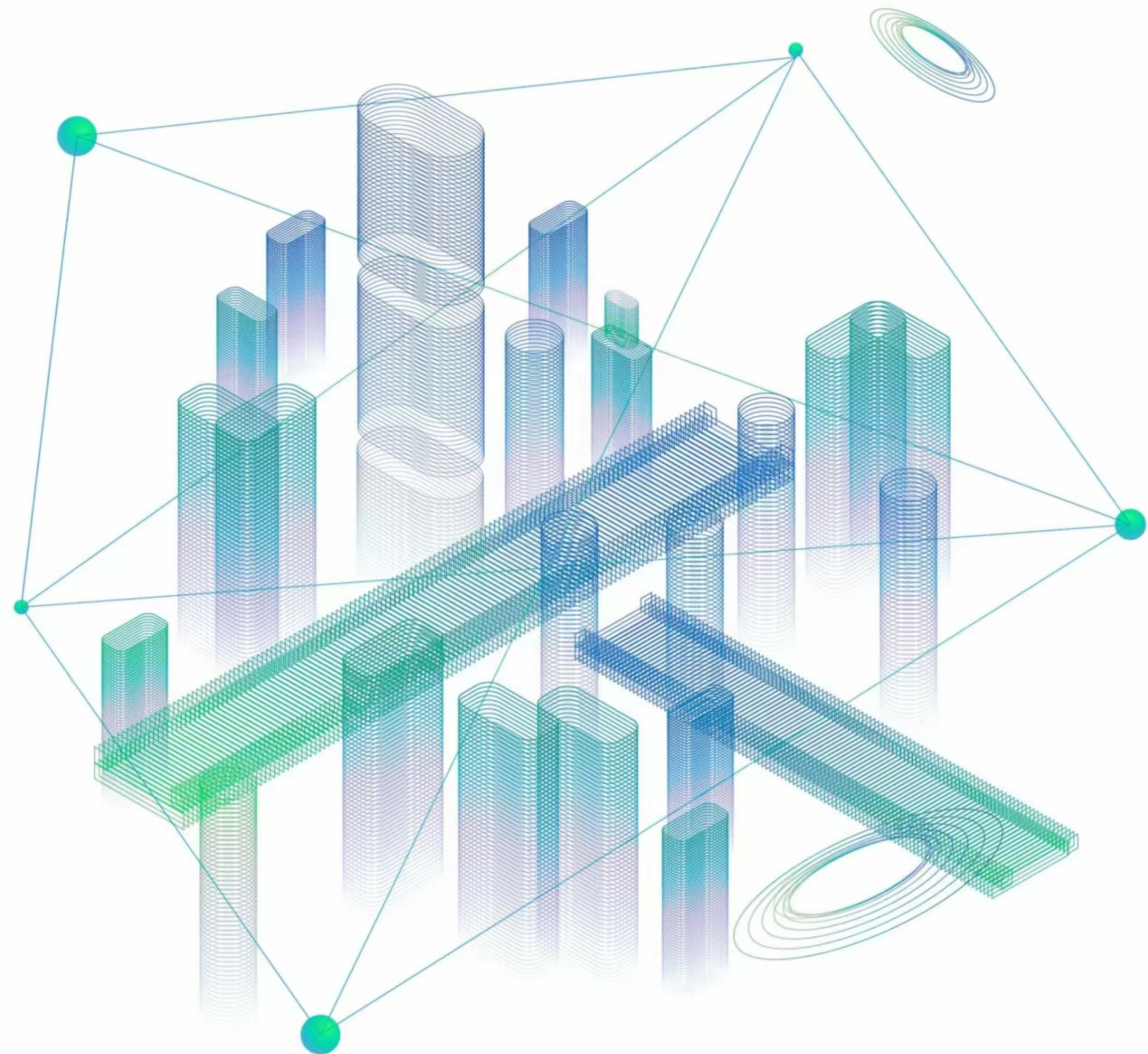




复旦大学大数据学院 本科生课程学习手册



2024

第一章 前言	2
第二章 大数据学院本科生培养模式	4
2.1 培养理念	4
2.2 大数据学院本科生培养模式	4
第三章 课程体系	5
3.1 数据科学与大数据技术专业	5
一、课程设置	5
二、课程体系与知识结构	9
3.2 人工智能专业	11
一、课程设置	11
二、课程体系与知识结构	13
3.3 卓越计划	14
第四章 主要课程简介	16
4.1 专业必修课程	16
4.2 专业选修课程	21
第五章 未来发展	26
5.1 未来深造	26
5.2 就业前景	26

第一章 前言

大数据伴随着信息技术革命应运而生，互联网、物联网、移动通讯、行业企业等数据的大量汇聚使得数据演化为重要的生产力，逐渐成为经济的新资源、发展的新引擎、信息的新矿山、科研的新依据、决策的新源泉。大数据的存取、交换、分析、应用对相关学科带来了诸多新挑战，在极大程度上改变了计算机科学、统计学和计算数学的内涵与外延：从硬件到软件、从存储到超算、从数据库到数据安全、从网络传输到并行计算、从数据分析到统计建模、从科学计算到优化方法等。

