

双教授线上科研项目-教授大纲

生物统计：

数据科学在生物医学中的应用

Big Data Science and Applications in Medical Science

生物统计，数据科学，公共健康，公共卫生，数据分析，统计建模

耶鲁大学 (美国) x 南开大学

Session Plan

Schedule	Topics
Professor Warm-Up	Course Introduction and Greetings
TA Preview	Course Preview and Introduction
Lecture 1	Objective: Key concepts of statistical modeling & Exploratory analysis 目标: 统计建模的关键概念&探索性分析 描述: 统计建模是使用数学模型和统计假设来生成样本数据, 并对现实世界进行预测。
中方教授课1	主题: 数据挖掘与生物信息的相关背景 描述: 讲授数据挖掘的背景与任务、生物信息学背景, 介绍数据的统计描述、可视化及预处理, 介绍Python编程。
Lecture 2	Objective: R programming & Regression analysis I (linear) 目标: R语言技术&回归分析I (线性) 描述: 回归分析是一组统计过程, 用于R编程和统计确定数据集变量之间的关系
中方教授课2	主题: 有监督分类算法与无监督聚类算法 描述: 讲授数据挖掘中的k邻近算法与神经网络等有监督数据分类算法, 以及k-Means与Louvain等无监督数据聚类算法。
Lecture 3	Objective: Regression analysis II (logistic) & Time series data analysis 目标: 回归分析II (逻辑) & 时间序列数据分析 描述: 时间序列分析是一种分析在一段时间内收集的一系列数据点的具体方法
中方教授课3	主题: 无监督降维算法与数据可视化方法 描述: 讲授数据挖掘中的主成分分析与自编码器等无监督数据降维方法, 以及t-SNE与UMAP等数据可视化方法。
Lecture 4	Objective: EHR and cancer omics data analysis I: characteristics, questions, analysis techniques 目标: 电子健康档案和癌症组学数据分析I: 特征, 问题, 分析技术 描述: 高通量技术的快速发展和电子健康档案(EHR)的广泛采用, 导致了经济数据和电子健康档案数据的快速积累。这些海量复杂的数据包含了精准医疗所需的丰富信息, 而大数据分析可以提取这些知识, 提高医疗质量。
中方教授课4	主题: 数据挖掘在基因组和转录组研究中的应用 描述: 讲授针对基因组和转录组的数据挖掘前沿研究, 包括卷积神经网络等算法以及调控元件、细胞类型的自动辨识等前沿研究案例。
Lecture 5	Objective: EHR and cancer omics data analysis II: case studies 目标: 电子健康档案和癌症组学数据分析II: 案例分析 描述: 同上节课, 额外增加了具体案例分析。

中方教授课5	主题：数据挖掘在表观组和蛋白组研究中的应用 描述：讲授针对表观组和蛋白组的数据挖掘前沿研究，包括表观数据不同分析任务的数学建模方法，并介绍AlphaFold2等前沿研究案例。
Final session	Summary, and Q&A 最终成果展示、课题答疑

Reading Materials

外方教授课题阅读材料	<ol style="list-style-type: none"> 1.生物统计的研究进展与挑战。统计研究。2016年6月。 2.Introduction to Biostatistics for Biomedical Research. Frank Harrell and James Slaughter. The first few chapters. 3.Information from social media on data science, big data, public health, biomedicine.
中方教授课题阅读材料	<ol style="list-style-type: none"> 1、数据挖掘导论（完整版），Pang-Ning Tan等，人民邮电出版社，ISBN：9787115241009，出版时间：2020-12-01（第一章） 2、机器学习，周志华，清华大学出版社，ISBN：9787302423287，出版时间：2016-01-01（第一章）