



TAIJI LABORATORY
FOR GRAVITATIONAL WAVE UNIVERSE



ICTP-AP
International Centre
for Theoretical Physics Asia-Pacific
国际理论物理中心-亚太地区



中国科学院大学
University of Chinese Academy of Sciences

引力波数据探索：编程与分析实战训练营

结营仪式与课程总结

主讲老师：王赫

ICTP-AP, UCAS

2024/01/14





致谢

- 主办单位

- 中国科学院大学·国际理论物理中心（亚太地区）
- 中国科学院大学·引力波宇宙太极实验室



United Nations
Educational, Scientific and
Cultural Organization



ICTP-AP
INTERNATIONAL CENTRE FOR
THEORETICAL PHYSICS ASIA-PACIFIC



TAIJI LABORATORY
FOR GRAVITATIONAL WAVE UNIVERSE

张敏 办公室主任

赵洁瑾 行政业务主管

孟繁梓 外联业务主管

刘颖慧 科研业务主管

张乐 学生业务主管

贾海洪 支撑业务主管

田昕晓 特邀嘉宾

赵俊杰 特邀嘉宾

高民权 特邀嘉宾

- 赞助单位

- 中科曙光



但易 活动策划+算力支持



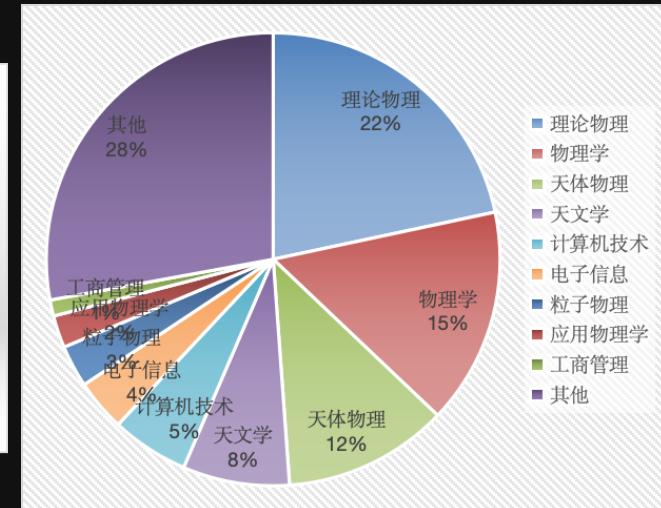
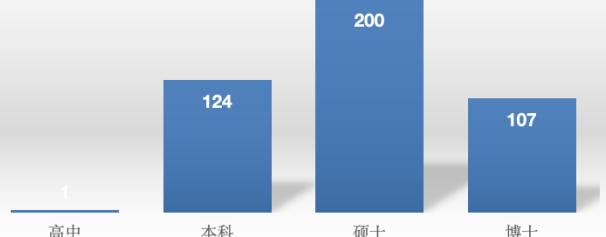


训练营课程回顾与总结

• 课程报名与参与人员情况

- 训练营报名初期，有效申请的学员人数共 432 人，来自各行各业。其中一半以上来自理论物理、物理学、天体物理和天文学等专业方向。

学历分布



- 在“其他”专业类别中，包含心理学、生物信息与医药、工商管理、材料与化工、地震与地球物理、金融、控制工程等等。



- 在学员填写的“个人研究的主要障碍”和“对本课程的期望”的词云分析中，“引力波”、“机器学习”、“数据处理”、“深度学习”、“能力”、“编程”等是最常见的关键词。



训练营课程回顾与总结

• 教学大纲与实战项目

Knowledge Framework

引力波数据分析：知识框架

- 理论基础：
 - 引力波物理 (npycbc, lalsuite, lisacode, bilby, ...)
 - 数学信号处理 (scipy, stat, ...)
 - 数学统计 (bilby, emcee, ptemcee, ptmcmc, ...)
- 编程基础：
 - Python (numpy, pandas, matplotlib; ...)
 - AI (sklearn, XGBoost, PyTorch, TensorFlow, JAX, ...)
 - Linux (docker, github, bash, vim, emacs ...)
- 硬件基础：
 - 主板、内存、GPUs、显存 ...
 - AI时代程序员都应该了解的GPU基础知识

Self-Mastery

通向自我实现之路：用心卷动每一步

- 对“卷”的理性认识
 - 人多，竞争激烈？
 - 奋斗、努力、上进？
 - 恶性的规则与制度。
 - 心之所向，行之力卷！

Self-Mastery

通向自我实现之路：学会请教问题

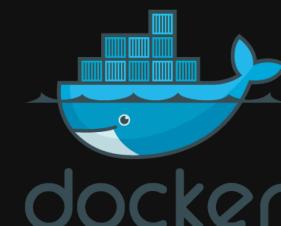
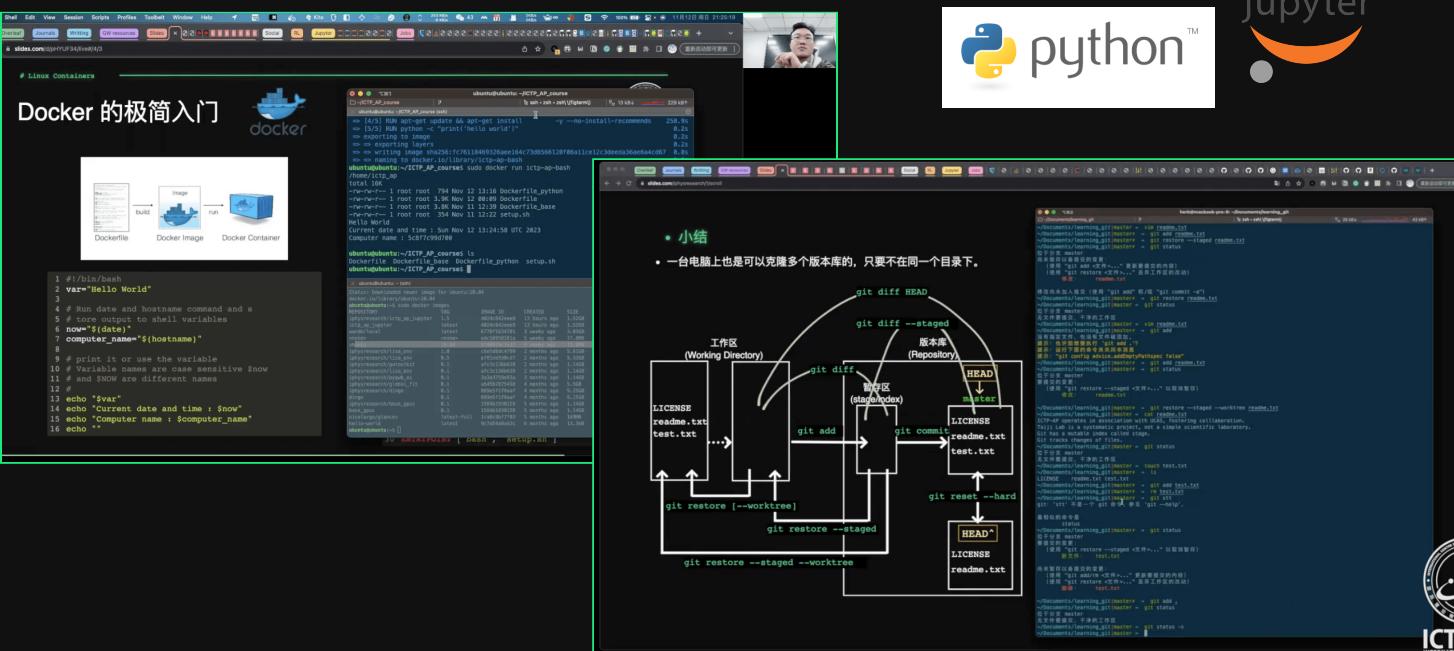
- 对“卷”的理性认识
 - 做研究所需要的范式转换
 - 做好笔记是自我学习的不二捷径
 - 自学路上必会之“上下求索”的技能
 - 自学路上必会之“提问的艺术”
- “提问的内容尽量包含”：
 1. 你为了解决这个问题，你是做了哪些操作？你希望这些操作是什么效果和实际的效果是什么？
 2. 你猜测这个问题的根源是什么，以及你认为解决该问题可能需要如何做？
 3. 请告知你的基本电脑环境：如何操作系统？在云平台？你的库版本是等~
 4. 如果你的操作步骤不同于课程中的notebook模板，请强调出来~
 5. 你收到的报错信息是什么？（很重要！）
 6. 你尝试过什么方法来解决这个问题？
 7. 代码和运行中遇到的返回结果不一致，请指出来~
 8. 你已经尝试的部分问题是什么做的？比方说你重启了电脑？重启了kernel？
 9. 你发觉到某些可行的解决方案，可以贴出来，如来自csdn、简书等~