大数据学院下设数据科学与大数据技术、人工智能两个本科专业。这些专业植根于数学、统计学、计算机科学等学科，但是在研究对象、方法论、学科体系等方面又与这些学科有显著不同。新兴的交叉学科专业的内涵包含了两个层次，第一个层次是以来源多样、结构各异、规模巨大、传输高速、应用广泛的大数据为研究对象，解决大数据在获取、处理、分析、展示与应用领域的理论与实践问题，如数据挖掘、机器学习、人工智能、数据库、统计计算等领域；第二个层次则是以大数据为研究手段的数据交叉科学，如生物信息、精准医疗、电子商务、大数据金融、智能电网、智慧城市等领域，大数据分析技术为这些学科提供了新的研究范式、也在解决这些学科计算复杂性问题的过程中获得近一步的发展。由此可见，数据科学与大数据技术专业的内涵已经超出了传统学科的范畴，而是通过将统计分析、系统计算、交叉科学等有机整合，形成一套面向大数据分析全流程、大数据应用全产业链的完整知识体系，培养大数据复合型人才。我国实施创新驱动战略需要加强创新型人才的培养，要能够积极应对全球工业 4.0 时代所特有的以数据为导向的制造模式、流通模式、消费模式、商业模式的变革，大数

据人才是面向这些新机遇、新挑战、新应用的高端专业数据人才。

数据科学与大数据技术专业和人工智能是复旦大学重点建设的特色专业，该专业推动本校传统优势学科如数学、计算机科学、统计学等的融合与提升，拓展数据科学与生物学、医学、经济学、管理学、社会学等学科的交叉，培养大数据分析、管理与应用的高层次复合型人才。复旦大学大数据学院人才培养坚持理论教育与技能培养相结合、坚持基础知识体系与应用知识拓展相结合，强调基础为先、夯实学生在数据统计与分析、系统与计算方面的基础理论知识体系、使学生具备进一步开展大数据科学研究及应用创新的核心技能；同时也注重交叉融合，以大数据分析为核心轴线，以计算数学、计算机科学、统计学为三大基础支撑性学科，适度拓展与理医工学、社会科学的交叉，建设面向大数据高水平研发和产业应用的复合型人才培养体系。

第二章 大数据学院本科生培养模式

2.1 培养理念

本科人才培养目标是培养德、智、体、美全面发展，具有良好的政治素质与道德修养，掌握数据科学与大数据技术完整理论知识体系、具备全面应用分析技能，能够从事大数据有关教学、科研、开发和应用的深层次、复合型人才。

学生将掌握面向大数据应用的数学、统计学、计算机科学基础理论和方法，熟练运用各种大数据分析技术和手段；在数据建模、数据管理和分析、统计推断的基本理论、方法和技能方面进行系统学习；同时具备自然科学和社会科学等领域中大数据的应用分析技能。本专业学生在系统的专业技术训练基础上，具备广泛的数据应用视野、能够胜任大数据分析挖掘、大数据系统开发等技术领域以及大数据商务与金融、大数据生物与医药、大数据传媒与公共管理等各类应用领域的多层次工作。

2.2 大数据学院本科生培养模式

根据大数据新工科的学科特点，数据科学与大数据技术专业的本科生培养采取“2+X”的“X”型模式，即本科生培养包含三个阶段，第一阶段为通识教育+基础教育阶段，在复旦学院（书院）和相关院系完成学业；第二阶段为专业教育阶段，在大数据学院完成；第三阶段为应用拓展与产学研阶段，由大数据学院与各相关院系通力合作，共同完成。为了确保大数据人才培养的交叉学科特色。学生的来源可来自于多个大类专业；在经过专业教育之后，也可以选择不同方向的模块课程，再对接各个专业的应用问题，成为具备应用视野的专业人才。

基于前期探索，继续优化培养方案，丰富课程修读方式，强化模块课程与各学科的交叉。以数据科学与大数据技术专业招收本科生，第一年培养将主要集中在数学、统计、计算机等基础科学知识以及通识教育；确保有志于数据科学方向、来自多个大类专业同学都能够合理的修读计划下修读基础课程。在进入本科二年级前的暑假，开设若干短期课程，以衔接各专业至数据科学与大数据技术专业时的转换。在完成大二、大三上的专业基础课的基础上，在大三下、大四进行专业课的学习，开设具有交叉学科融合特色的模块课，包括统计与分析模块、系统与数据挖掘模块、理医工学大数据分析模块和社会科学大数据分析模块等。通过数据科学专业课程与模块课程相结合，为学生打下扎实的专业基础，同时培养学生跨学科学习的能力和大数据视野。

同时大数据交叉学科人才的培养，仅仅四年本科教育是不够的，特别是要使这些学生能够具备应用技能，一定要实现本科生培养与研究生的无缝衔接，打造一批高层次大数据人才，更好地服务于社会需求。为此大数据学院在学校的支持下，正在建设各层次的硕士项目，大数据统计、统计机器学习的科学硕士、大数据统计、大数据商务与管理、金融科技等专业硕士将与本科培养有效衔接，打造“2+X”的高端新工科人才培养体系。

第三章 课程体系

3.1 数据科学与大数据技术专业

一、课程设置

（一）专业教育课程

专业必修课

数据结构	统计(机器)学习概论
高等线性代数	统计计算
概率论基础	人工智能
计算机原理	神经网络与深度学习
数值算法与案例分析 I	图像处理与可视化
统计学基础：方法、原理及 R 应用(I)	生产实习
数据库及实现	毕业论文
最优化方法	

(二) 专业选修课

I 大数据技术与应用方向专业进阶模块

1、 统计与分析模块：

数值算法与案例分析 II	线性规划
随机过程导论	多元统计分析
统计学基础 II：回归分析	计算方法
时间序列与空间统计	应用泛函分析
多模态数据同化	人工智能中的稀疏理论与应用
数学模型	因果推断与因果学习
随机分析	

