**《统计机器学习》教学大纲**

一、课程基本信息

课程代码：16163804

课程名称：统计机器学习

英文名称：Machine Learning

课程类别：专业课

学 时：64学时

学　　分：4学分

适用对象: 应用统计学专业

考核方式：考试

先修课程：数学分析，高等代数，概率论、数理统计

二、课程简介

统计机器学习是计算机科学以及信号信息处理领域中重要的热点研究领域。随着移动互联网、物联网等的快速发展，机器学习在多个领域（数据挖掘、大数据分析、视频技术、音频技术、智能机器人技术等）成为关键核心和支撑技术。

三、课程性质与教学目的

课程性质属于专业必修课，教学目的：主要讲授机器学习相关基础理论、主流机器学习思想和方法；介绍深度学习基础，为学生开展相关领域技术开发和科学研究奠定基础。

课程思政总体思路：统计机器学习是一门研究对各种分类问题与回归问题进行数据建模的课程，在课程讲授过程中，要结合知识内容培养学生整体观、系统观、联系观；同时，本课程也是一门与实际问题相结合的课程，要培养同学们理论与实践相结合、学以致用的正确学习观。

四、课程的内容与基本要求

**第一章 绪论**

1. 目的与要求

 1. 熟悉机器学习的基本概念、机器学习的一般模型。

2. 了解机器学习的主要内容。

1. 教学内容

 第一节 基本术语

 第二节 假设空间与归纳偏好

 第三节 发展历程与应用现状

（三）思政内容：

了解机器学习的发展历史，增强学生科技兴国的信念。

**第二章 模型评估与选择**

1. 目的与要求

 1. 掌握模型评估的方法。

2. 熟悉模型的性能度量。

3. 了解偏差与方差的关系。

1. 教学内容

 第一节 经验误差与过拟合

 第二节 评估方法

 第三节 性能度量

 第四节 比较检验

 第五节 偏差与方差

（三）思政内容：

通过深入讲解机器学习中模型评估与选择的的内容，与学生讨论科学与技术、理论与实际的辩证关系，培养学生既要认识理论的高度，也要重视掌握实践中解决问题的具体办法。

**第三章 线性模型**

1. 目的与要求

 1. 掌握线性模型的基本原理。

2. 熟悉多元线性回归模型。

3. 理解多分类学习的拆分方法。

4. 了解类别不平衡问题的解决。

1. 教学内容

 第一节 线性回归

 第二节 对数几率回归

 第三节 线性判别分析

 第四节 多分类学习

 第五节 类别不平衡问题

（三）思政内容：

通过线性模型的学习，引导学生进一步认识抽象理论的重要性，重视理论学习。

**第四章 决策树**

1. 目的与要求

 1. 掌握决策树的基本原理。

2. 熟悉划分选择的评价方法。

3. 理解剪枝的基本方法。

4. 了解连续与缺失值的处理方法。

1. 教学内容

 第一节 基本原理

 第二节 划分选择

 第三节 剪枝处理

 第四节 连续与缺失值

（三）思政内容：

通过深入讲解决策树模型，与学生讨论科学与技术、理论与实际的辩证关系，培养学生既要认识理论的高度，也要重视掌握实践中解决问题的具体办法。

**第五章 神经网络**

1. 目的与要求

 1. 掌握神经网络模型的基本原理。

2. 熟悉误差逆传播算法。

3. 理解全局最小与局部最小的差别。

4. 了解深度学习。

1. 教学内容

 第一节 神经元模型

 第二节 感知机与多层网络

 第三节 误差逆传播算法

第四节 全局最小与局部最小

 第五节 深度学习

（三）思政内容：

通过深入讲解当下最热门的神经网络模型与深度学习，呈现科技的魅力，增强学生学习科技的兴趣和热情。

**第六章 支持向量机**

1. 目的与要求

 1. 掌握支持向量机的原理。

2. 熟悉间隔的计算。

3. 理解核函数的作用。

1. 教学内容

 第一节 间隔与支持向量

 第二节 对偶问题

 第三节 核函数

 第四节 软间隔与正则化

第五节 支持向量回归

 第六节 核方法

（三）思政内容：

支持向量机必须依照所解决问题的具体关系进行建模，通过讲授这个核心道理，增强学生具体问题具体分析的思想和能力。

**第七章 贝叶斯分类器**

1. 目的与要求

 1. 掌握贝叶斯分类器的基本原理。

2. 熟悉极大似然估计的过程。

3. 理解EM算法。

1. 教学内容

 第一节 贝叶斯决策论

 第二节 极大似然估计

 第三节 朴素贝叶斯分类器

 第四节 半朴素贝叶斯分类器

 第五节 贝叶斯网

（三）思政内容：

通过讲解贝叶斯理论在分类问题上的应用，引导学生面对复杂问题的时候，要分清主次，好钢用在刀刃上，把关键资源用于解决关键问题。

**第八章 集成学习**

1. 目的与要求

 1. 掌握集成学习的基本思想。

2. 理解随机森林的详细过程。

1. 教学内容

 第一节 个体与集成

 第二节 Boosting

第三节 Bagging与随机森林

 第四节 结合策略

第五节 多样性

（三）思政内容：

通过讲解集成学习模型，培养学生从整体把握问题，再逐步细化的科研思想与方法。

**第九章 聚类**

1. 目的与要求

 1. 掌握聚类的基本原理。

2. 熟悉聚类的主要方法。

3. 理解距离的计算。

1. 教学内容

 第一节 聚类任务

 第二节 性能度量

 第三节 距离计算

（三）思政内容：

与当前热门的分类问题相结合，引导同学们做事情要抓重点分主次。

**第十章 降维与度量学习**

1. 目的与要求

 1. 掌握降维的基本原理。

2. 熟悉主成分分析的过程。

1. 教学内容

 第一节 k近邻学习

 第二节 低维嵌入与主成分分析

 第三节 核化线性降维

 第四节 流行学习

（三）思政内容：

通过对降维方法的讲解，引导同学们进一步认识计算机与人类的优缺点，引导学生使用计算机工具更正确的态度。

五、学时分配

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **教学环节****教学时数****课程内容** | **讲****课** | **习****题****课** | **讨****论****课** | **实验** | **其他教学环节** | **小****计** |
| 第一章 | 3 | 　 | 　 | 　 | 　 | 3 |
| 第二章 | 6 | 　 |  | 2 | 　 | 8 |
|  第三章 | 6 | 　 | 1 | 2 | 　 | 9 |
|  第四章 | 3 | 　 | 1 | 2 | 　 | 6 |
| 第五章 | 8 | 　 | 　 | 2 | 　 | 10 |
|  第六章 | 5 | 　 | 　 | 2 | 　 | 7 |
| 第七章 | 3 | 　 | 　 | 2 | 　 | 5 |
|  第八章 | 8 | 　 | 　 | 2 | 　 | 10 |
| 第九章 | 3 | 　 | 　 | 　 | 　 | 3 |
| 第十章 | 3 | 　 | 　 | 　 | 　 | 3 |
| 合计 | 48 |  | 2 | 14 |  | 64 |

1. 推荐教材和教学参考资源

 1、周志华，机器学习，清华大学出版社，2016

2、李航，统计学习方法，清华大学出版社，2012

 大纲修订人：叶伟杰 修订日期：2020年12月

 大纲审定人： 审定日期：