- 第0部分：打鸡血！
 - 通向自我实现之路
- 第1部分：编程开发环境与工作流
 - 基础运维技术
 - 容器化技术
 - 实战项目：Python / Jupyter 开发环境搭建 + 远程连接 VS Code
 - 实战项目：LALsuite / LISACode 的源码编译 (optional)
 - Git 分布式版本控制系统
- 【公开课】数据技术演进与现实应用 (特邀嘉宾：田昕晓)
- 第2部分：基于 Python 的数据分析基础
 - 数据科学语言 Python 从入门到熟悉
 - 数据分析实训之 Numpy / Pandas
 - 实战项目：GW Event Catalog 的探索性数据分析
 - 实战项目：股票数据分析案例 (optional)
 - 基于 Python 的数据可视化理论与实践之 Matplotlib / Seaborn
 - 实战项目：GWTC 论文中的 Figures
 - 实战项目：针对 GW150914 信号处理与匹配滤波数据分析
- 【公开课】贝叶斯推断在引力波科学中的应用 (特邀嘉宾：赵俊杰)
- 第3部分：机器学习基础
 - 机器学习算法之应用起步
 - 机器学习算法之应用进阶
 - 实战项目：基于 LIGO 的 Glitch 元数据完成多分类任务
 - 实战项目：基于 LIGO 的 Glitch 时频图数据实现聚类分析
- 第4部分：深度学习基础
 - 深度学习技术概述与神经网络基础
 - 实战项目：训练一个3层神经网络（手撸版）
 - 卷积神经网络与引力波信号探测
 - 实战项目：使用 CNN 识别双黑洞系统引力波信号
 - Kaggle 数据科学竞赛 (黑客马拉松): Can you find the GW signals?
- 【公开课】AI发展全景与GPT前沿解析 (特邀嘉宾：高民权)



训练营课程回顾与总结

• 教学大纲与实战项目



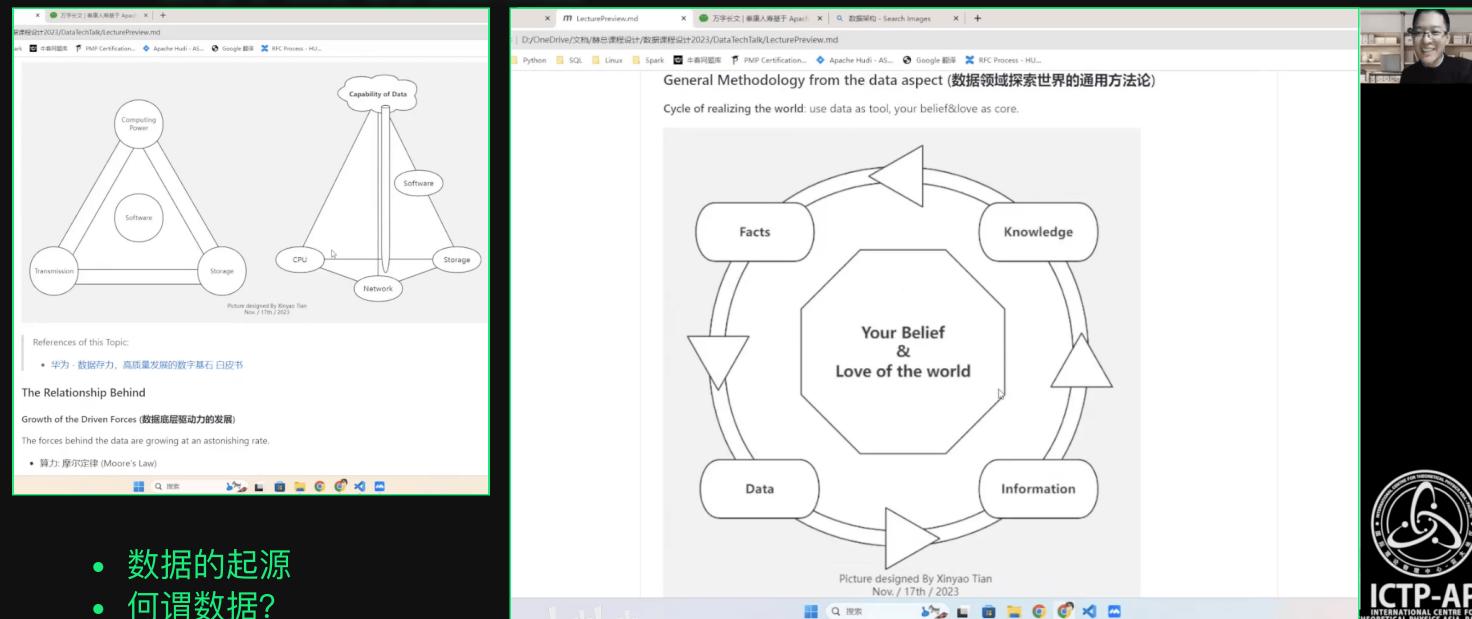
- 基础运维技术
 - 什么是Linux/Shell；新手必须掌握的Linux命令
 - 管道符、重定向与环境变量；SSH服务管理远端设备
 - 容器化技术
 - Linux 容器虚拟化；Docker 的极简入门
 - Git 分布式版本控制系统
 - Git 安装/创建版本库；工作区、暂存区、版本库
 - 远程仓库；分支管理；Git 可视化管理工具

