

数据挖掘-课堂练习

2021.4.16

杜逆索

此题未设置答案，请点击右侧设置按钮

1. 从数据库的大量数据中挖掘出有用的信息，是指

- ☐ A 知识发现
- ☐ B 机器学习
- ☐ C 数据挖掘
- ☐ D 人工智能

提交

2. 数据挖掘DM的全称是 [填空1] 。
- 数据库中的知识发现的简称是 [填空2] 。

正常使用填空题需3.0以上版本雨课堂

作答

此题未设置答案，请点击右侧设置按钮

3. 数据挖掘的研究内容主要包含的类别有

A

广义知识

B

关联知识

C

分类知识

D

预测型知识

E

偏差型知识

提交

此题未设置答案，请点击右侧设置按钮

4. 不符合挖掘算法的数据特征有

- ☐ A 不完整性
- ☐ B 杂乱性
- ☐ C 统一标准的定义
- ☐ D 含噪声

提交

此题未设置答案，请点击右侧设置按钮

5. “有噪音的” 数据，是指

- ☐ A 音频数据中的杂音
- ☐ B 存在错误或异常的数据
- ☐ C 存在一定意义的数据
- ☐ D 一类因其人烦躁、或音量过强而危害人体健康的声音

提交

此题未设置答案，请点击右侧设置按钮

6. 数据预处理的目的包含

- ☐ A 数据清理
- ☐ B 数据集成
- ☐ C 数据变换
- ☐ D 数据归约

提交

此题未设置答案，请点击右侧设置按钮

7. 下列**不属于**特征选择常用的算法是

- ☐ A 顺序前进法
- ☐ B 顺序后退法
- ☐ C 增r减1法
- ☐ D 增1减r法

提交

此题未设置答案，请点击右侧设置按钮

8. 以下哪个不是数据挖掘的研究内容

- ☐ A Web数据挖掘
- ☐ B 提升算法效率
- ☐ C 多模态数据分析
- ☐ D 核光滑处理

提交

此题未设置答案，请点击右侧设置按钮

9. 预测型知识是指

- ☐ A 可以用来作为预测的知识
- ☐ B 包含时间序列的知识
- ☐ C 可以用来预测类别、趋势
- ☐ D 具有普遍意义、广义用途的知识

提交

此题未设置答案，请点击右侧设置按钮

10. 数据挖掘的功能包含

- ☐ A 类和概念的描述
- ☐ B 关联分析
- ☐ C 分类和聚类
- ☐ D 演变分析

提交

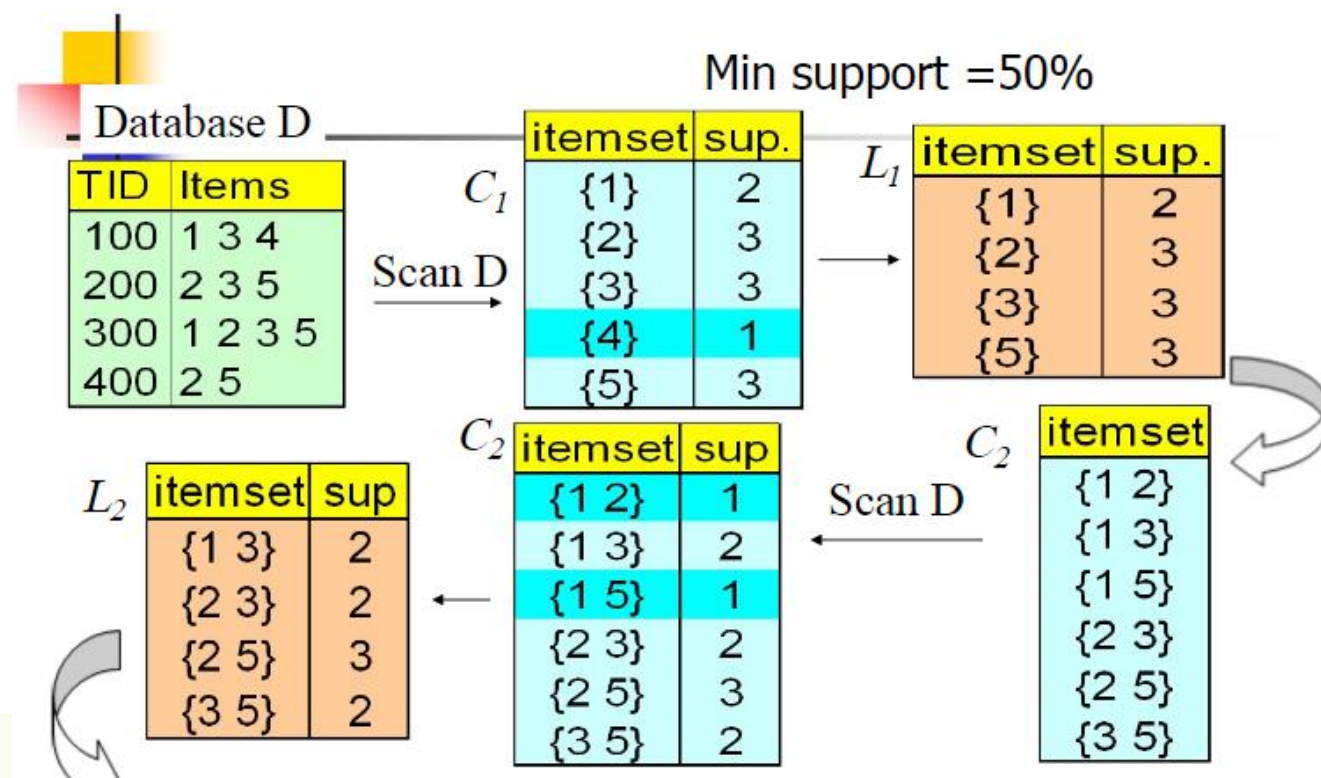
此题未设置答案，请点击右侧设置按钮

11. 下列描述中，_____是关联规则挖掘Apriori算法的优点。

- ☐ A 多次扫描事务数据库，需要很大的I/O负载
- ☐ B 适合稀疏数据集
- ☐ C 可能产生庞大的候选集
