“数据科学与工程”（iMDS）

选课建议

第一学年-秋冬学期

- ◆选择两门专业学位课《数据科学基础》和《决策与优化》
- ◆在专业选修课《数据治理技术导论》、《数据可视化》、《深度学习》、《数据建模案例分析》中选择1-2门。
或根据个人研究方向及兴趣选择其他课程1门。
- ◆在平台课程中选择3门
备注：选课时请注意每门课程所要求的先修课程
- ◆本学期建议修学分数：13-15
备注：专业学位6+专业选修4+平台5
或专业学位6+专业选修2+平台5

第一学年-春夏学期

- ◆选择两门专业学位课《现代统计方法》和《因果推断与机器学习》
- ◆在平台课程中选择2门
- ◆可根据个人研究方向及兴趣选择其他课程1门
备注：选课时请注意每门课程所要求的先修课程
- ◆第一学年结束时应该基本完成公共课程要求
- ◆本学期建议修学分数：9-10
备注：专业6+平台3-4

第二学年-秋冬学期

- ◆可选择 1 门专业选修课《数据治理技术导论》/《数据可视化》/《深度学习》/《数据建模案例分析》或根据个人研究方向及兴趣选择其他课程 1 门
- ◆在平台课程中选择 2 门
- ◆可根据个人研究方向及兴趣选择其他课程 1 门

备注:

- ◆本学期建议修学分数: 3-6
- ◆本学期结束时原则上应该完成所有课程要求

专业课程简介

1. **《数据科学基础》**: 为研究生介绍数据科学的基础知识和数据分析的基本方法。课程注重理论和实际应用相结合,旨在使学生掌握数据分析方法的基本思想和常见套路。课程内容主要有三个方面。(1) 概率论基础。介绍概率论基本概念,包括随机变量、分布、期望等,这些概念将有助于理解数据中不确定性及其对管理者、决策者面临复杂权衡环境的影响。课程还介绍与不确定性相关的关键概念,如方差、相关性和协方差。(2) 统计学基础与回归分析。介绍统计学中抽样及其分布、估计与显著性检验的基本方法及其原理,介绍进行基本统计分析和他人分析结果所需的技能。特别是,课程将介绍如何表达统计估计和检验的不确定性,根据有限的信息进行统计推断,并建立和理解回归模型。(3) 高级分

析。课程介绍更先进的现代数据量化分析方法，包括：非线性的预测分析方法，如神经网络；分类算法，如逻辑回归和树；聚类方法，以及金融和优化问题的模拟建模。这些方法已成为数据科学的重要组成部分。

2. **《决策与优化》**: 介绍管理与决策问题中的初等优化技术。课程着重以实例入手建立数学模型，探讨一些生产、管理和生活中比较实用的初等模型和方法。课程内容包括线性规划、对偶理论、灵敏度分析、运输问题、目标规划、整数规划、动态规划等。

3. **《现代统计方法》**: 介绍现代统计学研究中的重要方法。课程内容包括 bootstrap 方法和高维 multiple testing 等主流的现代统计推断方法，以及模型选择、变量选择，以及充分降维等高维统计方法。课程注重理论研究和实际应用相结合，旨在使同学们掌握现代统计方法的基本思想和常见思路。课程将同时使用教材、经典讲义、前沿学术论文，使同学们在巩固理论基础的同时，学习理解统计学的最新研究成果。

4. **《因果推断与机器学习》**: 本课程旨在数据科学层面上介绍因果学习的全貌。课程内容包括因果推断基础、双稳健估计、异质治疗效应、纵向数据的因果学习、工具变量、Proximal 因果学习、中介分析等。课程注重讲授因果推断的思维方式，介绍科学研究的思想和方法。

5. **《数据治理技术导论》**: 针对数据科学应用背景，介绍数据治理的概念、关键技术和优化方法。具体内容主要数据融

合与清洗、数据服务定价、时空数据/图数据/缺失数据管理、数据查询结果可用性分析、DB 和 AI 的融合。在学习的过程中，本课程同时讲解国内数据治理相关技术和产业的发展状况，运用数据治理技术解决实际问题，并展望未来数据治理研究在国家建设中所处的地位和作用，为进一步深入学习数据科学专业知识及从事相关实践工作奠定良好基础。

6. **《数据可视化》**: 本课程面向具有一定数学基础的硕士研究生同学。主要介绍数据可视化的定义、分类、基础知识、研究方法、研究内容、具体工具、算法和软件，以及研究的前沿问题。课程主要内容除了介绍具有基础性和通用性的背景知识和理论算法以外，分别根据来源于科学研究和生产生活的不同领域的不同需求，结合各自的背景情况和实际要求，有针对性的介绍科学可视化、地理数据可视化、数据库可视化、跨媒体可视化、文档可视化、复杂数据可视化以及常用可视化类库和软件平台。除了课本、参考书和文献内容以外，课程将介绍一定的计算机算法和软件操作。通过学习本课程，鼓励学生解决国内实际问题，促进学生在日常的学习工作生活中主动发现问题、解决问题。

7. **《深度学习》**: 主要包括计算机视觉和自然语言处理两部分内容。计算机视觉在图像理解、映射和视觉识别任务中有成功的应用，如图像分类、定位、检测和实例分割。语言翻译、广告、客户服务、总结、股票市场情感分析等，自然语言处理的应用也无处不在。将使用 Python 和 PyTorch 构建和训练神经网络，如 cnn、rnn、transformer 和许多其他神经网络。

络，以处理真实情况。

8. 《数据建模案例分析》: 主要介绍统计数据建模的思想和方法，分为三部分内容：线性模型及其推广、贝叶斯统计模型和案例分析。其中，线性模型及其推广从简单的线性模型出发讲解统计建模的基本思想和关键问题，贝叶斯统计模型部分则主要讲解贝叶斯模型的基本思想和基本的 MCMC 采样算法。最后一部分案例分析则以已发表的论文为题材讲解统计建模在数据分析中起到的关键作用。

其他选修课程

《金融数据分析与决策》

《自然语言分析与机器学习》

《区块链与数字货币》

《供应链管理》

《金融计量》