- 第 0 部分：打鸡血！
 - 通向自我实现之路
 - 第 1 部分：编程开发环境与工作流
 - 基础运维技术
 - 容器化技术
 - 实战项目：Python / Jupyter 开发环境搭建 + 远程连接 VS Code
 - 实战项目：LALsuite / LISACode 的源码编译 (optional)
 - Git 分布式版本控制系统
 - 【公开课】数据技术演进与现实应用 (特邀嘉宾：田昕峣)
 - 第 2 部分：基于 Python 的数据分析基础
 - 数据科学语言 Python 从入门到熟悉
 - 数据分析实训之 Numpy / Pandas
 - 实战项目：GW Event Catalog 的探索性数据分析
 - 实战项目：股票数据分析案例 (optional)
 - 基于 Python 的数据可视化理论与实践之 Matplotlib / Seaborn
 - 实战项目：GWTC 论文中的 Figures
 - 实战项目：针对 GW150914 信号处理与匹配滤波数据分析
 - 【公开课】贝叶斯推断在引力波科学中的应用 (特邀嘉宾：赵俊杰)
 - 第 3 部分：机器学习基础
 - 机器学习算法之应用起步
 - 机器学习算法之应用进阶
 - 实战项目：基于 LIGO 的 Glitch 元数据完成多分类任务
 - 实战项目：基于 LIGO 的 Glitch 时频图数据实现聚类分析
 - 第 4 部分：深度学习基础
 - 深度学习技术概述与神经网络基础
 - 实战项目：训练一个3层神经网络（手撸版）
 - 卷积神经网络与引力波信号探测
 - 实战项目：使用 CNN 识别双黑洞系统引力波信号
 - Kaggle 数据科学竞赛 (黑客马拉松): Can you find the GW signals?
 - 【公开课】AI发展全景与GPT前沿解析 (特邀嘉宾：高民权)



训练营课程回顾与总结

• 教学大纲与实战项目



- 数据的起源
- 何谓数据?
- 现代数据技术的发展脉络
- 当前主流数据技术
 1. 关系型数据库 (RDBMS)
 2. 非关系型数据库 (Not-only SQL (NoSQL) Database)
 3. 大数据 (Big Data)
 4. 数据仓库 (Data Warehouse)
 5. 流式计算 (Stream Processing)
 6. 数据湖 (Data Lake)
 7. 数据湖仓 (Data Lakehouse)
- 思考：从数据的角度认识世界
- 推荐阅读

- 第 0 部分：打鸡血！
 - 通向自我实现之路
- 第 1 部分：编程开发环境与工作流
 - 基础运维技术
 - 容器化技术
 - 实战项目：Python / Jupyter 开发环境搭建 + 远程连接 VS Code
 - 实战项目：LALsuite / LISACode 的源码编译 (optional)
 - Git 分布式版本控制系统
- 【公开课】数据技术演进与现实应用 (特邀嘉宾：田昕晓)
- 第 2 部分：基于 Python 的数据分析基础
 - 数据科学语言 Python 从入门到熟悉
 - 数据分析实训之 Numpy / Pandas
 - 实战项目：GW Event Catalog 的探索性数据分析
 - 实战项目：股票数据分析案例 (optional)
 - 基于 Python 的数据可视化理论与实践之 Matplotlib / Seaborn
 - 实战项目：GWTC 论文中的 Figures
 - 实战项目：针对 GW150914 信号处理与匹配滤波数据分析
- 【公开课】贝叶斯推断在引力波科学中的应用 (特邀嘉宾：赵俊杰)
- 第 3 部分：机器学习基础
 - 机器学习算法之应用起步
 - 机器学习算法之应用进阶
 - 实战项目：基于 LIGO 的 Glitch 元数据完成多分类任务
 - 实战项目：基于 LIGO 的 Glitch 时频图数据实现聚类分析
- 第 4 部分：深度学习基础
 - 深度学习技术概述与神经网络基础
 - 实战项目：训练一个3层神经网络（手撸版）
 - 卷积神经网络与引力波信号探测
 - 实战项目：使用 CNN 识别双黑洞系统引力波信号
 - Kaggle 数据科学竞赛 (黑客马拉松): Can you find the GW signals?
- 【公开课】AI发展全景与GPT前沿解析 (特邀嘉宾：高民权)



训练营课程回顾与总结

• 教学大纲与实战项目

数据可视化：模式与对象

• 定量数据 (Quantitative)
位置 > 长度/角度 > 面积 > 体积 > 密度 > 颜色
上述为优先关系，体现了维度优先的原则。

• 定序数据 (Ordinal)
位置 > 密度 > 颜色 > 连接 > 包含
其中的密度可以通过疏密程度来体现；颜色主要是通过深浅体现，避免视觉混淆；连接可以用箭头等隶属关系来体现有序性。

