大数据挖掘技术练习(试卷编号141)

1.[单选题]( ),用于显示树状结构数据。
A)矩形式树状结构图;
B)平行结构树
C)垂直结构树
答案:A
解析:

2.[单选题]下列（）不属于人工智能新突破取得的产品
A)科大讯飞的翻译器、记录仪
B)天猫精灵等智能AI音箱
C)佳能相机
D)某酒店通过人脸识别认证身份信息
答案:C
解析:

3.[单选题]SELECT命令中用于返回非重复记录的关键字是_____。
A)TOP
B)GROUP
C)DISTINCT
D)ORDER
答案:C
解析:

4.[单选题]寻呼过程是（）接口过程，MME通过向eNODEB发送寻呼消息来发起寻呼过程的。
A)Iub
B)Uu
C)S1
D)X2
答案:C
解析:

5.[单选题]weka系统汇集了最前沿的机器学习算法和数据预处理工具，提供的主要应用程序不包括
A)Explorer
B)KnowledgeFlow
C)Experimenter
D)Conclusion
答案:D
解析:

6.[单选题]SPSS最突出的特点是
A)处理效率高
B)界面友好
C)结果准确
D)操作方便
答案:B
解析:

7.[单选题]正则表达式“[a-z]”，不可以匹配下列的字符串为（ ）。
A)a
B)z
C)2
D)m
答案:C
解析:

8.[单选题]DBSCAN在最坏情况下的时间复杂度是（ ）。
A)O(m)
B)O(m2)
C)O(log m)
D)O(m*log m)
答案:B
解析:

9.[单选题]3GPP R8及以后的SGSN与MME之间的接口是（）
A)S3
B)S12
C)S6
D)S4
答案:A
解析:

10.[单选题]已知某企业第20期的模型参数a=91856-105，用二次指数平滑法预测第25期的销售量是()。
A)1023.5
B)1443.5
C)4697.5
D)5117.5
答案:B
解析:

11.[单选题]在DPI规范中，HTTP版本字段等于0x05代表（）

A)HTTP2.0
B)HTTP1.1
C)WAP1.0
D)WAP1.1
答案:C
解析:

12.[单选题]在基本DBSCAN的参数选择方法中，点到它的K个最近邻的距离中的K选作为哪一个参数（）
A)Eps
B)MinPts
C)质心
D)边界
答案:B
解析:

13.[单选题]有关数据抽取工具的叙述中正确的是( )
A)只能使用数据仓库开发工具所提供的数据抽取工具
B)只能使用开发人员自己开发的数据抽取工具
C)根据实际需要确定是否自己开发数据抽取工具
D)以上都不对
答案:C
解析:

14.[单选题]可以对按城市汇总的销售数据进行()，来观察按国家总的数据。
A)上卷
B)下钻
C)切片
D)切块
答案:A
解析:

15.[单选题]决策树算法是一种()数据挖掘算法。
A)关联分析
B)预测
C)分类
D)聚类
答案:C
解析:

16.[单选题]利用tree.DecisionTreeClassifier()训练模型时调用.fit()方法需要传递的第一个参数是( )。

A)样本特征X
B)样本标签Y
C)判断标准
D)设置结点的最小样本数量
答案:A
解析:

17.[单选题]下面关于数据粒度的描述不正确的是:
A)粒度是指数据仓库小数据单元的详细程度和级别;
B)数据越详细,粒度就越小,级别也就越高;
C)数据综合度越高,粒度也就越大,级别也就越高;
D)粒度的具体划分将直接影响数据仓库中的数据量以及查询质量.
答案:C
解析:

18.[单选题]以下属于可伸缩聚类算法的是( )。
A)CURE
B)DENCLUE
C)CLIQUE
D)OPOSSUM
答案:A
解析:

19.[单选题]在中移动的集中性能管理应用落地-物联网端到端业务质量分析手册中，其定界流程是基于:
A)八元六阶
B)七元五阶
C)六元四阶
D)五元三阶
答案:C
解析:

20.[单选题]关于统计学和大数据之间的关系,一下说法错误的是( )。
A)面临大数据,统计学的研究对象有所改变;
B)在大数据环境中,需要首先将未知的问题转化为可用的统计方法;
C)在大数据分析过程中,传统的统计分析过程“定量-定位-再定性”转变为“定量-定性”;
D)在大数据环境中,需要将统计研究的对象范围扩展到一切数据。
答案:A
解析:

21.[单选题]字典的_____方法返回字典的“键”列表
A)keys()

B)key()
C)values()
D)items()
答案:A
解析:

22.[单选题]假定用于分析的数据包含属性age。数据元组中age的值如下（按递增序）：13，15，16，16，19，20，20，21，22，22，25，25，25，30，33，33，35，35，36，40，45，46，52，70，问题：使用按箱平均值平滑方法对上述数据进行平滑，箱的深度为3。第二个箱子值为：
A)18.3
B)22.6
C)26.8
D)27.9
答案:A
解析:

23.[单选题]个人信息的收集、处理和利用应当遵循()的原则，不得违反法律、法规的规定和双方的约定收集、处理和利用个人信息。（ ）
A)正规、合法、必要
B)合法、正当、必要
C)合法、合规、正当
D)合法、合理、合规
答案:B
解析:

24.[单选题]以下哪个指标属于无线网络结构指标
A)即时通信响应成功率
B)重叠覆盖小区占比
C)4G占网时长占比
D)网络质量综合满意度
答案:B
解析:

25.[单选题]下述除哪个维度外，均有利于通过终端指标评估分析明确终端问题现象、场景，辅助终端问题的复现解决
A)芯片一致性
B)版本差异性
C)网络适配性
D)流程差异性
答案:D
解析:

26.[单选题]OnRetDW系统建模采用的是()。
A)星形模型
B)雪花模型
C)事实星座模型
D)关系数据库模型
答案:A
解析:

27.[单选题]NLTK最适用于哪种类型的任务
A)语言处理
B)图像处理
C)声音处理
D)文字处理
答案:A
解析:

28.[单选题]下列关于聚类挖掘技术的说法中,错误的是()
A)不预先设定数据归类类目,完全根据数据本身性质将数据聚合成不同类别
B)要求同类数据的内容相似度尽可能
C)要求不同类数据的内容相似度尽可能
D)与分类挖掘技术相似的是,都是要对数据进行分类处理
答案:B
解析:

29.[单选题]数据预处理的任务不包括（ ）。
A)数据清洗
B)数据规范化和离散化
C)数据分类
D)特征提取与特征选择
答案:C
解析:

30.[单选题]给定df是一个DataFrame对象，对df所有字段进行描述性统计，可以利用的方法为（ ）。
A)df.describe()
B)df.statistics()
C)df.mean()
D)df.summary()
答案:A
解析:

31.[单选题]( )将两个簇的邻近度定义为两个簇合并时导致的平方误差的增量,它是一种凝聚层次聚类技术。
A)MIN(单链)
B)MAX(全链)
C)组平均
D)Ward方法
答案:D
解析:

32.[单选题]原始的朴素贝叶斯只能处理离散数据，当x1，…，xn是连续变量时，我们可以使用（）完成分类任务
A)贝叶斯定理
B)半朴素贝叶斯
C)拉普拉斯平滑处理
D)高斯朴素贝叶斯
答案:D
解析:

33.[单选题]PageRank是一个函数 ， 它对 Web中的每个网页赋予一个实数值。它的意图在于网页的PageRank越高, 那么它就( )。
A)相关性越高
B)越不重要
C)相关性越低
D)越重要
答案:D
解析:

34.[单选题]在SELECT语句中,用来指定查询所用的表的子句是_____。
A)WHERE
B)GROUP BY
C)ORDER BY
D)FROM
答案:D
解析:

35.[单选题]若有频繁3-项集的集合:{1,2,3},{1,2,4},{1,2,5},{1,3,4},{1,3,5},(2,3,4),(2,3,5},{3,4,5},假定数据集中只有5个项,则产生的候选4-项集不包含( )。
A){1,2,3,4}
B){1,2,3,5}
C){1,2,4,5}
D)以上都不是

答案:C
解析:

36.[单选题]在SELECT查询语句中对字段排序的命令子句是_____。
A)ORDER BY
B)GROUP BY
C)INSERT
D)UPDATA
答案:A
解析:

37.[单选题]在哪里能下载到hadoop软件
A)apache网站 或者捐献给apache使用的公开服务器
B)oracle官网
C)hadoop公司官网
D)linux官网
答案:A
解析:

38.[单选题]实际接入的指标数占应接指标总数的比例,统计时又分线上和线下接入率称为( )
A)指标数据自动采集率 ;
B)指标数据接入率
C)指标数据接入及时率 ;
D)指标数据完整率
答案:B
解析:

39.[单选题]联机分析处理包括以下不是基本分析功能的为： ( )
A)聚类
B)切片
C)转轴
D)切块
答案:A
解析:

40.[单选题]LTE规划仿真中的详细规划不涉及的是()
A)业务分布
B)覆盖预测
C)参数规划
D)容量仿真
答案:A
解析:

41.[单选题]当用户发起附着时，如该用户的imsi号段信息在MME上并没有制作相应数据，则MME将
A)拒绝该用户附着请求
B)仍允许该用户附着请求
C)转发该用户附着请求至HLR/HSS进行鉴权
D)匹配默认规则
答案:A
解析:

42.[单选题]研究顾客是否想购买手机与年龄,性别,收入和工作地点的关系可以使用()
A)回归方法
B)分类方法
C)聚类方法
D)关联分析
答案:B
解析:

43.[单选题]（）打开了自动驾驶的天花板
A)驾驶员驾驶技术提升
B)汽车刹车性能增强
C)人工智能的理论和实践的突破
D)汽车制造商转型
答案:C
解析:

44.[单选题]MME附着用户数的指标来源是
A)网管系统
B)网优平台
C)集中性能平台
D)大数据平台
答案:A
解析:

45.[单选题]智能手机AI创新不包括（）
A)机器学习框架
B)3D结构光技术
C)手机芯片硬件层面推出了全新的NPU单元
D)手机材料更耐摔
答案:D
解析:

46.[单选题]噪声数据主要是包含错误数据、假数据和（）

A)异常数据
B)真实数据
C)污染数据
D)缺失数据
答案:A
解析:

47.[单选题]以下哪个聚类算法不属于基于网格的聚类算法（）
A)STING
B)MAFIA
C)BIRCH
D)WaveCluster
答案:C
解析:

48.[单选题]分析判断PGW是否存在业务受限的情况采用哪个指标
A)PGW承载容量平均利用率
B)PGW承载容量峰值利用率
C)PGW平均负荷
D)PGW数据吞吐容量利用率
答案:B
解析:

49.[单选题]有些数据挖掘算法，要求数据属性是标称类别，当数据中包含数值属性时，为了使用这些算法需要将数值属性转换成标称属性。通过采取各种方法将数值属性的值域划分成一些小的区间，并将这连续的小区间与离散的值关联起来，每个区间看作一个类别。例如，某个问题中的年龄属性一种可能的划分成类别操作是：[0…11]→儿童，[12…17]→青少年，[18…44]→青年，[45…69]→中年，[69…∞]→老年。这种将连续变量划分成不同类别的过程通常称为（ ）。
A)特征化
B)优化
C)标准化
D)离散化
答案:D
解析:

50.[单选题]KNN算法流程中不正确的有（）
A)计算已知类别数据集中的点与当前点之间的距离，按照距离递增次序排序；
B)选取与当前点距离最小的k个点；
C)确定前k个点所在类别对应的出现频率；
D)返回前k个点出现频率最低的类别作为当前点的预测分类。
答案:D
解析:

51.[单选题]设有如下所示的某商场购物记录集合，每个购物篮中包含若干商品：
现在要基于该数据集进行关联规则挖掘。如果设置最小支持度为 60%，则如下频繁项集中，符合条件的是（ ）

购物篮编号	商品
1	面包，牛奶
2	面包，啤酒，鸡蛋，尿布
3	牛奶，啤酒，尿布，可乐
4	面包，牛奶，啤酒，尿布
5	面包，牛奶，尿布，可乐（ ）

A)鸡蛋，尿布
B)面包，尿布，牛奶
C)面包，牛奶
D)面包，啤酒，尿布
答案:C
解析:

52.[单选题]终端支持的频段，在下列哪个流程中会得以体现
A)ATTACH
B)DETACH
C)切换流程
D)呼叫流程
答案:A
解析:

53.[单选题]如果说人工智能是一座高大上的房子，那么（）就是它的基石
A)新技术
B)资金
C)大数据
D)需求
答案:C
解析:

54.[单选题]下列对学生相关属性描述中，不是标称属性的是（ ）
A)身高

B)头发颜色
C)学号
D)婚姻状况
答案:A
解析:

55.[单选题]在LTE下，eNodeB通过（）接口连接MME。
A)S1-U
B)S1-MME
C)S6a
D)S1-MME
答案:B
解析:

56.[单选题]马云认为，（）是数据时代必须跨过的一个坎
A)数据隐私
B)数据服务
C)数据获取
D)数据应用
答案:A
解析:

57.[单选题]描述一组对称(或正态)分布数据的离散程度时,最适宜选择的指标是()
A)极差
B)标准差
C)均值
D)变异系数
答案:B
解析:

58.[单选题]下面哪个不是Python Requests库提供的方法?
A)head()
B)post()
C)push()
D)get()
答案:C
解析:题型：

59.[单选题]在数据分析和处理方面具有分析方法丰富、分析模型扩展强、数据挖掘能力强等特点的分析工具是( )。
A)Weka
B)SPSS

C)SAS
D)R
答案:D
解析:

60.[单选题]( )算法是最广泛使用的聚类算法,算法简单,易于理解和操作。
A)gglomerative
B)C.URE
C)K-means
D)k-中心点算法
答案:C
解析:

61.[单选题]如果一个匹配中,任何一个节点都不同时是两条或多条边的端点,也称作( )
A)极大匹配
B)二分匹配
C)完美匹配
D)极小匹配
答案:C
解析:

62.[单选题]下列selenium库的方法中，通过元素名称进行单元素定位的是（）
A)find_element_by_name
B)find_elements_by_name
C)find_elements_by_id
D)find_elements_by_class_name
答案:A
解析:

63.[单选题]依照《中华人民共和国数据安全法》和有关法律、行政法规的规定,( )负责统筹协调网络数据安全和相关监管工作。
A)工业和信息化部
B)国家安全部门
C)国家网信部门
D)通信主管部门
答案:C
解析:

64.[单选题]要求满足连接条件的记录,以及连接条件左侧表中的记录都包含在结果中,应使用_____。
A)左连接
B)右连接

C)内部连接
D)完全连接
答案:A
解析:

65.[单选题]提升决策树法训练效率的措施包括
A)增加树的深度
B)减少数的深度
C)增加学习率
D)减少树的个数
答案:B
解析:

66.[单选题]OLAP的核心是()
A)对用户的快速响应
B)互操作性
C)多维数据分析
D)以上都不是
答案:C
解析:

67.[单选题]使用python处理缺失值的方法中叙述错误的是( )。
A)fillna()填充缺失值
B)dropna()删除缺失值
C)isnull()判断缺失值
D)interpolate()使用中位数填充缺失值
答案:D
解析:

68.[单选题]下列不属于数据质量问题的是( )。
A)缺失值
B)不一致的值
C)重复数据
D)非结构数据
答案:D
解析:

69.[单选题]贝叶斯信念网络由两部分组成，分别是网络结构和()。
A)条件概率
B)先验概率
C)后验概率
D)条件概率表

答案:D
解析:

70.[单选题]在进行数据挖掘任务的时候，通常面临样本数据特征过多的情况，我们可以通过Filter 过滤法选择
那些对我们分析任务更有帮助的特征，下列方法哪个不是用来做特征过滤的（ ）
A)卡方检验
B)F 检验
C)互信息法
D)奇异值分解
答案:D
解析:

71.[单选题]通过建立一个模型来实现已知变量值来预测其他某个变量值属于数据挖掘的哪类任务
A)内容检索
B)建模描述
C)预测建模
D)寻找模式和规则
答案:C
解析:

72.[单选题]DPI采集中，需要获取LTE切换信息，不需要采集的接口是（）
A)S1-MME
B)X2
C)S11
D)S6a
答案:D
解析:

73.[单选题]下面关于因子分析的说法正确的是( )
A)因子分析就是主成分分析
B)因子之间可相关也可不相关
C)因子受量纲的影响
D)可以对因子进行旋转,使其意义更明显
答案:D
解析:

74.[单选题]如下表所示,X={butter, cheese},Y={beer},则置信度confidence(X→Y)=( )。
交易号(TID)
商品(Items)
1beer, diaper, nuts
2beer, biscuit, diaper

3bread, butter, cheese
4beer, cheese, diaper, nuts
5beer, butter, cheese, nuts
A)2/5
B)1/3
C)1/2
D)1/4
答案:C
解析:

75.[单选题]在Numpy包中,计算中位数的函数为( )。
A)numpy.median()
B)numpy.var()
C)numpy.std()
D)numpy.mean()
答案:A
解析:

76.[单选题]以下哪个算法是无监督学习算法（ ）
A)DBSCAN
B)RandomForestRegressor
C)KNN
D)SVC
答案:A
解析:

77.[单选题]当所有观测值都落在回归直线上,则这两个变量之间的相关系数为( )
A)1
B)-1
C)+1 或-1
D)0
答案:C
解析:

78.[单选题]对回归问题和分类问题的评价最常用的指标是
A)准确率
B)召回率
C)误差
D)方差
答案:C
解析:

79.[单选题]关于SQL量词叙述正确的是_____。
A)ANY和ALL是同义词
B)ANY和SOME是同义词
C)ALL和SOME是同义词
D)ALL和EXISTS是同义词
答案:B
解析:

80.[单选题]DBSCAN最大时间复杂度的是
A)O(m)
B)O(m2)
C)O(log m)
D)O(m*log m)
答案:B
解析:

81.[单选题]()提供的支撑技术,有效解决了大数据分析、研发的问题,比如虚拟化技术、并行计算、海量存储和海量管理等。
A)点计算
B)线计算
C)云计算
D)面计算
答案:C
解析:

82.[单选题]EXCEL 中,求标准差的函数是()
A)AVERAGE
B)MEDIAN
C)MODE
D)STDEV
答案:D
解析:

83.[单选题]当时间序列的环比增长速度大体相同时,适宜拟合( )
A)指数曲线
B)抛物线
C)直线
D)对数曲线
答案:A
解析:

84.[单选题]面对人工智能可能存在高度风险，暗藏危机，我们应（）

A)因人类的生物进化速度相当有限，终将被人工智能淘汰，所以要阻断人工智能发展
B)人工智能对人类造成威胁论调没有科研依据，人工智能可以随意发展
C)以尽力发展为前提，拥抱AI技术的同时，多多考虑如何避免损害人类
D)不必去面对此类问题
答案:C
解析:

85.[单选题]下列选项中，属于结构化数据的是____。
A)图像
B)文本
C)办公文档
D)JSON
答案:D
解析:

86.[单选题]有一条关联规则为 A →
B,此规则的信心水平(confidence) 为 60% ,则代表( )
A)买 B 商品的顾客中,有 60% 的顾客会同时购买 A
B)同时购买
A,B 两商品的顾客,占所有顾客的 60%
C)买 A 商品的顾客中,有 60%的顾客会同时购买 B
D)两商品
A,B 在交易数据库中同时被购买的机率为 60%
答案:C
解析:

87.[单选题]在潜在购机用户挖掘时，与以下哪个因素无关
A)上一次购机时间
B)用户偏好的APP使用情况
C)终端品牌
D)套餐消费情况
答案:B
解析:

88.[单选题]Hadoop集群是（ ），通常情况下（ ）。
A)完全开放，可以从互联网进行操作
B)半开放，特殊情况下可以从互联网进行操作
C)半开放，无法从互联网进行操作
D)完全隔离，无法从互联网进行操作
答案:D
解析:

89.[单选题]在比较模型的拟合效果时,甲、乙、丙、丁四个模型的决定系数R^2的值分别约为0.96、0.85、0.80和0.7,则拟合效果好的模型是( )。
A)丁
B)乙
C)丙
D)甲
答案:D
解析:

90.[单选题]假设 {BCE} 为一频繁项目集 (Frequent Itemset),则根据 Apriori Principle 以下何者不是子频繁项目?
A)BC
B)CE
C)C
D)CD
答案:D
解析:

91.[单选题]某牙膏厂原来生产两面针药物牙膏,现在又增加牙刷生产,这属于( )
A)同心多元化
B)水平多元化
C)集团多元化
D)相关多元化
答案:B
解析:

92.[单选题]某文本分类需求，存在一定的数据缺失情况且数据规模较小，能做增量式训练要求的是哪种算法
A)贝叶斯
B)决策树
C)SVM
D)逻辑回归
答案:A
解析:

93.[单选题]下列哪些选项能表示序数属性的数据集中趋势度量( )。
A)四分位数
B)标准差
C)众数
D)均值
答案:C
解析: