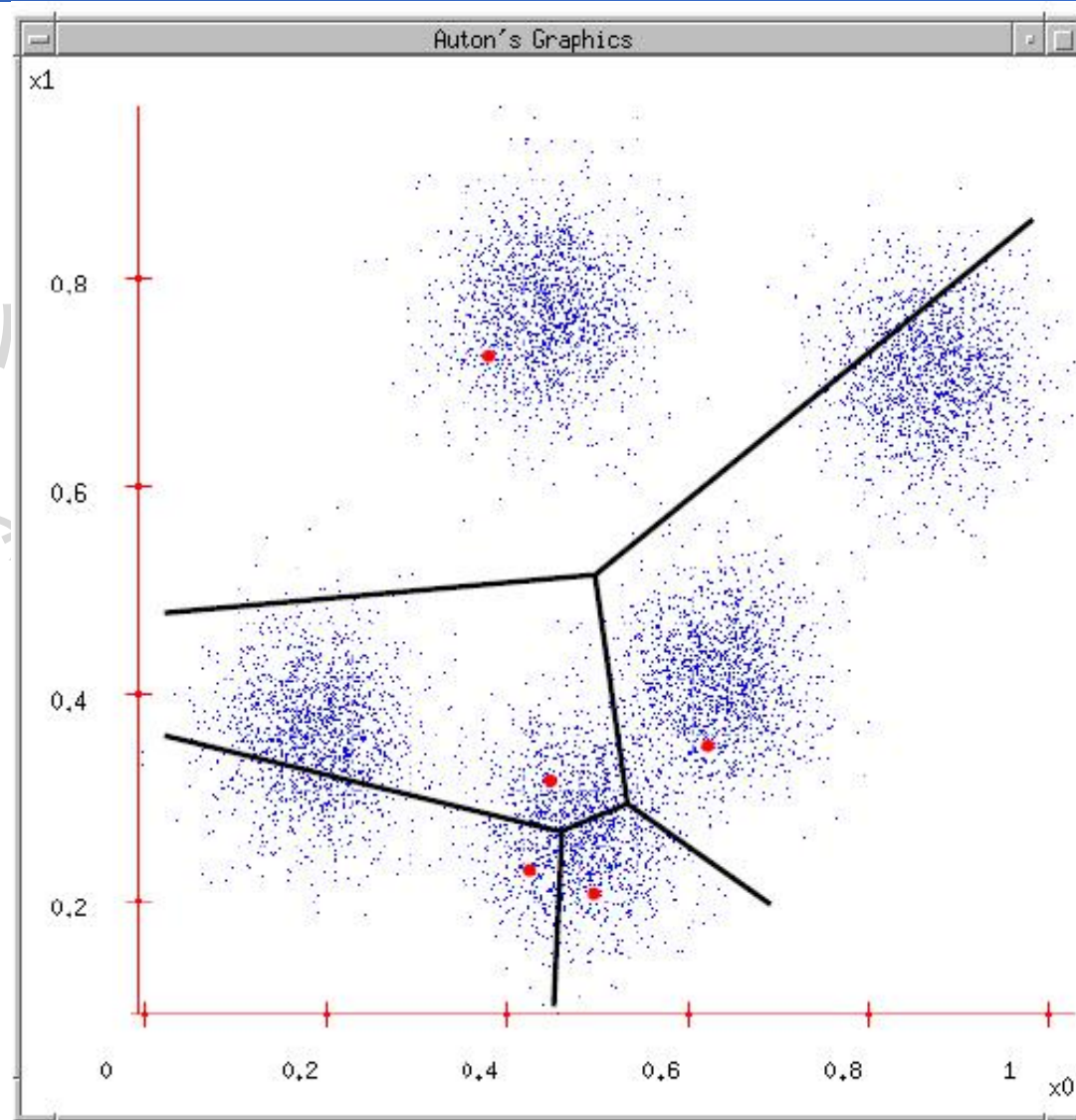
确定类别个数 $K=5$

随机选择5个初始类中心



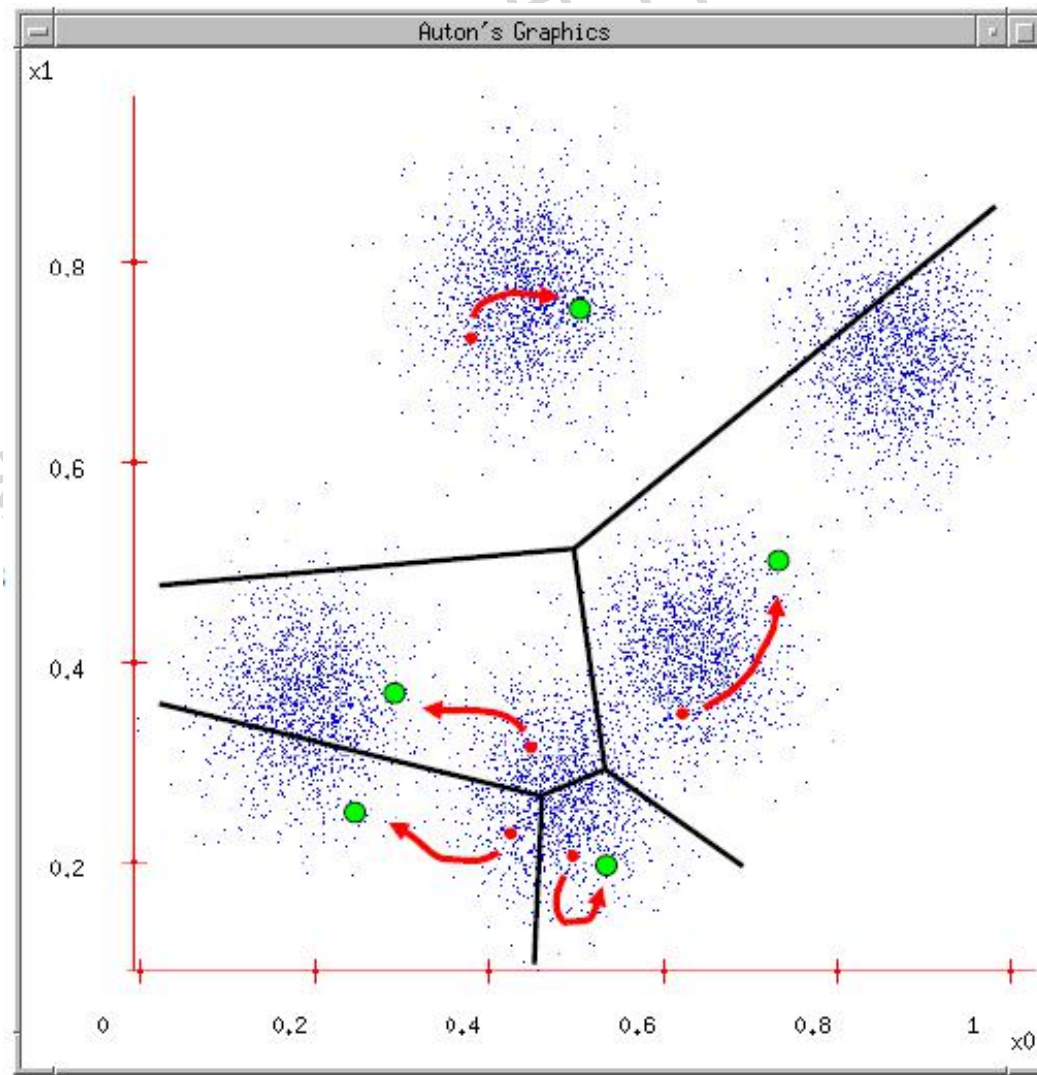
聚类分析

对于每个样本点
计算它与5个类中心的距离
然后将其分配到最近的一类
里面



聚类分析

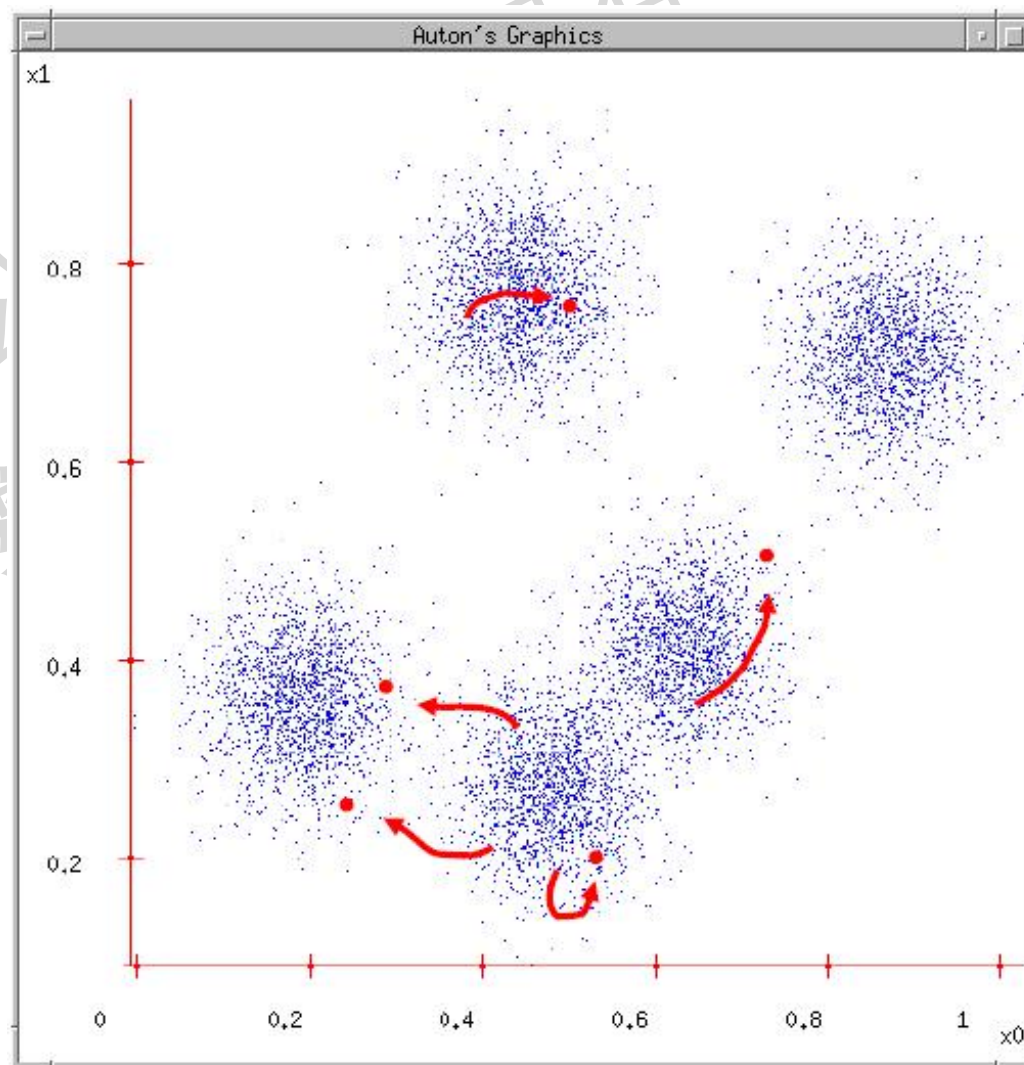
寻找当前每个类别的类中心



聚类分析

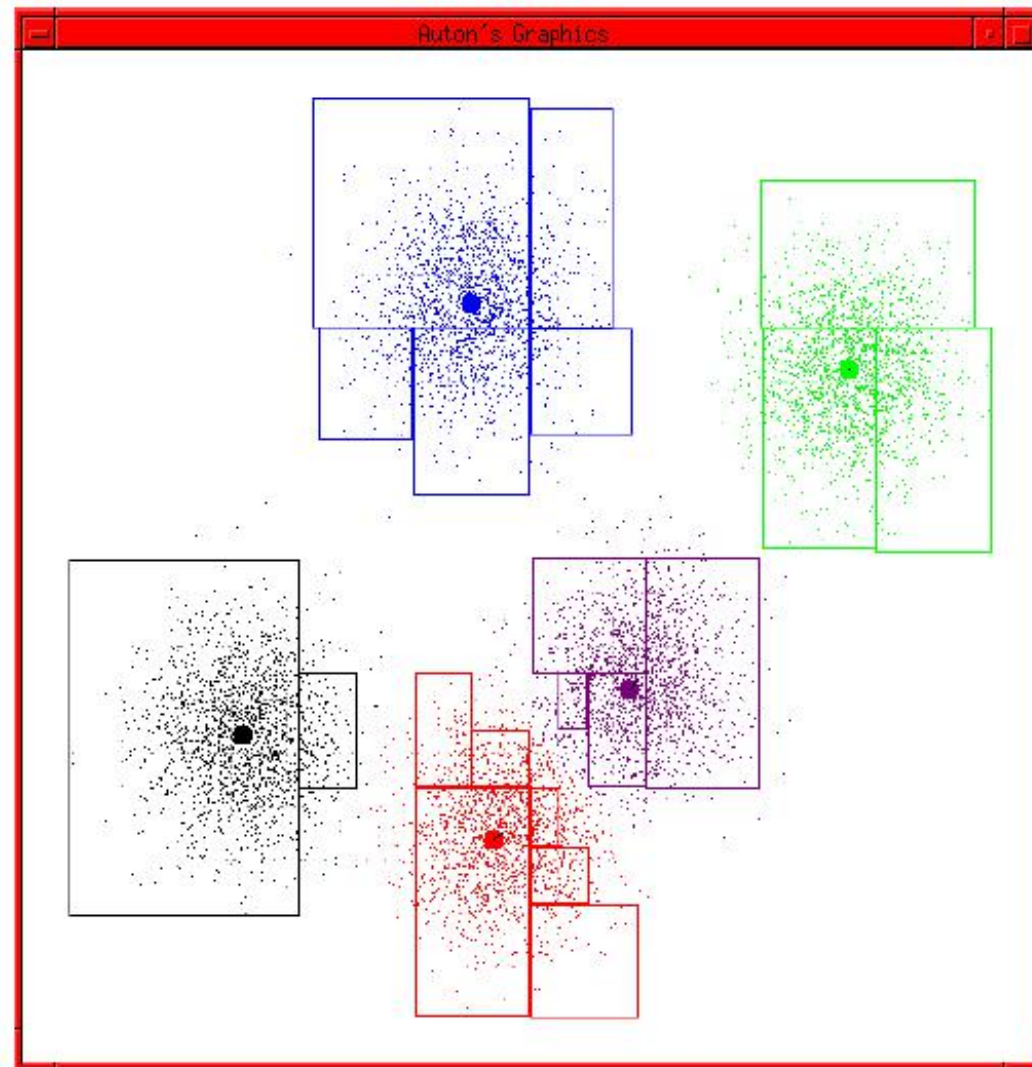
重复之前的步骤

直到类中心的改变很小或达到循环次数



聚类分析

收敛之后的结果



聚类分析

聚类该聚成几类？

- 首先应该明确自己的分析目标，初步制定初始分类个数，即需要将样本划分为几类。如做欺诈聚类，那么可以把客户分为2类：有欺诈倾向/无欺诈倾向；如做营销数据分析，可以把客户分为优质/良好/一般/较差等。
