

单选题 (20分)

第01题

第02题

第03题

第04题

第05题

第06题

第07题

第08题

第09题

第10题

1. 【单选题】用来转换为数据框的函数和用来判断变量是否是逻辑型的函数分别是_____。

- (A) is.data.frame().as.type()
- (B) as.data.frame().is.logical()
- (C) as.data.frame().is.type()
- (D) is.data.frame().as.logical()

2. 【单选题】关于向量、数据框和矩阵的描述正确的是_____。

- (A) 矩阵中各列元素类型可以不同
- (B) 以上说法均不正确
- (C) 数据框中各列元素类型可以不同
- (D) 向量中各元素类型可以不同

3. 【单选题】以下哪些有可能是状态转移矩阵_____。

$$\begin{bmatrix} 0.7 & 0.5 & 0.1 & 0.1 \\ 0.2 & 0.4 & 0.3 & 0.1 \\ 0.3 & 0.6 & 0.5 & 0.1 \\ 0.2 & 0.2 & 0.1 & 0.3 \end{bmatrix}$$

①

$$\begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.35 & 0.25 & 0.4 \\ 0.15 & 0.25 & 0.6 \end{bmatrix}$$

②

$$\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.15 & 0.1625 & 0.4375 \\ 0 & 0 & 0.125 & 0.875 \\ 0 & 0 & 0.1094 & 0.8906 \end{bmatrix}$$

③

$$\begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.45 & 0.15 & 0.4 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

④

- (A) ②③④
- (B) ②③
- (C) ③④
- (D) ①③

单选题 (20分)

- 第01题
- 第02题
- 第03题
- 第04题
- 第05题
- 第06题
- 第07题
- 第08题
- 第09题
- 第10题

4. 【单选题】设最小支持度为55%，从下面购物篮提取的项集中，属于频繁1-项集的是_____。

ID	购买项
1	牛奶, 饼干, 黄油
2	面包, 尿布, 啤酒
3	尿布, 牛奶, 饼干
4	啤酒, 面包, 尿布
5	黄油, 啤酒, 牛奶

- (A) 牛奶, 啤酒, 尿布
- (B) 牛奶, 尿布、面包、啤酒
- (C) 饼干、黄油、尿布、啤酒
- (D) 牛奶、饼干、黄油、面包、尿布、啤酒

5. 【单选题】以下说法正确的是_____。

- ①向量是一维数据，同一对象不允许多种类型
- ②矩阵是二维数据，同一对象不允许多种类型
- ③数组是三维数据，同一对象不允许多种类型
- ④数据框是二维数据，同一对象不允许多种类型

- (A) ①②
- (B) ①②③④
- (C) ①②③
- (D) ①④

6. 【单选题】关于信息熵和信息增益，以下说法正确的是_____。

- ①熵值越大，信息量越大
- ②熵值越小，信息量越大
- ③ID3算法中属性的信息增益越大，会使得类的不确定性程度减少得越大
- ④ID3算法中属性的信息增益越大，会使得类的不确定性程度减少得越小

单选题 (20分)

- 第01题
- 第02题
- 第03题
- 第04题
- 第05题
- 第06题**
- 第07题
- 第08题
- 第09题
- 第10题

6. 【单选题】关于信息熵和信息增益,以下说法正确的是_____。

- ① 熵值越大,信息量越大
- ② 熵值越小,信息量越大
- ③ ID3算法中属性的信息增益越大,会使得类的不确定性程度减少得越大
- ④ ID3算法中属性的信息增益越大,会使得类的不确定性程度减少得越小

- (A) ①③
- (B) ①④
- (C) ②④
- (D) ②③

7. 【单选题】在Rstudio中用read.csv读入文件,读出的结果是_____。

- (A) 逻辑型
- (B) 向量
- (C) 数据框
- (D) 矩阵

8. 【单选题】下列有关聚类、分类的说法正确的是_____。

- (A) 分类是一种无监督学习算法,直接把样本根据某种度量划分为某个类别
- (B) 聚类常常需要先训练,是有监督学习算法
- (C) 分类是有监督学习算法,需要先训练一个分类器然后再预测
- (D) 常见的聚类算法有kmeans、层次聚类、决策树等

9. 【单选题】关于马尔科夫链,以下说法错误的是_____。

- (A) 状态转移概率矩阵为正则概率矩阵时,如稳态分布存在则平稳分布存在且唯一

单选题 (20分)

- 第01题
- 第02题
- 第03题
- 第04题
- 第05题
- 第06题
- 第07题
- 第08题
- 第09题
- 第10题

- (B) 聚类常常需要先训练，是有监督学习算法
- (C) 分类是有监督学习算法，需要先训练一个分类器然后再预测
- (D) 常见的聚类算法有kmeans、层次聚类、决策树等

9. 【单选题】关于马尔科夫链，以下说法错误的是_____。

- (A) 状态转移概率矩阵为正则概率矩阵时，如稳态分布存在则平稳分布存在且唯一
- (B) N个状态的马尔科夫链，稳态分布必定存在
- (C) N个状态的马尔科夫链稳态分布如存在，则平稳分布必然存在且唯一
- (D) 状态转移矩阵为正则概率矩阵时，平稳分布必定存在

10. 【单选题】线性回归方法的因变量是_____数据。

- (A) 其他各项都可以
- (B) 数值型
- (C) 逻辑型
- (D) 字符型

数据分析题 (60...)

