



3.1 机器学习助力 客户分群(下)

HUAWEI TECHNOLOGIES CO., LTD.



www.huawei.com

目录

Contents

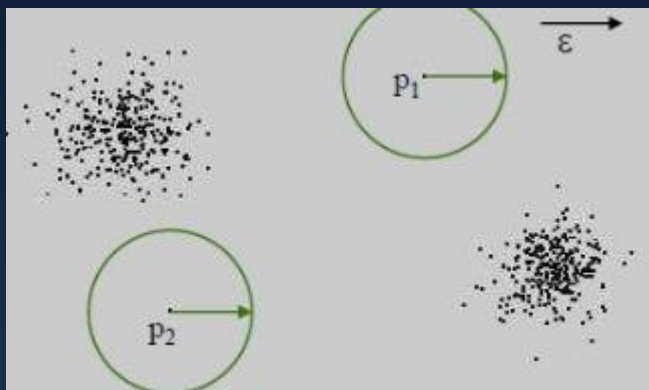
1

离群点

2

K-means算法初始化策略

离群点



离群点概率定义：离群点是少数异常于正常数据集的数据对象，在概率分布模型中，具有低概率。

顾名思义，离群点就是几乎“远离”绝大部分数据点的数据。

离群点的影响：

- 1、会影响分群簇心的变动，使得簇心远离了正常数据。
- 2、会影响群内数据属性的分析，从而不能很好的定义每个群。

目录

Contents

1

离群点

2

K-means算法初始化策略

k-means 初始化方法

K值设定：

提供一些基于经验的k值，或者一个k值的范围，通过比较由不同的k得到的聚类结果，来确定最佳的k值。

初始簇心选择：两种方式

- 1、随机初始簇心：k个簇的簇心的初始化都是随机的，随机选择k个样本作为簇心。
- 2、有限最大距离簇心：
 - 1) 随机选择1个样本作为第1个簇心
 - 2) 随机选择m个样本，计算这些样本和当前所有簇心的距离dist
 - 3) 在2的计算结果中选择dist最大的样本作为下1个簇心。
 - 4) 循环2和3，直到选出k个簇心。

终止规则：两种方式

- 1、完全终止，即在当前k值设定下，簇完全不发生更新时停止
- 2、设定一个次数，当簇的更新次数达到此值时即停止

K-means优化

- 1、检测离群点，删除离群点，优化簇的聚集程度。
- 2、尝试不同的k-means初始化方法，以寻找更好的簇。

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