- 其次可以围绕初始类别个数进行上下浮动，并挑选几组极具有业务含义的特征进行散点图和概率密度分布图的观测，从中选取业务意义最好的分类个数。



聚类分析

案例演示

目的：在手机广告中，根据**购买意愿**、**说服力**和**收入水平**对用户进行聚类分析

方法：利用Credamo的建模工具，在聚类分析中聚类变量选择**说服力**和**购买意愿**，标签展示变量选择**收入水平**，系统即可自动生成聚类模型。

The screenshot shows the Credamo modeling tool interface with the following configuration:

- 分析方法** (Analysis Method): 聚类分析 (Clustering Analysis) [修改]
- 分析名称** (Analysis Name): 聚类分析1
- 分析配置** (Analysis Configuration): 聚类变量仅支持数值变量，标签展示变量仅支持分类变量
- 聚类变量** (Clustering Variables):
 - 作答时长
 - 趣味性2
 - 信息性1
 - 信息性3
 - 态度2
 - 信任度1
 - 信任度3
 - 说服力2
 - 购买意愿1
 - 趣味性
 - 态度
 - 说服力
- 聚类变量** (Clustering Variables):
 - 趣味性1
 - 趣味性3
 - 信息性2
 - 态度1
 - 态度3
 - 信任度2
 - 说服力1
 - 说服力3
 - 购买意愿2
 - 信息性
 - 信任度
 - 购买意愿
- 标签展示变量** (Label Display Variables):
 - 作答ID
 - 用户名
 - 结束时间
 - 发布ID
 - IP
 - 纬度
 - 城市
 - 操作系统类型
 - 屏幕分辨率
 - 性别
 - 职业
 - Q5
 - 信息型
 - Q9
 - 用户ID
 - 开始时间
 - 作答渠道
 - 问卷发布名称
 - 经度
 - 省份
 - 设备类型
 - 浏览器类型
 - 随机元素
 - 年龄
 - 月收入
 - Q6
 - 情感型
 - Q10

聚类分析

案例演示一结果分析（一）

- 在自动生成的结果中，可以明显看到样本被聚为三类，三类样本所占的百分比和频数也由图可知，可得类别_1占比最大，达到47.5%。

聚类结果汇总表

下载数据

聚类类别	频数	百分比 (%)
类别_1	95	47.5
类别_2	20	10
类别_3	85	42.5

聚类分析

案例演示一结果分析（二）

- 由两个变量的p值均小于0.05可知分析结果显著，三类聚类类别都具有显著的差异性。
- 由具体的变量均值可将聚类后的三类样本类别进行定义和评估。

聚类类别差异对比和方差分析结果

下载数据

变量	变量均值±标准差			F	p
	聚类类别1 (n=95)	聚类类别2 (n=20)	聚类类别3 (n=85)		
说服力	5.04±0.61	2.53±1.13	6.21±0.4	311.5844	0.0000
购买意愿	4.98±0.68	2.2±0.88	6.29±0.48	361.6607	0.0000

