填空题（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】A=matrix(c(1,2,3,4,5,6),2,3)，则 A[2,2] = _____

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】有状态转移概率矩阵，它 _____ 稳态分布。(有/没有)

5. 【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	复发	不复发
数量	43	37

本题到此为止，以下内容为空白

考试客户端 V11.6--Microsoft Internet Explorer

【姓名】姓名【性别】男【考号】113053【证件】113053【科目】大数据分析与可视化(计算机通过性考试)【院系】不详(****系)

1:53:35

交卷

单选择题 10　填空题 5　数据分析题 4

单选择题（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

在每小题备选答案中选出一个正确的答案。

1. 【单选题】执行下述两个语句：a<- matrix(seq(2,18,by=2),3,3)；a[-2,]，则a的值是_____。

(A) $\begin{bmatrix} 2 & 6 \\ 8 & 12 \\ 14 & 18 \end{bmatrix}$

(B) $\begin{bmatrix} 2 & 4 & 6 \\ 14 & 16 & 18 \end{bmatrix}$

(C) $\begin{bmatrix} 2 & 8 & 14 \\ 4 & 10 & 16 \\ 6 & 12 & 18 \end{bmatrix}$

(D) $\begin{bmatrix} 2 & 8 & 14 \\ 6 & 12 & 18 \end{bmatrix}$

2. 【单选题】有如下格式的数据medicine.txt,现在需要进行关联规则分析，请问下列读取数据命令中正确的是_____。

卡马西平片,内戊酸钠缓释片
奥卡西平片,茴拉西坦分散片
奥卡西平片,丙戊酸钠口服液
丙戊酸钠缓释片,奥卡西平片,茴拉西坦分散片
丙戊酸钠缓释片,奥卡西平片
丙戊酸钠缓释片,奥卡西平片,卡马西平片
...

(A) read.transactions("medicine.txt", format="single", cols=c(1,2), sep=",")
(B) read.transactions("medicine.txt", format=" basket ", sep=",")
(C) read.transactions("medicine.txt", format="basket", cols=c(1,2), sep=",")
(D) read.transactions("medicine.txt", format="single", sep=",")

3. 【单选题】购物篮分析属于

(D) read.transactions("medicine.txt", format="single", sep=",")

3. 【单选题】购物篮分析属于_____。

(A) 聚类分析
(B) 描述性统计
(C) 分类与预测
(D) 关联规则

4. 【单选题】下列关于向量、矩阵、数据框、因子说法正确的是_____。

(A) 矩阵是二维结构，可以使用$符号按名称索引列数据
(B) 向量中的元素可以具有不同的类型
(C) 因子类型常常用来存储连续型变量
(D) 数据框可以包含不同类型的数据

5 【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) as.data.frame();is.type()
(B) is.data.frame();as.logical()
(C) as.data.frame();is.logical()
(D) is.data.frame();as.type()

6. 【单选题】关于信息熵和信息增益，以下说法正确的是_____。

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，信息越确定
②熵值越小，信息越确定
③ID3算法选择信息增益最大的作为根节点
④ID3算法选择信息增益最小的作为根节点

(A) ②③
(B) ①④
(C) ②④
(D) ①③

7. 【单选题】线性回归方法的因变量是_____数据。

(A) 字符型
(B) 数值型
(C) 其他各项都可以
(D) 逻辑型

8. 【单选题】下列有关聚类、分类的说法正确的是_____。

(A) 聚类常常需要先训练，是有监督学习算法
(B) 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
(C) 常见的聚类算法有kmeans、层次聚类、决策树等
(D) 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

9. 【单选题】若有限状态马尔可夫链的状态转移概率矩阵P如下图所示，则其状态概率平稳分布为_____。

9. 【单选题】若有限状态马尔可夫链的状态转移概率矩阵P如下图所示，则其状态概率平稳分布为_____。

$$\begin{pmatrix} 1 & 0 \\ 0.5 & 0.5 \end{pmatrix}$$

(A) 存在平稳分布但不唯一
(B) 无法判断
(C) 存在唯一平稳分布
(D) 没有平稳分布

10. 【单选题】读取文件时，不能在装入内存时被重新赋值的文件格式为_____。

(A) CSV
(B) rdata
(C) txt
(D) rds

考试客户端 V11.6--Microsoft Internet Explorer

【姓名】姓名 【科目】大数据分析与可视化【考号】113053 1:52:21

单选择题 10 填空题 5 数据分析题 4 交卷

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\2720049文件夹下。
打开脚本文件R1008.r，按下面要求补充下划线处的缺失代码。
[要求]:
本题进行马尔可夫预测，根据要求完成下列问题：
（1）补全代码，在R中输入如下图所示一步状态转移概率矩阵；
（2）求出两步状态转移概率矩阵；
（3）如果初始概率向量P0=c(0.5,0.5)，预测2个周期后的概率向量，并在下划线处回答相应问题。

$$\begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$$

恢复本题

2．以下程序必须保存在C:\SRC\2720048文件夹下。
打开脚本文件R1007.r，按下面要求补充下划线处的缺失代码。
[要求]:
文件夹中有，请按要求完成以下并回答下列问题：
（1）读取“iris.csv”文件并删除缺失值；
（2）计算距离矩阵，并采用平均距离对数据进行层次聚类；
（3）若按2类进行划分时，在下划线处回答题目中相应问题。

恢复本题

考试客户端 V11.6--Microsoft Internet Explorer

【姓名】姓名 【科目】大数据分析与可视化 【考号】113053

1:52:01

单选择题 10 填空题 5 数据分析题 4 交卷

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\2720046文件夹下。
打开脚本文件R1005.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iris可直接引用，请使用ID3算法建立决策树模型对其Species属性字段进行分类分析。（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）
(1)加载rpart包；
(2)随机种子设为201，提取75%数据作为训练集，剩余的25%数据作为测试集；
(3)用ID3算法和训练集建立决策树模型；
(4)查看决策树，在题目指定位置处填写正确根节点的属性名；
(5)利用建立好的决策树对测试集数据进行预测，生成混淆矩阵并计算预测准确率填写在指定位置。

恢复本题

4．以下程序必须保存在C:\SRC\2720047文件夹下。
打开脚本文件R1006.r，按下面要求补充下划线处的缺失代码。
[要求]:
文件夹中有longley.rdata数据集，记录着16个年份的宏观经济数据，包含GNP.deflator（GNP平减指数）、GNP（国民生产总值）、Unemployed（失业率）、Armed.Forces（武装力量）、Population（人口）、Year（年份）、Employed（就业率）七个属

4．以下程序必须保存在C:\SRC\2720047文件夹下。

打开脚本文件R1006.r，按下面要求补充下划线处的缺失代码。

[要求]:

文件夹中有longley.rdata数据集，记录着16个年份的宏观经济数据，包含GNP.deflator（GNP平减指数）、GNP（国民生产总值）、Unemployed（失业率）、Armed.Forces（武装力量）、Population（人口）、Year（年份）、Employed（就业率）七个属性。请按照要求完成代码：

（1）读取longley.rdata数据集，删除Year（年份）属性，并在下划线处回答相应问题；

（2）对数据进行z-score规范化；

（3）以GNP.deflator为因变量，建立它与其他所有变量之间的多元线性回归模型；(多个自变量请按照数据集中列的顺序填写，顺序错不给分)

（4）对模型进行逐步法筛选出最优回归方程，并在下划线处回答相应问题；

（5）如果GNP=-1.3，Unemployed=-0.9，Armed.Forces=-1.45，Population=-1.46，Employed=-1.4，那么根据最优模型预测GNP.deflator的值。（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