东北大学秦皇岛分校本科课程教学大纲

一、课程基本信息

<table>
<tr><td>开课单位</td><td colspan="3">悉尼智能科技学院</td><td>课程代码</td><td colspan="5">3060513056</td></tr>
<tr><td>中文名称</td><td colspan="8">机器学习</td></tr>
<tr><td>英文名称</td><td colspan="8">Machine Learning</td></tr>
<tr><td>总学时</td><td>44</td><td rowspan="2">理论学时</td><td rowspan="2">32</td><td rowspan="2">实验学时</td><td rowspan="2">12</td><td rowspan="2">上机学时</td><td rowspan="2">0</td><td rowspan="2">设计学时</td><td rowspan="2">0</td><td rowspan="2">课外学时</td><td rowspan="2">0</td></tr>
<tr><td>学　分</td><td>2.5</td></tr>
<tr><td>课程类型</td><td colspan="11">☐理论课程　☑理论（含实验）课程　☐实践类课程</td></tr>
<tr><td>课程类别</td><td colspan="11">☐通识类 ☐学科基础类 ☑专业方向类 ☐实践类</td></tr>
<tr><td>适用专业</td><td colspan="11">应用统计学</td></tr>
<tr><td>先修课程</td><td colspan="11">概率论与数理统计（含随机过程）</td></tr>
<tr><td>建议开课学期</td><td colspan="11">☐1-1 ☐1-2 ☐2-1 ☐2-2 ☑3-1 ☐3-2
☐4-1 ☐4-2 ☐5-1 ☐5-2 ☐*（代表每学期都开设）</td></tr>
<tr><td>考核方式</td><td colspan="4">☐考试☑考查</td><td>成绩分制</td><td colspan="5">☑百分制 ☐五级制
☐二级制</td></tr>
<tr><td>使用教材</td><td colspan="11">1、周志华，《机器学习》，第 2 版，清华大学出版社，2020 年</td></tr>
<tr><td>参考书及教学资源</td><td colspan="11">1、谢文睿、秦州、贾彬彬，《机器学习公式详解》，第 2 版，人民邮电出版社，2023 年
2、奥雷利安·杰龙，《机器学习实战：基于 Scikit-Learn、Keras 和 TensorFlow》第 2 版，机械工业出版社，2020</td></tr>
<tr><td>其他说明</td><td colspan="11">无</td></tr>
</table>

二、课程简介

中文描述

本课程是应用统计学专业的选修课程。机器学习课程涵盖概率论知识，统计学、复杂算法知识等，通过本课程的学习，学生掌握一些经典而且常用的机器学习方法，例如：逻辑回归、决策树、随机森林、人工神经网络、贝叶斯学习等。机器学习是人工智能及模式识别领域的共同研究热点，其理论和方法已被广泛应用于解决工程应用和科学领域的复杂问题。

English Description

This course is an elective course for majors in Applied Statistics. The machine learning course covers knowledge of probability theory, statistics, complex algorithms, etc. Through this course, students will master some classic and commonly used machine learning methods, such as logistic regression, decision trees, random forests, artificial neural networks, Bayesian learning, etc. Machine learning is a common research hotspot in the fields of artificial intelligence and pattern recognition, and its theories and methods have been widely applied to solve complex problems in engineering applications and scientific fields.

