



Business Analytics Homework 3

Instructor: Zach Zhizhong ZHOU

1. 在 D 盘建立目录: D:/BA/Homework/HW03, 将附件中的数据放入该目录中, 以该目录做为工作目录。文件 EastWestAirlinesCluster.xls 包含了一家航空公司的将近 4000 个持有常旅客卡乘客的记录。对于每个乘客, 数据库记录了他们的累积里程和可用于消费的里程信息。我们的目标是对乘客做聚类。打开 EastWest.R, 在 R Studio 中逐行运行代码, 注意在 R Studio 的代码栏窗口左边显示了 R Script 的行号。
 - 1.1) 这个 R Script 使用了哪种聚类算法? 是层次聚类还是 kMeans 聚类? 使用了哪些指标 (measurements) 进行聚类?
 - 1.2) 解释第 21、22 行做了什么工作。
 - 1.3) 解释第 24 行是什么意思, 目的是什么。
 - 1.4) 运行第 26、27 行能得到什么信息?
 - 1.5) 解释第 29-31 行画了什么图, 从这几张图上可以看出什么信息?
 - 1.6) 解释第 33-41 行画了什么图, 这张图可以表达什么信息?
 - 1.7) 使用另外一种聚类指标: ward, 而不是 average (见 45 行), 其他不变。先将 EastWestModel 的数据聚成 4 类, 然后再聚成 2 类 (类似 12 行-27 行的做法)。最后画出类似 29 行-31 行的 3 个图。从图上看, 得到的聚类结果和使用 average 指标得到的聚类结果哪个看起来更好一些, 为什么?
 - 1.8) 本题是附加题: 根据数据分析结果向航空公司提出合理建议帮助航空公司改进目前的里程奖励计划。你可以自己写新的代码对数据做进一步分析。
2. 数据文件 Cosmetics.csv 记录了 1000 条女士购买化妆品的记录。一行表示一条消费记录。列名表示化妆品, 取值 yes 代表购买了该化妆品, 取值 no 则代表未购买。我们希望了解以下问题:
 - 2.1) 如果某个女士购买了 Nail.Polish, 那么在支持度 support 至少 0.1, 信心水平 confidence 至少 0.5 的情况下该女士还会买什么化妆品?
 - 2.2) 如果某个女士购买了 Mascara, 那么在支持度至少 0.2, 信心水平至少 0.5 的情况下该女士还会买什么化妆品?
 - 2.3) 利用以上得到的任意一条规则解释支持度和信心水平是什么意思, 应该如何计算?
 - 2.4) 问题 (2.1) 和 (2.2) 如果只让你分别给出最有用的一条规则, 那么它们是什么?
3. 以下是某投资银行人事部门 2 人的对话:

人事部门干事李先生: 郑总, 今年有 2000 人通过在线求职平台申请我们的 5 个岗位。我刚来公司, 不太清楚我们这边如何筛选面试对象。

人事部门主管郑女士: 你手头不是有去年的数据吗? 去年也有 2000 人申请 5 个岗位, 你分析一下去年的数据应该就知道我们怎么选人了。

人事部门干事李先生: 但问题是去年的 5 个岗位和今年不同, 而且今年我们提高了薪水, 申请者的平均水平和去年相比有很大提升, 比如有不少上海交大毕业的学生。

人事部门主管郑女士: 那你能不能帮我先初步处理一下数据, 我下午再过来看一下申请我们公司都有些什么样的人, 不过我可不想去把那 2000 份申请材料全都读一遍。

 - 3.1) 如果按照郑女士的建议去分析去年的数据, 应该使用什么方法? 预期得到什么结果?
 - 3.2) 李先生要准备下午给郑女士阅读的报告, 应该使用什么方法? 预期得到什么结果?



4. 数据文件 Segmentation.csv 被用来进行市场细分 (Market Segmentation), 其中 Class 列表示消费者类型, 其余 58 列被用来预测消费者类型。请使用现有数据建立一个决策树模型。要求: 每个树叶节点至少包含 100 个数据才能继续长出枝叶, 把决策树画出来, 写出决策树给出的至少 3 条规则。

提交作业的截止时间: **6月11日周日晚上20:00**。

应该通过电子邮件提交到 zachzhoucourse@hotmail.com。提交作业的邮件标题是: “BA HW03 你的名字”。

提交文件列表:

- 1、对题目的解答, 将解答放在一个 Word 或者 PDF 文档里面, 将文档命名为 HW03-你的名字。
- 2、解答题目的完整代码。

以上代码确保放在以下目录当中可以成功执行: D:/BA/Homework/HW03。如果你修改工作目录, 导致代码无法在上面的目录中执行, TA 将不会检查你的代码而是直接给你 0 分!