

# 《大数据处理与智能决策

(Big data processing and intelligent decision making) 》

## 教学大纲

制定时间：2024 年 4 月

### 一、课程基本信息

- (一) **适用专业**：本科计算机、物联网、软件专业
- (二) **课程代码**：3DX1054A
- (三) **学分/课内学时**：2 学分/32 学时
- (四) **课程类别**：专业教育
- (五) **课程性质**：必修、选修/理论课
- (六) **先修课程**：《C 语言程序设计》、《数据可视化》、《概率论与数理统计》
- (七) **后续课程**：《深度学习》、《毕业设计（论文）》

### 二、课程教学目标

本课程是一门新兴的交叉性学科，涵盖了数据库、机器学习、统计学、模式识别、人工智能以及高性能计算等技术。开设本课程的目的，是使学生全面而深入地掌握大数据处理的基本概念和原理，掌握常用的大数据处理算法，了解大数据处理的最新发展、前沿的大数据处理研究领域、以及大数据处理技术在不同学科中的应用。为学生学习后续物联网新技术、毕业设计打下基础。

课程目标及能力要求具体如下：

#### (一) 具体目标

**目标 1**：能够设计并实现大数据平台下的大数据处理系统。了解由工程问题，到建模、再到大数据处理算法设计的问题求解思维模式。具有将大数据处理算法应用于具体工程的能力。掌握大数据预处理、关联规则、分类以及聚类技术，并能够在主流大数据平台上实现。（对应毕业要求指标点 4.2）

**目标 2**：具备较强的学习最新大数据处理领域研究成果的能力；能够分析和评价现有研究成果的问题与不足，并能够提出自己独立见解的能力。能够撰写系统设计方案和阶段性技术报告，能够组织和协调项目组的工作，与成员进行交流与沟通。能分析和理解现有案例的软、硬件设计与实现，并能对其进行变化或改进。（对应毕业要求指标点 4.3）

**目标 3**：能够针对物联网领域的复杂工程问题，分析、选择与使用恰当的技术、资源、现代工程工具和信息技术工具，实现对复杂工程问题的预测与模拟，

并能够理解其局限性，选用合适的云平台和人工智能方法，应用大数据工具实现大数据分析。（对应毕业要求指标点 4.3）

## 二) 课程目标与毕业要求的对应关系

毕业要求	毕业要求指标点	课程目标	教学单元	评价方式
4.能够基于科学原理，采用适当的工程方法对物联网领域的复杂工程问题进行研究，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。	4.2 能够正确采集和整理实验数据，对实验结果进行分析和解释，获取有效结论；	目标 1	大数据处理与智能决策概论、大数据预处理	实验
	4.3 能够运用云计算、大数据相关知识开展物联网领域的科学研究和社会服务。	目标 2	线性回归、聚类算法、分类算法	实验； 期末考试（非标试卷考核）
		目标 3	应用大数据工具实现大数据分析	实验； 期末考试（非标试卷考核）

## 三、教学内容与方法

### (一) 教学内容及要求

序号	教学单元	教学内容	学习产出要求	推荐学时	推荐教学方式	支撑课程目标	备注
1	大数据处理与智能决策概论	了解大数据处理和智能决策概念及其发展过程，在各行业中的典型应用，以及大数据处理的过程。	理解大数据处理的概念，掌握各类识别方法。了解各种大数据处理技术和发展趋势。	2	讲授	目标 1	
2	大数据处理与智能决策绪论	大数据背景、大数据处理的业务需求、机器学习概述。 重点：机器学习的背景 难点：主要的机器学习算法分类	了解大数据处理的业务需求，机器学习的起源与算法分类。	2	讲授	目标 1	作业

3	大数据预处理	<p>数据预处理的概念，为什么要进行数据预处理，评价数据质量的指标，数据清洗和数据填充的方法，数据平滑的方法。</p> <p>重点：数据填充方法，数据平滑方法</p> <p>难点：数据填充方法，数据平滑方法</p>	了解数据预处理的概念，数据预处理的方法；掌握数据填充和数据平滑的方法。	4	讲授	目标1	作业
4	机器学习算法-线性回归	<p>介绍机器学习中回归的基本概念，接着介绍回归算法中最简单的一元线性回归的概念，理解线性回归中误差函数使用平方和的原因，能使用梯度下降法求解误差函数。</p> <p>重点：误差函数概念</p> <p>难点：梯度下降法的推导</p>	理解回归分析的基本思想；会求经验回归系数，掌握一元线性回归方程的求法，会将一元非线性模型转化为一元线性模型。	4	讲授	目标2	
5	实验一：一元线性回归算法实现	<p>一元线性回归编程实现。</p> <p>重点：一元线性回归数学模型建立</p> <p>难点：一元线性回归模型建立与预测</p>	掌握使用编程软件对数据进行预处理，掌握使用一元线性回归对数据进行建模和预测。	2	操作	目标1、2	实验
6	聚类算法概论	掌握相似性和相异性度量，了解聚类算法的分类。	掌握聚类中相异性度量的概念、掌握核心的聚类算法，了解不同算法的优缺点。	2	讲授	目标1	
7	聚类算法基础理论	掌握无监督学习的概念，讲述邻居的概念，掌握欧式距离、数据特征和样本间的相似度的概念。	了解聚类算法的基础概念。	2	讲授	目标2、3	
8	聚类算法典型算法	掌握K-means算法的思想与流程，掌握K-means算法的伪代码。了解熵的概念，以及基于熵的聚类算法。掌握聚类算法的衡量指标。	了解K-means算法的原理，对其数学机理熟练掌握。	4	讲授	目标2、3	

