

原

测试数据科学家聚类技术的40个问题（能力测验和答案）（上）

2017年03月24日 12:13:06

阅读数：343

介绍

创造出具有自我学习能力的机器——人们的研究已经被这个想法推动了十几年。如果要实现这个梦想的话，无监督学习和聚类将会起到关键性作用。但是，无监督学习在带来许多灵活性的同时，也带来了更多的挑战。

在从尚未被标记的数据中得出见解的过程中，聚类扮演着很重要的角色。它将相似的数据进行分类，通过元理解来提供相应的各种商业决策。

在这次能力测试中，我们在社区中提供了聚类的测试，总计有1566人注册参与过该测试。如果你还没有测试过，通过阅读下面的文章，你可以统计一下自己能正确答对多少道题。

总结果

下面是分数的分布情况，可以帮你评估你的表现：



你也可以通过访问[这里](#)来查看自己的成绩。超过390个人参加了测试，最高分数是33分。下面是对分数分布的部分统计。

总分布：

平均分：15.11  
中位数：15  
模型分数：16

相关资源：

- [An Introduction to Clustering and different methods of clustering](#)
- [Getting your clustering right \(Part I\)](#)
- [Getting your clustering right \(Part II\)](#)

Questions & Answers

Q1. 电影推荐系统是以下哪些的应用实例：

1. 分类
2. 聚类
3. 强化学习
4. 回归

选项：

- C. 1和3
- D. 2和3
- E. 1 2 3
- F. 1 2 3 4

答案：E

一般来说，电影推荐系统会基于用户过往的活动和资料，将用户聚集在有限数量的相似组中。然后，从根本上来说，对同一集群的用户进行相似的推荐。

在某些情况下，电影推荐系统也可以归为分类问题，将最适当的某类电影分配给特定用户组的用户。与此同时，电影推荐系统也可以视为增强学习问题，即通过先前的推荐来改进以后的电影推荐。

Q2. 情感分析是以下哪些的实例：

1. 回归
2. 分类
3. 聚类
4. 强化学习

- A. 只有2
- B. 1和2
- C. 1和3
- D. 2和3
- E. 1 2 3
- F. 1 2 3 4

答案：E

在基本水平上的情感分析可以被认为是将图像、文本或语音中表示的情感，分类成一些情感的集合，如快乐、悲伤、兴奋、积极、消极等。同时，它也可以被视为对相应的图像、文本或语音按照从1到10的情感分数进行回归。

另一种方式则是从强化学习的角度来思考，算法不断地从过去的情感分析的准确性上进行学习，以此提高未来的表现。

Q3. 决策树可以用来执行聚类吗？

- A. 能
- B. 不能

答案：A

决策树还可以用在数据中的聚类分析，但是聚类常常生成自然集群，并且不依赖于任何目标函数。

Q4. 在进行聚类分析之前，给出少于所需数据的数据点，下面哪种方法最适合用于数据清理？

1. 限制和增加变量
2. 去除异常值

选项：

- A. 1
- B. 2
- C. 1和2
- D. 都不能

答案：A

在数据点相对较少的时候，不推荐去除异常值，在一些情况下，对变量进行剔除或增加更合适。

Q5. 执行聚类时，最少要有多少个变量或属性？

- A. 0
- B. 1
- C. 2
- D. 3

-----

HTML5高薪有多高

Java高薪有多高

师徒1对1

Python薪资多少钱

数据分析师培训

心理抑郁测试

**Q6. 运行过两次的K均值聚类，是否可以得到相同的聚类结果？**[登录](#)[注册](#)[×](#)

- A. 是
- B. 否

**答案：B**

K均值聚类算法通常会对局部最小值进行转换，个别时候这个局部最小值也是全局最小值，但这种情况比较少。因此，更建议在绘制集群的推断之前，多次运行K均值算法。

然而，每次运行K均值时设置相同的种子值是有可能得出相同的聚类结果的，但是这样做只是通过对每次的运行设置相同的随机值来进行简单的算法选择。

**Q7. 在K均值的连续迭代中，对簇的观测值的分配没有发生改变。这种可能性是否存在？**

- A. 是
- B. 否
- C. 不好说
- D. 以上都不对

**答案：A**

当K均值算法达到全局或局部最小值时，两次连续迭代所产生的数据点到簇的分配不会发生变化。

**Q8. 以下哪项可能成为K均值的终止条件？**

1. 对固定数量的迭代。
2. 在局部最小值不是特别差的情况下，在迭代中对簇观测值的分配不发生变化。
3. 在连续迭代中质心不发生变化。
4. 当 RRS 下降到阈值以下时终止。

选项：

- A. 1 3 4
- B. 1 2 3
- C. 1 2 4
- D. 全部都是

**答案：D**

这四种条件都可能成为K均值聚类的终止条件：

1. 这个条件限制了聚类算法的运行时间，但是在一些情况下，由于迭代次数不足，聚类的质量会很差。
2. 在局部最小值不是特别差的情况下，会产生良好的聚类，但是运行时间可能相当长。
3. 这种条件要确保算法已经收敛在最小值以内。
4. 在 RRS 下降到阈值以下时终止，可以确保之后聚类的质量。实际上，这是一个很好的做法，在结合迭代次数的同时保证了K均值的终止。