• 定类数据 (Categorical)
位置 > 颜色 > 连接 > 包含 > 形状
表现类之间的关系，确保元素清晰。

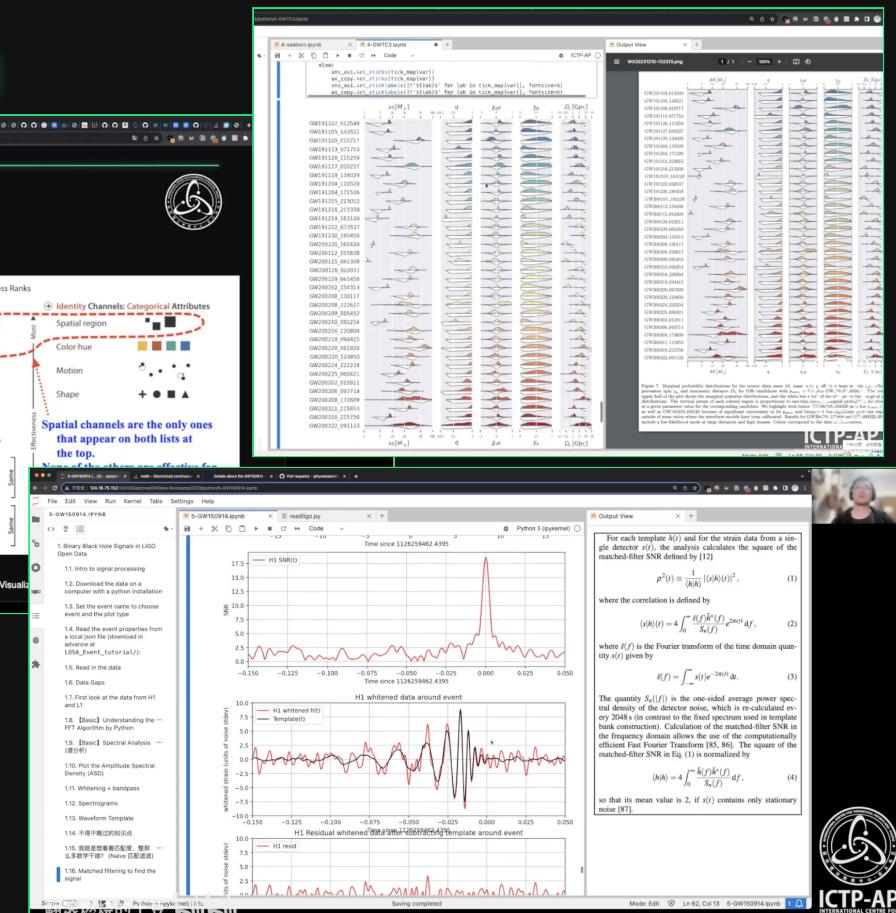
Munzner T. Visualization

翻滚燃烧的土豆

- Python 基本数据类型：
 - 数值精度
 - 可变对象 vs 不可变对象
- Python 流程控制与异常处理
- Python 面向对象编程
- Python 中的内存回收机制
- Numpy 中的数据结构
- Numpy 中的数据切片
- Numpy 中的广播机制
- Numpy 中的线性代数与数理统计
- Pandas 中的数据结构
- DataFrame 的增删改查
- DataFrame 的组合与聚合&透视表



- 数据化思维与数据可视化
- 数据可视化：目的与原则
- 数据可视化：模式与对象
- 数据可视化：基本处理流程



- LIGO Open Data
- FFT by Scratch
- Spectral Analysis by Scratch
- Data analysis on GW150914
- Matched filtering to find the signal



- 第 0 部分：打鸡血！
 - 通向自我实现之路
- 第 1 部分：编程开发环境与工作流
 - 基础运维技术
 - 容器化技术
 - 实战项目：Python / Jupyter 开发环境搭建 + 远程连接 VS Code
 - 实战项目：LALsuite / LISACode 的源码编译 (optional)
 - Git 分布式版本控制系统
- 【公开课】数据技术演进与现实应用 (特邀嘉宾：田昕晓)
- 第 2 部分：基于 Python 的数据分析基础
 - 数据科学语言 Python 从入门到熟悉
 - 数据分析实训之 Numpy / Pandas
 - 实战项目：GW Event Catalog 的探索性数据分析
 - 实战项目：股票数据分析案例 (optional)
 - 基于 Python 的数据可视化理论与实践之 Matplotlib / Seaborn
 - 实战项目：GWTC 论文中的 Figures
 - 实战项目：针对 GW150914 信号处理与匹配滤波数据分析
- 【公开课】贝叶斯推断在引力波科学中的应用 (特邀嘉宾：赵俊杰)
- 第 3 部分：机器学习基础
 - 机器学习算法之应用起步
 - 机器学习算法之应用进阶
 - 实战项目：基于 LIGO 的 Glitch 元数据完成多分类任务
 - 实战项目：基于 LIGO 的 Glitch 时频图数据实现聚类分析
- 第 4 部分：深度学习基础
 - 深度学习技术概述与神经网络基础
 - 实战项目：训练一个3层神经网络（手撸版）
 - 卷积神经网络与引力波信号探测
 - 实战项目：使用 CNN 识别双黑洞系统引力波信号
 - Kaggle 数据科学竞赛 (黑客马拉松): Can you find the GW signals?
- 【公开课】AI发展全景与GPT前沿解析 (特邀嘉宾：高民权)



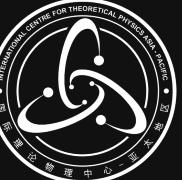
训练营课程回顾与总结

• 教学大纲与实战项目



- Brief introduction to gravitational wave (引力波简要介绍)
- Part I: Bayesian inference (贝叶斯推断)
- Part II: Bayesian computation (贝叶斯计算方法)
 - Markov Chain Monte Carlo (MCMC; 马尔可夫链-蒙特卡罗方法)
 - Nested sampling (嵌套采样)
- Part III: All in gravitational-wave data (一切尽在引力波数据中)
 - Use Bilby & Parallel Bilby in the GW data analysis
 - Show the complete pipeline for the data analysis
- The AMAZING Thomas Bayes (为美好的世界献上"贝叶斯定理")

- 第 0 部分: 打鸡血!
 - 通向自我实现之路
- 第 1 部分: 编程开发环境与工作流
 - 基础运维技术
 - 容器化技术
 - 实战项目: Python / Jupyter 开发环境搭建 + 远程连接 VS Code
 - 实战项目: LALsuite / LISACode 的源码编译 (optional)
 - Git 分布式版本控制系统
- 【公开课】数据技术演进与现实应用 (特邀嘉宾: 田昕晓)
- 第 2 部分: 基于 Python 的数据分析基础
 - 数据科学语言 Python 从入门到熟悉
 - 数据分析实训之 Numpy / Pandas
 - 实战项目: GW Event Catalog 的探索性数据分析
 - 实战项目: 股票数据分析案例 (optional)
 - 基于 Python 的数据可视化理论与实践之 Matplotlib / Seaborn
 - 实战项目: GWTC 论文中的 Figures
 - 实战项目: 针对 GW150914 信号处理与匹配滤波数据分析
- 【公开课】贝叶斯推断在引力波科学中的应用 (特邀嘉宾: 赵俊杰)
- 第 3 部分: 机器学习基础
 - 机器学习算法之应用起步
 - 机器学习算法之应用进阶
 - 实战项目: 基于 LIGO 的 Glitch 元数据完成多分类任务
 - 实战项目: 基于 LIGO 的 Glitch 时频图数据实现聚类分析
- 第 4 部分: 深度学习基础
 - 深度学习技术概述与神经网络基础
 - 实战项目: 训练一个3层神经网络 (手撸版)
 - 卷积神经网络与引力波信号探测
 - 实战项目: 使用 CNN 识别双黑洞系统引力波信号
 - Kaggle 数据科学竞赛 (黑客马拉松): Can you find the GW signals?
- 【公开课】AI发展全景与GPT前沿解析 (特邀嘉宾: 高民权)



训练营课程回顾与总结

• 教学大纲与实战项目

机器学习模型的分类

注：有些时候数据没有提供足够信息来事先假设分布、或者问题本身没有明显的分布特性

- 根据 **数据分布** 分类：参数 vs 非参数模型
 - 这里的“参数”并不是模型中的参数，而是数据分布的参数
- 参数模型：**
 - 对数据分布进行假设，待求解的数据模式/映射可以用一组有限且固定数目的模型参数进行刻画
- 非参数模型：**
 - 不对数据分布进行假设，数据的所有统计特性都来源于数据本身
 - 非参数模型的时空复杂度一般比参数模型大得多
 - 非参数模型是自适应数据的，模型参数随样本变化而变化

