**《机器学习基础》课程教学大纲**

1. **课程名称：** **机器学习基础**
2. **课程代码:**  **CST30104**
3. **课程英文名称：**Basic of Machine Learning
4. **课程负责人：** 邹东升
5. **学时与学分**： 48学分，总学时48：理论课40学时，实验课16学时
6. **课程性质：** 专业课
7. **课程类型:**  选修
8. **适用专业**： 计算机科学与技术/信息安全/物联网工程
9. **选课对象**： 本科生
10. **预修课程**： 高等数学、线性代数、概率论与数理统计，程序设计基础、数据结构、算法分析与设计
11. **使用教材：**

[1] 周志华. 机器学习[M], 清华大学出版社，2016.

[2] 李航. 统计学习方法[M]. 清华大学出版社, 2012.

[3] Mitchell, T. M. McGraw-Hill .Machine Learning[M], 机械工业出版社，1997.

教材选用说明：其中教材[1]作为课程主要用书，教材[2-3]作为课外阅读资料。

1. **参考书目：**

[1] 机器学习之路： 阿布等，电子工业出版社，2017

[2] 数学之美：吴军，人民邮电出版社，2014.

1. **开课单位**： 计算机学院
2. **课程描述**

机器学习研究计算机程序如何随着经验积累自动提高性能，是一门多领域交叉学科，涉及概率论、统计学、凸优化、算法复杂度理论等多门学科，目前在很多领域取得成功应用，包括计算机视觉、自然语言处理、生物特征识别、搜索引擎、医学诊断、语音识别等等。

《机器学习基础》是一门面向计算机类专业大学三年级学生开设的专业选修课程，是计算机类专业学生学习人工智能领域知识的核心课程。本课程的主要目标是通过本课程的理论学习与实践，使学生掌握经典的机器学习算法的思想、原理；掌握线性模型、决策树、神经网络、支持向量机、贝叶斯、集成学习、聚类等不同算法、模型的评价指标；能够编程实践解决分类、回归、聚类等实际工程问题。课程的教学内容包括模型评估与选择、分类算法、聚类算法、降维、特征选择等机器学习核心内容。

1. **课程支撑的毕业要求指标点及课程目标**

**1.本课程支撑的毕业要求指标点**

（1）指标点R2.2：能够应用计算机科学基本原理，借助文献研究，识别、表达、分析计算机系统和计算机应用系统中的复杂工程问题，并获得有效结论

（2）指标点R4.3：能够搜索和整合各类资源（文献、软/硬件、数据等），探索归纳本专业前沿技术和发展趋势，得到合理有效的结论

（3）指标点R10.2：对计算机科学与技术领域及其相关行业的国际状况有基本了解，具备专业英语交流能力，能够在跨文化背景下进行沟通和交流。

**2.本课程拟达到的教学目标**

通过本课程的理论学习与实践，使学生掌握经典的机器学习算法的思想、原理；掌握线性模型、决策树、神经网络、支持向量机、贝叶斯、集成学习、聚类等不同算法、模型的评价指标；能够编程实践解决分类、回归、聚类等实际工程问题。培养学生的计算思维能力，为他们进一步开展相关领域的学习和科研打下良好基础。具体来说，主要教学目标包括：

（1）理解和掌握分类、回归、聚类算法的原理和方法；

（2）熟练使用开发工具、编程语言将算法用于工程实践中的机器学习问题

（3）查阅最新文献，了解机器学习、深度学习相关研究的最新进展。

**3.课程教学目标与毕业要求指标点的关系**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 课程教学目标 | 毕业要求二级指标点 | | | |
| R2.2 | R4.3 | R10.2 |  |
| 子目标1 | ✓ |  |  |  |
| 子目标2 |  | ✓ |  |  |
| 子目标3 |  |  | ✓ |  |
|  |  |  |  |  |

课程教学子目标1是培养学生能够理解和掌握分类、回归、聚类算法的原理和方法，具体包括：线性模型、决策树、神经网络、支持向量机、贝叶斯、集成学习、聚类等不同算法。本课程通过课后编程实践作业等环节来培养、考核学生此方面的能力，因此可以对毕业要求指标点R2.2的达成提供支撑。