2、 系统与数据挖掘模块：

大规模分布式系统	计算机视觉
----------	-------

高级大数据解析	图数据管理与挖掘
自然语言处理	强化学习算法与理论基础
计算理论	认知智能前沿技术与实践
数字图像处理	算法设计与分析

3、 理医工学大数据分析模块：

医疗大数据统计学	从生物学和统计学视角看人类疾病
医学图像分析	卫生统计学 A
生物统计学	心理统计学（一）
组学数据的统计分析和挖掘	心理统计学（二）
生物医学工程学基础	

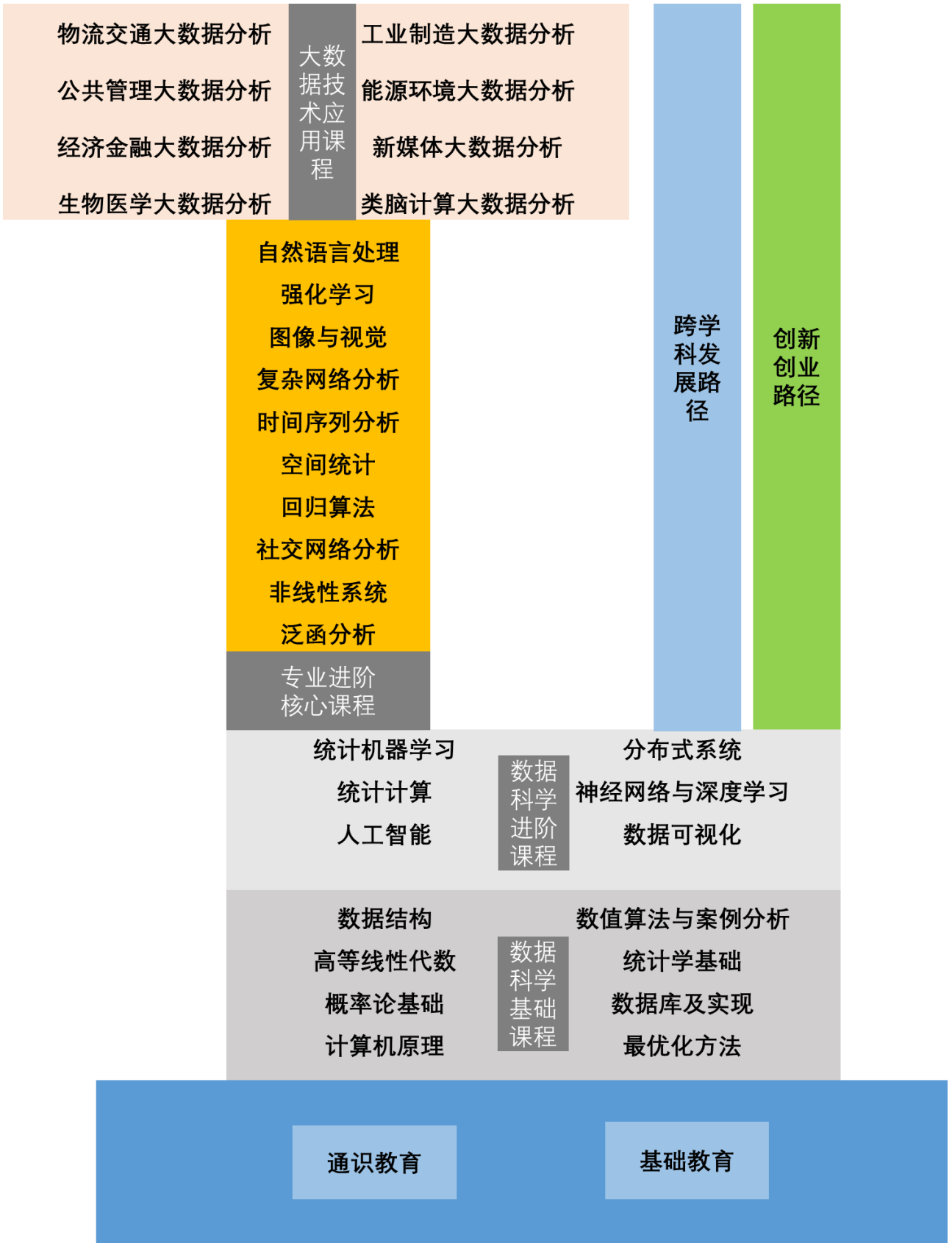
4、 社会科学大数据分析模块：

社交网络挖掘	社会数据管理与分析
金融计量学	决策理论
商务分析	金融风险管埋
大数据传播与新媒体分析	社会科学方法论
金融工程	

II 类脑计算方向专业进阶模块

非线性系统	泛函分析
医疗大数据统计学	生物医学工程学基础
大规模分布式系统	多模态数据同化
随机过程导论	生物数学

二、课程体系与知识结构



学期	大类基础课+专业核心课				X路径
1	<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">数学分析BI</div> <div style="border: 1px solid black; border-radius: 10px; padding: 5px;">线性代数</div>	<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">大学物理B（上）</div> <div style="border: 1px solid black; border-radius: 10px; padding: 5px;">大学物理B（下）</div>	<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">程序设计</div>		专业进阶模块I
2	<div style="border: 1px solid black; border-radius: 10px; padding: 5px;">数学分析BII</div>				专业进阶模块II
3	<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">数值算法与案例分析I</div> <div style="border: 1px solid black; border-radius: 10px; padding: 5px;">高等线性代数</div>	<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">概率论基础</div>	<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">数据结构</div> <div style="border: 1px solid black; border-radius: 10px; padding: 5px;">计算机原理</div>		荣誉项目路径
4	<div style="border: 1px solid black; border-radius: 10px; padding: 5px;">最优化方法</div>	<div style="border: 1px solid black; border-radius: 10px; padding: 5px;">统计学基础：原理方法及R应用（I）</div>	<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">数据库及实现</div>		跨学科发展路径
5	<div style="border: 1px solid black; border-radius: 10px; padding: 5px;">统计（机器）学习概论</div>	<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">统计计算</div>	<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">人工智能</div> <div style="border: 1px solid black; border-radius: 10px; padding: 5px;">图像处理与可视化</div>		辅修学士学位路径
6			<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">神经网络与深度学习</div>		创新创业路径