线性回归模型

条件概率 $P(Y|X)$ 属于高斯分布

$y = mx + b$

K 近邻模型

如：随机森林、朴素贝叶斯、SVM、神经网络

优点：函数可变性、模型强大程度少、拟合性好

局限性：数据需求量大、速度慢、易过拟合、预测解释性不高

滚动燃烧的土豆 Bilibili

- 人工智能 > 机器学习 > 深度学习
- 机器学习的定义、目标和过程
- 机器学习的常见类型
- 机器学习模型的分类
- 机器学习项目开发规划与准备及其开发步骤
- scikit-learn 机器学习库：分类+回归
- 机器学习中的模型调优与模型融合
 - 特征工程
 - 交叉验证
 - 网格搜索
 - 集成学习

-
- 第0部分：打鸡血！
 - 通向自我实现之路
 - 第1部分：编程开发环境与工作流
 - 基础运维技术
 - 容器化技术
 - 实战项目：Python / Jupyter 开发环境搭建 + 远程连接 VS Code
 - 实战项目：LALsuite / LISAcode 的源码编译 (optional)
 - Git 分布式版本控制系统
 - 【公开课】数据技术演进与现实应用 (特邀嘉宾：田昕峣)
 - 第2部分：基于 Python 的数据分析基础
 - 数据科学语言 Python 从入门到熟悉
 - 数据分析实训之 Numpy / Pandas
 - 实战项目：GW Event Catalog 的探索性数据分析
 - 实战项目：股票数据分析案例 (optional)
 - 基于 Python 的数据可视化理论与实践之 Matplotlib / Seaborn
 - 实战项目：GWTC 论文中的 Figures
 - 实战项目：针对 GW150914 信号处理与匹配滤波数据分析
 - 【公开课】贝叶斯推断在引力波科学中的应用 (特邀嘉宾：赵俊杰)
 - 第3部分：机器学习基础
 - 机器学习算法之应用起步
 - 机器学习算法之应用进阶
 - 实战项目：基于 LIGO 的 Glitch 元数据完成多分类任务
 - 实战项目：基于 LIGO 的 Glitch 时频图数据实现聚类分析
 - 第4部分：深度学习基础
 - 深度学习技术概述与神经网络基础
 - 实战项目：训练一个3层神经网络（手撸版）
 - 卷积神经网络与引力波信号探测
 - 实战项目：使用 CNN 识别双黑洞系统引力波信号
 - Kaggle数据科学竞赛 (黑客马拉松): Can you find the GW signals?
 - 【公开课】AI发展全景与GPT前沿解析 (特邀嘉宾：高民权)



训练营课程回顾与总结

• 教学大纲与实战项目

深度学习技术的发展

- D. Rumelhart & BP Algorithm
 - 神经网络被 Minsky 诟病的问题：巨大的计算量：XOR 问题；
 - 传统的感知器所谓“梯度下降”的算法纠错时，耗费的计算量和神经元数目的平方成正比，当神经元数目增多，庞大的计算量是当时的硬件无法胜任的。
 - 1982年，美国加州理工物理学家 J.J.Hopfield 提出了 Hopfield 神经网格模型，引入了计算能量概念，给出了网络稳定性判断。
 - 1986 年 7 月，Hinton 和 David Rumelhart 合作在 Nature 杂志上发表论文：“Learning Representations by Back-propagating Errors.”第一次系统简洁地阐述 BP 算法及其应用：
 - 反向传播算法把纠错的运算量降到只和神经元数目本身成正比；
 - BP 算法通过在神经网络里增加一个所谓 **隐层**（hidden layer），解决了 XOR 难题
 - 使用了 **BP 算法**的神经网络在做如形状识别之类的简单工作时，效率比感知器大大提高，八十年代末计算机的运行速度，也比二十年前高了 **几千数量级**；
 - 神经网络及其应用的研究 [开始复习](#)！
- $h_j = \text{Sgn}(\sum_{i=1}^n w_{ji}x_i - \theta_j)$
- $y = \text{Sgn}(\sum_{j=1}^m w_jh_j - \theta)$

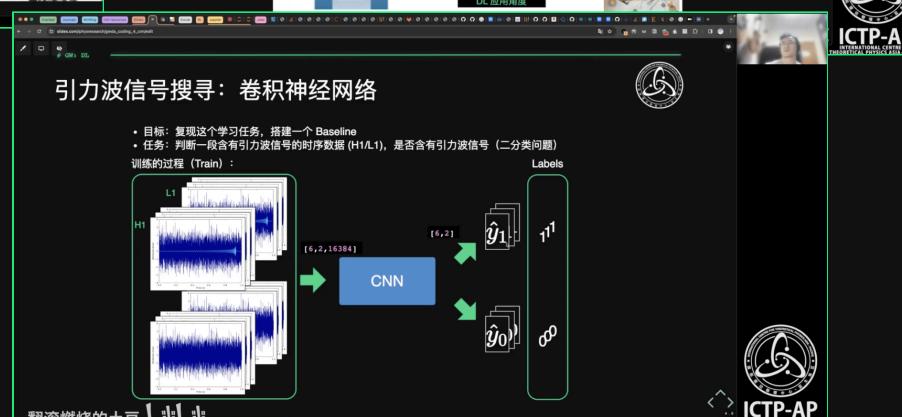
翻滚燃烧的土豆 bilibili

机器学习与深度学习技术的学习材料

工程角度、无需高等数学背景
参数非参数+频率贝叶斯角度
讲理论、不会讲堆量
DL 应用角度
统计角度
贝叶斯角度
设计方法集大成的书
The Nature of Statistical Learning Theory
Understanding Machine Learning
大量数学推导

PyTorch

- 深度学习技术的起源、应用、特点
- 人工智能技术自学材料推荐
- 深度学习技术的“不能”
- 深度学习：神经网络基础
- 模型性能评估与测试调优
- 一维、二维卷积
- 引力波信号搜寻：卷积神经网络



kaggle™

Can you find the GW signals?