第01题

第02题

第03题

第04题

1. 以下程序必须保存在C:\SRC\2720042文件夹下。

打开脚本文件R1001.r, 按下面要求补充下划线处的缺失代码。

[要求]:

R语言有内置数据集mtcars, 该数据集包含汽车模型的数据, 这些数据取自《美国汽车》杂志(1974年《汽车趋势》杂志)。对于每辆汽车, 有11个属性, 以不同的单位(美国单位)表示, 它们如下(按实际程序运行的小数位填写, 不要四舍五入):

- (1) 查看数据集mtcars的维度信息;
- (2) 以油耗为Y轴, 排量为X轴, 画散点图;
- (3) 以油耗为因变量, 排量为自变量, 建立一元线性回归模型;
- (4) 查看回归模型结果, 根据显示结果和题目的文字说明填写相应答案在下划线处。

1	mpg	油耗(每加仑英里)
2	cyl	汽缸数
3	disp	排量(立方英寸)
4	hp	总马力
5	drat	后轴比率
6	wt	重量(1000 磅)
7	qsec	1/4 英里时间
8	vs	发动机缸体
9	am	变速箱 (自动为 0, 手动 1)
10	gear	前进挡的数量
11	carb	化油器数量

数据分析题 (60...)

第01题

第02题

第03题

第04题

2. 以下程序必须保存在C:\SRC\2720043文件夹下。
打开脚本文件R1002.r, 按下面要求补充下划线处的缺失代码。

[要求]:

读取文件夹中的wine.rdata数据集,wine 数据集包含来自 3 种不同起源的葡萄酒, 13 个属性是葡萄酒的 13 种化学成分, 13个属性分别是酒精、苹果酸、灰、灰分的碱度、镁、总酚、黄酮类化合物、非黄烷类酚类、原花色素、颜色强度、色调、稀释葡萄酒的OD280/OD315、脯氨酸, 通过化学分析可以推断葡萄酒的起源。

- (1) 读取该数据集,查看3种葡萄酒各有多少条记录;
- (2) 删除wine第一列数据并保存wine中;
- (3) 加载factoextra程序包, 使用k均值聚类, 根据fviz_nbclust () 函数平均轮廓系数来确定簇数;
- (4) 设置种子数为119, 根据第3步确定的簇数进行k均值聚类, 聚类结果保存在kjl变量中;
- (5) 显示kjl模型的结果, 根据显示结果和题目的文字说明填写相应答案在下划线处。(按实际程序运行的小数位填写, 不要四舍五入)

恢复本题

3. 以下程序必须保存在C:\SRC\2720045文件夹下。
打开脚本文件R1004.r, 按下面要求补充下划线处的缺失代码。

[要求]:

读取文件夹中的wine.csv数据集, 进行主成分分析, 根据要求完成:

数据分析题 (60...)

第01题

第02题

第03题

第04题

3. 以下程序必须保存在C:\SRC\2720045文件夹下。

打开脚本文件R1004.r, 按下面要求补充下划线处的缺失代码。

[要求]:

读取文件夹中的wine.csv数据集, 进行主成分分析, 根据要求完成:

- (1) 读取wine.csv文件到wine变量中;
- (2) 利用相关系数矩阵计算特征值和特征值向量, 存入变量y中;
- (3) 计算前5个主成分的累计贡献率, 按照计算结果将答案填写在相应下划线处。(按实际程序运行的小数位填写, 不要四舍五入, 也不要转化为百分数)

恢复本题

4. 以下程序必须保存在C:\SRC\2720044文件夹下。

打开脚本文件R1003.r, 按下面要求补充下划线处的缺失代码。

[要求]:

在arules包里面有数据集Epub, 记录了2003-2008年间从维也纳经济和工商管理大学电子出版平台下载文档的历史数据。请根据要求完成以下问题:

- (1) 加载arules包, 获取数据并显示数据前6条;
- (2) 计算每个属性的支持度, 并按降序排序;
- (3) 根据第(2)步显示结果把出现最频繁的项填写在题目下划线处;
- (4) 设置支持度为0.001, 置信度为0.1, 进行关联规则挖掘;
- (5) 查看关联规则结果, 根据显示结果和题目的文字说明填写相应答案在下划线处;
- (6) 查看规则的前10条, 在下划线处填写第6条规则的提升度。(按实际程序运行的小数位填写, 不要四舍五入)

R语言 (20分)

要求根据题目的描述在空格内填写正确的答案

1. 【填空题】R语言有自带数据集airmiles, 请计算其1/4分位数: _____.

2. 【填空题】有如下统计结果, 则关于类别的信息熵为 _____ . (写成小数形式, 四舍五入保留4位小数)

类别	复友	不复友
数量	43	37

3. 【填空题】有如下销售数据, 则状态转移概率矩阵中P(畅销→一般)= _____.

月份	1月	2月	3月	4月	5月
销售情况	畅销	畅销	畅销	一般	一般
月份	6月	7月	8月	9月	10月
销售情况	一般	畅销	畅销	一般	一般

4. 【填空题】 $A = \text{matrix}(c(1,2,3,4,5,6), 2, 3)$, 则 $A[2,2] =$ _____.

5. 【填空题】有状态转移率矩阵, 它 _____ 稳态分布。(有/没有)

本题到此为止, 以下内容为空白

第01题
第02题
第03题
第04题
第05题