课程教学子目标2是培养学生能够熟练使用开发工具、编程语言将算法用于工程实践中的机器学习问题,学生需要调研、选择业界主流的Python、Pycharm、Anaconda等编程语言和工具将机器学习算法应用于工程实践项目。课程通过课后编程实践作业等环节来培养考核学生此方面的能力，因此可以对毕业要求指标点R4.3的达成提供支撑。

课程教学子目标3是培养学生查阅最新文献，了解机器学习、深度学习相关研究的最新进展能力，在神经网络等重要章节部分，要求学生自行阅读最新国际前沿科研文献，追踪最新进展和动态，并在课程上进行讲解分享来培养和考核学生此方面的能力。因此可以对毕业要求指标点R10.2的达成提供支撑。

1. **课程教学方式/方法**

对于课程教学目标1，主要采用课堂教学的方式，通过知识教授、问题讨论等途径，使学生能够理解和掌握机器学习算法的思想、原理。

对于课程教学目标2，主要采用课后实践项目作业的方式，通过合理布置的分类和聚类算法及应用的任务，使学生能够熟练使用开发工具、编程语言将算法用于工程实践中的机器学习问题。

对于课程教学目标3，主要通过课堂讨论、翻转教学的方式来训练学生。通过布置相应的主题，要求学生独立查阅最新文献，了解机器学习、深度学习相关研究的最新进展，并在课堂上公开讨论分享，通过这种训练能够为学生打下良好的基础。

1. **课程教学内容与学时**

**1.课堂教学（40学时）**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **一级知识点** | **二级知识点** | **三级知识点** | **掌握层次** | **学时数分配** |
| 绪论 | 机器学习基本概念 | 机器学习基本术语 | 记忆 | 2 |
| 假设空间 | 理解 |
| 归纳偏好 | 理解 |
| 机器学习研究现状与趋势 | 机器学习研究现状 | 理解 | 2 |
| 机器学习发展趋势 | 理解 |
| 线性模型 | 线性模型 | 线性回归、线性判别分析 | 理解 | 2 |
| 多分类学习 | 多分类 学习 | 理解 | 2 |
| 决策树 | 算法原理 | 基本流程、划分选择 | 理解 | 3 |
| 剪枝处理 | 应用 |
| 连续与缺失值处理 | 应用 |
| 多变量决策树 | 多变量决策树概念及原理 | 理解 | 2 |
| 神经网络 | 基本概念 | 神经元模型、感知机与多层网络 | 理解 | 2 |
| 算法原理 | 误差逆传播算法、全局最小与局部 极小 | 应用 | 2 |
| 深度学习 | 概念、主要模型、应用现状 | 应用 | 2 |
| 支持向量机 | SVM模型 | 间隔与支持向量机、对偶问题、核函数、软间隔与正则化 | 应用 | 3 |
| 贝叶斯分类 | 基本原理 | 贝叶斯决策论、极大似然估计 | 理解 | 2 |
| 分类器 | 朴素、半朴素贝叶斯分类器 | 应用 | 2 |
| EM算法 | EM优化算法 | 应用 | 2 |
| 集成学习 | 基本概念 | 个体与集成Boosting Bagging | 理解 | 2 |
| 随机森林 | RandomForest算法 | 应用 | 2 |
| 聚类 | 基本原理 | 聚类任务、性能度量、距离计算、 | 理解 | 2 |
| 聚类算法 | 原型聚类、层次聚类、密度聚类 | 应用 | 2 |
| 降维及度量学习 | 降维 | 主成分分析 | 应用 | 2 |
| 特征选择与稀疏学 习 | 特征选择 | 过滤式选择、包裹式选择、嵌入式选择与L1正则化 | 应用 | 2 |

注释1：课程内容按照相对独立性用层次结构列出1、2、3级知识点单元。

注释2：对知识学习的要求一般分为“记忆，理解，应用”三个层次。

**十七、 能承担此课的教师**

邹东升、何中市、陈自郁、谢洪

教学大纲制订者：邹东升

教学大纲审定者：葛亮