**Q9. 以下哪种算法会受到局部最优的聚焦问题的影响？**

1. K均值聚类算法
2. 层次聚类算法
3. 期望-最大化聚类算法
4. 多样聚类算法

选项：

- A. 1
- B. 2 3
- C. 2 4
- D. 1 3
- E. 1 2 4
- F. 以上都是

**答案：D**

在K均值聚类中，局部最优的聚焦问题是一个常见问题，它会导致算法收敛到一个局部最优解，而不是全局最优解。层次聚类算法和期望-最大化聚类算法也会受到局部最优问题的影响，因为它们都是基于贪心策略的。多样聚类算法则不会受到局部最优问题的影响，因为它是一种基于模型的聚类方法。

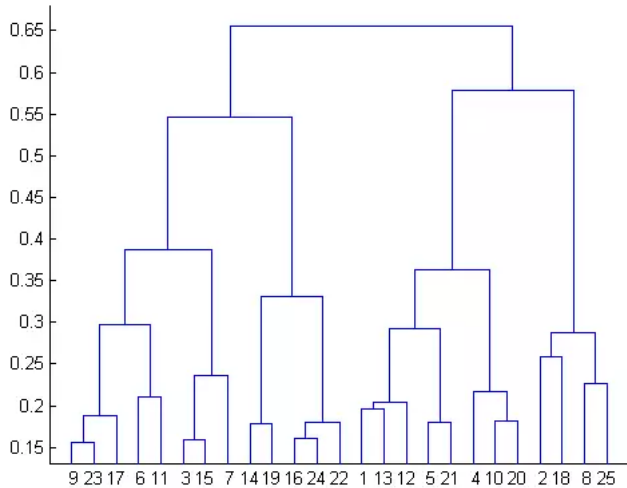
[HTML5高薪有多高](#)[Java高薪有多高](#)[师徒1对1](#)[Python薪资多少钱](#)[数据分析师培训](#)[心理抑郁测试](#)

- A. K均值聚类算法
- B. K中位数聚类算法
- C. K模型聚类算法
- D. K中心点聚类算法

答案：A

在上面给出的选项中，K均值聚类算法对离群值最敏感，因为它使用集群数据点的平均值来查找集群的中心。

Q11. 在对数据集执行K均值聚类分析以后，你得到了下面的树形图。从树形图中可以得出那些结论？



- A. 在聚类分析中有28个数据点
- B. 被分析的数据点里最佳聚类数是4
- C. 使用的接近函数是平均链路聚类
- D. 对于上面树形图的解释不能用于K均值聚类分析

答案：D

树形图不可能用于聚类分析。但是可以根据K聚类分析的结果来创建一个簇状图。

Q12. 如何使用聚类（无监督学习）来提高线性回归模型（监督学习）的准确性：

1. 为不同的集群组创建不同的模型。
2. 将集群的id设置为输入要素，并将其作为序数变量。
3. 将集群的质心设置为输入要素，并将其作为连续变量。
4. 将集群的大小设置为输入要素，并将其作为连续变量。

选项：

- A. 1
- B. 1 2
- C. 1 4
- D. 3
- E. 2 4
- F. 以上都是

答案：F

将集群的 id 设置为序数变量和将集群的质心设置为连续变量，这两项可能不会为多维数据的回归模型提供更多的相关信息。但是当在一个维度上进行聚类分析时，上面给出的所有方法都有望为多维数据的回归模型提供有意义的信息。举个例子，根据头发的长度将人们分成两组，将聚类 ID 存储为叙述变量，将聚类质心存储为连续变量，这样一来，多维数据的回归模型将会得到有用的信息。

