

北京大学医学部研究生课程教学大纲

课程名称	健康数据科学的 Python 语言编程基础					
	Basis of Python Programming in Health Data Science					
开课单位	医学技术研究院			课程代码		
学分	2	总学时		36	教学周数	9
课程性质	基础理论课					
先修课程	无要求					
课程负责人及授课教师						
课程中文简介						
Python 语言编程是开展健康数据科学、人工智能、机器学习等相关领域研究的重要基础，它具有通用灵活、语法简洁、社区支持强大等众多优点。讲授健康数据科学的 Python 语言编程基础，能够使学生掌握模块化、流程化、抽象化等编程思想，并通过实践加深对健康数据科学的相关分析算法原理的认知，提升解决实际问题的动手操作能力，从而提高研究生的数据科学素养，激发科研兴趣。本课程属于基础理论课，面向无编程基础的学生。						
课程英文简介						
Python language programming is one of the most important tools for researchers in health data science, artificial intelligence, machine learning related areas. Python has many advantages, such as universal flexibility, concise syntax, strong community support, etc. Learning Python language programming can enable students to master modular, procedural, abstract and other programming ideas. This course is a method course for students with no requirement of existing programming basis, includes basic concepts and practical tools for data analytics. It can deepen students' practical understanding of analysis algorithms in health data science.						
教学目的						
本课程属于基础理论课，面向无编程基础的硕博研究生。希望学生掌握 Python 语言编程基础和编程思想，以及运用 Python 数据分析工具解决问题的能力。包括掌握 Python 的基础语法、运算符和语句的原理；掌握字符串、列表、元组、字典、日期时间等基本数据类型的原理；掌握通过函数进行代码模块化；了解模块的原理；掌握常见数据文件的解析、读取和写出；了解正则表达式；了解面向对象的编程思想；掌握并运用 numpy 进行科学计算的原理；掌握并运用 pandas 进行表格清洗；掌握并运用 statsmodels 进行统计分析的原理；掌握并运用 sklearn 进行机器学习建模的原理；了解 matplotlib 作图；了解 pytorch 深度学习的原理。培养学生结合健康数据科学理论与实践的能力，并通过真实医疗数据分析的实际项目，深入了解编程和机器学习的相关原理，以及锻炼解决问题的实际动手能力。						
教学内容及学时分配						
主要内容						
周	课程内容（模块+标题）			学时	授课教师	
1	Python 通用基础（一）：基础语法、运算符和语句			4	洪申达	
2	Python 通用基础（二）：数据类型、函数和模块			4	洪申达	
3	Python 通用基础（三）：文件输入输出、解析与正则表达式			4	洪申达	
4	Python 通用基础（四）：面向对象的编程			4	洪申达	
5	Python 数据分析（一）：numpy 科学计算			4	洪申达	
6	Python 数据分析（二）：pandas 表格清洗与 statsmodels 统计分析			4	洪申达	
7	Python 数据分析（三）：sklearn 机器学习与 matplotlib 作图			4	洪申达	
8	Python 数据分析（四）：pytorch 深度学习（上）			4	洪申达	
9	Python 数据分析（五）：pytorch 深度学习（下）			4	洪申达	
1. Python 通用基础（一）：基础语法、运算符和语句（4 学时）						
1.1 简介：Python 发展历史、生态圈						
1.2 环境搭建：Anaconda 的下载与安装，Jupyter、Spyder 等界面及功能介绍						
1.3 基础语法：代码与命令行，缩进与注释，变量						
1.4 运算符与运算符优先级						
1.5 语句：条件，循环						
1.6 第一个程序 Hello World!						
2. Python 通用基础（二）：数据类型、函数和模块（4 学时）						
2.1 数据类型：字符串、列表、元组、字典、时间日期						
2.2 函数：定义与调用，参数及作用域						
2.3 模块：定义与调用						
3. Python 通用基础（三）：文件输入输出、解析与正则表达式（4 学时）						
3.1 文件输入输出：txt、csv、xlsx						
3.2 复杂文件解析：xml、json						
3.3 正则表达式						
4. Python 通用基础（四）：面向对象的编程（4 学时）						
4.1 面向对象的编程思想：面向对象与面向过程						
4.2 类与实例，变量及作用域，成员函数						
4.3 继承与重写						
5. Python 数据分析（一）：numpy 科学计算（4 学时）						
5.1 线性代数简介						
5.2 使用 numpy 工具包进行科学计算						
6. Python 数据分析（二）：pandas 表格清洗与 statsmodels 统计分析（4 学时）						
6.1 使用 pandas 清洗表格数据						
6.2 使用 statsmodels 进行统计分析						
7. Python 数据分析（三）：sklearn 机器学习与 matplotlib 作图（4 学时）						
7.1 机器学习简介						
7.2 使用 sklearn 预处理数据，建立、训练、评估、筛选机器学习模型						
7.3 使用 matplotlib 作图						
8. Python 数据分析（四）：pytorch 深度学习（上）（4 学时）						
8.1 深度学习简介						

8.2 pytorch 环境搭建 8.3 pytorch 语法基础 9. Python 数据分析（五）：pytorch 深度学习（下）（4 学时） 9.1 使用 pytorch 预处理数据，建立、训练、评估深度学习模型 9.2 部署机器学习模型					
教学方式					
线下课堂讲授，鼓励学生课后多动手练习。					
课程考核及成绩评定方式					
包括平时作业与大作业。大作业：根据本课程所学知识，以单人或 2-3 人分组合作的方式：在 kaggle 平台上参加一项数据科学相关的比赛，完成提交，进行赛题解决方案报告，包括问题描述、解决方案、结果及讨论等。 总评成绩 = 课堂表现 20% + 平时作业 40% + 大作业 40%					
课程教材及主要参考书目、文献及相关资料					
教材	书名	作者		出版机构、ISBN 号	出版日期
参考书	书名	作者		出版机构、ISBN 号	出版日期
	Deep learning	Goodfellow, Ian, Yoshua Bengio, and Aaron Courville		MIT press	2016-01-01
	Kaggle	数据比赛平台		https://www.kaggle.com/	2022-11-16
	Scikit-learn: Machine learning in Python	Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al.		Journal of machine Learning research 12 (2011): 2825-2830	2011-10-11
	Learn Python 3	在线课程		https://www.codecademy.com/learn/learn-python-3	2022-11-16
	Array programming with NumPy	Harris, Charles R., K. Jarrod Millman, Stéfan J. Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser et al.		Nature 585, no. 7825 (2020): 357-362	2020-09-16
	pandas: a foundational Python library for data analysis and statistics	McKinney, Wes.		Python for high performance and scientific computing 14, no. 9 (2011): 1-9	2011-01-01
	Pytorch: An imperative style, high-performance deep learning library	Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen et al		Advances in neural information processing systems 32 (2019)	2019-12-03
	机器学习	周志华		清华大学出版社，9787302423287	2016-01-01
其他参考材料（文献、网站等）					
授课团队简介					
课程负责人	姓名		出生年月		来校工作时间
	办公地点		3	办公时间	
	联系电话			邮箱	ho . .
课程负责人及授课教师简历					
研					