聚类分析

案例演示—结果分析（三）

- 生成的智能解读与之前的结果分析相符合，得到结论聚得的三类聚类变量具有显著的差异性。

智能解读

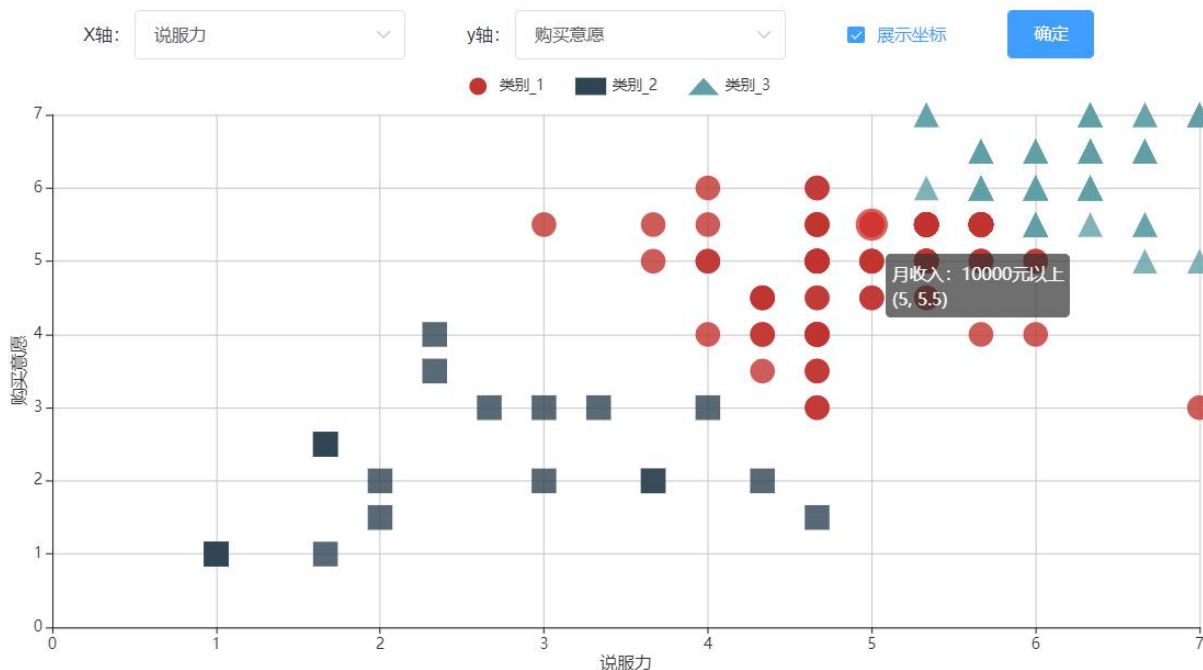
通过K-means法对样本进行聚类分析，将样本处理为3类。对聚类类别和各聚类变量进行方差分析，由表中可知，方差分析结果全部显著，说明聚类得到的3类群体在所有聚类变量上都具有显著的差异性。

聚类分析

案例演示一结果分析（四）

- 结果图将聚类结果进行了可视化。各个聚类类别的样本点以不同的形状和颜色展示在图上，可以将鼠标移动到样本点上，以查看该样本的标签展示变量信息。

聚类结果图





CBC联合分析

第十二届市桥杯赛公益培训课件
(Credamo 版权所有)

产品组合



价格
屏幕大小
颜色
性能
...

