

# 《机器学习》

## 课程教学大纲

### 一、课程基本信息

|        |  |   |    |    |   |
|--------|--|---|----|----|---|
| 课程类型   | 总学时为学时数  | <input checked="" type="checkbox"/> 理论课（含上机、实验学时）                                       |    |    |   |
|        | 总学时为周数   | <input type="checkbox"/> 实习 <input type="checkbox"/> 课程设计 <input type="checkbox"/> 毕业设计 |    |    |   |
| 课程编码   | 7293531  | 总学时   | 64 | 学分 | 4 |
| 课程名称   | 机器学习   |   |    |    |   |
| 课程英文名称 | Machine Learning                                 |   |    |    |   |
| 适用专业   | 数据科学与大数据技术                                       |   |    |    |   |
| 先修课程   | (7030701) 高等数学、(7101201) 线性代数、(7029501) 概率论与数理统计 |   |    |    |   |
| 开课部门   | 信息学院人工智能系（大数据）                                   |   |    |    |   |

### 二、课程性质与目标

本课程为数据科学与大数据技术专业必修课。本课程主要讲授机器学习相关知识，包括有关机器学习的基本概念，模型评估与选择方法，线性分类模型的基本形式与算法设计，浅层学习与深度学习的典型算法框架，支持向量机、贝叶斯等经典方法原理，集成学习的思想，特征选择与稀疏学习方法以及机器学习在信号处理中的典型应用等内容。通过基本思想、基本算法的引入、推导并配以习题进行分析，使学生能够对机器学习相关算法有较为全面的理解，并能够应用于实践中，学以致用。

课程目标 1：学生应掌握机器学习的相关基础知识，熟悉机器学习在信号处理中的典型应用，了解机器学习的发展趋势及面临的问题。

课程目标 2：学生应能通过利用一些机器学习方法构建实际问题的解决方案，锻炼灵活运用各种算法的能力；同时通过对经典算法的发展历程、思想演变等的学习，培养发现问题、解决问题的能力以及创新思维。

课程思政目标：作为人工智能时代新工科的核心课程之一，具有时代性和应用性强的特点，将从科技强国战略、科技应用方向和敬业精神方面开展，全方位将社会主义核心价值观引导于知识传授和能力培养之中，在提高学生专业能力的同时，帮助学生树立正确的学习理念、思考方法和民族自豪感。

### 三、 课程教学基本内容与要求

#### 第一章 绪论

##### （一）教学基本内容

- a) 基本术语
- b) 假设空间
- c) 归纳偏好
- d) 发展历程
- e) 应用现状

##### （二）教学基本要求：

- a) 掌握：机器学习的基本术语和相关概念。
- b) 理解：归纳偏好的原理。
- c) 了解：机器学习的发展历程和应用现状。

#### 第二章 模型评估与选择

##### （一）教学基本内容：

- 2.1 经验误差与过拟合
- 2.2 评估方法
- 2.3 性能度量
- 2.4 比较检验
- 2.5 偏差与方差

##### （二）教学基本要求：

- 1、掌握：评估方法和性能度量的定义，比较检验的方式和偏差与方差的权衡方法。
- 2、理解：经验误差与过拟合的含义。
- 3、了解：评估方法的几种选择。

#### 第三章 线性模型

##### （一）教学基本内容：

- 3.1 基本形式
- 3.2 线性回归
- 3.3 最小二乘
- 3.4 线性判别分析

##### （二）教学基本要求：

- 1、掌握：线性回归模型、最小二乘、线性判别分析算法流程。
- 2、理解：线性模型的相关理论。

- 3、了解：线性模型的基本形式。

## 第四章 决策树

### （一）教学基本内容：

- 4.1 基本流程
- 4.2 划分选择
- 4.3 剪枝处理
- 4.4 连续与缺失值
- 4.5 多变量决策树

### （二）教学基本要求：

- 1、掌握：决策树学习的基本流程和划分选择方法。
- 2、理解：剪枝处理策略。
- 3、了解：连续与缺失值处理方法和多变量决策树的思想。

## 第五章 神经网络

### （一）教学基本内容：

- 5.1 神经元模型
- 5.2 误差逆传播算法
- 5.3 全局最小与局部最小
- 5.4 随机梯度下降法
- 5.5 随机向量泛函连接网络
- 5.6 深度学习