- ☐ D 在频繁项目集长度变大的情况下,运算时间显著增加

提交

12. 下列关联规则挖掘算法-Apriori算法的最终结果：
itemset是{ [填空1] }, sup是 [填空2] 。



此题未设置答案，请点击右侧设置按钮

13. Apriori算法的应用领域包含

- ☐ A 商业
- ☐ B 网络安全领域
- ☐ C 高校管理
- ☐ D 移动通信领域

提交

此题未设置答案，请点击右侧设置按钮

14. 决策树的关键点是

- ☐ A 剪枝
- ☐ B 寻找类别分类
- ☐ C 确定分支结点
- ☐ D 计算信息熵

提交

此题未设置答案，请点击右侧设置按钮

15. 决策树ID3算法的特点是

- ☐ A 不容易过拟合
- ☐ B 偏向于对特征数据多的属性
- ☐ C 多变量决策树过于复杂
- ☐ D 递增算法

提交

此题未设置答案，请点击右侧设置按钮

16. 下列关于决策树C4.5算法描述**错误**的是

- ☐ A 分支指标采用绝对增益水平
- ☐ B k次迭代交叉验证
- ☐ C 可解决数值属性
- ☐ D 决策树可产生if-then规则

提交

此题未设置答案，请点击右侧设置按钮

17. 决策树中经典的属性划分方法有

- ☐ A 信息增益
- ☐ B 增益率
- ☐ C 基尼指数
- ☐ D 拉格朗日系数

提交

18. 决策树中信息熵的值越 [填空1] , 则数据样本的纯度就越高。

正常使用填空题需3.0以上版本雨课堂

作答

此题未设置答案，请点击右侧设置按钮

19. 信息增益擅长处理的数据类型

- ☐ A 取值数目较多的属性
- ☐ B 取值数目较少的属性

提交

20. 一般而言，信息增益越 [填空1] ，则意味着使用属性 a 来进行划分所获得的“纯度提升”越大。

此题未设置答案，请点击右侧设置按钮

21. 增益率擅长处理的数据类型

- ☐ A 取值数目较多的属性
- ☐ B 取值数目较少的属性

提交

此题未设置答案，请点击右侧设置按钮

22. 下列关于基尼指数的描述**正确**的是

- ☐ A 基尼指数越大，数据集的纯度越高
- ☐ B 基尼指数越小，数据集的纯度越高
- ☐ C 基尼指数越大，数据集的复杂度越高
- ☐ D 基尼指数越小，数据集的复杂度越高

提交

此题未设置答案，请点击右侧设置按钮

23. 关于过拟合的描述**正确**的是

- ☐ A 近乎完美的预测/区分了训练集上的所有的数据
- ☐ B 不依赖于现有训练数据集
- ☐ C 损失函数值可以降得很低
- ☐ D 在新的测试集上表现平平