三、课程目标

1、知识目标：

1-1：通过理论课程学习掌握假设空间、经验误差与过拟合、性能度量等机器学习的基础知识。

1-2：掌握对数几率回归、决策树、神经网络、支持向量机等经典而常用的机器学习算法。

1-3：能利用相关软件实现机器学习算法，学会用计算机求解科学技术问题。

2、能力目标：

2-1：具有演绎推理，数学演算，分析归纳、数学建模等科学研究能力。

2-2：具有良好的计算机编程能力。

2-3：具有综合运用机器学习算法解决人工智能领域实际问题的工程能力。

3、德育/素养目标：

3-1：通过大国工匠故事、创新作品展示等典型案例学习，增强家国情怀和强国意识，坚定勤学、励志、力行精神，提升创新思维和创新设计能力。

3-2：通过实验课程实际操作训练及团队创新项目作品设计，领悟创新精神内涵，提升创新创业实践和团队协作能力。

表 1 课程目标与毕业要求的对应关系

毕业要求	指标点	课程目标	关联性（H/M/L）
2、问题分析：能够借助应用统计学的基本原理、方法和手段，识别、表达、并通过文献研究分析复杂实际问题，以获得有效结论。	2.2 能够借助应用统计学的基本原理、方法和手段，针对复杂实际问题设计针对性的方案，并综合运用文献、科学理论和技术手段予以解决。。	1-1、1-2、1-3、2-1、2-2、2-3、3-1、3-2	H
4、研究：能够基于科学原理并采用科学方法对复杂实际问题进行研究，包括设计实验、分析与解释数据、并通过信息综合得到合理有效的结论。	4.1 能够基于科学原理并采用科学方法，在本专业相关理论指导下对复杂实际问题设计实验进行研究。	1-3、2-1、2-2、2-3、3-1、3-2	M
	4.2 能够结合本专业知识对实验数据进行分析与解释，设计并优化实验方案，并通过信息综合得到合理有效的结论。	1-3、2-1、2-2、2-3、3-1、3-2	M

四、主要教学内容

（一）理论教学

第 1 章 绪论、模型评估与选择（支撑课程目标 1-1、2-1、3-1）

【1】知识点：

（1）基本术语、假设空间、归纳偏好、发展历程。

（2）经验误差与过拟合、评估方法、性能度量。

（3）比较检验、偏差与方差。

【2】学习目标：

机器学习的产生与发展，掌握经验误差与过拟合、偏差与方差。

【3】重点内容：

经验误差与过拟合、偏差与方差。

【4】难点内容：

偏差与方差。

【5】思政案例：

讲解机器学习的基础知识时引入课程思政内容，包括讲解过拟合与欠拟合时引入矛盾的普遍性和特殊性、讲解偏差与方差时引入共享理念、讲解多分类学习时引入对外开放、共同发展、合作与竞争的辩证关系等。在进行“课程思政”教育时，培养学生具备科学素养，培育学生的创新能力，将科学精神传授给学生，通过学习机器学习的应用技术，让学生全身心地投入到世界科技强国建设的过程中去。

第 2 章：线性模型（支撑课程目标 1-2、1-3、2-1、2-2、3-1、3-2）

【1】知识点：

（1）线性回归、对数几率回归、线性判别分析。

（2）多分类学习、类别不平衡问题。

【2】学习目标：

理解：多分类学习、类别不平衡问题。

掌握：线性回归、对数几率回归、线性判别分析。

【3】重点内容：

对数几率回归、线性判别分析。

【4】难点内容：

对数几率回归、多分类 MvM 技术。

第 3 章：经典的机器学习方法（支撑课程目标 1-2、1-3、2-1、2-2、3-1、3-2）

【1】知识点：

（1）决策树。

（2）神经网络。

（3）支持向量机。

（4）贝叶斯分类器。

【2】学习目标：

理解：各种机器学习方法的算法原理与理论推导。

掌握：各种机器学习方法的应用及编程实现。

【3】重点内容：

（1）决策树的基本流程与剪枝处理。

（2）神经网络基本原理及卷积神经网络。

（3）支持向量机的对偶问题与核函数。

（4）贝叶斯决策论与朴素贝叶斯分类器。

【4】难点内容：

（1）卷积神经网络。

（2）对偶问题与核函数。

【5】思政案例：

讲解神经网络知识点时，介绍“深度学习三巨头”Yoshua Bengio、Yann LeCun、Geoffrey Hinton 共同获得了 2018 年的图灵奖，这是图灵奖 1966 年建立以来少有的一年颁奖给三位获奖者，结合“工匠精神”重点介绍 Geoffrey Hinton 的科研历程。融入工匠精神，让学生在掌握专业知识的同时，增强“四个意识”，坚定“四个自信”，做到“两个维护”，做好职业生涯规划，树立正确的价值观，进而推动高校立德树人工作纵深式发展。

