



K均值聚类算法（附例子）-b站：笔记鲨

1. 快速理解
2. 算法步骤
3. 实例讲解

低价高质

多年经验

名校硕博

多元背景



孙梁
机器学习
算法

不想学代码？
不想装环境？
我们可以帮您解决：

机器学习
神经网络
统计分析
自动化程序
量化研究

Spark/Hadoop

哥伦比亚大学

南加州大学

南开大学



数据科学

统计科学

金融工程

关注**笔记鲨**的微信公众号
点击左侧“建立联系”加微信交流

- K均值聚类算法 K-Means Clustering Algorithm

- ▼ 快速理解：

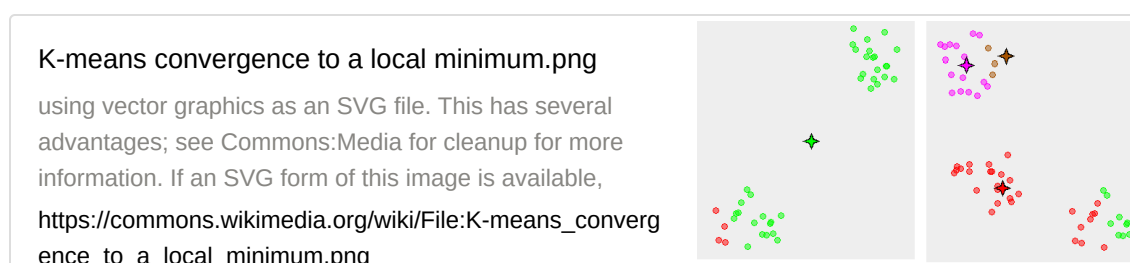
1. 有四个牧师去郊区布道，一开始牧师们随意选了几个布道点，并且把这几个布道点的情况公告给了郊区所有的居民，于是每个居民到离自己家最近的布道点去

听课。

2.听课之后，大家觉得距离太远了，于是每个牧师统计了一下自己的课上所有的居民的地址，搬到了所有地址的中心地带，并且在海报上更新了自己的布道点的位置。

3.牧师每一次移动不可能离所有人都更近，有的人发现A牧师移动以后自己还不如去B牧师处听课更近，于是每个居民又去了离自己最近的布道点.....就这样，牧师每个礼拜更新自己的位置，居民根据自己的情况选择布道点，最终稳定了下来。

图片理解：



。 步骤

- 先定义总共有多少个类/簇 (**cluster**)
- 将每个簇心 (**cluster centers**) 随机定在一个点上
- 将每个数据点关联到最近簇中心所属的簇上
- 对于每一个簇找到其所有关联点的中心点 (取每一个点坐标的平均值)
- 将上述点变为新的**簇心**
- 不停重复，直到每个簇所拥有的点**不变**

▼ 例子

题目：有以下6个点，将A3和A4作为两个簇的初始簇心。问最后的簇的所属情况

	X	Y
A1	1	2
A2	1	4
A3	3	1
A4	3	5
A5	5	2
A6	5	4

1.计算每个点到簇心距离（根据距离公式），将距离近的点归为一类

	X	Y	G1 distance	G2 distance
A1	1	2	2.24	3.61
A2	1	4	3.61	2.24
A3	3	1	0.00	4.00
A4	3	5	4.00	0.00
A5	5	2	2.24	3.61
A6	5	4	3.61	2.24

2.将蓝色每个点，和紫色每个点的X,Y值分别求平均。获得新的簇心

	x	y
new M1	3.00	1.67
new M2	3.00	4.33

3.计算每个点到簇心的新距离，将距离近的点归为一类

	X	Y	G1 distance	G2 distance
A1	1	2	2.03	3.07
A2	1	4	3.07	2.03
A3	3	1	0.67	3.33
A4	3	5	3.33	0.67
A5	5	2	2.03	3.07
A6	5	4	3.07	2.03

4.由于关联点没有变化，所以之后的计算结果不会改变。停止计算。

5. 蓝色簇:A1,A3,A5。紫色簇：A2,A4,A6。

低价高质

多年经验

计算机科学

名校硕博

多元背景

不想学代码?
不想装环境?
我们可以帮您解决:

机器学习
神经网络
统计分析
自动化程序
量化研究
Spark/Hadoop

哥伦比亚大学
南加州大学
南开大学

APACHE
Spark™

数据科学
统计科学
金融工程

关注笔记鲨的微信公众号
点击左侧“建立联系”加微信交流