提交

此题未设置答案，请点击右侧设置按钮

24. 关于预剪枝的优点，**正确**的是

- ☐ A 降低过拟合风险
- ☐ B 降低欠拟合风险
- ☐ C 减少测试时间
- ☐ D 减少训练时间

提交

此题未设置答案，请点击右侧设置按钮

25. 关于后剪枝的优点，**错误**的是

- ☐ A 降低欠拟合风险
- ☐ B 泛化性能好
- ☐ C 训练时间开销大
- ☐ D 减少测试时间

提交

26. 聚类和分类，最核心的差别，就是进行处理之前，后者有 [填空1] 。

正常使用填空题需3.0以上版本雨课堂

作答

此题未设置答案，请点击右侧设置按钮

27. 下列选项中，哪些是聚类的距离计算属性所包含

- ☐ A 连续属性
- ☐ B 离散属性
- ☐ C 有序属性
- ☐ D 无序属性

提交

28. 聚类性能度量的指标有 [填空1] 指标和 [填空2] 指标。

正常使用填空题需3.0以上版本雨课堂

作答

29. 内部指标中的DB指数是越 [填空1] 越好。

30. 内部指标中的Dunn指数是越 [填空2] 越好。

正常使用填空题需3.0以上版本雨课堂

作答

此题未设置答案，请点击右侧设置按钮

性能度量中，属于外部指标的是

- ☐ A Jaccard系数
- ☐ B FM指数
- ☐ C Cluster指数
- ☐ D Rand指数

提交

此题未设置答案，请点击右侧设置按钮

31. 下列选项中**不属于**聚类中包含的类型是

- ☐ A 原型聚类
- ☐ B 密度聚类
- ☐ C 质量聚类
- ☐ D 层次聚类

提交

此题未设置答案，请点击右侧设置按钮

32. 下列不属于原型聚类算法的是

- ☐ A k均值算法
- ☐ B k近邻算法
- ☐ C 学习向量量化算法
- ☐ D 高斯混合聚类算法

提交

此题未设置答案，请点击右侧设置按钮

33. 近邻算法属于什么类型的学习模型

- ☐ A 有监督学习
- ☐ B 半监督学习
- ☐ C 无监督学习

提交

此题未设置答案，请点击右侧设置按钮

34. 课堂上介绍的最近邻法中常用距离测度有

- ☐ A 欧氏距离
- ☐ B 马氏距离
- ☐ C 一般化的明氏距离
- ☐ D 角度相似性函数

提交

此题未设置答案，请点击右侧设置按钮

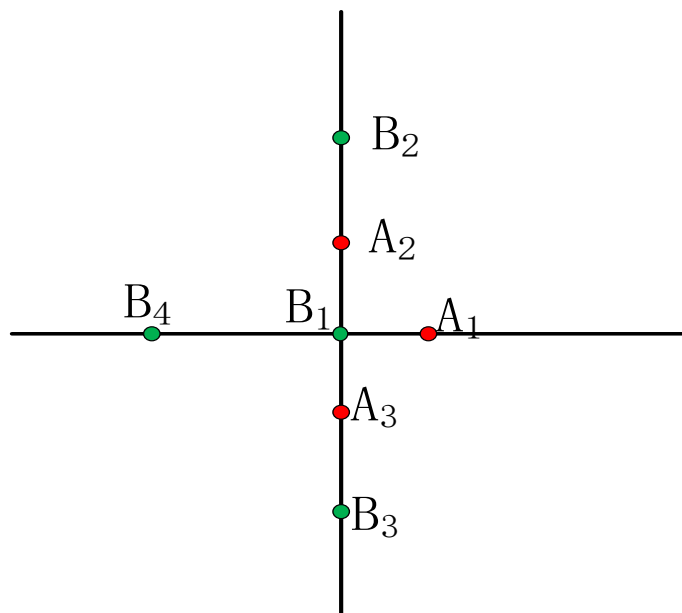
35. K近邻算法的基本要素**不包含**

- ☐ A K 值的选择
- ☐ B 类别的数量
- ☐ C 距离度量
- ☐ D 分类决策规则

提交

36. 设在一个二维空间，A类有三个训练样本，图中用红点表示，B类四个样本，图中用绿点表示。

问题：接近邻法分类实际用到的分界面有 [填空1] 条。



正常使用填空题需3.0以上版本雨课堂

作答

此题未设置答案，请点击右侧设置按钮

37. 下列关于人工神经网络算法的说法**错误**的是

- ☐ A 模仿了生物神经系统运作
- ☐ B 包含了兴奋和抑制状态
- ☐ C 可以实现脉冲与点位的转换
- ☐ D 主要对输入信号进行线性处理

提交

此题未设置答案，请点击右侧设置按钮

38. 课堂中介绍过的常用的激活函数有

- ☐ A 阈值型函数
- ☐ B 分段线性函数
- ☐ C Sigmoid函数
- ☐ D 高斯函数

提交