Q13. 使用层次聚类算法对同一个数据集进行分析，生成两个不同的树形图有哪些可能的原因：

- A. 使用了接近函数

HTML5高薪有多高

Java高薪有多高

师徒1对1

Python薪资多少钱

数据分析师培训

心理抑郁测试

- D. 只有B和C
- E. 以上都有

登录

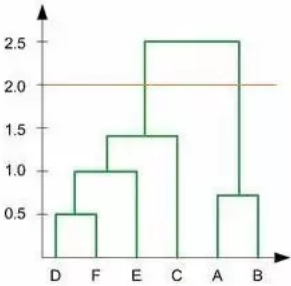
注册

×

答案：E

接近函数、数据点、变量，无论其中哪一项的改变都可能使聚类分析产生不同的结果，并产生不同的树状图。

Q14. 在下面的图中，如果在y轴上绘制一条y=2的水平线，将产生多少簇？

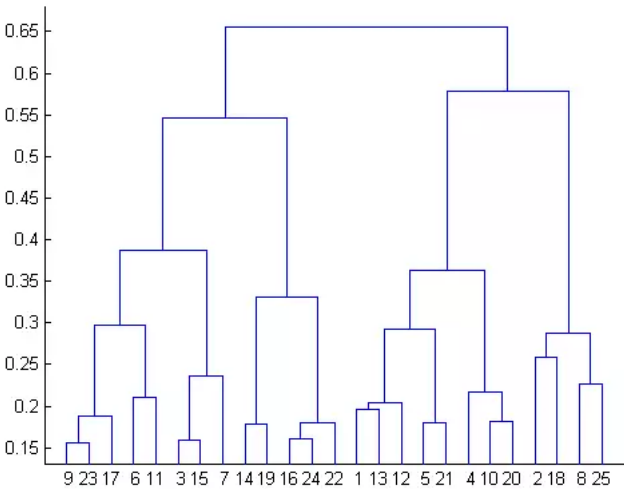


- A. 1
- B. 2
- C. 3
- D. 4

答案：B

因为在树状图中，与 y=2 红色水平线相交的垂直线有两条，因此将形成两个簇。

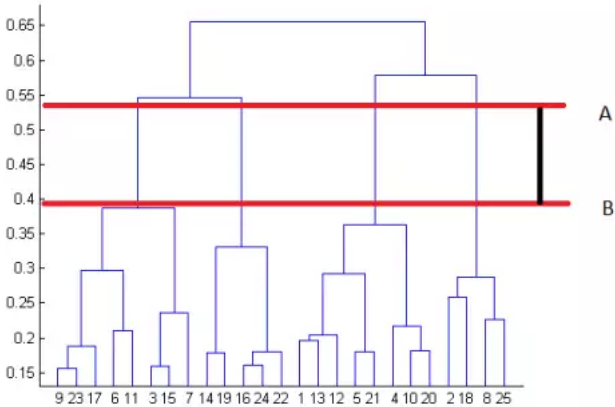
Q15. 根据下面的树形图，数据点所产生的簇数最可能是？



- A. 2
- B. 4
- C. 6
- D. 8

答案：B

通过观察树状图，可以很好的判断出不同组的簇数。根据下图，水平线贯穿过的树状图中垂直线的数量将是簇数的最佳选择，这条线保证了垂直横穿最大距离并且不与簇相交。



在上面的例子中，簇的数量最佳选择是4，因为红色水平线涵盖了最大的垂直距离AB。

**Q16. K均值聚类分析在下面哪种情况下无法得出好的结果？**

1. 具有异常值的数据点
2. 具有不同密度的数据点
3. 具有非环形的数据点
4. 具有非凹形的数据点

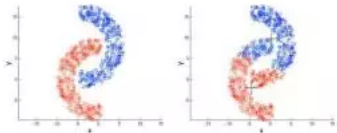
选项：

- A. 1 2
- B. 2 3
- C. 2 4
- D. 1 2 4
- E. 1 2 3 4

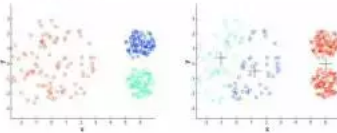
**答案：D**

在数据包含异常值、数据点在数据空间上的密度扩展具有差异、数据点为非凹形状的情况下，K均值聚类算法的运行结果不佳。

Non-convex/non-round-shaped clusters: Standard *K*-means fails!



