

《机器学习》大纲

一、课程名称：机器学习

二、课程性质：选修、理论+实践课

三、学时与学分：40 学时（理论学时：28，实验学时：12）， 2.5 学分：

四、课程先导课：

线性代数（或矩阵论），概率论与数理统计，微积分，C 语言程序设计（或 Python 语言设计），算法设计与分析

五、课程介绍

“机器学习”是一门理论性、技术性和实践性很强的专业选修课程，为计算机科学中人工智能相关课程的核心和基础。课程主要讲授机器学习的基础理论、概念和方法，主要内容包括：监督学习、无监督学习的简介，k 近邻和 k-means、感知机、贝叶斯与概率估计、朴素贝叶斯法、逻辑回归、梯度下降、线性回归、支持向量机、经验风险最小化、方差与偏差估计、核方法、高斯过程、KD 树和决策树、集成方法等。目标让学生掌握机器学习的基本理论与算法，为进一步学习人工智能相关课程和从事相关科研与实践打下坚实的基础。

六、课程目标

《机器学习》的具体目标包括：

目标 1:掌握机器学习的基本思想原理、流行算法和技术。为毕业要求 1 提供支持。

目标 2:使学生理解概率论、统计、算法等基础交叉学科知识和思想在计算机相关问题的建模发挥的作用，理解相关模型的思想本质，学生对机器学习领域的研究对象、研究目标和研究方法有全局性的了解；为毕业要求 1 提供支持。

目标 3:能认识到计算机科学与技术、人工智能科学与技术日新月异的发展特点，以机器学习相关学习任务为例，使学生认同自主学习和终身学习的必要性。为毕业要求 12 提供支持。

目标 4:通过习题讨论、实验和课程设计，使学生巩固理论思想和知识，培养学生分析模型、动手实现机器学习中不同任务和应用，以及解决科研与实际问题的能力。这需要通过不断学习掌握新技术和新方法才能做，从而锻炼自主学习的能力。为毕业要求 12 提供支持。

七、课程目标对毕业要求的支撑关系

支撑的毕业要求二级指标点	课程目标
1.1 能将数学、自然科学和信息科学的语言工具用于计算机复杂工程问题的表述。	目标 1-2
1.2 能针对计算机复杂工程问题的具体对象进行建模和求解。	目标 1-2
12.1 能认识到计算机技术日新月异的发展特点，认同自主学习和终身学习的必要性	目标 3
12.2 具备自主学习能力，能通过多种途径拓展自己的知识和能力，包括理解能力，归纳总结的能力和提出问题的能力等	目标 4

八、教学设计及对课程目标的支持

第一章 引言

本章主要知识点包括机器学习概述，发展历史，机器学习算法的主要类型。包括机器学习是什么，用来做什么；机器学习方法的分类与综述；机器学习的基本概念。

1. 教学目标

- 1) 了解机器学习主要应用
- 2) 了解机器学习基本技术方法

本章教学支持课程目标 1 和课程目标 4。

2. 教学重点

- 1) 机器学习基本概念
- 2) 机器学习算法的主要类型

3. 教学难点

- 1) 学习概率论中相关知识思想，应用分析解决实际机器学习中问题。

4. 教学环节设计

结合现代机器学习的发展，开展文献阅读与讨论，较全面地介绍自然语言处理领域的核心概念、任务和方法。

5. 思政元素

从人工智能的发展历史以及三次浪潮，说明科学发展乃至事物发展具有前进性和曲折性。培养学生从事人工智能领域机器学习基础研究的兴趣，树立对国家人工智能发展战略的理想信念。

以 AlphaGo 与 AlphaGoZero 为例，说明自我学习和终身学习的主动学习精神。

第二章 基本数学及信息学理论(学生自看*)

本章的主要知识点包括概率论基础及信息论基础以及线性代数,最大似然估计等机器学习知识与方法。

- 概率 Probability, 条件概率 Conditional Probability, 期望 Expectation, 方差 Variance, 标准差 Deviation, 协方差 Covariance, 二项分布 Binomial Distribution, 高斯分布 Gaussian Distribution 等;
- 熵 Entropy, 联合熵 Joint Entropy, 条件熵 Conditional Entropy, 互信息 Mutual Information, KL 散度 KL-Divergence, 交叉熵 Cross Entropy
- 最大似然估计 Maximum Likelihood Estimation;
- 梯度下降 Gradient Descent;
- 参数学习 Parameter Estimation;

