

## 中国社会科学院大学《大数据分析》课程大纲

课程基本信息 (Course Information)					
课程编号 (Course ID)		*学时 (Credit Hours)	32	*学分 (Credits)	2
*课程名称 (Course Name)	大数据分析				
	Big Data Analysis				
先修课程 (Prerequisite Courses)					
*课程简介 (Description)	<p>本课程的开课目的是介绍大数据的基本概念和相应的应用技术，课程从底层逻辑出发，由浅入深的介绍数据科学的概念、所需的数学基础、数据的整理清洗方法、数据分析方法（包括统计分析，可视化、数学模型、线性回归、logistic 回归，机器学习算法等）以及大数据分析的方法。针对本专业的学科特色，构建了对应的数据分析模式和分析路线，并给出了重要的实践案例。学生通过本课程的学习可以更快更好的掌握大数据分析技术，更好的服务于本专业的问题研究。</p>				
*课程简介 (Description)	<p>The purpose of this course is to introduce the basic concepts and corresponding application technologies of big data. Starting from the underlying logic, the course introduces the concepts of data science, the required mathematical foundations, data sorting and cleaning methods, data analysis methods (including statistical analysis, visualization, mathematical models, linear regression, logistic regression, machine learning algorithms, etc.), and big data analysis methods from shallow to deep. Based on the disciplinary characteristics of this major, corresponding data analysis models and analysis routes have been constructed, and important practical cases have been provided. Through the study of this course, students can quickly and better master the analysis technology of big data, and better serve the problem research of their major.</p>				
*教材 (Textbooks)	大数据分析，史蒂文，机械工业出版社，2022/4/1，1，978711170347				
参考资料 (Other References)					
*课程类别 (Course Category)	<input type="checkbox"/> 公共基础课/全校公共必修课 <input type="checkbox"/> 通识教育课 <input type="checkbox"/> 专业基础课 <input type="checkbox"/> 专业核心课/专业必修课 <input checked="" type="checkbox"/> 专业拓展课/专业选修课 <input type="checkbox"/> 其他_____				
*授课对象 (Target Students)	经济学、国贸专业本科生	*授课模式 (Mode of Instruction)	<input type="checkbox"/> 线上，教学平台_____ <input checked="" type="checkbox"/> 线下 <input type="checkbox"/> 混合式 <input type="checkbox"/> 其他 <input type="checkbox"/> 实践类（70%以上学时深入基层）_		
*开课院系 (School)	应用经济学院	*授课语言 (Language of Instruction)	<input type="checkbox"/> 中文 <input type="checkbox"/> 全外语_____ <input type="checkbox"/> 双语：中文+_____（外语讲授不低于 50%）		

*授课教师信息 (Teacher Information)	课程负责人 姓名及简介						
	团队成员 姓名及简介						
学习目标 (Learning Outcomes)	<p>1.了解大数据技术的基本概念、基本特征、基本技术组成</p> <p>2.掌握大数据技术涉及的数学理论基础。</p> <p>3.掌握常用大数据技术的常用分析方法、分析模型和分析过程。通过课程的学习，学生们可以更加敏捷的对海量数据变化进行感知，做出有效的分析和判断，并在复杂多样的信息中去伪存真，找到正确的观点和方法，并正式和准确的分析国家当前的经济局势搭建扎实的基础。</p>						
*考核方式 (Grading)	课程的成绩由平时成绩、期末成绩组成总评成绩。平时成绩占总评成绩的 30%，期末考评采用课下设计的方式进行。						
<b>*课程教学计划 (Teaching Plan) 填写规范化要求见附件</b>							
周次	周学时	其中					教学内容摘要 (必含章节名称、讲述的内容提要、实验的名称、教学方法、课堂讨论的题目、阅读文献参考书目及作业等)
		讲授	实验课	习题课	课程讨论	其他环节	
第一周	3	3	0				<p>第一章：大数据</p> <p>第一节 大数据概述 大数据的介绍、特征、技术应用和基础。</p> <p>第二节 大数据的意义 发展大数据的国家战略意义、大数据分析对企业的意义、我过的大数据分析现状。</p> <p>第三节 大数据现状分析 大数据技术分析和大数据运行分析</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业：使用 json 格式定义数据格式，并编写代码读取数据</p>
第二周	3	3	0				<p>第二章：数学基础</p> <p>第一节 概率 概率与统计、复合事件与独立事件、条件概率、概率分布</p> <p>第二节 统计性描述 中心性度量、变异性度量、解释方差、描述分布</p> <p>第三节 相关性分析 相关系数：皮尔逊和斯皮尔曼秩、相关的强弱和显著性</p> <p>第四节 对数 对数与乘法概率、对数和比率、对数与正态分布。</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业：完成概率相关概念练习题</p>