Gravitational Wave Data Exploration: A Practical Training in Programming and Analysis (2023)

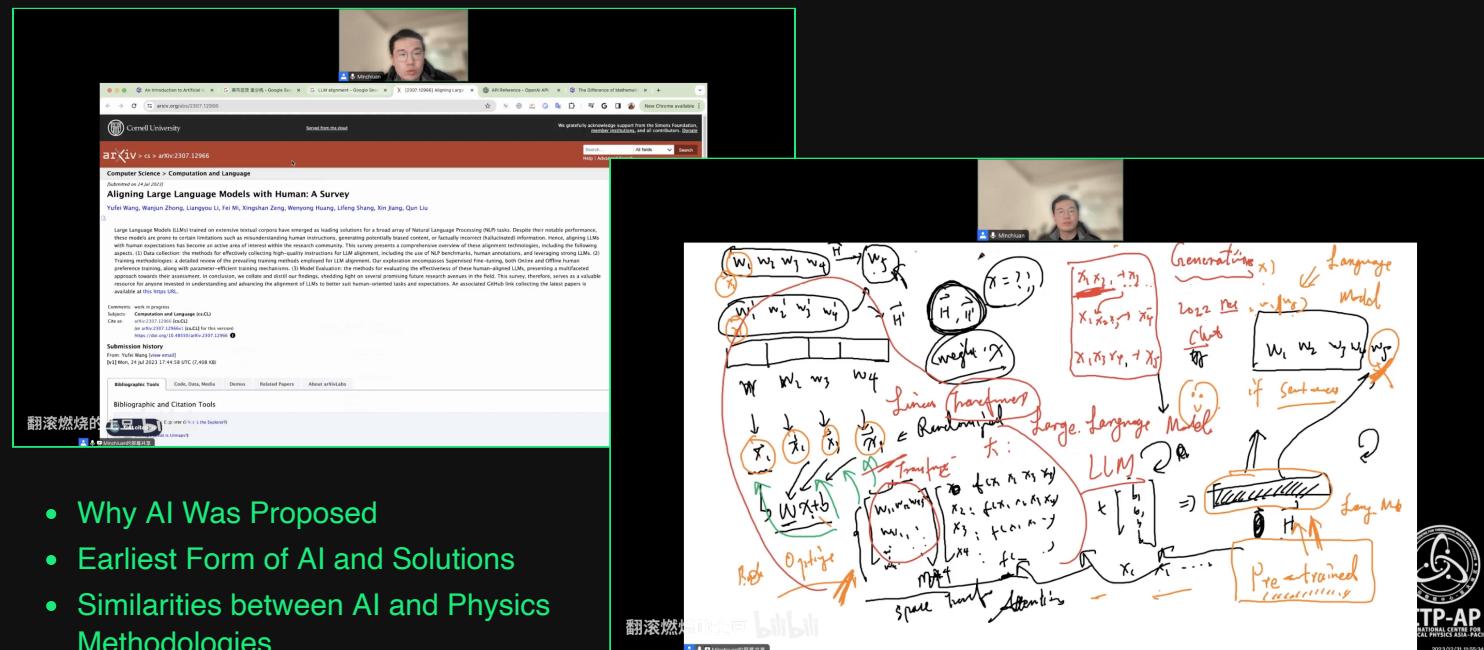


- 第 0 部分：打鸡血！
 - 通向自我实现之路
- 第 1 部分：编程开发环境与工作流
 - 基础运维技术
 - 容器化技术
 - **实战项目**：Python / Jupyter 开发环境搭建 + 远程连接 VS Code
 - **实战项目**：LALsuite / LISACode 的源码编译 (optional)
 - Git 分布式版本控制系统
- 【公开课】数据技术演进与现实应用 (特邀嘉宾：田昕晓)
- 第 2 部分：基于 Python 的数据分析基础
 - 数据科学语言 Python 从入门到熟悉
 - 数据分析实训之 Numpy / Pandas
 - **实战项目**：GW Event Catalog 的探索性数据分析
 - **实战项目**：股票数据分析案例 (optional)
 - 基于 Python 的数据可视化理论与实践之 Matplotlib / Seaborn
 - **实战项目**：GWTC 论文中的 Figures
 - **实战项目**：针对 GW150914 信号处理与匹配滤波数据分析
- 【公开课】贝叶斯推断在引力波科学中的应用 (特邀嘉宾：赵俊杰)
- 第 3 部分：机器学习基础
 - 机器学习算法之应用起步
 - 机器学习算法之应用进阶
 - **实战项目**：基于 LIGO 的 Glitch 元数据完成多分类任务
 - **实战项目**：基于 LIGO 的 Glitch 时频图数据实现聚类分析
- 第 4 部分：深度学习基础
 - 深度学习技术概述与神经网络基础
 - **实战项目**：训练一个3层神经网络（手撸版）
 - 卷积神经网络与引力波信号探测
 - **实战项目**：使用 CNN 识别双黑洞系统引力波信号
 - Kaggle 数据科学竞赛 (黑客马拉松): Can you find the GW signals?
- 【公开课】AI发展全景与GPT前沿解析 (特邀嘉宾：高民权)



训练营课程回顾与总结

• 教学大纲与实战项目



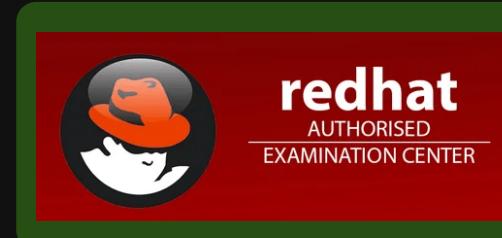
- Why AI Was Proposed
 - Earliest Form of AI and Solutions
 - Similarities between AI and Physics Methodologies
 - From Symbolic Systems to Machine Learning
 - Principles of Deep Learning
 - Breakthroughs Brought by Deep Learning
 - Typical Deep Learning Scenarios
 - Pre-trained Models and Large Models
 - Principles of GPT
 - Breakthroughs in AIGC (AI Generated Content)
 - Current Challenges in AI
 - Frontiers of AI Research

-
- 第0部分：打鸡血！
 - 通向自我实现之路
 - 第1部分：编程开发环境与工作流
 - 基础运维技术
 - 容器化技术
 - 实战项目：Python / Jupyter 开发环境搭建 + 远程连接 VS Code
 - 实战项目：LALsuite / LISAcode 的源码编译 (optional)
 - Git 分布式版本控制系统
 - 【公开课】数据技术演进与现实应用 (特邀嘉宾：田昕峣)
 - 第2部分：基于 Python 的数据分析基础
 - 数据科学语言 Python 从入门到熟悉
 - 数据分析实训之 Numpy / Pandas
 - 实战项目：GW Event Catalog 的探索性数据分析
 - 实战项目：股票数据分析案例 (optional)
 - 基于 Python 的数据可视化理论与实践之 Matplotlib / Seaborn
 - 实战项目：GWTC 论文中的 Figures
 - 实战项目：针对 GW150914 信号处理与匹配滤波数据分析
 - 【公开课】贝叶斯推断在引力波科学中的应用 (特邀嘉宾：赵俊杰)
 - 第3部分：机器学习基础
 - 机器学习算法之应用起步
 - 机器学习算法之应用进阶
 - 实战项目：基于 LIGO 的 Glitch 元数据完成多分类任务
 - 实战项目：基于 LIGO 的 Glitch 时频图数据实现聚类分析
 - 第4部分：深度学习基础
 - 深度学习技术概述与神经网络基础
 - 实战项目：训练一个3层神经网络（手撸版）
 - 卷积神经网络与引力波信号探测
 - 实战项目：使用 CNN 识别双黑洞系统引力波信号
 - Kaggle数据科学竞赛 (黑客马拉松): Can you find the GW signals?
 - 【公开课】AI发展全景与GPT前沿解析 (特邀嘉宾：高民权)