7				<div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin-bottom: 5px;">生产实习</div>	
8				<div style="border: 1px solid black; border-radius: 10px; padding: 5px;">毕业论文</div>	

3.2 人工智能专业

一、课程设置

(一) 专业教育课程

专业必修课

人工智能的编程基础	数据库引论
信息论、控制论、系统论基础	人工智能的伦理与治理
概率论与数理统计	模式识别与机器学习
数据结构	操作系统
集合与图论	并行分布式计算
人工智能导论	本科生科研实践
最优化方法	毕业论文
计算机组成与体系结构	

(二) 专业选修课

专业进阶模块 I

	课程名称		课程名称
大数据方向 (认知计算 与算法智 能)	高等微积分	计算机方向 (数据计算 与智能)	人工智能安全
	常微分方程		自然语言处理和大语言模型
	随机过程导论		计算机视觉

	应用随机过程		大数据分析技术
	数值算法与案例分析 I		计算机图形学 A
	统计学基础 II: 回归分析		
	多模态数据同化		
	计算系统生物学		
	图像处理与可视化		
	生物统计学		

专业进阶模块 II

智能算法与 计算	脑科学与类脑系统	计算机方向 (数据计算 与智能)	代数结构与数理逻辑
	神经网络与深度学习		社会计算导论
	计算机视觉		计算理论基础
	认知智能前沿技术与实践		类脑智能与人工通用智能
	应用泛函分析		计算机网络
	复杂系统理论与仿真		编译
	限定专业选修课程 (不高于 3 学分 (从另一个模块方向的专业进阶 I 或者 II 课程中选择))		软件工程
		数字图象处理	
		虚拟现实技术导引	
		机器学习系统	
		人机交互	
		数据挖掘技术	
		社交网络挖掘	

			信息检索导论
			不高于 3 学分（从另一个模块方向的专业进阶 I 或者 II 课程中选择）

二、课程体系与知识结构

学期	大类基础课+专业核心课				X路径
1	数学分析BI 线性代数	大学物理B（上） 大学物理B（下）	程序设计		专业进阶模块I
2	数学分析BII	基础物理实验 信息论、控制论、 系统论基础	电路基础 人工智能的编程 基础		专业进阶模块II
3	概率论与数理统计	集合图论 数据结构			跨学科发展路径
4	最优化方法	人工智能导论	计算机组成与体系 结构 数据库引论	人工智能的伦理与 治理	辅修学士学位 路径
5		模式识别与机器学习 操作系统			创新创业路径
6					
7		并行分布式计算		本科生科研实践	
8				毕业论文	

3.3 卓越计划

3.2.1 专业导师制

1、培养周期

学院所有本科生从进入学院起配备专业导师，直至毕本科四年级结束。实现学院本科生专业导师制的全覆盖。

2、分配原则

根据师生比例进行平均分配，分配方法遵循随机原则。

3、导师工作职责

导师对其学术、科研、日常学习生活等给予及时指导关心。主要教育活动方式有：学生单独面谈、seminar 讨论、导师午餐会、学术科研项目指导、学术讲座、咨询导航、学术论坛、班会互动等师生对话交流，充分发挥导师人格魅力和学术精神在全人教育理念中的作用。具体要求如下：

- 1) 定期与学生进行学术交流，每月与每位学生单独见面不少于 1 次；
- 2) 定期组织所有指导学生进行 seminar 讨论，每季度 seminar 次数不少于 1 次；
- 3) 定期参加学院组织的导师午餐会，为学生与导师相互交流提供平台和场所，导师午餐会每学期不少于 1 次；
- 4) 为学生提供科创项目、社会实践项目、挑战杯项目、学术论文投稿等相关科研指导；
- 5) 组织并参加相关学术讲座、论坛等学术活动；
- 6) 参加日常班级班会，与学生全方位互动；
- 7) 每学期规定 office hour，专门用于接待学生，指导日常学术、科研和思想生活。

3.2.2 第二课堂学习

学院强调第二课堂对第一课堂的支持支撑作用，鼓励和支持学生各类学术科研项目，并形成各级各类学术竞赛和实践创新项目的孵化激励体系。

按照不同级别、等第的获奖或发表情况，给予学生不同额度的奖励，鼓励学生积极参与学术生产，投身创新创业培育，孵化一批有质量的学术成果。

3.2.3 学术交流活动

学生自费参加国内外的学术会议，可申请学院资助。申请人应得到主办方正式邀请，提交研究论文，并在活动中公开宣读论文。申请人应为论文第一作者；如与学院教师合作，应由合作教师证明其为论文主要贡献人。论文应为学术性研究论文，长度不低于 4000 字（英文）或 8000 字（中文或日文）。