Clusters with different densities



**Q17. 通过以下哪些指标我们可以在层次聚类中寻找两个集群之间的差异？**

1. 单链
2. 完全链接
3. 平均链接

选项：

- A. 1 2
- B. 1 3
- C. 2 3

HTML5高薪有多高

Java高薪有多高

师徒1对1

Python薪资多少钱

数据分析师培训

心理抑郁测试

答案：D

通过单链接、完全链接、平均链接这三种方法，我们可以在层次聚类中找到两个集群的差异。

Q18. 下面哪些是正确的？

1. 特征性多重共线性对聚类分析有负面效应
2. 异方差性对聚类分析有负面效应

选项：

- A. 1
- B. 2
- C. 1 2
- D. 以上都不是

答案：A

聚类分析不会受到异方差性的负面影响，但是聚类中使用的特征/变量多重共线性会对结果有负面的影响，因为相关的特征/变量会在距离计算中占据很高的权重。

Q19. 给定具有以下属性的六个点：

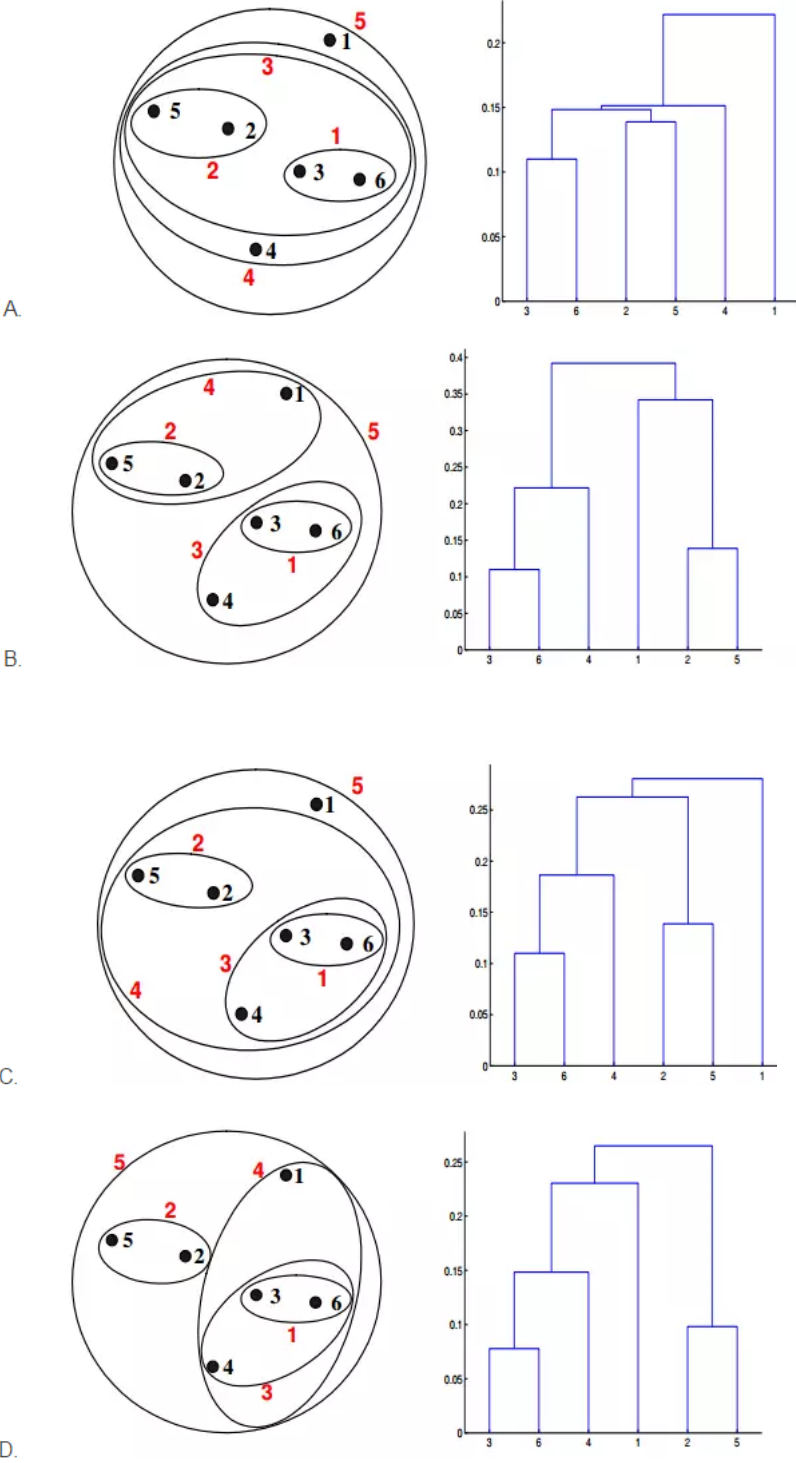
point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

Table : X-Y coordinates of six points.

	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points

如果在层次聚类中使用最小值或单链接接近函数，可以通过下面哪些聚类表示和树形图来描述？



答案：A

对于层级聚类的单链路或者最小化，两个簇的接近度指的是不同簇中任何两个点之间的距离的最小值。例如，我们可以从图中看出点3和点6之间的距离是0.11，这正是他们在树状图中连接而成的簇的高度。再举一个例子，簇{3，6}和{2，5}之间的距离这样计算： $\text{dist}(\{3, 6\}, \{2, 5\}) = \min(\text{dist}(3, 2), \text{dist}(6, 2), \text{dist}(3, 5), \text{dist}(6, 5)) = \min(0.1483, 0.2540, 0.2843, 0.3921) = 0.1483$ 。

Q20. 给定具有以下属性的六个点：



point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

Table : X-Y coordinates of six points.