训练营课程回顾与总结

• 如何登峰造极? (自我提升的途径)



- 第 0 部分: 打鸡血!
 - 通向自我实现之路
- 第 1 部分: 编程开发环境与工作流
 - 基础运维技术
 - 容器化技术
 - 实战项目: Python / Jupyter 开发环境搭建 + 远程连接 VS Code
 - 实战项目: LALsuite / LISACode 的源码编译 (optional)
 - Git 分布式版本控制系统
- 【公开课】数据技术演进与现实应用 (特邀嘉宾: 田昕晓)
- 第 2 部分: 基于 Python 的数据分析基础
 - 数据科学语言 Python 从入门到熟悉
 - 数据分析实训之 Numpy / Pandas
 - 实战项目: GW Event Catalog 的探索性数据分析
 - 实战项目: 股票数据分析案例 (optional)
 - 基于 Python 的数据可视化理论与实践之 Matplotlib / Seaborn
 - 实战项目: GWTC 论文中的 Figures
 - 实战项目: 针对 GW150914 信号处理与匹配滤波数据分析
- 【公开课】贝叶斯推断在引力波科学中的应用 (特邀嘉宾: 赵俊杰)
- 第 3 部分: 机器学习基础
 - 机器学习算法之应用起步
 - 机器学习算法之应用进阶
 - 实战项目: 基于 LIGO 的 Glitch 元数据完成多分类任务
 - 实战项目: 基于 LIGO 的 Glitch 时频图数据实现聚类分析
- 第 4 部分: 深度学习基础
 - 深度学习技术概述与神经网络基础
 - 实战项目: 训练一个3层神经网络 (手撸版)
 - 卷积神经网络与引力波信号探测
 - 实战项目: 使用 CNN 识别双黑洞系统引力波信号
 - Kaggle 数据科学竞赛 (黑客马拉松): Can you find the GW signals?
- 【公开课】AI发展全景与GPT前沿解析 (特邀嘉宾: 高民权)



训练营课程回顾与总结

• 作业完成情况

Screenshot of the GitHub repository iphysresearch / GWData-Bootcamp showing the Pull requests tab. The list includes various pull requests related to assignments and homework, such as Matplotlib and Seaborn Homework, Machine Learning assignments, and Python homework. Most pull requests are labeled as 'reviewed' and approved.

Category	Count
Open Pull Requests	33
Closed Pull Requests	214

• Git / GitHub: Pull Request

- 第 0 部分：打鸡血！
 - 通向自我实现之路
- 第 1 部分：编程开发环境与工作流
 - 基础运维技术
 - 容器化技术
 - 实战项目：Python / Jupyter 开发环境搭建 + 远程连接 VS Code
 - 实战项目：LALsuite / LISACode 的源码编译 (optional)
 - Git 分布式版本控制系统

- Python: 108 quizzes
- Numpy: 10 quizzes
- Pandas: 12 quizzes
- LeetCode: 5 problems
- Matplotlib: 4 datasets
- Seaborn: 4 datasets

- Credit Scoring dataset
 - Modeling
 - Finetune
- Kaggle competition
 - Can you find the GW signal?

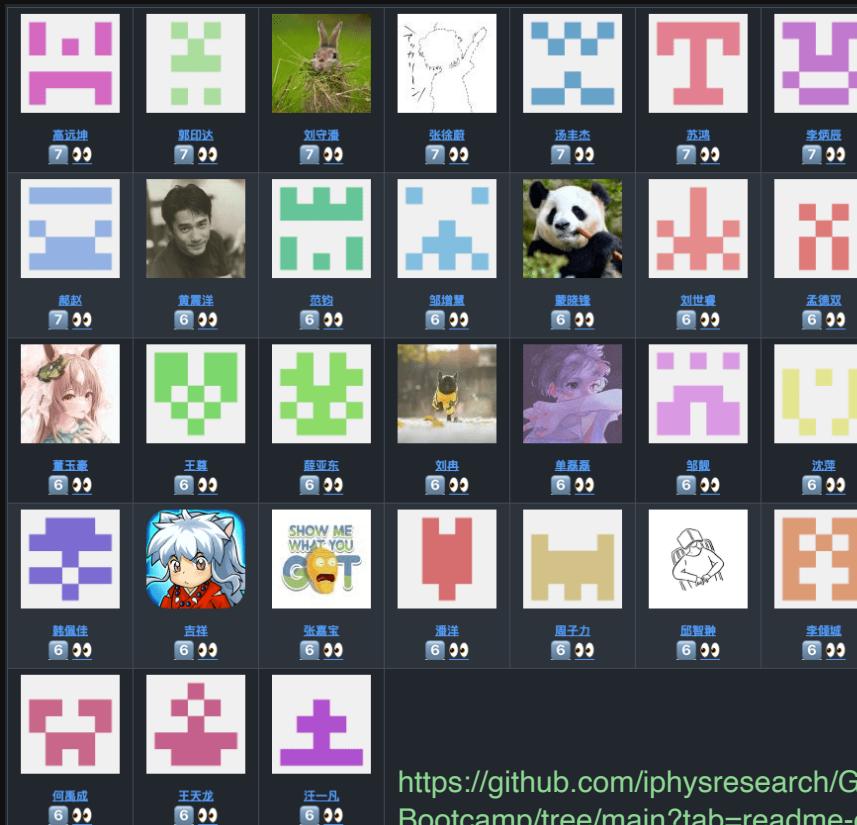
- 【公开课】数据技术演进与现实应用 (特邀嘉宾：田昕晓)
- 第 2 部分：基于 Python 的数据分析基础
 - 数据科学语言 Python 从入门到熟悉
 - 数据分析实训之 Numpy / Pandas
 - 实战项目：GW Event Catalog 的探索性数据分析
 - 实战项目：股票数据分析案例 (optional)
 - 基于 Python 的数据可视化理论与实践之 Matplotlib / Seaborn
 - 实战项目：GWTC 论文中的 Figures
 - 实战项目：针对 GW150914 信号处理与匹配滤波数据分析
- 【公开课】贝叶斯推断在引力波科学中的应用 (特邀嘉宾：赵俊杰)
- 第 3 部分：机器学习基础
 - 机器学习算法之应用起步
 - 机器学习算法之应用进阶
 - 实战项目：基于 LIGO 的 Glitch 元数据完成多分类任务
 - 实战项目：基于 LIGO 的 Glitch 时频图数据实现聚类分析
- 第 4 部分：深度学习基础
 - 深度学习技术概述与神经网络基础
 - 实战项目：训练一个3层神经网络（手撸版）
 - 卷积神经网络与引力波信号探测
 - 实战项目：使用 CNN 识别双黑洞系统引力波信号
 - Kaggle 数据科学竞赛 (黑客马拉松): Can you find the GW signals?
- 【公开课】AI发展全景与GPT前沿解析 (特邀嘉宾：高民权)



训练营课程回顾与总结

• 作业完成情况

总得分	1	2	3	4	5	6	7
频数	4	5	6	10	7	23	8
前#百分比排名	100.00%	93.65%	85.71%	76.19%	60.32%	49.21%	12.70%



部分获奖同学:





训练营课程回顾与总结

• 竞赛排名情况

Can you find the GW signals?