3.2. 社会实践、国内外挂职实习

学院设立学生境外交流学习资助项目，鼓励学生根据校际、院际协议，申请赴境外高校、研究机构或组织，参加一学期（4 或 6 个月）及以上的交流学习。

4、荣誉项目

学院自 2019 级开设本科生荣誉项目，具体要求参见《大数据学院本科“荣誉项目”实施方案（试行）》

第四章 主要课程简介

(具体课程设置详见培养方案)

4.1 专业必修课程

一、课程基本情况		
课程代码	DATA130005	
课程名称	统计学基础：原理、方法及 R 应用 (I)	
	Statistics: Principles, Methods and R (I)	
课程类别	专业必修	
开课时间	春季	
学分	3	
总学时	54	
先修课程		
教学方式	课堂讲授	
课时分配	18*3	
考核方式	随堂测试+期中考试+期末考试	
教材及参考资料	1. Wasserman, L. (2004) All of Statistics. Springer (Chapters 1-10) 2. Knight, K. (2000). Mathematical Statistics. Chapman & Hall/CRC	
二、教学目的和基本要求		
本课程覆盖概率和统计方法及原理的基础知识。统计软件 R 贯穿于课程教学之中，通过对概率论，统计软件，函数等的学习以提高学生在模拟和数据分析方面的动手能力。		
三、课程大纲和知识点		
章节顺序	章节名称	知识点
1	R 语言及描述统计学简介	R 语言入门
2	概率及组合	概率定义
3	条件概率，预期数额，贝叶斯公式	概率论基础知识原理
4	伯努利和二项式分布，几何分布	随机变量及分布入门
5	连续随机变量，正态分布，卡方分布	随机变量及分布分析
6	概率图，多元分布，独立性（或相关性）	概率独立性，相关性
7	矩母函数	矩母函数知识原理
8	概率不等式	概率论进阶
9	相互关系	知识点相互关系
10	样本分布，点估计值	点分布
11	极大似然估计法	极大似然估计法原理
12	期中	期中测验
13	弱大数法则 (WLLN)	大数定律
14	强大数法则 (SLLN)	大数定律
15	中心极限定理 (CLT)	中心极限定理原理
16	费雪信息量	费雪信息量原理

一、课程基本情况		
课程代码	DATA130002	
课程名称	数值算法与案例算法 (I)	
	Numerical Algorithms with case studies I	
课程类别	专业必修	
开课时间	春季	
学分	3	
总学时	54	
先修课程		
教学方式	授课	
课时分配	18*3	
考核方式	随堂测试+期中考试+期末考试	
教材及参考资料	1. D.P. O'Leary, Scientific Computing with case studies, SIAM, 2009 2. C. Moler, Numerical Computing with MATLAB (revised reprint), SIAM, 2008	
二、教学目的和基本要求		
通过本课程学习,使学生了解数值算法基础知识,掌握数值分析技巧,判断数值算法的稳定性和复杂度,并能与数值案例相结合。学会误差分析、稠密矩阵分解、基本优化方法和 Monte Carlo 算法,熟练使用 MATLAB 语言作数值模拟。		
三、课程大纲和知识点		
章节顺序	章节名称	知识点
1	贸易简介与工具	贸易入门
2	贸易误差与算法工具	浮点计算误差分析
3	调节与稳定性	基本原理知识点讲解
4	计算机存储与计算机编程	计算机程序
5	稠密矩阵计算	计算机程序与算法
6	LU, Cholesky, QR, 特征值和奇异值分解	稠密矩阵的 LU、QR 和 SVD 分解
7	案例分析	低阶近似,主成分分析法等的运用
8	优化和数据拟合	优化与数据拟合入门
9	线性搜索,二次修正模型法,信赖域,共轭梯度	数据拟合和优化的方法
10	线性等式约束的可行方向算法,优化浮点的内点法	无约束优化和约束优化
11	案例分析	拟合指数,数据分析
12	蒙特卡罗算法	算法介绍
13	随机与伪随机数,基本统计	随机数
14	积分,最小化,线性代数	算法统计基本知识
15	案例分析	马尔可夫链,感染模型

一、课程基本情况		
课程代码	DATA130004	
课程名称	统计计算	
	Computational Statistics	
课程类别	专业必修	
开课时间	秋季	
学分	3	
总学时	54	
先修课程	概率论，统计学基础	
教学方式	课堂讲授	
课时分配	18*3	
考核方式	平时成绩+期中期末考试	
教材及参考资料	1. Maria Rizzo (2007). Statistical Computing with R, CRC 2. Geof Givens and Jennifer Hoeting (2012). Computational Statistics, Wiley 3. Norman Matloff (2011). The Art of R Programming: A Tour of Statistical Software Design, No Starch Press	
二、教学目的和基本要求		
面向数学、统计以及计算机专业高年级本科生，介绍统计编程语言 R 基础，以及现代统计计算方法。学生将通过实际编程操作，加深对统计理论基础的 understanding，进而具备在分析实际数据时，熟练、正确地使用统计方法的能力。		
三、课程大纲和知识点		
章节顺序	章节名称	知识点
1	R 语言基础	入门，数据类型，函数，矩阵运算，基本绘图，R package
2	概率论基础	连续随机变量，随机向量，多元分布
3	统计理论基础	古典频率学派，贝叶斯统计简介
4	随机数、随机变量生成方法	逆变换法，接收-拒绝法，特殊变换法
5	数据可视化	二维、三维绘图方法，梯度图
6	蒙特卡洛积分	数值积分近似，
7	方差缩减	对偶变量法，控制变量，重要性抽样，分层抽样，分层重要性抽样
8	参数估计中的蒙特卡洛方法	均值，方差，平均误差，置信区间
9	假设检验中的蒙特卡洛方法	均值、多样本假设检验
10	Bootstrap 方法	理论基础，应用：参数估计，置信区间构造
11	Jackknife 方法	理论基础，应用：参数估计，置信区间构造，与 Bootstrap 方法比较，交叉验证
12	MCMC 方法 I	Metropolis-Hastings 算法
13	MCMC 方法 II	Gibbs sampler, 收敛性判断
14	EM 算法	理论及高斯混合模型的应用
15	置换检验	非参数检验简介，独立性检验的应用

