《机器学习》教学大纲

**课程名称：**机器学习

**课程类别：**必修

**适用专业：**大数据技术类相关专业

**总学时：**48学时（其中理论36学时，实验12学时）

**总学分：**3学分

# 课程的性质

大数据时代已经到来，在广告营销、文化管理、商业经济等领域中基于数据和分析去发现问题并做出科学、客观的决策越来越重要。本课程主要面向非计算机类专业学生，介绍如何利用大数据分析方法来实现常见数据分析任务，侧重于方法的应用和问题的解决，注重案例结合和实际操作的学习，强调学生掌握具体数据分析方法并可以自主开展各种数据分析活动。为了推动我国大数据，云计算，人工智能行业的发展，面向社会数据分析人才需求，开设大数据分析与应用课程。课程任务是通过本课程的学习，使学生学会使用Python进行数据质量校验、可视化绘图、数据处理、特征工程、构建有/无监督和智能推荐模型，并详细拆解学习回归、分类、聚类和智能推荐4个企业案例和一个综合案例，将理论与实践相结合，为将来从事机器学习研究、工作奠定基础，同时提高学生的自我学习能力和创新能力。

# 课程学时分配

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 教学内容 | 理论学时 | 实验学时 | 其它 |
| 1 | 第1章 机器学习概述 | 2 | 0 |  |
| 2 | 第2章 数据准备 | 2 | 2 |  |
| 3 | 第3章 特征工程 | 2 | 2 |  |
| 4 | 第4章 有监督学习 | 8 | 4 |  |
| 5 | 第5章 无监督学习 | 6 | 2 |  |
| 6 | 第6章 智能推荐 | 4 | 2 |  |
| 7 | 第7章 市财政收入分析 | 2 | 4 |  |
| 8 | 第8章 基于非侵入式电力负荷监测与分解的电力分析 | 2 | 4 |  |
| 9 | 第9章 航空公司客户价值分析 | 2 | 4 |  |
| 10 | 第10章 广电大数据营销推荐 | 2 | 4 |  |
| 11 | 第11章 基于TipDM数据挖掘建模平台实现航空公司客户价值 | 2 | 2 |  |
|  | 总计 | 34 | 30 |  |

# 教学内容及学时安排

### 理论教学

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **章节名称** | **主要内容** | **教学目标** | **学时** |
| 1 | 机器学习概述 | 1. 掌握机器学习的概念 2. 掌握机器学习通用流程 3. 了解机器学习的应用场景 4. 了解机器学习的常用工具 | 1. 掌握机器学习的概念、流程与应用场景 2. 了解机器学习的常用工具 | 2 |
| 2 | 数据准备 | 1. 数据质量较检 2. 数据分析趋势探查 3. 数据清洗 4. 数据合并 | 1. 掌握数据质量校检 2. 掌握数据分析趋势探查 3. 掌握数据清洗与合并 | 2 |
| 3 | 特征工程 | 1. 特征变换 2. 特征选择 | 1. 掌握特征变换 2. 掌握特征选择 | 2 |
| 4 | 有监督学习 | 1. 了解有监督学习概念 2. 性能度量 3. 线性模型 4. K近邻分类 5. 决策树 6. 支向量机 7. 神经网络 8. 集成学习 | 1. 掌握性能度量 2. 掌握线性模型 3. 掌握K邻近分类 4. 掌握决策树 5. 掌握神经网络 | 8 |
| 5 | 无监督学习 | 1. 了解无监督学习概念 2. PCA降维 3. 核化线性降维 4. 原型聚类 5. 密度聚类 6. 层次聚类 | 1. 掌握降维 2. 掌握聚类 | 6 |
| 6 | 智能推荐 | 1. 了解智能推荐的概念 2. 性能度量 3. 关联规则 4. 协同过滤 | 1. 掌握性能度量 2. 掌握关联规则 3. 掌握协同过滤 | 4 |
| 7 | 市财政收入分析 | 1. 分析财政收入预测背景 2. 了解财政收入预测的方法 3. 熟悉财政收入预测的步骤与流程 4. 了解相关性分析 5. 分析计算结果 6. 了解Lasso回归方法 7. 分析Lasso回归结果 8. 了解灰色预测算法 9. 了解SVR算法 10. 分析预测结果 | 1. 熟悉财政收入预测的步骤和流程 2. 掌握相关性分析方法与应用 3. 掌握使用Lasso模型选取特征的方法 4. 掌握灰色预测的原理与应用 5. 掌握支持向量回归算法的基本原理与应用 | 2 |
| 8 | 基于非侵入式电力负荷监测与分解的电力分析 | 1. 分析电力分项计量的背景 2. 了解设备分类预测的方法 3. 熟悉设备分类预测的步骤与流程 4. 处理缺失值 5. 合并数据与构建特征 6. 了解K近邻算法 7. 分析预测结果 | 1. 熟悉设备分类预测的步骤和流程 2. 掌握处理缺失值的方法 3. 掌握合并数据与构建特征的方法 4. 掌握K近邻的原理与应用 | 2 |
| 9 | 航空公司客户价值分析 | 1. 分析航空公司现状 2. 认识客户价值分析 3. 熟悉航空客户价值分析的步骤与流程 4. 处理缺失值与异常值 5. 构建航空客户价值分析关键特征 6. 标准化LRFMC 5个特征 7. 了解K-Means聚类算法 8. 分析聚类结果 | 1. 熟悉航空客户价值分析的步骤和流程 2. 了解RFM模型的基本原理 3. 掌握K-Means算法的基本原理与使用方法 4. 比较不同类别客户的客户价值，制定相应的营销策略 | 2 |
| 10 | 广电大数据营销推荐 | 1. 目标分析 2. 数据准备 3. 特征工程 4. 模型构建 5. 性能度量 6. 结果分析 | 1. 掌握目标分析 2. 掌握模型构建 | 2 |
| 11 | 基于TipDM数据挖掘建模平台实现航空公司客户价值 | 1. 平台简介 2. 案例应用 | 1. 掌握平台简介 2. 掌握案例分析 | 2 |
| 学 时 合 计 | | | | 34 |