Gravitational Wave Data Exploration: A Practical Training in Programming and Analysis (2023)

- 概述
 - 欢迎参加“引力波数据探索：编程与分析实战训练营”系列课程的最后挑战 - “你能找到引力波信号吗？”Kaggle数据科学竞赛（黑客马拉松）！这个竞赛旨在应用你在整个课程中学到的知识和技能，重点关注引力波数据分析和研究。
- 任务目标
 - 本次竞赛的目标是开发一个能够准确识别引力波信号的模型。我们将提供一个包含噪声和引力波信号的数据集。你的任务是开发一个能够准确区分两者的模型。
- 时间线（7天）
 - 本竞赛将于北京时间 2023年12月29日22:00 开始，并于北京时间 2024年1月6日23:59 结束。请确保在截止日期前提交你的解决方案。

Rank	Team	Members	Score	Rank	Team	Members	Score
1	XAO	黄震洋 张徐蔚	0.86173	16	Yuanhao Zhang	张渊峰	0.83317
2	UCAS Li Jiahao	李嘉豪	0.86160	17	B4rRY_G	郭意扬 赖景祺	0.82729
3	UCAS_212x2	刘洋洋	0.86157	18	Shao dong zhao	赵少东	0.82723
4	sophiainshao	沈萍	0.86145	19	Sparkle79		0.82595
5	Haihao SHI	史海浩	0.85954	20	Shoupan Liu	刘守潘	0.82526
6	Yinda Guo	郭印达	0.85832	21	Capoo Cat	孙文博	0.82387
7	deslenlir	温怡馨	0.85753	22	Tian_Jun	田军	0.82204
8	MengXiaofeng-UCAS	董晓峰	0.85188	23	Zhao_Hao	郝赵	0.82047
9	1500! ! !	刘冉 刘世豪 王天龙	0.84868	24	douking	王尊	0.81871
10	Qinglin Yan	王霆湖 国庆琳	0.84530	25	JunFan	范钩	0.8141
11	Zhiqing Zhu	朱智清	0.84527	26	Phi267	秦戈宇	0.81304
12	knnbenn		0.84023	27	tastonlyjust		0.81038
13	HIAS	苏鸿 张景瑞 汤丰杰	0.84023	28	junda zhou	周均达	0.80846
14	HanPeijia	韩佩佳	0.84003	29	SCU_CTP	曹旭 高鸿飞 李志威	0.80705
15	Zenghui Zou	邹增慧 李倾城	0.83883	30	DESHUANGMeng	孟德双	0.80462



训练营课程回顾与总结

- 作为本次竞赛的冠军，XAO团队展现了非凡的实力。现在让我们有请这个团队的黄震洋同学作为代表，来分享一下他们背后的故事和解题策略。show time!

● 竞赛排名情况

<https://public.flourish.studio/visualisation/16361996/>

<https://flo.uri.sh/visualisation/16361996/embed>

https://github.com/iphysresearch/GWData-Bootcamp/blob/main/2023/deep_learning/baseline/plot_kaggle_ranking.ipynb

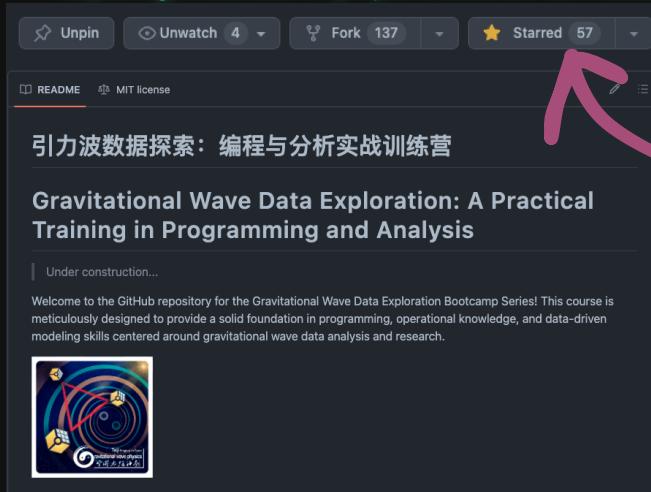
Rank	Team	Members	Score	Rank	Team	Members	Score
1	XAO	黄震洋 张徐蔚	0.86173	16	Yuanhao Zhang	张渊峰	0.83317
2	UCAS Li Jiahao	李嘉豪	0.86160	17	B4rRY_G	郭意扬 赖景祺	0.82729
3	UCAS_212x2	刘洋洋	0.86157	18	Shao dong zhao	赵少东	0.82723
4	sophiainshao	沈萍	0.86145	19	Sparkle79		0.82595
5	Haihao SHI	史海浩	0.85954	20	Shoupan Liu	刘守潘	0.82526
6	Yinda Guo	郭印达	0.85832	21	Capoo Cat	孙文博	0.82387
7	deslenlir	温怡馨	0.85753	22	Tian_Jun	田军	0.82204
8	MengXiaofeng-UCAS	董晓峰	0.85188	23	Zhao_Hao	郝赵	0.82047
9	1500! ! !	刘冉 刘世豪 王天龙	0.84868	24	douking	王尊	0.81871
10	Qinglin Yan	王霆湖 国庆琳	0.84530	25	JunFan	范钩	0.8141
11	Zhiqing Zhu	朱智清	0.84527	26	Phi267	秦戈宇	0.81304
12	knnbenn		0.84023	27	tastonlyjust		0.81038
13	HIAS	苏鸿 张景瑞 汤丰杰	0.84023	28	junda zhou	周均达	0.80846
14	HanPeijia	韩佩佳	0.84003	29	SCU_CTP	曹旭 高鸿飞 李志威	0.80705
15	Zenghui Zou	邹增慧 李倾城	0.83883	30	DESHUANGMeng	孟德双	0.80462



训练营后记

• 训练营课程网址

- <https://github.com/iphysresearch/GWData-Bootcamp>



• 欢迎各类反馈与建议:

- 太极计划引力波物理的最新发展
- 深入数理统计与数据分析的世界
- 全面了解PyTorch系统性课程
- 发现高效科研工具的秘密
- 掌握现代科研方法与学术写作的技巧
- 改善与同学和导师的人际关系
- 学术圈生存的实用指南
- 探索卷的动力学原理
- 寻求自我实现的途径
- 如何找到适合自己的人生方向
- 见到我，记得和我打招呼，期待交流与反馈

• 课程录屏已上线 Bilibili

- <https://space.bilibili.com/76060243/channel/seriesdetail?sid=3896245>