	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points

如果在层次聚类中使用最大值或完全链接接近函数，可以通过下面哪些聚类表示和树形图来描述？

  
0

  
写评论

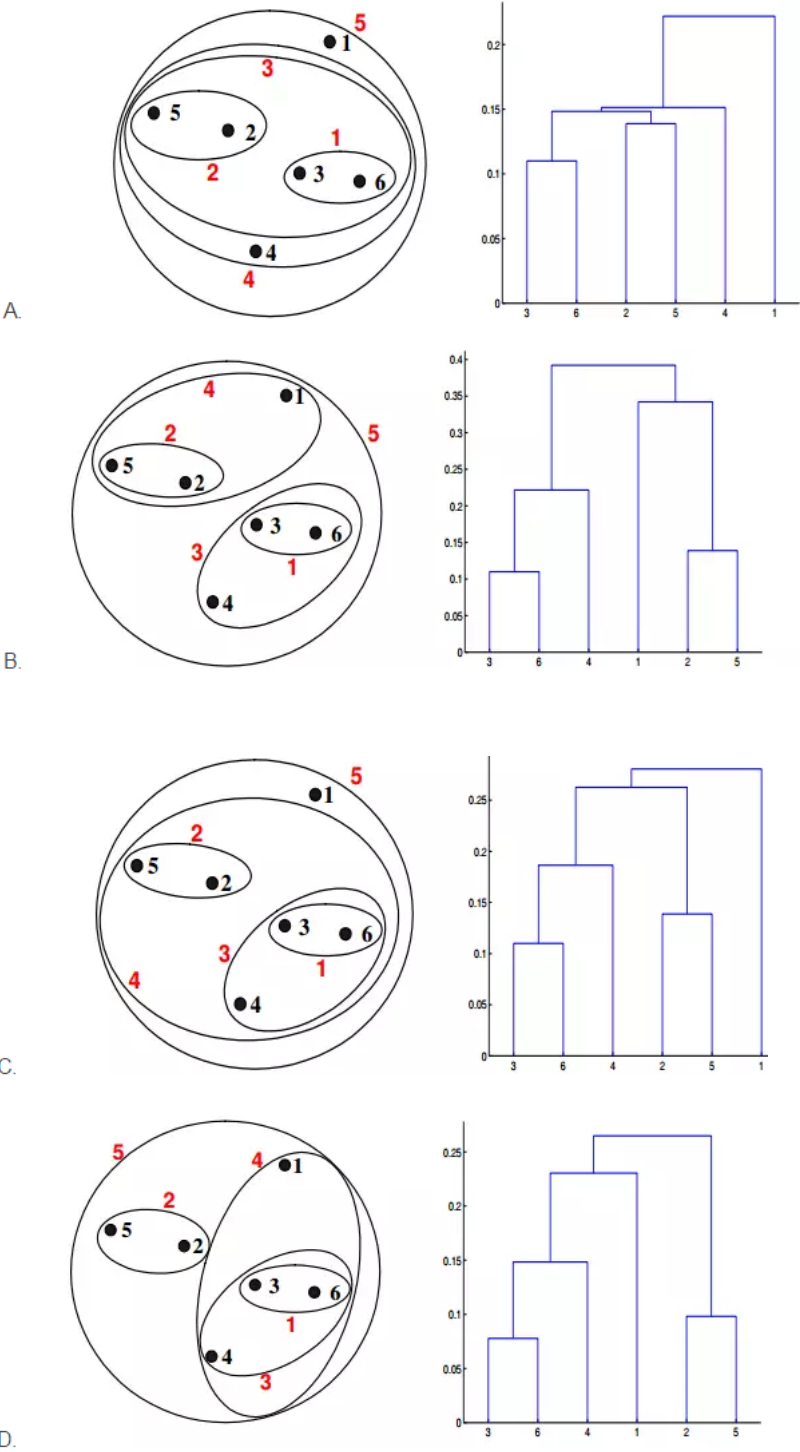
  
目录

  
收藏

  
微信

  
微博

  
QQ



答案：B

对于层级聚类的单链路或者最大值，两个簇的接近度指的是不同簇中任何两个点之间的距离的最大值。同样，点3和点6合并在了一起，但是{3，6}没有和{2，5}合并，而是和{4}合并在了一起。这是因为 $\text{dist}(\{3, 6\}, \{4\}) = \max(\text{dist}(3, 4), \text{dist}(6, 4)) = \max(0.1513, 0.2216) = 0.2216$ ，它小于 $\text{dist}(\{3, 6\}, \{2, 5\}) = \max(\text{dist}(3, 2), \text{dist}(6, 2), \text{dist}(3, 5), \text{dist}(6, 5)) = \max(0.1483, 0.2540, 0.2843, 0.3921) = 0.3921$  and  $\text{dist}(\{3, 6\}, \{1\}) = \max(\text{dist}(3, 1), \text{dist}(6, 1)) = \max(0.2218, 0.2347) = 0.2347$ 。

本文作者 Saurav Kaushik 是数据科学爱好者，还有一年他就从新德里 MAIT 毕业了，喜欢使用机器学习和分析来解决复杂的数据问题。

编译：AI100

原文链接：40 Questions to test a Data Scientist on Clustering Techniques (Skill test Solution)

扫码关注

CSDN AI 公众号「人工智能头条」

AI 热点案例跟踪 • TOP 100 人物专访 • 最新技术全面解析 • 每日博客万里挑一

扫码加我进群

CSDN AI 干货分享交流

名家大师、千余位业界同行 • CSDN福利、资料秒送达 • 线上线下活动优先报名

加群请注明：公司+职位+姓名

版权声明：本文为博主原创文章，未经博主允许不得转载。https://blog.csdn.net/heyc861221/article/details/80128726

文章标签：数据科学 聚类 线性回归

相关关键词：pat测试数据 nyoj测试数据 php测试数据 ai测试数据 dfs的测试数据

上一篇

程序编写程序：泛用人工智能领域的一颗明珠

下一篇

用深度学习（CNN RNN Attention）解决大规模文本分类问题 - 综述...