一、课程基本情况		
课程代码	DATA130003	
课程名称	统计（机器）学习概论	
	Introduction to Statistical Learning and Machine Learning	
开课时间	秋季	
学分	3	
总学时	54	
先修课程	统计学基础：原理、方法及R应用（I）	
教学方式	课堂讲授、上机操作	
课时分配	18*3	
考核方式	课程实验+平时作业+项目设计	
教材及参考资料	James, Witten, Hastie and Tibshirani (2013) An Introduction to Statistical Learning, with applications in R. Springer; Hastie, T., Tibshurani, R. and Friedman, J. (2011) The Elements of Statistical Learning, data mining, inference and Prediction, 2nd Edition. Springer.	
二、教学目的和基本要求		
这是统计学习与机器学习的入门课程，课程的重点在于培养学生以统计与机器学习的思想和方法处理并解决实际大数据问题的能力。本课程设计目的是让学生全面了解当前统计机器学习研究用到的方法论、技巧、数学和算法。		
三、课程大纲和知识点		
章节顺序	章节名称	知识点
1	课程概览、实用入门	统计机器学习概览；基本概念；高数回顾
2	线性回归（1）：	最小二乘法；偏差-方差的权衡
3	线性回归（2）：	子集选择法；收缩方法（岭回归、Lasso，偏最小二乘法等）；贝叶斯线性回归：最大似然估计、最大后验估计
4	线性分类：	线性判别分析；逻辑回归；K近邻方法
5	支持向量机与核方法（1）	核平滑；核密度估计；最大间隔分类方法
6	支持向量机与核方法（2）	核技巧；对偶表示；径向基函数和内核
7	计算学习理论：	过拟合与弱可学习性；VC维；PAC学习理论基础
8	正规化和基本的神经网络：	正规化；损失函数反向传播误差
9	无监督学习与降维：	主成分分析；K聚类方法；高斯混合模型
10	决策树方法：	决策树基本概念；随机数、随机森林等
11	有向图模型简介	隐含马尔科夫模型、贝叶斯网络；
12	半监督学习方法	流形学习（LLE，ISOMAP等）；统计排序算法；直推式学习
13	近似推断（1）：	变分法
14	近似推断（2）：	吉布斯采样

一、课程基本情况		
课程代码	DATA130008	
课程名称	人工智能	
	Introduction to Artificial Intelligence	
课程类别	专业必修	
开课时间	春季	
学分	3	
总学时	54	
先修课程	数据结构, 程序设计	
教学方式	课堂讲授	
课时分配	18*3	
考核方式	平时作业+项目设计	
教材及参考资料	Stuart J. Russell, Peter Norvig (2009) Artificial Intelligence A Modern Approach, 3rd Edition. 2009, Prentice Hall	
二、教学目的和基本要求		
<p>本课程要求学生具备使用 Python 进行简单程序编写的能力, 或者短时间掌握 Python 使用的能力。通过学习本课程, 学生会了解人工智能领域的一些核心问题和应用, 并且掌握相关的原理和算法。另外, 学生会掌握编写, 维护和测试 Python 语言的能力, 并且能够使用 Python 解决现实的问题。</p>		
三、课程大纲和知识点		
章节顺序	章节名称	知识点
1	人工智能简介	背景介绍
2	搜索问题 - I	无信息搜索; 有信息搜索
3	搜索问题 - II	动态规划; 局部搜索
4	马尔可夫决策过程 - I	马尔可夫决策过程简介
5	马尔可夫决策过程 - II	概率, 马尔可夫模型
6	马尔可夫决策过程 - III	隐马尔可夫模型及其应用
7	增强学习	增强学习方法
8	对抗搜索	游戏树搜索
9	博弈论	博弈论简介
10	限制满足问题	限制满足问题介绍
11	贝叶斯网络 - I	贝叶斯网络的表示和独立性
12	贝叶斯网络 - II	贝叶斯网络推理和采样
13	贝叶斯网络 - III	决策图
14	逻辑表示 - I	谓词逻辑; 一阶逻辑
15	逻辑表示 - II	高阶逻辑; 马尔可夫逻辑; 语义分析
16	课程回顾和总结	课程回顾