1. 教学目标

- 1) 了解线性代数等基本数学知识
- 2) 了解概率论基础
- 3) 了解信息论基础

本章教学支持课程目标 1 和目标 2。

2. 教学重点

统计概率论基础。

3. 教学难点

经典概率论知识: 条件概率, , 期望, 方差, 标准差, 协方差, 二项分布, 高斯分布等。

4. 教学环节设计

针对教学内容展开讨论, 重视理解和掌握基本原理和方法, 从数理基础、方法与应用、前沿专题三个层次培养学生的科学思维方法。

第三章 监督学习

1. 教学目标

- 4) 监督学习的概念;
- 5) 监督学习的基本元素。

本章教学支持课程目标 1、目标 2 和课程目标 3。

2. 教学重点

- 1) 监督学习的基本元素;

3. 教学难点

- 1) 监督学习与无监督学习的概念;

2) 监督学习的基本元素;

4. 教学环节设计

示例空间、目标函数、训练集、测试集、假设空间、损失函数、泛化等监督学习的基本元素的学习

第四章 K 近邻算法

本章主要知识点包括 K 近邻算法。

1. 教学目标

K 近邻算法的概念和要素。

本章教学支持课程目标 1。

2. 教学重点

掌握 K 近邻算法在高维情形下的求解难度以及高维流形下的可用性。

3. 教学难点

K 近邻算法在高维情形下的求解难度以及高维流形下的可用性。

4. 教学环节设计

对应设计课堂讨论，在图像数据上的应用。

5. 思政元素

通过 K 近邻算法，说明团结与协作，集体精神的重要性。通过对此类算法的学习，将个体竞争上升到群体协作，培养学生正确认识个人与集体、个人与国家、合作与竞争等辩证统一关系，形成和树立竞争中合作、合作中竞争的意识，深刻理解集思广益、汇集众智的实质内涵，为贯彻构建人类命运共同体战略贡献更多智慧。

第五章 感知机

本章主要知识点包括线性分类、感知机算法。

1. 教学目标

感知机的概念和要素。

本章教学支持课程目标 1。

2. 教学重点

掌握感知机算法的求解过程和适用性。

3. 教学难点

感知机算法的求解过程和适用性。

4. 教学环节设计

对应设计课堂讨论，在真实数据上的应用。

5. 思政元素

通过感知机模型的发现及历史，说明坚持理想信念、持之以恒对于科研工作是非常重要的。从唯物史观的角度，分析和认识科学真理与谬误的转化过程，科学技术总是在实践中得到发展，从知识驱动到数据驱动再到联合驱动，就是科学技术在不同认知阶段的发展规律，是推动科学真理和谬误的不断转化过程，正是三者之间对立统一的螺旋式上升，推动了科学理论的伟大变革。