CBC联合分析



价格: 1000 2000 3000 4000

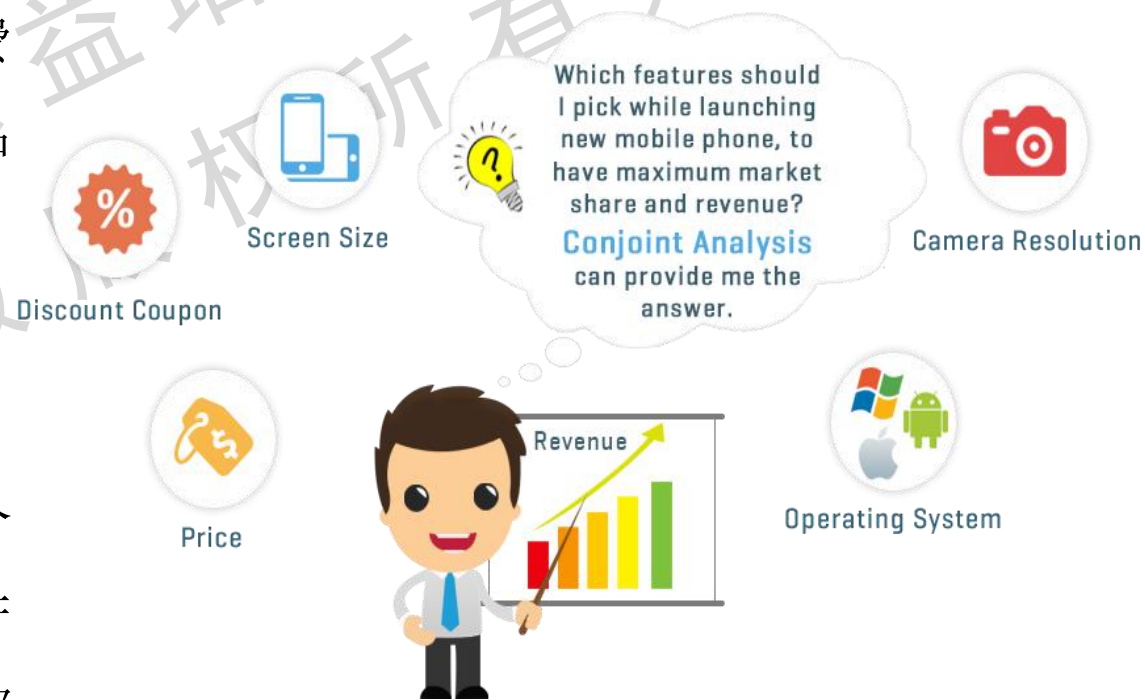
屏幕大小: 4.5 5.5 6.5

内存: 64M 128M 256M

颜色: 白色 黑色 红色 玫瑰金

CBC联合分析

- **联合分析法** (Choice-Based Conjoint analysis) 是一种多元的统计分析方法，是一种定量研究消费者选择偏好的方法，用于估测消费者对特定的产品属性组合的相对重要性或效用的评价。
- **应用：**联合分析已经广泛应用于消费品、工业品、金融以及其它服务等领域。在现代市场研究的各个方面，如新产品的概念筛选、开发，竞争分析，产品定价，市场细分，广告，分销，品牌等领域，都可见到联合分析的应用。



基本假定

- 消费者是根据构成产品/服务的多个属性来进行感知和作偏好判断。也就是说，消费者对产品/服务的偏好并不是基于一个因素，而是基于几个因素的结合来判断的。单独或直接的询问消费者各个属性的重要性，是不现实，甚至是无用的。
- 所以在市场研究中，需要模拟“消费者在面对不同的产品 / 服务及其组合时，参考多种属性的结合效应之后做出选择”的过程，通过提供给消费者以不同的属性水平组合形成的产品，并请消费者做出判断，因此需要消费者回答的问题更加贴近消费者真实的消费决策。

CBC联合分析

主要步骤

- **1. 确定产品特征与特征水平：**联合分析首先要对产品或服务的特征进行识别，这些特征与特征水平必须是显著影响消费者购买的因素。
- **2. 产品模拟：**联合分析将产品的所有特征与特征水平通盘考虑，并采用正交设计的方法将这些特征与特征水平进行组合，生成一系列虚拟产品。
- **3. 数据收集：**请受访者对虚拟产品进行评价，通过打分、排序等方法调查受访者对虚拟产品的喜好、购买的可能性等。
- **4. 计算特征的效用：**从收集的信息中分离出消费者对每一特征以及特征水平的偏好值，这些偏好值也就是该特征的“效用”。
- **5. 市场预测：**利用效用值来预测消费者将如何在不同产品中进行选择，从而决定应该采取的措施。

CBC联合分析

功能与特征

- 联合分析主要解决了三个问题：
 - (1) 测量消费者在多属性选择时的偏好。
 - (2) 确定消费者在多属性产品之间作出利弊选择。
 - (3) 模拟市场，开发消费者最喜欢的产品/服务。
- 基于选择的工作类似购买者在市场上的行为，从一组产品中选择喜欢的产品，是每个人可以理解的简单、自然的工作。
- 如果在一组概念中没有受访者喜欢的产品，CBC允许受访者选择“none”，从而更加真实的模拟了消费者购买产品的过程。



CBC联合分析

案例演示

目的：通过联合分析得到更具消费者吸引力的手机属性特征组合。

方法：利用Credamo的建模工具，在应用列表中选择联合分析，将问卷导入后可创建联合分析项目，而后在右上角的查看报告中查看分析结果。

个人资料 我的标签 账户信息 交易记录 问卷记录 订单管理 发票信息 **应用列表** 账户升级

数据模型

- 高级账户专享
结构方程模型
使用结构方程模型完成多元数据分析
查看帮助 立即使用
- 高级账户专享
联合分析
基于CBC联合分析精准 规划营销方案
查看帮助 立即使用
- 高级账户专享
语音识别
将录音文件智能识别为文字
查看帮助 立即使用