### （二）教学基本要求：

- 1、掌握：神经元模型结构，误差逆传播算法思想，RVFL网络算法。
- 2、理解：全局最小与局部最小、随机梯度下降等参数寻优方法。
- 3、了解：神经网络的基本思想和深度学习框架。

## 第六章 支持向量机

### （一）教学基本内容：

- 6.1 间隔与支持向量
- 6.2 对偶问题
- 6.3 核函数
- 6.4 软间隔与正则化
- 5.5 支持向量回归
- 6.6 核方法

### （二）教学基本要求：

- 1、掌握：支持向量机算法原理。

2、理解：核函数在支持向量机中的作用。

3、了解：支持向量回归。

## 第七章 贝叶斯分类器

(一) 教学基本内容：

7.1 贝叶斯决策论

7.2 极大似然估计

7.3 朴素贝叶斯分类器

7.4 贝叶斯网

7.5 EM 算法

(二) 教学基本要求：

1、掌握：朴素贝叶斯分类器原理。

2、理解：极大似然估计和贝叶斯决策论思想。

3、了解：EM 算法思想。

## 第八章 集成学习

(一) 教学基本内容：

8.1 个体与集成

8.2 Boosting

8.3 Bagging 与随机森林

8.4 结合策略

8.5 多样性

(二) 教学基本要求：

1、掌握：Boosting 与 Bagging 原理。

2、理解：随机森林算法。

3、了解：其他的结合策略。

## 第九章 聚类

(一) 教学基本内容：

9.1 聚类任务

9.2 性能度量

9.3 距离计算

9.4 原型聚类

9.5 密度聚类

9.6 层次聚类

(二) 教学基本要求：

1、掌握：原型聚类原理。

- 2、理解：距离计算的方法。
- 3、了解：密度聚类 and 层次聚类。

## 第十章 降维与度量学习

### （一）教学基本内容：

- 10.1 k 近邻学习
- 10.2 低维嵌入
- 10.3 主成分分析
- 10.4 核化线性降维
- 10.5 流形学习
- 10.6 度量学习

### （二）教学基本要求：

- 1、掌握：k 近邻学习的工作机制和主成分分析的原理。
- 2、理解：低维嵌入的基本概念和降维方法。
- 3、了解：流形学习和度量学习的概念。

## 第十一章 特征选择与稀疏学习

### （一）教学基本内容：

- 11.1 子集搜索与评价
- 11.2 过滤式选择
- 11.3 包裹式选择
- 11.4 嵌入式选择与 L1 正则化
- 11.5 稀疏表示与字典学习
- 11.6 压缩感知

### （二）教学基本要求：

- 1、掌握：特征选择的概念和 L1 正则化。
- 2、理解：特征选择方法的原理。
- 3、了解：稀疏学习与压缩感知的概念。

## 第十二章 计算学习理论

### （一）教学基本内容：

- 12.1 基础知识
- 12.2 PAC 学习
- 12.3 有限假设空间
- 12.4 VC 维

### （二）教学基本要求：

- 1、掌握：VC 维的意义与用法，PAC 的概念与算法流程。

- 2、理解：有限假设空间相关理论。
- 3、了解：计算学习理论的基础知识。

### 第十三章 半监督学习

#### （一）教学基本内容：

- 13.1 未标记样本
- 13.2 生成式方法
- 13.3 半监督 SVM
- 13.4 图半监督学习
- 13.5 基于分歧的方法
- 13.6 半监督聚类

#### （二）教学基本要求：

- 1、掌握：半监督学习的相关概念。
- 2、理解：半监督学习的常用算法。
- 3、了解：协同训练算法。

### 第十四章 概率图模型

#### （一）教学基本内容：

- 14.1 隐马尔可夫模型
- 14.2 马尔可夫随机场
- 14.3 条件随机场
- 14.4 学习与推断
- 14.5 近似推断
- 14.6 话题模型

#### （二）教学基本要求：

- 1、掌握：隐马尔可夫模型的结构。
- 2、理解：隐马尔可夫模型的基本问题和学习与推断的方法。
- 3、了解：话题模型的相关概念。

### 第十五章 规则学习

#### （一）教学基本内容：

- 15.1 基本概念
- 15.2 序贯覆盖
- 15.3 剪枝优化
- 15.4 一阶规则学习
- 15.5 归纳逻辑程序设计

#### （二）教学基本要求：

- 1、掌握：规则学习的基本概念和目标。
- 2、理解：RIPPER 算法基本流程。
- 3、了解：归纳逻辑程序设计思想。

## 第十六章 强化学习

### （一）教学基本内容：

- 16.1 任务与奖赏
- 16.2 K-摇臂赌博机
- 16.3 有模型学习
- 16.4 免模型学习
- 16.5 值函数近似
- 16.6 模仿学习

### （二）教学基本要求：

- 1、掌握：强化学习的策略。
- 2、理解：强化学习与监督学习的联系与区别。
- 3、了解： $\epsilon$ -贪心算法。

## 四、课程学时分配

| 教学内容          | 讲授 | 实验 | 上机 | 课内<br>学时<br>小计 | 课外<br>学时 |
|---------------|----|----|----|----------------|----------|
| 1. 绪论         | 2  | 0  | 0  | 2              | 0        |
| 2. 模型评估与选择    | 4  | 0  | 0  | 4              | 0        |
| 3. 线性模型       | 4  | 0  | 2  | 6              | 0        |
| 4. 决策树        | 4  | 0  | 0  | 4              | 0        |
| 5. 神经网络       | 4  | 0  | 4  | 8              | 0        |
| 6. 支持向量机      | 4  | 0  | 2  | 6              | 0        |
| 7. 贝叶斯分类器     | 4  | 0  | 0  | 4              | 0        |
| 8. 集成学习       | 4  | 0  | 0  | 4              | 0        |
| 9. 聚类         | 4  | 0  | 2  | 6              | 0        |
| 10. 降维与度量学习   | 2  | 0  | 2  | 4              | 0        |
| 11. 特征选择与稀疏学习 | 2  | 0  | 2  | 4              | 0        |
| 12. 计算学习理论    | 2  | 0  | 0  | 2              | 0        |

|           |    |   |    |    |   |
|-----------|----|---|----|----|---|
| 13. 半监督学习 | 2  | 0 | 2  | 4  | 0 |
| 14. 概率图模型 | 2  | 0 | 0  | 2  | 0 |
| 15. 规则学习  | 2  | 0 | 0  | 2  | 0 |
| 16. 强化学习  | 2  | 0 | 0  | 2  | 0 |
| 合 计       | 48 | 0 | 16 | 64 | 0 |

## 五、 教学设计与教学组织

- 1、本课程以课堂讲授为主，辅以一定的讨论环节。
- 2、使用 MS PowerPoint 幻灯片作为主要教学辅助工具，主要方法用软件演示其效果。

## 六、 教材与参考资料

### 1、教材

《统计学习方法》，李航著，清华大学出版社，2012 年第 1 版，9787302275954

### 2、参考资料

《机器学习》，周志华著，清华大学出版社，2016 年第 1 版，9787302423287

## 七、 课程考核方式与成绩评定标准

本课程成绩由平时成绩及期末考试成绩二部分组成。课程成绩以百分制计算，平时成绩占 30%（其中出勤成绩占 10%，作业成绩占 20%），期末考试成绩占 70%。

## 八、 大纲制(修)订说明

无。

大纲执笔人：曲洪权

大纲审核人：曲洪权

开课系主任：王彦平

开课学院教学副院长：宋威

制（修）订日期：2022 年 2 月