4.2 专业选修课程

一、课程基本情况		
课程代码	DATA130006.01	
课程名称	文本数据管理与分析	
	Text Data Management and Analysis	
课程类别	专业选修	
开课时间	秋季	
学分	3	
总学时	54	
先修课程		
教学方式	讲授	
课时分配	18*3	
考核方式	课程实践+实践报告	
教材及参考资料	https://textprocessing.github.io/	
二、教学目的和基本要求		
<p>本课程讲授文本数据管理与分析的主要方法，包括文本相似度计算、检索、语言模型基础技术以及机器学习、自然语言处理、文本挖掘等相关知识的系统性介绍，并且通过课程项目使得学生掌握文本数据管理和分析的主要方法。</p> <p>基本要求：熟悉 Java 或 Python 等编程语言，了解概率论基础知识</p>		
三、课程大纲和知识点		
章节顺序	章节名称	知识点
1	概述	
2	文本处理基础	正则表达式
3	文本表示与相似度	词袋模型等
4	文本表示与相似度	PCA、LDA
5	文本表示与相似度	概率模型
6	统计语言模型	n-gram
7	文本存储与检索	倒排表、检索模型
8	基于机器学习的文本处理技术	基础模型
9	基于机器学习的文本处理技术	基础模型
10	基于机器学习的文本处理技术	文本分类、聚类
11	基于机器学习的文本处理技术	序列标注
12	基于深度学习的文本处理技术	基础模型 CNN/RNN
13	基于深度学习的文本处理技术	词嵌入
14	基于深度学习的文本处理技术	语言模型
15	基于深度学习的文本处理技术	语言表示
16	基于深度学习的文本处理技术	应用

一、课程基本情况		
课程代码	DATA130007	
课程名称	社交网络挖掘	
	Social Network Mining	
课程类别	专业选修	
开课时间	春季	
学分	3	
总学时	54	
先修课程	离散数学、线性代数	
教学方式	课堂教学+上机实践	
课时分配	18*3	
考核方式	笔试+完成项目设计	
教材及参考资料	《社会网络挖掘》	
二、教学目的和基本要求		
<p>本课程的教学目标是让学生学习并掌握社交网络分析与挖掘中所涉及的与社会学、经济学、传播学、统计学、计算机科学等领域有关的原理和方法，了解现今流行的社交网络数据采集和分析技术，并能够运用所学知识和技能开展相关实践项目的应用，让学生掌握社会管理与大数据挖掘等综合技能与素质。</p>		
三、课程大纲和知识点		
章节顺序	章节名称	知识点
1	概论	社交网络概念、历史、研究目标、方法
2	图论基础	欧拉回路、连通分支、网络最大流算法、最小生成树算法
3	网络度量	中心性度量、传递性、结构平衡
4	网络模型	随机图模型、小世界模型
5	社区挖掘	基于成员/群组的社区发现
6	社交网络信息传播	羊群效应、信息级联模型、传染病模型
7	社会影响与同质性	同质性度量
8	数据挖掘概论	监督/非监督学习、分类、聚类
9	社交网络数据采集	爬虫程序设计
10	用户行为分析	用户相似度分析
11	社交媒体推荐	冷启动、协同过滤
12	社交网络挖掘应用案例	

一、课程基本情况		
课程代码	DATA130010	
课程名称	数值算法与案例分析 II	
	Numerical Algorithms with case studies II	
课程类别	专业选修	
开课时间	秋季	
学分	3	
总学时	54	
先修课程	数值算法与案例分析 I	
教学方式	课堂讲授	
课时分配	18*3	
考核方式	随堂测试+期中考试+期末考试	
教材及参考资料	1. D.P. O'Leary, Scientific Computing with case studies, SIAM, 2009 2. C. Moler, Numerical Computing with MATLAB (revised reprint), SIAM, 2008	
二、教学目的和基本要求		
<p>通过本课程学习,使学生了解数值算法基础知识,掌握数值分析技巧,判断数值算法的稳定性和复杂度,并能与数值案例相结合。学会稀疏矩阵计算、约束优化、非线性方程组和随机算法,熟练使用 MATLAB 语言作数值模拟。</p>		
三、课程大纲和知识点		
章节顺序	章节名称	知识点
1	稀疏矩阵计算	稀疏矩阵计算入门
2	稀疏线性方程组的直接算法和迭代法	稀疏矩阵计算进阶
3	案例分析	快速求解, Sylvester 方程, 特征值的应用
4	约束优化, 非线性方程和连续法	约束优化
5	KKT 条件, 非线性最小平方估计法	非线性方程组
6	案例分析	盲解卷算法的应用
7	随机算法	随机算法原理
8	矩阵草图	数据模拟
9	随机梯度下降法	随机梯度下降法原理
10	方差缩减	方差缩减原理
11	案例分析	矩阵乘法, ISVD 的应用

一、课程基本情况		
课程代码	DATA130009.01	
课程名称	统计学基础：原理、方法及R应用(II)	
	Statistics: Principles, Methods and R (II)	
课程类别	专业选修	
开课时间	春季	
学分	3	
总学时	54	
先修课程	统计学基础：原理、方法及R应用(I)	
教学方式	课堂讲授	
课时分配	18*3	
考核方式	随堂测试+期中考试+期末考试	
教材及参考资料	All of Statistics	
二、教学目的和基本要求		
本课程覆盖概率和统计方法及原理的基础知识。统计软件R贯穿于课程教学之中，以提高学生在模拟和数据分析方面的动手能力。		
三、课程大纲和知识点		
章节顺序	章节名称	知识点
1	经验分布	经验分布函数及性质、统计泛函估计
2	Bootstrap方法	Bootstrap方法、方差估计、置信区间
3	EM算法及混合模型	EM算法、指数家族定义、混合模型
4	统计稳健性	错误指定模型下的统计推断
5	贝叶斯推断	贝叶斯方法、性质及先验分布选取
6	蒙特卡罗方法	蒙特卡罗方法、积分
7	MCMC算法	重要抽样、Metropolis-Hastings算法、蒙特卡洛仿真
8	线性回归	线性回归模型、最小二乘估计及其性质
9	多变量回归	线性回归模型选择、多变量回归
10	logistic回归	logistic回归、随机向量、协方差估计
11	多元正态分布	多元正态分布、独立性推断
12	因果推断	因果推断、混淆变量、辛普森悖论
13	图模型	DAG图模型、图模型估计
14	分类法	贝叶斯分类器、高斯分类器、logistic回归分类、支持向量机
15	非参数曲线估计	核密度估计、非参数回归
16	随机过程简介	马尔科夫链、泊松过程

