

课程名称：数据挖掘及其应用

一、课程编码：1700118

课内学时： 48 学分： 3

二、适用学科专业： 统计专业

三、先修课程： 概率论与数理统计、数学分析

四、教学目标

数据挖掘是统计学专业的专业课程，本课程以数据挖掘为主要内容，讲述实现数据挖掘的主要挖掘算法和应用方法，教授大型数据挖掘软件 SPSS Modeler 的使用。通过本课程的学习，使学生理解数据挖掘的基本概念和方法，学习和掌握 SPSS Modeler 软件的使用，并通过对实际数据的分析提升学生数据分析和处理的能力。

五、教学方式

以多媒体教学手段为主要形式的课堂教学。

六、主要内容及学时分配

- | | |
|--------------------------|------|
| 1 数据挖掘和 modeler 使用概述 | 3 学时 |
| 1.1 数据挖掘的产生背景 | |
| 1.2 什么是数据挖掘 | |
| 1.3 modeler 软件概述 | |
| 2 modeler 的数据读入和数据集成 | 4 学时 |
| 2.1 变量类型 | |
| 2.2 读入数据 | |
| 2.3 生成数据流 | |
| 2.4 数据集成 | |
| 3 modeler 的数据理解 | 3 学时 |
| 3.1 变量说明 | |
| 3.2 数据质量的评估和调整 | |
| 3.3 数据的排序 | |
| 3.4 数据的分类汇总 | |
| 4 modeler 的数据准备 | 4 学时 |
| 4.1 变量变换 | |
| 4.2 变量派生 | |
| 4.3 数据精简 | |
| 4.4 数据筛选 | |
| 4.5 数据准备的其他工作 | |
| 5 modeler 的基本分析 | 4 学时 |
| 5.1 数值型变量的基本分析 | |
| 5.2 基本描述分析 | |
| 6 modeler 的图形绘制 | 2 学时 |
| 6.1 绘制散点图、直方图等 | |
| 7 分类预测：modeler 的决策树 | 5 学时 |
| 7.1 决策树算法概述 | |
| 7.2 modeler 的 c5.0 算法及应用 | |

- 7.3 modeler 的分类回归树及应用
- 7.4 modeler 的 chaid 算法及应用
- 7.5 modeler 的 quest 算法及应用
- 7.6 模型的对比分析
- 8 分类预测：modeler 的人工神经网络 4 学时
 - 8.1 神经网络算法概述
 - 8.2 modeler 的 b-p 反向传播网络
 - 8.3 modeler 的 b-p 反向传播网络的应用
 - 8.4 modeler 的径向基函数网络及应用
- 9 分类预测：modeler 的支持向量机 5 学时
 - 9.1 支持向量分类的基本思路
 - 9.2 支持向量分类的基本原理
 - 9.3 支持向量回归
 - 9.4 支持向量机的应用
- 10 分类预测：modeler 的贝叶斯网络 5 学时
 - 10.1 贝叶斯方法基础
 - 10.2 贝叶斯网络概述
 - 10.3 tan 贝叶斯网络
 - 10.4 马尔科夫毯网络
 - 10.5 贝叶斯网络的应用
- 11 探索内部结构：modeler 的聚类分析 5 学时
 - 11.1 聚类分析的一般问题
 - 11.2 modeler 的 k-means 聚类及应用
 - 11.3 modeler 的两步聚类及应用
 - 11.4 modeler 的 kohonen 网络聚类及应用
 - 11.5 基于聚类分析的离群点探索
- 12 探索内部结构：modeler 的关联分析 4 学时
 - 12.1 简单关联规则及其有效性
 - 12.2 modeler 的 apriori 算法及应用
 - 12.3 modeler 的序列关联及应用

七、考核与成绩评定

成绩评定：平时作业、测验及日常考核等占 30%，期末数据分析大报告考试占 70%，按百分制给出最终成绩。

八、参考书及学生必读参考资料

1. 薛薇. 基于 SPSS Modeler 的数据挖掘(第 2 版)[M]. 北京: 中国人民大学出版社, 2014.
2. Jiawei Han、Micheline Kamber, 范明、孟小峰等译. 数据挖掘概念与技术. 北京: 机械工业出版社. 2012.

九、大纲撰写人：黄宝胜