第三周	3	3	0			<p>第三章：数据清洗</p> <p>第一节 大数据存储</p> <p>第二节 数据清洗 数据清洗概述、数据清洗的原理、数据清洗的流程、数据清洗的工具。</p> <p>第三节 pandas 语法与使用</p> <p>第四节 pandas 的读取与数据清洗</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业：使用 pandas 清洗网络数据。</p>
第四周	3	3	0			<p>第四章：得分与排名</p> <p>第一节 评分系统概述 黄金标准和代理、排名与得分、评分函数</p> <p>第二节 Z 得分和归一化</p> <p>第三节 高级排名技术 Elo 排名、合并排名、基于有向图的排名、PageRank</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业：使用 Elo 评分系统对篮球比赛中的球队进行排名。</p>
第五周	3	3	0			<p>第五章：统计分析</p> <p>第一节 统计分布 二项分布、正态分布、泊松分布、</p> <p>第二节 采样 Gibs 采样、蒙特卡罗方法。</p> <p>第三节 显著性 显著性的意义、t 检验、Bonferroni 矫正。</p> <p>第四节 置换检验与 P 值</p> <p>第五节 贝叶斯定理</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业：完成相关概率分布问题</p>
第六周	3	3	0			<p>第六章数据可视化</p> <p>第一节 数据可视化 数据可视化概述、数据可视化工具、数据可视化图表</p> <p>第二节 Matplotlib 基础 Numpy 库、Matplotlib 库</p> <p>第三节 Matplotlib 的操作 线性图、柱状图、直方图、散点图、极坐标图、饼图。</p> <p>第四节 pycharts 可视化应用</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业： 作业：完成作业中指定的可视化任务。</p>
第七周	3	3	0			<p>第七章：数据模型</p> <p>第一节 模型分类 线性模型与非线性模型、黑盒与描述性模型、第一原理与数据驱动、随机模型与确定性模型、平面模型与分层模型。</p> <p>第二节 基准模型 分类的基准模型、价值预测基准模型。</p> <p>第三节 评估模型</p>

						<p>评估分类器、受试者工作特征曲线、评估多类系统、评估价值预测模型。</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业： 使用课程所学完成堵住推算问题分析。</p>
第八周	3	3	0			<p>第八章：线性代数</p> <p>第一节 线性代数的作用 集合和向量、</p> <p>第二节 矩阵运算 矩阵的加法、乘法、单位矩阵和求逆、矩阵的秩</p> <p>第三节 矩阵的分解 LU 分解和 SVD 分解</p> <p>第四节 特征值和特征向量</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业： 根据实验指导书完成矩阵计算练习。</p>
第九周	3	3	0			<p>第九章：线性回归和 logistic 回归</p> <p>第一节 线性回归 线性回归与对偶、线性回归误差、最优拟合、</p> <p>第二节 参数回归 参数空间、梯度下降法、随机梯度下降法</p> <p>第三节 正则简化模型 岭回归、Lasso 回归</p> <p>第四节 分类与 logistic 回归 分类回归、决策边界、logistic 回归</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业： 根据实验指导书完成回归练习。</p>
第十周	3	3	0			<p>第十章：聚类</p> <p>第一节 距离测量 距离度量、维度平均</p> <p>第二节 最近邻分类 K 紧邻、局部敏感哈希</p> <p>第三节 图、网络和距离 加权图和诱导网络</p> <p>第四节 聚类 K 均值聚类、凝聚聚类、相似图和基于切割的聚类</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业：聚类问题习题集</p>
第十一周	2	2	0			<p>第十一章：机器学习</p> <p>第一节 朴素贝叶斯 公式说明，零计数问题</p> <p>第二节 Boosting 和集成学习 分类投票器和 Boosting 算法</p> <p>第三节 支撑向量机 线性支撑向量机、非线性支撑向量机</p>

						<p>第四节 特征学习</p> <p>监督学习、无监督学习、半监督学习、特征工程</p> <p>第五节 深度学习</p> <p>网络和深度、反向传播、Keras 框架简介</p> <p>教学方法：以课堂讲授为主，穿插给出练习案例。</p> <p>作业：基于深度学习的数据分类</p>
总计	32	32	0			课程教学方法包括课堂教授、课下作业、经典案例分享、期末课程实践
备注 (Notes)						