一、课程基本情况		
课程代码	DATA130014	
课程名称	高级大数据解析	
	Advanced Big Data Analytics	
开课时间	春季	
学分	3	
总学时	54	
先修课程	统计机器学习	
教学方式	课堂讲授、上机实验授课相结合	
课时分配	18*3	
考核方式	课程实验及平时作业、期末考试	
教材及参考资料	Big Data Analytics, From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph, D. Loshin, 2013, Press: Morgan Kaufmann; [2] Learning Spark Lightning-Fast Big Data Analysis, Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, 2015, Press: O'Reilly Media Hastie, T., Tibshurani, R. and Friedman, J. (2011) The Elements of Statistical Learning, data mining, inference and Prediction, 2nd Edition. Springer.	
二、教学目的和基本要求		
大数据分析这门课会首先介绍大数据整体的应用，市场趋势，以及需要学习的知识。同时介绍基础平台和其他工具，以及一些数据储存、上传，分离和处理这些数据的方法。课程还会介绍如何在不同平台上运用不同算法。		
三、课程大纲和知识点		
章节顺序	章节名称	知识点
1	课程介绍	课程概览
2	大数据分析平台简介	Hadoop 平台简介；linux 平台
3	大数据存储与分析技术	Spark 与数据分析方法；与 Hadoop 平台比较
4	实验：介绍大数据分析基本平台	熟悉 spark, linux 等
5	大数据分析的机器学习算法 (I)	大数据分析的聚类算法；大数据分析分类算法
6	大数据分析的机器学习算法 (II)	逻辑回归；支持向量机；随机森林
7	在大数据平台上实现机器学习算法	实现基本的机器学习算法
8	Hadoop and Spark 的机器学习库	Hadoop 和 Spark 流数据分析
9	实验：高级大数据机器学习算法	大数据平台上的高级机器学习算法
10	期中回顾	期中回顾与考试
11	社会与认知分析 (I)	众筹的概念；图上的情感分析
12	社会与认知分析(II)	网络分析；推荐系统概览
13	实验：大数据的数据挖掘、情感分析	大数据的数据挖掘、情感分析案例
14	期末项目提案	小组讨论期末项目
15	GPU 和 CUDA 编程简介	数据并行化处理
16	实验：熟悉 GPU 与 CUDA	实验：熟悉 GPU 与 CUDA

第五章 未来发展

5.1 未来深造

一、本校直研要求

（一）报名条件

凡思想品德表现良好，学术研究兴趣浓厚，学业成绩优秀，具有较强的创新意识、创新能力和专业能力，达到当年学校要求的，均可报名申请本校推免生推荐资格。

二、出国深造

在海外，许多一流大学已经开设了与大数据相关的课程。诸如哈佛大学、斯坦福大学、剑桥大学、牛津大学等知名高校。不少高校对数据科学（Data Science）和商业分析（Business Analytics）等专业相当重视，例如美国哥伦比亚大学建立了独立的数据科学研究院，并为该专业学生提供了包括媒体、金融、信息技术、医疗卫生、公共部门等多行业的数据分析岗位实习资源；而商业分析专业更是近年来受到了学术界、业界和政府的高度重视——新加坡、澳大利亚政府甚至直接拨款赞助开设大数据分析的相关专业。

相较于传统的学科，这些以数据分析为核心的新兴专业都有一个共同的特点——结合了数学、计算机和商业、医学等诸多学科，而这种学科的交叉性将更加适用于当下的社会发展。

5.2 就业前景

大数据学院坚持“国家意识、人文情怀、科学精神、专业能力、国际视野”的育人理念，为国家经济和社会发展培养输送创新拔尖人才。

学院立足党、团、班架构，全面开展思政教育和学生自治。形式多样的主题教育教育活动与党建、团学工作相结合，重视理想信念教育、学风学养教育、心理健康教育和生涯发展教育。以“学生学术科创”为重心，建设学习型集体，促进第二课堂建设和师生融合，推进全员育人。实行本科导师制，推出“导师下午茶”；开展“悦读”活动，鼓励学生读经典、看前沿、重沉淀；开展高端学术会议志愿服务，带领学生走近学术前沿；在徐汇区康健街道、杨浦区定海街道设立社区实践基地，探索深入社区的学术科创实践。

大数据学院本科毕业生以升学为主，多名学生进入哥伦比亚大学、东京大学、卡内基梅隆大学等著名学府继续深造。研究生毕业生在就业市场深受用人单位欢迎，在阿里巴巴、腾讯、百度等一线互联网企业，华为等通讯企业研发部门，以及基金、证券和银行、保险等知名金融单位，从事算法、人工智能、大数据研究等工作。