## 作业2

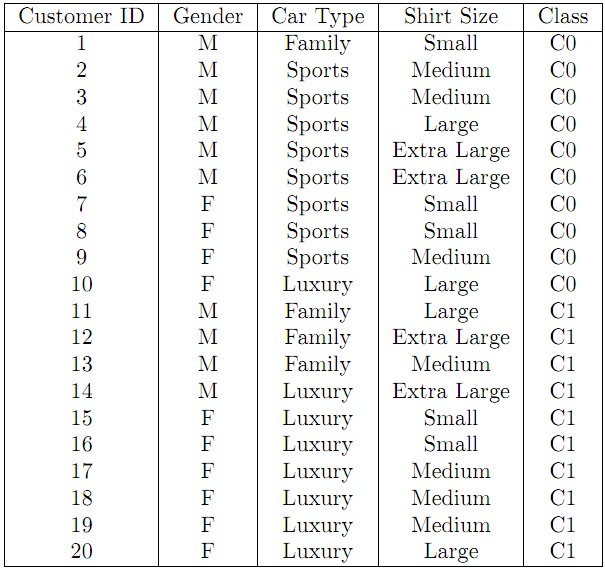
**截止日期：11月. 11，**  **2020**

**提交要求：**

**请 将您的解决方案提交到我们的班级网站。**

**第一部分： 书面部分：**

1.



1. 计算性别、汽车类型和衬衫尺寸的信息增益。
2. 使用信息增益构建决策树。
3. （a）D 为第 1 季度数据集的多层前馈神经网络（一个隐藏 层）签名。标记输入层和输出图层中的节点。
4. 使用上述神经网络，在返回传播算法的一次迭代后显示权重值，给定训练实例"（M、家庭、小"）。指示初始权重值和偏差以及使用的学习率。

3.根据Q1中的*训练*数据集对未知样本 Z 进行分析：

*Z* = （性别 = F， 汽车 Type = 家庭、 衬衫尺寸= 大）。天真的贝叶斯分类器会对 Z 进行分类吗？

4. 假设数据挖掘任务是将以下十个点（x、y、z）表示位置）聚类到三个群集中：

A1（4，2，5），A2（10，5，2），A3（5，8，7），B1（1，1，1），B2（2，3，2），B 3（3，6，9），C1（11，9，2），C2（1，4，6），C3（9，1，7），C4（5，6，7）

距离函数为欧几里德距离。假设最初我们分别将 A1、B1、C1 分配给每个群集的中心。使用 K-Means 算法仅显示

1. 第一轮执行后 三个群集中心
2. 最后三个群集

**第二部分： 实验室部分**

**问题1.**假设超市想推广意大利面pasta。使用"交易"中的数据作为训练数据来构建决策树 （C5.0 算法） 模型，以预测客户是否会购买面食。

1. 使用数据集"事务"构建决策s树，该数据集预测面食是其他字段的函数。将每个字段的"类型"设置为"标志"，将"面食pasta"的"方向"设置为"out"，将COD的"类型"设置为"无类型"，选择"专家"，将"修剪严重性"设置为65，将"每个子分支的最小记录"设置为95。. **手拉手**：显示你树的图。
2. 使用模型（由克莱门汀在上面的步骤1中生成的完整树）对"推出"数据中的20个客户进行预测，以确定客户 是否会购买意大利面。 **交手：** 您对 20 个客户的预测。

**手接**：从决策树（至第 五级 ）确定的面食购买正（是）预测规则。根被视为级别 1）。

**成绩预测**问题**22：**  **考试**

通过对某在线培训系统的标注数据集进行建模，预测其它会员期末考试的结果。 数据集来自在线培训系统的日志，数据包括每个会员的在线学习行为。 请尝试多种不同的模型、不同的参数，建立高质量的预测模型。

训练集有873条记录，测试集有461条记录。 训练集和测试集包含如下变量：

·伊德

在线总时长（分钟）

在线阅读时长（分钟）

在线测试时长（分钟）

全文阅读次数

智能阅读次数

知识点阅读次数

试题阅读次数

回溯原文次数

题库测试次数

仿真考试次数

仿真考试优秀次数

仿真考试良好次数

仿真考试合格次数

仿真考试不合格次数

**Class:** 期末考试结果及格**=1,** 期末考试结果不及格**=0**

1. 对训练数据集执行决策树分类。选择除 ID 以外的所有输入变量。将类的"方向"设置为"out"，将"类型"设置为"标志"。然后，您可以指定"每个子分支的最小记录"和"修剪严重性"，然后单击"使用全局修剪"。 **将测试数据的混乱矩阵交到。** （为"每个子 分支的最小记录"和"修剪严重性"提供最佳参数。

2. 使用默认设置对训练数据集执行神经网络。同样，选择除 ID 以外的所有输入变量。 **将测试数据的混淆矩阵交到。**

3. 使用默认设置对训练数据集执行逻辑回归。同样，选择除 ID 以外的所有输入变量。 **将测试数据的混淆矩阵交到。**

**4. 使用混淆矩阵，对决策树、神经网络和逻辑回归的模型质量进行处理。**