基本设置

项目标题 手机广告

内容简介 通过联合分析探究手机消费者在手机广告中的受影响情况

问卷 手机广告测试研究4.0 选择问卷 + 创建新问卷

确定创建 取消创建

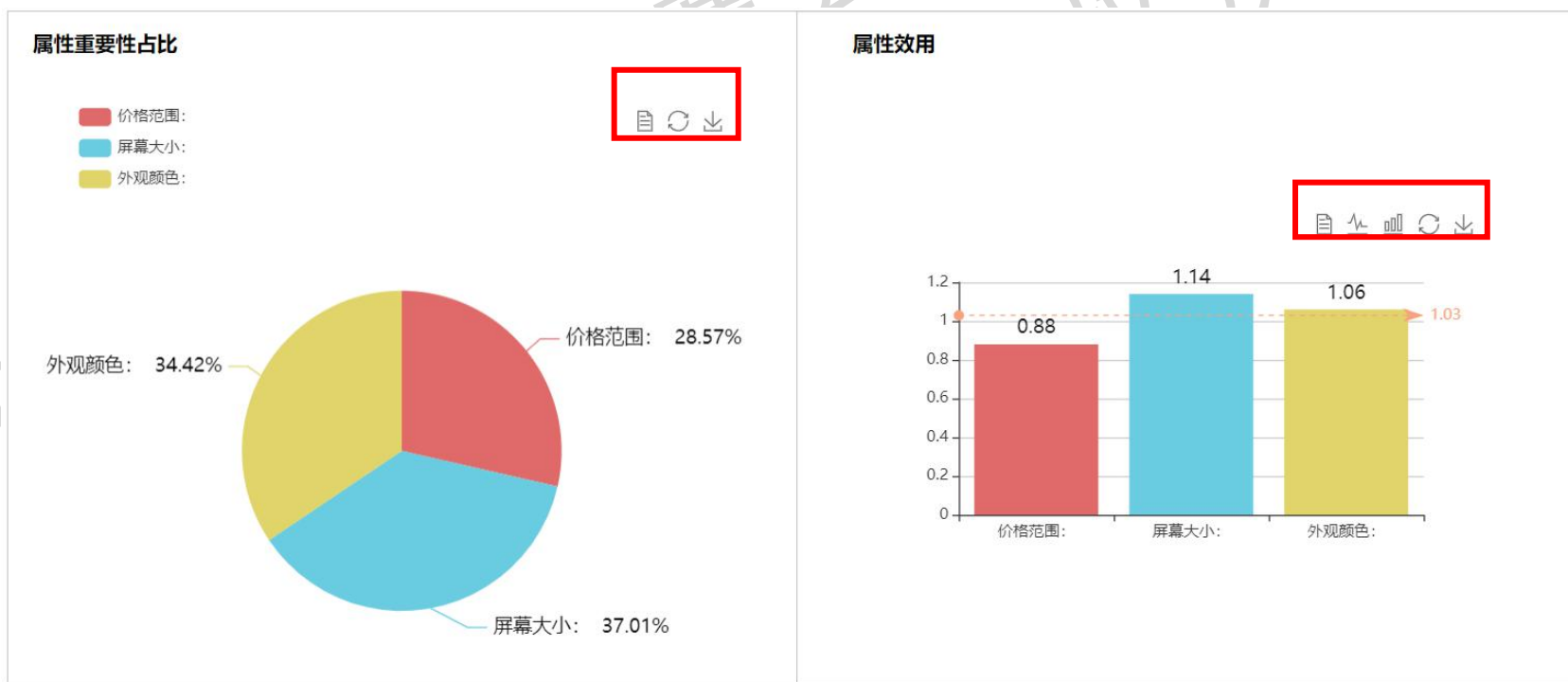
CBC联合分析

- 如果需要平衡性别、地域等被试特征，可选择多次发布

编号	发布渠道	计划发布量
1	数据集市	800
2	数据集市	400
3	数据集市	1200
4	数据集市	50

案例演示—属性分析

- 在联合分析的属性分析中，可以看到智能分析得到的属性重要性占比、属性效用和属性评分，也可通过图示右上角的功能框选择改变图示类型或保存图片。

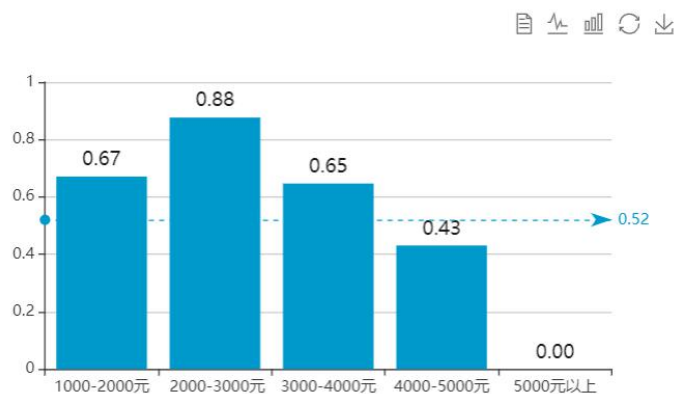


CBC联合分析

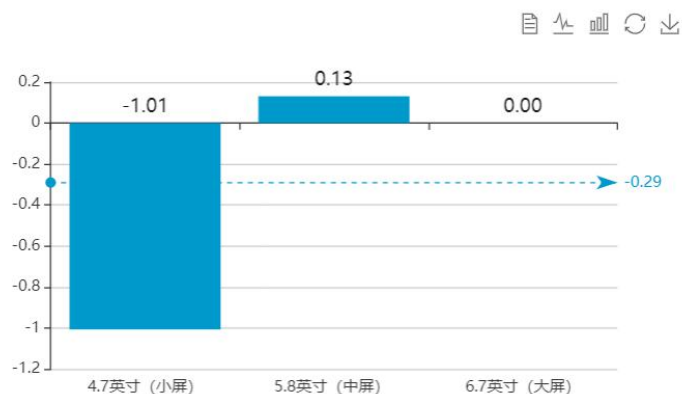
案例演示—水平分析

- 在联合分析的属性分析中，可以看到各属性不同类别间的横向效用对比情况。以最后一个类别为基准，其他条件不变时，从一个类别更换到基准类别时被调查者选择的几率会随之发生改变，变动数值与类别效用相关。虚线所代表的是各类别的平均效用。