区块链以太坊DApp开发为什么人才稀少？薪资到底有多高？

区块链开发平均薪资是多少？快速学习区块链开发，掌握以太坊开发和智能合约应用。

想对作者说点什么？

我来说两句

数据科学家的修炼之道（完整版）

2017年03月15日 41.54MB 下载

PDF

测试数据科学家聚类技术的40个问题（附答案和分析）

介绍 创造出具有自我学习能力的机器——人们的研究已经被这个想法推动了十几年。如果要实现这个梦想的话，无监督学习和聚类...

1788

测试数据科学家聚类技术的40个问题(附答案和分析) - CSDN博客

它将相似的数据进行分类,通过元理解来提供相应的各种商业决策。在这次能力测试中,我们在社区中提供了聚类的测试,总计有1566人注...

7-16

属性重要性的举例(定性数据聚类) - CSDN博客

近日在解决一组定性数据的聚类问题,看文献时觉得论文中“基于属性重要性的定性数据聚类方法”举例不太容易理解(尤其是对我这种不...

5-22

数据可视化

百度广告

聚类总结（上）——划分聚类

概述聚类指根据一定的准则，把一份事物按照这个准则归纳成互不重合的几份。机器学习中，聚类指按照一个标准，这个标准通常是...

4167

聚类分析(二):图团体检测 - CSDN博客

接着上篇介绍聚类算法,本篇介绍图团体检测算法。图团体检测 当我们的样本以及样本之间的关系可以被表示为一个网络或图(graph)时...

6-29

HTML5高薪有多高

Java高薪有多高

师徒1对1

Python薪资多少钱

数据分析师培训

心理抑郁测

聚类(一) - CSDN博客

一.概念性介绍若样本的标记信息未知,我们称这样的问题为“无监督学习”(unsupervised learning)。针对于无监督学习我们常用的方法是...

数据挖掘聚类分析技术实验常用真实数据集

2016年03月19日 98KB 下载



Kmeans聚类算法在python下的实现--附测试数据



435

Kmeans聚类算法在python下的实现--附测试数据

四种聚类方法之比较 - CSDN博客

结果显示对该测试类型数据,FCM和k-means都具有较高...的FCM算法通过通用测试数据集进行聚类效果的比较和...组织在2维网格上的...

6-29

数据挖掘笔试题 - CSDN博客

1. 某超市研究销售纪录数据后发现,买啤酒的人很大概率也会购买尿布,这种属于数据挖掘的哪类问题?(A) A. 关联规则发现 B. 聚类 C. 分...

6-6

文本分类与聚类 ( text categorization and clustering )



1888

原文地址 文本分类与聚类(text categorization and clustering) 1. 概述广义的分类 ( classification 或者 categorization ) 有两种含义 : ...

58元解决男人问题，教你一招解决，千万要记住这个方法！

一优 · 顶新

四种聚类算法的比较 - CSDN博客

同样的数据尝试多种算法,以发现数据可能揭示的结果。...的FCM算法通过通用测试数据集进行聚类效果的比较和...组织在2维网格上的...

3-4

ML:教你聚类并构建学习模型处理数据(附数据集) - CSDN博客

翻译: 王雨桐校对: 顾佳妮本文约1500字,建议阅读6分钟。本文以Ames住房数据集为例,对数据进行聚类,并构建回归模型。摘要本文将...

5-27

中国数据科学家年薪有多少？



4262

近日，数据科学社群Kaggle发布了数据科学 / 机器学习业界现状调查报告。这份调查问卷的受访者囊括了全球50多个国家的16,000...

如何成为一名数据科学家（附学习资源）



1083

原文：Becoming a Data Scientist 作者：Roger Huang 翻译：Vincent 译者注：本文是出自Springboard上面一篇文章的摘录，介绍...

四种聚类方法之比较 - CSDN博客

同样的数据尝试多种算法,以发现数据可能揭示的结果。...的FCM算法通过通用测试数据集进行聚类效果的比较和...组织在2维网格上的...

11-17

当今世界最NB的25位大数据科学家



3.2万

在大数据技术飞速发展的今天，谁才是我们大数据科研与工业界中最有威望的科学家呢？下面我们来梳理，共罗列了25位当今...

如何成为一名数据科学家



4374

在回答这个问题之前，希望你先想想另外一个问题：为什么要成为数据科学家？当然，如果你是为了10万美元的年薪也无可厚非，...

77个数据科学家常见面试题



1081

随着大数据概念的火热，数据科学家这一职位应时而出，那么成为数据科学家要满足什么条件？或许我们可以从国外的数据科学家面...

美国数据科学家带你看看大数据的未来



1657

从SGI的首席科学家John R. Masey在1998年提出大数据概念，到大数据分析技术广泛应用于社会的各个领域，已经走过了17年的时...