### 实验教学

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **实验项目名称** | **实验要求** | **学时** |
| 1 | 数据准备 | 1. 掌握缺失值与异常值的分析与处理 2. 掌握数据分布与趋势探查 3. 掌握数据合并 | 2 |
| 2 | 特征工程 | 1. 掌握特征变换 2. 掌握特征选择 | 2 |
| 3 | 有监督学习 | 1. 掌握性能度量 2. 掌握线性模型 3. 掌握K近邻分类 4. 掌握决策树 5. 掌握神经网络 | 4 |
| 4 | 无监督学习 | 1. 掌握PCA降维 2. 掌握核化线性降维 3. 掌握原型聚类 4. 掌握密度聚类 5. 掌握层次聚类 | 2 |
| 5 | 智能推荐 | 1. 构建关联规则模型 2. 构建智能推荐模型 3. 评价智能推荐模型 | 2 |
| 6 | 市财政收入分析 | 1. 分析财政收入数据特征的相关性 2. 使用Lasso回归选取财政收入预测的关键特征 3. 使用灰色预测和SVR构建财政收入预测模型 4. 评价SVR模型 | 4 |
| 7 | 基于非侵入式电力负荷监测与分解的电力分析 | 1. 处理数据缺失值 2. 合并设备数据 3. 构建周波数据特征 4. 构建K近邻分类预测模型 5. 评价K近邻分类预测模型 | 4 |
| 8 | 航空公司客户价值分析 | 1. 处理数据缺失值与异常值 2. 构建航空客户价值分析的关键特征 3. 标准化LRFMC 5个特征 4. 构建K-Means聚类模型 5. 评价K-Means聚类模型 | 4 |
| 9 | 广电大数据营销推荐 | 1. 掌握广电大数据营销推荐目标分析 2. 掌握广电大数据营销推荐模型构建 | 4 |
| 10 | 基于TipDM数据挖掘建模平台实现航空公司客户价值 | 1. 掌握TipDM平台简介 2. 掌握TipDM平台应用案例分析 | 2 |
| 学 时 合 计 | | | 30 |

# 考核方式

突出学生解决实际问题的能力，加强过程性考核。课程考核的成绩构成 = 平时作业（10%）+ 课堂参与（20%）+ 期末考核（70%），期末考试建议采用开卷形式，试题应包括基本概念、绘图、分组聚合、数据合并、数据清洗、特征工程、模型构建等部分，题型可采用判断题、选择、简答、应用题等方式。

# 教材与参考资料

### 教材

何伟，张良均．机器学习原理与实战[M]．北京：人民邮电出版社．2021．

### 参考资料

[1] 张良均．Python数据分析与挖掘实战[M]．北京：机械工业出版社．2015．

[2] 张良均．Python与数据挖掘 [M]．北京：机械工业出版社．2016．

[3] 黄红梅，张良均．Python数据分析与应用[M]．北京：人民邮电出版社．2018．

[4] 林耀进，张良均．Python机器学习编程与实战[M]．北京：人民邮电出版社．2020．