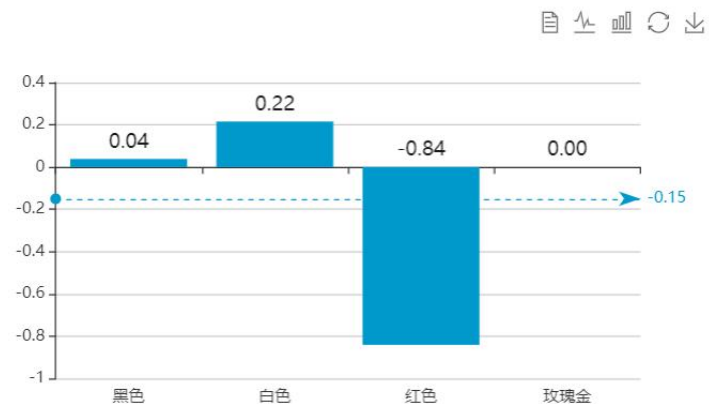
属性: 价格范围:



属性: 屏幕大小:



属性: 外观颜色:

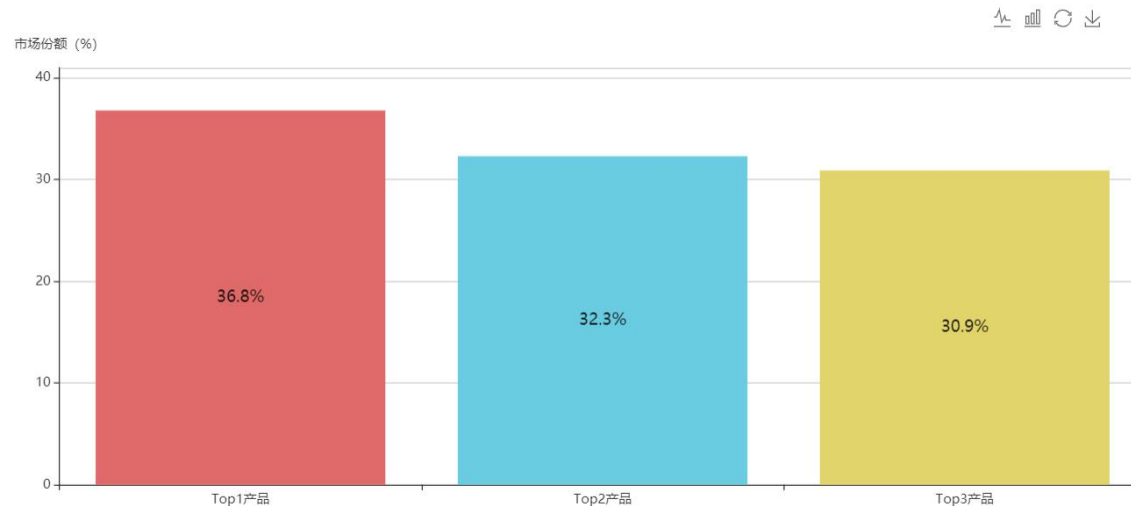


案例演示—市场模拟

- 在联合分析的市场模拟中，Credamo会自动生成市场模拟得到的top3推荐产品，并预测各产品依次能占领的市场份额，每一件产品对应各自不同的属性。也可在右上角的修改产品信息中改变推荐的产品数量或是选择自定义产品。

市场模拟结果

修改产品信息



案例演示—精准营销

- 在联合分析的精准营销项中，会根据市场模拟得到的top3产品来进行精准营销。同时可通过右上角的选择市场细分变量，来根据业务情况细分所需要的变量，达到精准触达目标人群的营销效果。

精准营销结果

选择市场细分变量

细分市场 (Q: N/A) 的相对规模为 100.0%



选择市场细分变量

请选择市场细分变量 (至多2个)

- Q
- 手机品牌
- 性别
- 年龄
- 职业
- 收入水平
- 甄别题
- 推荐与否

取消

确定

谢谢大家

关注右边公众号



及时了解大赛资讯和进程

随时学习大赛公益培训

中国商业统计学会

官方网址：<http://www.china-cssc.org>

公众号：



Credamo见数

官方网址：www.credamo.com

公众号：