第 4 章：进阶知识（支撑课程目标 1-2、1-3、2-1、2-2、3-1、3-2）

【1】知识点：

（1）集成学习。

（2）聚类分析。

【2】学习目标：

理解：个体与集成，聚类无监督机器学习方法。

掌握：并行与串行的集成学习方法，原型聚类与密度聚类。

【3】重点内容：

（1）Bagging 与 Boosting 集成算法。

（2）K 均值聚类、高斯混合聚类、DBSCAN 聚类、层次聚类。

【4】难点内容：

（1）AdaBoost 集成学习。

（2）高斯混合聚类。

（二）实验教学

实验教学部分以对数几率回归、决策树、人工神经网络、支持向量机、贝叶斯分类和聚类六个实验项目为主线，以学生实际操作为主、教师传授为辅的教学模式，充分利用各种软件资源，挖掘学生潜能，打好操作技能基础，提升学生的计算机力和综合用机器学习算法解决人工智能领域实际问题的工程能力，实验教学的项目名称及学时分配相关信息如表 2 所示。

表 2　实验教学信息表

序号	实验项目名称	实验学时	每组人数	实验类型	实验性质
1	对数几率回归	2	1	综合性	必做
2	决策树	2	1	综合性	必做
3	人工神经网络	2	1	综合性	必做
4	支持向量机	2	1	综合性	必做
5	贝叶斯分类	2	1	综合性	必做
6	聚类	2	1	综合性	必做

实验项目一：编程实现对数几率回归算法（支撑教学目标 1-2、1-3、2-2、2-3、3-1、3-2）

【1】实验内容：

选用 Python 编程语言实现对数几率回归算法。

【2】实验目的：

熟悉 Python 编程语言的基本操作，自选数据输入实现对数几率回归算法。

【3】教学要求：

通过实验，学生熟悉 Python 编程语言的特点，了解 Python 编程的基本过程，掌握数据文件的输入与输出，学生能编写或调用函数包编程实现对数几率回归算法，并对程序输出结果给出统计意义上的解释。

【4】实验方式：利用 Python 语言编程实现机器学习算法。

【5】实验场地与设备：机房，计算机。

实验项目二：编程实现决策树算法（支撑教学目标 1-2、1-3、2-2、2-3、3-1、3-2）

【1】实验内容：

利用 ID3、ID4.5，CART 决策树算法进行数据预测。

【2】实验目的：

利用 Python 语言编程实现决策树算法。

【3】教学要求：

通过实验，学生能编写 ID3、ID4.5，CART 决策树程序代码，并对软件的输出结果给出统计意义上的解释。

【4】实验方式：利用 Python 语言编程实现机器学习算法。

【5】实验场地与设备：机房，计算机。

【6】思政案例：

各种决策树算法各有优缺点，且互补，不同场合采用不同方法。在实际应用中本着具体问题具体分析的原则，也可将多种算法进行比较，找到解决问题的最佳途径。针对实际问题合理选择方法。理解具体问题具体分析理论的灵魂。

实验项目三：编程实现人工神经网络算法（支撑教学目标 1-2、1-3、2-2、2-3、3-1、3-2）

【1】实验内容：

利用 BP、CNN 神经网络进行数据预测。

【2】实验目的：

利用 Python 语言编程实现人工神经网络算法。

【3】教学要求：

通过实验，学生能编写 BP、CNN 神经网络程序代码，并对软件的输出结果给出统计意义上的解释。

【4】实验方式：利用 Python 语言编程实现机器学习算法。

【5】实验场地与设备：机房，计算机。

实验项目四：编程实现支持向量机算法（支撑教学目标 1-2、1-3、2-2、2-3、3-1、3-2）

【1】实验内容：

利用 SVM 进行数据预测。

【2】实验目的：

利用 Python 语言编程实现支持向量机算法。

【3】教学要求：

通过实验，学生能编写 SVM 程序代码，并对软件的输出结果给出统计意义上的解释。

【4】实验方式：利用 Python 语言编程实现机器学习算法。

【5】实验场地与设备：机房，计算机。

实验项目五：编程实现贝叶斯分类器（支撑教学目标 1-2、1-3、2-2、2-3、3-1、3-2）

【1】实验内容：

利用朴素贝叶斯分类器、贝叶斯网进行数据预测。

【2】实验目的：

利用 Python 语言编程实现贝叶斯分类器算法。

【3】教学要求：

通过实验，学生能编写贝叶斯分类器程序代码，并对软件的输出结果给出统计意义上的解释。

【4】实验方式：利用 Python 语言编程实现机器学习算法。

【5】实验场地与设备：机房，计算机。

实验项目六：编程实现聚类算法（支撑教学目标 1-2、1-3、2-2、2-3、3-1、3-2）

【1】实验内容：

利用 K-means、GMM、DBSCAN 算法进行数据分类预测。

【2】实验目的：

利用 K-means、GMM、DBSCAN 算法进行数据分类预测。

【3】教学要求：

通过实验，学生能编写 K-means、GMM、DBSCAN 算法程序代码，并对软件的输出结果给出统计意义上的解释。

【4】实验方式：利用 Python 语言编程实现机器学习算法。

【5】实验场地与设备：机房，计算机。

五、教学安排与学时分配

本课程以课堂教学为主，辅以实验教学形式。课堂教学围绕本课程的基本概念、基本理论、基本原理和实践操作进行授课。课程围绕立德树人根本任务，以培养德智体美劳全面发展社会主义建设者和接班人为总目标，以机器学习经典算法基础原理为主线，发挥理论知识传授、实验实

践操作及价值引领作用，扎实基础知识，提升实践技能和创新精神，树立正确的人生观、世界观和价值观。线下面授课堂的学时分配如表 3 所示。

表 3 学时分配表

讲　　课　　内　　容		学时				
		讲课	实验	上机	设计	课外
第 1 章	绪论、模型评估与选择	2	0	0	0	0
第 2 章	线性模型	4	2	0	0	0
第 3 章	经典的机器学习方法	20	8	0	0	0
第 4 章	进阶知识	6	2	0	0	0
小计		32	12	0	0	0
总计	44					

六、教学方法

1、讲授基本原理，理论联系实际，强调机器学习方法的设计应用，培养学生创新能力。

2、采用多媒体课件和传统教学相结合的教学方式，赋予课堂生机。

3、理论教学、讨论形式相结合，调动学生的学习积极性，培养学生的自学能力和解决实际问题的能力。

七、成绩评定

1、本门课程成绩评定方式为百分制。课程考核包括 3 个部分，分别为平时、实验和结课考核。具体要求及评分方法如下。

（1）平时成绩占 10%，主要包括课堂交流、作业、线上资料学习等综合评定。

（2）实验成绩占 30%，主要为上机实验练习、作业和测评。

（3）期末成绩占 60%，主要为结课考核。

2、所缺实验、作业等超过应完成量 1/3 者，或缺课学时超过计划学时 1/3 者，或实验考核成绩不合格者，取消该课程考核资格。

八、评价及持续改进

1、教学大纲

考核周期：4 年，修订周期：4 年。

改进措施：课程负责人组织课程团队所有教师讨论后提出改进意见，经教研室主任审核后由教学院长批准。

2、成绩评定考核

考核周期：1 学年，评价依据：学生最终成绩综合分析。

改进措施：根据学生最终成绩的分布，进行相应的内容调整和优化，并改进教学方法。

撰写：　张琨，陆杰

审核：　韩鹏

批准：　史文博

日期：　2023.07.05