数据挖掘本科课程教学大纲

编制人：刘建华

审定人：吴宜勇

开课部门：商学院

编制时间：2019.6.30

一、课程基本信息

|  |  |  |
| --- | --- | --- |
| 课程名称 | 中文 | 数据挖掘 |
| 英文 | Data Mining |
| 课程代码 |  | 课程性质 | 专业必修课程 |
| 课程学分 | 2 | 课程学时 | 32 |
| 适用专业 | 经济统计学 | 课程组负责人 | 刘建华 |
| 课程组成员 |  |
| 先修课程 | 计算机应用基础、统计学、计量经济学、多元统计分析 |
| 选用教材 | R语言数据挖掘，薛薇，中国人民大学出版社，2018 |
| 参考书目 | 1.R语言实战，卡巴科弗【美】，人民邮电出版社。 2. 统计学习导论：基于R应用，加雷斯.詹姆斯【美】，机械工业出版社 |
| 推荐教材 | R语言数据挖掘，薛薇，中国人民大学出版社，2018 |

二、课程目标

**（一）课程具体目标**

|  |  |
| --- | --- |
| **序号** | **课程具体目标** |
| 课程目标1 | 掌握数据挖掘的常用方法的基本思想、基本原理、适用场合。 |
| 课程目标2 | 掌握用R软件，用R软件实现各种数据挖掘方法 |

**（二）课程目标与毕业要求的关系**

|  |  |  |
| --- | --- | --- |
| **课程目标** | **支撑的毕业要求** | **支撑的毕业要求指标点** |
| 课程目标1 | 毕业要求1:知识要求 | 1.2 专业知识：牢固掌握本专业基础知识、基本理论，了解数据挖掘理论发展，系统掌握各种常用的数据挖掘方法。 |
| 毕业要求2:能力要求 | 能力要求：理解各种数据挖掘方法所包含的统计思想，将专业理论与知识融会贯通，能综合运用专业知识分析和解决问题。 |
| 毕业要求3：素质要求 | 素质要求：具有良好的统计职业素养，有完整的统计学知识结构和良好的统计素养，具有正确的人生观、价值观和啥结果按 |
| 课程目标2 | 毕业要求1:工具性要求 | 1.1 工具性知识。熟练掌握计算机和R语言编程技术，能用运用现代信息技术进行数据处理、研究分析。 |
| 毕业要求2:能力要求 | 能力要求：理解各种数据挖掘方法所包含的统计思想，将专业理论与知识融会贯通，能综合运用专业知识分析和解决问题。 |

三、课程教学要求与重难点

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **课程内容框架** | **教学要求** | **教学重点** | **教学难点** |
| 1 | 第 1 章 数据挖掘与R语言概述 | 要求了解什么是数据挖掘、数据挖掘方法的特点，掌握R语言入门基础，了解Rstudio工作环境 | 理解什么是数据挖掘、数据挖掘能做什么，数据挖掘方法的特点，了解Rstudio工作环境 | 什么是数据挖掘，数据挖掘能做什么，Rstudio工作环境 |
| 2 | 第2章R的数据组织和整理 | 理解R的数据对象，掌握R的数据对象的创建、访问和互换，掌握数据导入、掌握R语言程序设计基础 | R的数据对象、数据对象的创建、访问和互换、数据导入 | 数据对象的创建、访问和互换、数据导入 |
| 3 | 第3章 R的数据可视化 | 掌握R语言的绘图基础，掌握分布特征的可视化、变量间相关性的可视化、文本数据的可视化 | R语言的绘图基础、分布特征可视化、变量相关性可视化、文本数据可视化 | 绘图基础、分布特征可视化、文本数据可视化 |
| 4 | 第4章 R的近邻分析 | 掌握KNN方法的基本原理，掌握KNN方法的R语言实现 | KNN方法的基本原理，用R语言实现KNN方法 | 用R语言实现KNN方法 |
| 5 | 第5章R的决策树 | 掌握决策树算法原理，理解分类回归树的生长和剪枝，掌握用R语言实现决策树的建立与剪枝，掌握分类回归树的组合预测模型及其R实现，掌握随机森林及其R实现 | 决策树算法和原理、分类回归树的生长与剪枝、回归分类树的组合预测模型、随机森林 | 决策树算法和原理、分类回归树的组合预测模型及其R实现、随机森林及其R实现 |
| 6 | 第 6 章 R的人工神经网络 | 掌握人工神经网络基本原理，算法的基本思想，掌握BP反向传播网络的R语言实现 | 人工神经网络的算法、BP反向传播网络的R语言实现 | 人工神经网络的算法、BP反向传播网络的R语言实现 |
| 7 | 第7 章R的支持向量机 | 掌握支持向量机的原理、算法，掌握不同情况下支持向量机的R语言实现 | 支持向量机的原理、算法、支持向量机分类的R语言实现 | 支持向量机原理、算法、支持向量机分类的R语言实现 |
| 8 | 第8章 R的一般聚类 | 掌握聚类分析的原理、算法，掌握K-Means聚类、层次聚类和EM聚类的R语言实现 | 聚类分析的原理、算法，K-Means聚类、层次聚类和EM聚类的R语言实现 | 聚类分析的原理、算法，K-Means聚类、层次聚类和EM聚类的R语言实现 |
| 9 | 第10 章R的关联分析 | 理解关联规则，掌握关联规则的测度，掌握Apriori和Eclat算法及其R语言实现 | 关联规则的测度， Apriori和Eclat算法及其R语言实现 | Apriori和Eclat算法及其R语言实现 |