9	实验二： k-means 算法实现	利用自带工具箱的 k-means 算法，设计一个简单的聚类算法。	了解编程软件 k-means 工具箱的使用方法。	2	讲授 实操	目标 2	实验
10	分类 算法 介绍	分类算法简介、分类与聚类的区别、分类的基本流程、常用的分类算法介绍。	了解分类的概念、特点、常用的分类算法、熟悉分类的基本流程。	2	讲授	目标 1、 2	预习
11	KNN 分类 算法 介绍	KNN 算法概述、有监督学习与无监督学习、欧氏距离和曼哈顿距离、KNN 算法原理介绍、KNN 算法的实现步骤。	了解 KNN 分类的概念、特点。熟悉欧氏距离和曼哈顿距离、掌握 KNN 算法原理介绍、KNN 算法的实现步骤。	2	讲授	目标 1、2	练习
12	实验 三、 实验 四： KNN 分类 算法 实现	KNN 算法实现的详细步骤、KNN 算法的数学模型、通过 MATLAB 编程实现 KNN 分类算法的基本实现案例、KNN 算法关键参数的详解。	掌握 KNN 算法实现的详细步骤、可以通过编程实现 KNN 分类算法案例，了解算法相关的参数调优。	4	讲授 演示 实操	目标 3	实验
总计				32		0	

## (二) 教学方法

### 1.课堂讲授

(1) 采用启发式教学，激发学生主动学习的兴趣，培养学生独立思考、分析问题和解决问题的能力，引导学生主动通过实践和自学获得自己想学到的知识。

(2) 在教学内容上，本课程的理论知识比较多，也比较细，除上课听讲以外，还可以利用视频资源学习。作业应在理解的复习理解的基础上独立完成。系统地讲授大数据概论、聚类方法与分类方法，对当前最前沿的应用场景进行剖析，为学生学习后续课程深度学习、毕业设计打下基础。

(3) 在教学过程中采用电子教案，多媒体教学与传统板书、教具教学相结合，提高课堂教学信息量，增强教学的直观性。

(4) 理论教学与工程实践相结合，引导学生应用数学、自然科学和工程科学的基本原理，采用现代设计方法和手段，进行机构分析、综合与仿真，培养其识别、表达和解决专业相关工程问题的思维方法和实践能力。

### 2.实验教学

实验教学是大数据处理与智能决策课程中重要的实践环节，目的是培养学生在掌握基本的知识、方法和技巧后，根据教师的要求和自己的设想，对现有案例进行改进和完善，增强自己的独立工作能力。课程必做实验 4 个，各实验按照实验指导书的要求学生独立或分组完成，并提交实验报告。

#### 四、考核及成绩评定

##### (一) 考核内容及成绩构成

课程考核以考核学生能力培养目标的达成为主要目的，以检查学生对各知识点的掌握程度和应用能力为重要内容，包括平时实验和期末非标准试卷两个部分。各课程目标的考核内容、成绩评定方式、目标分值建议如下：

考核方式在课程目标的比例			
课程目标	考核评定方式 (占比)		总成绩 (占比)
	实验 (A)	期末考试 (B)	
目标 1	10	20	30
目标 2	20	-	20
目标 3	10	40	50
按占比合计	40	60	100

课程目标	考核内容	成绩评定方式	成绩占总评分比例	学生当次考核平均得分	课程目标总成绩占比	目标达成情况计算公式
目标 1: 能够设计并实现大数据平台下的大数据处理系统。了解由工程问题，到建模、再到大数据处理算法设计的问题求解思维模式。具有将大数据处理算法应用于具体工程的能力。掌握大数据预处理、关联规则、分类以及聚类技术，并能够在主流大数据平台上实现。	利用自带工具箱，设计一个简单的一元线性回归算法的实现。 掌握 k-means 算法，设计一个简单的聚类算法，能够显示聚类过程以及聚类结果。 KNN 算法实现的详细步骤、KNN 算法的数学模型、通过编程实现 KNN 分类算法的基本实现案例、KNN 算法关键参数的详解。	实验	10%	A1	30	$(A1 + B1) / 30$

课程目标	考核内容	成绩评定方式	成绩占总评分比例	学生当次考核平均得分	课程目标总成绩占比	目标达成情况计算公式
	分析和评价现有大数据处理领域研究成果的问题与不足	期末考试(非标试卷考核)	10%	B1		
目标 2: 具备较强的学习最新大数据处理领域研究成果的能力; 能够分析和评价现有研究成果的问题与不足, 并能够提出自己独立见解的能力。能够撰写系统设计方案和阶段性技术报告, 能够组织和协调项目组的工作, 与成员进行交流与沟通。能分析和理解现有案例的软、硬件设计与实现, 并能对其进行变化或改进。	掌握 k-means 算法, 设计一个简单的聚类算法, 能够显示聚类过程以及聚类结果; KNN 算法实现的详细步骤、KNN 算法的数学模型、通过编程实现 KNN 分类算法的基本实现案例、KNN 算法关键参数的详解。	实验	20%	B2	20	B2/20
目标 3: 能够针对物联网领域的复杂工程问题, 分析、选择与使用恰当的技术、资源、现代工程工具和信息技术工具, 实现对复杂工程问题的预测与模拟, 并能够理解其局限性, 选用合适的云平台和人工智能方法, 应用大数据工具实现大数据分析。	综合几种算法, 设计一个数据分析系统并实现	实验	10%	A3	50	(A3 + B3) /50
	通过讨论与合作完成任务, 能够撰写调试报告、实验报告等软件工程技术文件。能够提出自己独立见解的能力。	期末考试(非标试卷考核)	40%	B3		
按占比合计			100%		100	

## (二) 平时考核成绩评定

**实验：**必做实验 4 次。支撑目标 1、2、3，共占总评分 40%。对应目标的评分标准如下：

<b>对应目标</b>	目标 1:能够设计并实现大数据平台下的大数据处理系统。了解由工程问题,到建模、再到大数据处理算法设计的问题求解思维模式。具有将大数据处理算法应用于具体工程的能力。掌握大数据预处理、关联规则、分类以及聚类技术,并能够在主流大数据平台上实现。	目标 2: 具备较强的学习最新大数据处理领域研究成果的能力;能够分析和评价现有研究成果的问题与不足,并能够提出自己独立见解的能力。能够撰写系统设计方案和阶段性技术报告,能够组织和协调项目组的工作,与成员进行交流与沟通。能分析和理解现有案例的软、硬件设计与实现,并能对其进行变化或改进。	目标 3: 能够针对物联网领域的复杂工程问题,分析、选择与使用恰当的技术、资源、现代工程工具和信息技术工具,实现对复杂工程问题的预测与模拟,并能够理解其局限性,选用合适的云平台 and 人工智能方法,应用大数据工具实现大数据分析。	
<b>考查点</b>	基本原理理解, 实验操作内容与掌握情况。	系统设计能力、分析和评价现有大数据处理领域研究成果的问题与不足。	综合几种算法, 设计一个数据分析系统并实现	
<b>总评分占比</b>	25%	50%	25%	
<b>评分标准</b>	100% 至 90%	了解基本知识, 对技术有深刻理解, 给出专业评述。	熟练分析聚类算法、分类算法的应用场景。完全掌握开发环境的使用, 能独自建立、调试、仿真、编译、运行工程项目。	熟练应用数据分析方法。完成复杂工程问题的应用分析, 并能对其进行变化或改进。
	89.9% 至 80%	了解基本知识, 对技术有较好理解, 给出较为专业评述。	较熟练分析聚类算法、分类算法的应用场景。较好地掌握开发环境的使用, 能建立、调试、仿真、编译、运行工程项目。	熟练应用数据分析方法。完成复杂工程问题的应用分析并能选择正确数据分析方案。
	79.9% 至 70%	了解基本知识, 对技术有理解, 给出评述。	能应用开发环境建立、调试、仿真、编译、运行工程项目。理解数据分析方法。	能对复杂工程问题的进行分析并选择数据分析方法进行应用。
	69.9% 至 60%	了解基础知识, 对技术部分有理解。	会使用公开数据集和识别算法。在他人帮助下建立、调试、仿真、编译、运行工程项目。	对复杂工程问题的进行分析。
	59.9% 至 0	不了解或只能浅显了解相关技术。	掌握数据分析算法原理不熟悉。对开发环境的使用不熟悉, 不能建立、调试、仿真、编译、运行工程项目。	不能对复杂工程问题的进行分析, 无法选择相应正确的数据分析方法。

## (三) 期末考核成绩评定

课程期末考试为非标准答案考试, 占总成绩的 60%, 支撑目标 1、目标 2, 根据课程目标设计相关试题。每次考试试题不同, 根据每次考试题目设计试卷评分标准。

## 五、参考学习资料

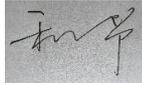
推荐教材 1:《大数据分析挖掘》,石胜飞等,人民邮电出版社,2018,第 1 版,ISBN: 9787115483058。

参考资料 1:《人工智能》,王东等,清华大学出版社,2018,第 1 版,ISBN: 9787307073104。

参考资料 2:《大数据架构详解:从数据获取到深度学习》,朱洁,罗华霖,北京:电子工业出版社,2016,第 1 版,ISBN: 9787121300004。

参考资料 3:《大数据技术原理与应用(第 2 版)》,林子雨,人民邮电出版社,2017,第 2 版,ISBN: 9787115443304。

制订人:



审核人:



制定时间: 2024 年 4 月