HTML5高薪有多高

Java高薪有多高

师徒1对1

Python薪资多少钱

数据分析师培训

心理抑郁测试

由于最近一年大数据火爆异常，各种培训班开课广告满天飞，很多做开发的朋友也想转到大数据这一行，在投递简历的时候

登录

注册

×

50万码农评论：英语对于程序员有多重要！

不背单词和语法，老司机教你一个数学公式秒懂天下英语



3551

524

1366

9140

2.3万

分类与聚类的区别

分类与聚类的区别

聚类 -- 数据挖掘

作者：林子 链接：https://blog.csdn.net/leoch007/article/details/80027056 欢迎转载，记得注明出处 聚类 聚类，属于无监督的机器...

聚类（下）

1.原型聚类 原型聚类亦称“基于原型的聚类”（prototype-based clustering），此类算法假设聚类结构能通过一组原型刻画，在现实聚类...

常见聚类算法分类

聚类划分：（1）划分聚类 k-means、k-medoids、k-modes、k-medians、kernel k-means （2）层次聚类 Agglomerative、divisiv...

让你看懂聚类分析

目录1.聚类分析概述 2.各种距离的定义 2.1 样本相似性度量 2.2 类与类间的相似性度量 2.3 变量间的相似性度量 3.划分聚类 4.层次...

舆情监测平台

军犬网络舆情监测



聚类（一）

一.概念性介绍若样本的标记信息未知，我们称这样的问题为“无监督学习”（unsupervised learning）。针对于无监督学习我们常用的...

聚类算法K-Means, K-Medoids, GMM, Spectral clustering, Ncut

聚类算法是ML中一个重要分支，一般采用unsupervised learning进行学习，本文根据常见聚类算法分类讲解K-Means, K-Medoids, G...

聚类

聚类的任务：聚类属于“无监督学习”，目标是通过对于无标记训练样本的学习来揭示数据的内在的性质和规律，为进一步的数据分析提...

四种聚类方法之比较

摘要: 介绍了较为常见的k-means、层次聚类、SOM、FCM等四种聚类算法，阐述了各自的原理和使用步骤，利用国际通用测试数...

浅谈聚类&层次聚类

浅谈聚类 引言 聚类分析的目标就是在相似的基础上收集数据来分类。即聚类是我们在面对大量数据时，所常采用的一种数...

50万码农评论：英语对于程序员有多重要！

不背单词和语法，老司机教你一个数学公式秒懂天下英语



937

3.7万

聚类 - 1 - 聚类介绍

本总结是个人为防止遗忘而作，不得转载和商用。 很多算法，比如“线性回归、Logistics/softmax回归、SVM”等都是给出了...

聚类分析

HTML5高薪有多高

Java高薪有多高

师徒1对1

Python薪资多少钱

数据分析师培训

心理抑郁测试

### 四种聚类算法

聚类分析是一种重要的人类行为，早在孩提时代，一个人就通过不断改进下意识中的聚类模式来学会如何区分猫狗、动物植物。目前...

### 聚类(clustering)

聚类(clustering): 人类具有归纳与总结的能力，让计算机按照人们思考问题的方式将训练集中的数据进行有差异性的分组，将有一类...

### 聚类分析常用算法原理：KMeans,DBSCAN, 层次聚类

聚类分析是非监督学习的很重要的领域。所谓非监督学习，就是数据是没有类别标记的，算法要从对原始数据的探索中提取出一定的...

### 一点点加盟

加盟一点点奶茶需要多少钱



### Python机器学习——聚类的有效性指标

聚类有效性指标 ( Cluster Validity Index , CVI ) 用于度量聚类的效果。很显然，希望彼此相似的样本在一个簇，彼此不相似的样本尽...

### 分类聚类区别及聚类概述

在初学分类聚类时，对这两个概念不是很了解。随着深入的了解，现有了一些基本的认识。现对聚类进行个人理解上的总结，欢迎大...

### 聚类分析matlab

感觉自己的业余时间好像越来越少了，实习和学校的工作看来必然会带来一定的冲突，需要适应这个节奏的变化啊，加油。 一、理...

### 机器学习笔记(九)聚类

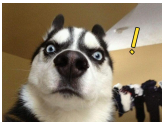
9.聚类 有必要回顾下前文所涉及的机器学习主流分类，有监督学习中根据预测结果离散和连续属性分为分类和回归两大类，常见的...

### 聚类总结

聚类总结

### 程序猿不会英语怎么行？英语文档都看不懂！

不背单词和语法，一个公式教你读懂天下英文→



### 回归、分类与聚类：三大方向剖解机器学习算法的优缺点

回归、分类与聚类：三大方向剖解机器学习算法的优缺点 2017-05-20 13:56:14 机器学习 数学 3 0 0 在本教程中，作者对现代...

### 聚类分析2

1.基于分层的分类 ( matlab ) 实现 基本套路： a=[1,0;1,1;3,2;4,3;2,5]; Y=pdist(a,'cityblock');%求距离矩阵 yc=squareform(Y)%将距...