四、课程教学内容、教学方式、学时分配及对课程目标的支撑情况

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **课程内容框架** | **教学内容** | **教学方式** | **学时** | **支撑的****课程目标** |
| 1 | 第1章数据挖掘与R语言概述 | 1.1什么是数据挖掘1.2数据挖掘的结果1.3数据挖掘能做什么1.4数据挖掘方法的特点 | 讲授 | 1学时 | 课程目标1 |
| 1.5数据挖掘的典型应用1.6R语言入门必备1.7Rstudio简介 | 讲授、上机实验 | 1学时 | 课程目标2 |
| 2 | 第2章R的数据组织和整理 | 2.1R的数据对象2.2向量的创建和访问2.3矩阵的创建和访问2.4数据框的创建和访问 | 讲授、上机实验 | 1学时 | 课程目标2 |
| 2.5数组和列表的创建和访问2.6数据对象的相互访问2.7导入外部数据 | 讲授、上机实验 | 1学时 | 课程目标2 |
| 2.8R语言程序设计基础2.9R语言数据整理和程序设计综合应用 | 讲授、上机实验 | 1学时 | 课程目标2 |
| 3 | 第3章R的数据可视化 | 3.1绘图基础3.2单变量分布特征的可视化3.3多变量分布特征的可视化 | 讲授、上机实验 | 1学时 | 课程目标2 |
| 3.4变量间相关性的可视化3.5GIS数据的可视化3.6文本词频数据的可视化 | 讲授、上机实验 | 1学时 | 课程目标2 |
| 4 | 第4章R的近邻分析 | 4.1近邻分析：K近邻法 | 讲授、上机实验 | 2学时 | 课程目标1课程目标2 |
| 4.2基于变量重要性的加权K近邻法 | 讲授、上机实验 | 1学时 | 课程目标1课程目标2 |
| 4.3基于观测相似性的加权K近邻法 | 讲授、上机实验 | 1学时 | 课程目标1课程目标2 |
| 5 | 第5章R的决策树 | 5.1决策树算法概述 | 讲授 | 0.5学时 | 课程目标1 |
| 5.2分类回归树的生长过程 | 讲授 | 1.5学时 | 课程目标1 |
| 5.3分类回归树的剪枝 | 讲授 | 0.5学时 | 课程目标1 |
| 5.4分类回归树的R函数和应用示例 | 讲授、上机实验 | 1.5学时 | 课程目标2 |
| 5.5建立分类回归树的组合预测模型 | 讲授、上机实验 | 2学时 | 课程目标1课程目标2 |
| 5.6随机森林 | 讲授、上机实验 | 1学时 | 课程目标1课程目标2 |
| 6 | 第6章R的人工神经网络 | 6.1人工神经网络概述 | 讲授 | 0.5学时 | 课程目标1 |
| 6.2BP反向传播网络 | 讲授 | 1.5学时 | 课程目标1 |
| 6.3BP反向传播网络的R函数和应用示例 | 讲授、上机实验 | 1学时 | 课程目标2 |
| 7 | 第7章R的支持向量机 | 7.1支持向量机分类概述 | 讲授 | 0.5学时 | 课程目标1 |
| 7.2线性可分问题下的支持向量分类 | 讲授 | 1.5学时 | 课程目标1 |
| 7.3广义线性可分问题下的支持向量分类7.4线性不可分问题下的支持向量分类 | 讲授 | 1学时 | 课程目标1 |
| 7.5多分类的支持向量分类7.6支持向量回归 | 讲授 | 1学时 | 课程目标1 |
| 7.7R的支持向量机及其应用示例 | 讲授、上机实验 | 1学时 | 课程目标2 |
| 8 | 第8章R的一般聚类 | 8.1聚类分析概述8.2基于质心的聚类模型：K-Means聚类 | 讲授、上机实验 | 1学时 | 课程目标1课程目标2 |
| 8.3基于质心的聚类模型：PAM聚类8.4基于连通性的聚类模型：层次聚类 | 讲授、上机实验 | 1学时 | 课程目标1课程目标2 |
| 8.5基于统计分布的聚类模型：EM聚类 | 讲授、上机实验 | 1学时 | 课程目标1课程目标2 |
| 9 | 第10章R的关联分析 | 10.1简单关联规则及其测度10.2Apriori算法及其应用示例 | 讲授 | 1学时 | 课程目标1 |
| 10.3Eclat算法及其应用示例10.4简单关联分析的应用示例 | 讲授、上机实验 | 1学时 | 课程目标1课程目标2 |
| 10.5序列关联分析及SPADE算法 | 讲授 | 1学时 | 课程目标1课程目标2 |

五、课程目标与考核内容

|  |  |
| --- | --- |
| **课程目标** | **考核内容** |
| 课程目标1 | 数据挖掘方法的适用场合的掌握，对数据挖掘算法和基本思想思想的理解。 |
| 课程目标2 | 用R语言实现各种数据挖掘方法 |

六、考核方式与评价细则

|  |  |  |
| --- | --- | --- |
| **考核方式** | **比例** | **考核/评价细则** |
| 课堂表现 | 10% | 考勤记录和课堂表现情况加分、扣分记录 |
| 作业 | 30% | 数据挖掘作业完成情况汇总记录表 |
| 期末考试 | 60% | 上机考试。用R语言进行编程，实现对数据集的数据挖掘分析。以正确性、规范性和简洁性为评价依据 |