第六章 贝叶斯和概率统计

本章主要知识点包括联合概率分布、最大先验和最大后验。

1. 教学目标

联合概率分布、最大先验和最大后验。

本章教学支持课程目标 1。

2. 教学重点

掌握最大先验和最大后验。

3. 教学难点

最大先验和最大后验。

4. 教学环节设计

对应设计课堂讨论，在真实数据上的应用。

第七章 逻辑回归、线性回归

本章主要知识点包括逻辑回归、通用的梯度下降算法、线性回归。

1. 教学目标

逻辑回归、梯度下降、线性回归的概念和算法。

本章教学支持课程目标 1。

2. 教学重点

掌握逻辑回归算法、线性回归以及适用范围。

3. 教学难点

逻辑回归算法、线性回归模型与算法以及适用范围。

4. 教学环节设计

对应设计课堂讨论、实验，在真实数据上的应用。

第八章 支持向量机

本章主要知识点包括支持向量机的模型、算法。

1. 教学目标

线性可分与不可分情形下的支持向量机的模型与算法、对偶优化问题。

本章教学支持课程目标 1、3。

2. 教学重点

掌握线性可分与不可分情形下的支持向量机的模型与算法。

3. 教学难点

线性可分与不可分情形下的支持向量机的对偶优化问题、主问题和对偶问题的关系。

4. 教学环节设计

对应设计课堂讨论、实验和在真实数据上的应用。

第九章 经验风险最小化原则、偏差与方差估计

本章主要知识点包括经验风险最小化原则、偏差与方差估计。

1. 教学目标

经验风险最小化原则、偏差与方差估计。

本章教学支持课程目标 1、4。

2. 教学重点

掌握经验风险最小化原则、偏差与方差估计。

3. 教学难点

经验风险最小化原则、偏差与方差估计。

4. 教学环节设计

对应设计课堂讨论。

第十章 核方法

本章主要知识点包括特征设计、核的设计、通用的核方法、良定义的核、核线性回归和基于核的支持向量机。

1. 教学目标

核方法的基本理论和方法。

本章教学支持课程目标 1、3、4。

2. 教学重点

掌握通用的核方法、核线性回归和基于核的支持向量机。

3. 教学难点

核线性回归和基于核的支持向量机。

4. 教学环节设计

对应设计课堂讨论和实验。

第十一章 高斯过程

本章主要知识点包括高斯过程的基本理论与方法。

1. 教学目标

高斯过程的基本理论与方法。

本章教学支持课程目标 1。

2. 教学重点

掌握高斯过程的基本理论与方法以及适用范围。

3. 教学难点

高斯过程的基本理论与方法以及适用范围。

4. 教学环节设计

对应设计课堂讨论，在真实数据上的应用。

第十二章 KD 树与决策树

本章主要知识点包括 KD 树、决策树的基本理论与方法。

1. 教学目标

KD 树、决策树的基本理论与方法。

本章教学支持课程目标 1、3。

2. 教学重点

掌握 KD 树、决策树的基本理论与方法以及适用范围。

3. 教学难点

KD 树、决策树的基本理论与方法以及适用范围。

4. 教学环节设计

对应设计课堂讨论，在真实数据上的应用。

第十三章 人工神经网络、多层感知机

本章主要知识点包括人工神经网络的基本结构和优化方法、多层感知机的基本理论与方法。

1. 教学目标

人工神经网络的基本模型。

本章教学支持课程目标 1。

2. 教学重点

掌握人工神经网络基础、BP、多层感知机、卷积神经网络的基本模型。

3. 教学难点

BP 过程。

4. 教学环节设计

对应设计课堂讨论，在真实数据上的应用。

5. 思政元素

深度学习作为一种实现机器学习的技术，其主要是利用复杂的深度神经网络，使模型对数据的理解更加深入。其中，神经元以及由神经元广泛互联所构成的神经网络，是需要掌握的基础与核心，而在对神经元与神经网络的认识中，可引入对立统一的哲学思想，培养应对和解决困难的能力。

第十四章 集成方法(选讲*)

本章主要知识点包括集成的思想、Bagging、Boosting 的基本理论与方法。

1. 教学目标

Bagging、Boosting 的基本理论与方法。

本章教学支持课程目标 1、3。

2. 教学重点

掌握 Bagging、Boosting 的基本理论与方法以及适用范围。

3. 教学难点

Bagging、Boosting 的基本理论与方法以及适用范围。

4. 教学环节设计

对应设计课堂讨论，在真实数据上的应用。

九、实验内容

序号	教学内容	教学目标	教学方式	课程目标
1	K 近邻	掌握 KNN 基本原理；能使用 KNN 解决部分机器学习基本问题。	教师演示、实验+教师指导+检查	目标 1 目标 2 目标 3
2	逻辑回归	掌握逻辑回归基本理论与数学推导；理解逻辑回归的适用场景；能基于模型训练后的实验现象提出相应的理论解释。	教师演示、实验+教师指导+检查	目标 1 目标 2 目标 3
3	朴素贝叶斯	掌握朴素贝叶斯基本理论与数学推导；理解朴素贝叶斯的三种类型和适用场景；能基于模型训练后的实验现象提出相应的理论解释。	教师演示、实验+教师指导+检查	目标 1 目标 2 目标 3
4	支持向量机	理解支持向量机模型的理论基础与数学推导；理解支持向量机的不同求解方法；具有编写支持向量机模型代码的能力。	教师演示、实验+教师指导+检查	目标 1 目标 2 目标 3
5	基于核的支持向量机	理解核函数在支持向量机中的作用；掌握基本公式推导；能区分不同核函数。	教师演示、实验+教师指导+检查	目标 1 目标 2 目标 3
6	高斯过程	理解高斯过程的理论基础与数学推导；理解高斯过程的求解方法；具有编写高斯过程模型代码的能力。	教师演示、实验+教师指导+检查	目标 1 目标 2 目标 3

十、教与学

本课程采用课堂讲授和上机实践相结合的教学模式。教学过程中采用电子教案演示讲解、现场实际演示和共同讨论等方法。

1.教学方法

1) 通过课堂讲授，让学生了解如何运用机器学习求解问题的基本方法；

2) 通过合理的实验设计，着力培养学生使用不同机器学习方法解决问题的工程实践能力；

3) 以提升完成度为导向的过程管理，执行严格的考勤管理，努力营造良好学习氛围，要求学生在规定地点、时间内集中进行实验；

4) 营造以学习者为中心的学习环境。每个班配备一名指导教师和助教，对实验过程中出现的技术问题及时与学生沟通与交流，引导学生主动分析与解决问题，指导及时、到位。

2.学习方法

1) 学生要努力实践，通过具体的实践，加深对课堂所学内容的理解，真正达到掌握本课程所学内容的目的；

2) 掌握有关的基本原理和方法，注重理论指导下的工程实践；

3) 实验中遇到问题要积极思考，训练发现问题、分析问题、解决问题的能力；同学之间也要主动交流与讨论,遇到难题也可及时与任课老师交流；

4) 注意对实验结果进行工程化的分析和评价。

3.思政元素

通过实验，激发学生投身国产 IT 生态自主可控事业的使命感，同时增强学生肩负起时代责任担当的信心。

十一、学时分配

序号	主要内容	学时分配
1	引言	课内 2
2	基本数学及信息学理论*	课内 0+课外 4
3	监督学习	课内 2+课外 4
4	K 近邻算法	课内 2+课外 4
5	感知机	课内 2+课外 4
6	贝叶斯和概率统计	课内 2+课外 4
7	逻辑回归、线性回归	课内 2+课外 4
8	支持向量机	课内 4+课外 4
9	经验风险最小化原则、偏差与方差估计	课内 2+课外 4
10	核方法	课内 4+课外 4
11	高斯过程	课内 2+课外 4
12	KD 树与决策树	课内 2+课外 4
13	人工神经网络、多层感知机	课内 2+课外 4

14	集成方法*	课内 0+课外 4
总计		40

十二、课程考核与成绩评定

1.课程成绩构成

课程最终成绩由考勤、课后作业与阅读笔记报告、实验及课堂展示成绩、课程报告综合而成，各部分成绩的比例如下：

- 1) 考勤和课堂平时表现：10%。
- 2) 三次实验：30%。
- 3) 课程报告：60%。

课程考核成绩评定如表 1 所示：

表 1 机器学习课程考核与成绩评定

课程目标	毕业要求	考核与评价方式及成绩比例		成绩占比(约)
		考勤与实验	课程报告	
1	课程目标 1	20	40	60%
2	课程目标 2	10	30	40%

2.考核与评价标准

1) 作业与实验成绩考核与评价标准

表 2 机器学习课程考勤与实验考核与成绩评定

课程目标	评价标准			
	优秀	良好	及格	不及格
目标 1 目标 2	按时参加课堂学习，在规定的时间内完成全部实验任务，报告撰写优秀。	按时参加课堂学习，能完成全部实验任务，报告撰写良好。	按时参加课堂学习，能完成实验任务，结果存在少量问题，报告撰写基本满足要求。	未按时参加课堂学习，实验任务完成情况差，结果存在较多问题，报告撰写有问题。

2) 课程考核与成绩评定

根据期末考试的报告评分标准进行评定。

机器学习课程组

2015年6月制定;2019年12月修订;
2021年10月修订