### 聚类综述

聚类算法分类划分聚类 层次聚类 基于密度聚类 基于表格聚类

### 聚类之层次聚类与密度聚类

注：本文中所有公式和思路来自于邹博先生的《机器学习升级版》，我只是为了加深记忆和理解写的本文。讲密度聚类之前，想先...

### 聚类系列-概述

本系列要讲解一下本人这一段时间工作中 使用聚类来做的一些事情，第一次写博客大家不喜勿喷，有错误请及时指出，本人一定改...

大数据分析工具

关于大数据分析的8大工具



登录

注册

×

聚类算法应用场景实例十则

本文整理了5个天池、DataCastle、CCF中可使用回归算法处理的问题场景。 1 基于用户位置信息的商业选址 随着信息技术的...

1.1万

文本挖掘、文本聚类、热词发现流程图

1102

聚类分析中几种算法的比较

将数据库中的对象进行聚类是聚类分析的基本操作，其准则是使属于同一类的个体间距离尽可能小，而不同类个体间距离尽可能大，...

2232

聚类算法（基础及核心概念）

聚类分析是数据挖掘中比较常见和基础的方法；本篇文章简要介绍了聚类分析算法的基础知识以及核心概念~~....

921

聚类算法

一、串行聚类算法 1.1 划分方法（partitioning method）划分方法首先根据给定要构建划分的数目k创建一个初始划分，然后采用一...

7285

50万码农评论：英语对于程序员有多重要！

不背单词和语法，老司机教你一个数学公式秒懂天下英语



机器学习第六课：聚类

上一篇记载了高斯混合模型，这个模型在机器学习中的很多地方都会用到。既然大家对高斯混合模型这么熟了，我们就从这里开始讲...

811

机器学习：Python实现聚类算法(三)之总结

http://www.cnblogs.com/lc1217/p/6963687.html 考虑到学习知识的顺序及效率问题，所以后续的几种聚类方法不再详细讲解原理...

1万

聚类算法综述（7）

原文地址：聚类算法综述（7）作者：hyman3.2.2 基于粒度的聚类方法 从表面上看，聚类和分类有很大差异——聚类是无导师的...

1768

聚类算法的评价指标

ARI，AMI

8803

【聚类】性能度量

聚类的性能是很难度量，因为它从数据中学习，并归类。数据的特性的多方面的，那么归类的结果也会是多种。比如苹果，可以从...

1864

crm客户管理系统

crm系统



深入理解K-Means聚类算法

概述什么是聚类分析聚类分析是在数据中发现数据对象之间的关系，将数据进行分组，组内的相似性越大，组间的差别越大，则聚类...

6.5万

matalb k-means函数使用方法（附一个很好的k-means参考博客）

首先得说，虽然matlab提供了KMeans函数，但是matlab的KMeans聚类好慢，真的好慢！ K-means算法是硬聚类算法，是典型...

3.9万

HTML5高薪有多高

Java高薪有多高

师徒1对1

Python薪资多少钱


数据分析师培训

心理抑郁测试



登录注册

个人资料



何永灿CSDN

关注

原创340

粉丝843

喜欢69

评论23









等级：博客5

访问：24万+

积分：2354

排名：2万+

### 抑郁有哪些症状



#### 最新文章

好未来发布“WISROOM” 基于“AI+脑科学”打造智慧课堂

资源 | 5月Python好文TOP 10新鲜出炉，精选自1000篇文章，你都看了吗？

我们从1400篇机器学习文章中挑出了Top 10

用Python更加了解微信好友

实战 | 手把手教你搭一个机器翻译模型

#### 个人分类

信息无障碍	8篇
基于UAAG2.0 Reference的移动无障碍案例	4篇
活动	5篇
区块链	1篇
人工智能	1篇

#### 归档

2018年7月	1篇
2018年5月	41篇
2018年4月	16篇
2018年3月	5篇
2018年2月	4篇

展开

#### 热门文章

《程序员》11月精彩内容：AI 工程师职业指南  
阅读量：21517

Android——实现无障碍  
阅读量：7462

如何入门Python与机器学习



登录

注册

解决大

和实践

阅读量：6059

无障碍小技巧

阅读量：5553

最新评论

林轩田《机器学习基石》资源汇总（视...  
ed35007：云盘地址失效了，帮贴下最新的云盘地...  
址吧

如何入门Python与机器学习  
jiangjunshow：写得不错！我最近也在写一系列人...  
工智能教程，通俗易懂，无需高等数学基础，教程...  
也力求风趣幽默。点击我的...

分布式深度学习（I）：分布式训练神...  
jiangjunshow：写得不错！我最近也在写一系列人...  
工智能教程，通俗易懂，无需高等数学基础，教程...  
也力求风趣幽默。点击我的...

自然语言处理技术（NLP）在推荐系...  
DKhadoop：博主您好，是否方便加一下您的微...  
信，想向您请教一些大数据方面的技术问题，可有...  
偿回答问题！

大疆RoboMaster技术总监：...  
xiaodaozi99：感谢 大神