

QG工作室2026寒假训练营培训内容

Tips

- 遇到问题一定要先自己想办法解决灵活运用搜索引擎
 - chatgpt: 学会如何提问, 如何使用, 但是不能依赖
- 建议每天记笔记并将笔记以及代码同步上传至github, 完成项目实战需要上传笔记文件和代码到github仓库
- 每日学习任务基本上分为基础学习与练习, 基础学习需确保全部完成, 而练习为加分项, 需上传笔记以及代码
- 具体学习进度可以根据自身进度安排
- 可以借鉴别人的代码, 但是一定要理解, 不能直接照抄
- 建议提前学习数据结构, 为开学后的考核做准备
- 如果你对[挖掘数据背后的惊人真相、破解人工智能的底层逻辑、挑战未知世界的复杂谜题]充满热情与兴趣, 并且具备一定的团队精神, 能够胜任团队工作, 那么QG AI组欢迎你的到来!

基本工具安装 (2.2-2.3)

Day1-2

体验AI工程师工作流程: [Orange Data Mining](#)

- 借助该平台了解AI工作流程、目的, 明确个人目标、未来方向
- 欢迎各位体验的同学与导师交流自己的游玩、使用体验!

找到以下工具的官方网站, 下载并安装正确版本;

也可以在群文件直接下载安装包

配置环境会很折磨人, 但配置环境是第一步, 不要被劝退!

- **anaconda, pycharm, jupyter的安装与使用**
 - anaconda: 官网下载
 - pycharm: 账号可以使用学生邮箱免费申请, 或者直接上淘宝购买
 - jupyter: 可以在anaconda下载, 轻量级的网页编程, 便于调试分析
- **markdown语法**
 - 编辑器安装 (推荐使用 typora)
 - 图床配置 (PicGo+阿里云OSS): 便于将图片添加到markdown文件中
- **学术资源加速工具**
 - 需科学上网, 有需要私聊导师
- **Git工具链**
 - 了解[什么是git](#), 访问并探索[GitHub](#), 欢迎各位同学和导师讨论你探索的优质开源项目!

1、基础任务

基础任务需确保全部完成

☐ 熟悉git流程

创建自己的GitHub账号，建立仓库用于存放寒假训练营的代码与笔记

☐ 学习基础markdown语法

在GitHub的仓库上传自己编写的markdown文件；

GitHub desktop 或 Git 的使用（都是上传文件到GitHub的工具，选择其中适合自己的一个熟练使用即可）

☐ 初识Python工具链

创建好python环境，在pycharm软件、Jupyter中输出"hello world!"

2、练习

- 了解如何利用git多人协作，探索GitHub项目，尝试提交issue、fork仓库等操作
- 利用markdown做一个简单、炫酷的个人简介上传到GitHub，从中学习markdown语法与git操作流程，欢迎与导师交流分享！
- 体验uv的工作效率与极致运行速度，体验前先安装：pip install uv，体验说明例如 `uv pip install numpy`

Python知识学习（2.4-2.7）

参考资料

- 基础学习类
 - [【尚学堂】2025最新版Python400集](#)
- 基于项目类
 - [基于项目的python学习](#)
 - [runoob网站python教程](#)
 - [Python - 100天从新手到大师](#)
- 学到会用的程度就行了

Day3（2.4）

基础学习，了解规范

1.基础学习

- python语言编码规范
 - [python语言规范中文网站](#)；希望所有同学严肃学习python语言规范！这不仅仅是审美问题，更是你未来团队协作、职业生涯的素养门面！Python之父 Guido van Rossum有言：“代码主要不是为了给人阅读的，只是顺便给机器执行。”
- 简单/复合数据类型
 - int float bool complex

- 通过代码对这些数据类型进行运算学习数据类型（字符串 列表 字典 元组 集合）

2.练习

- 情报解密

- 截获一条杂乱情报: `" Agent:007_Bond; Coords:(40,74); Items:gun,money,gun; Mission:2025-RESCUE-X "`
- 请运用所有所学知识清洗数据: 利用 **String** 方法去除干扰空格; 利用 **Set** 帮特工去除重复装备; 利用 **Slicing** 截取核心任务代号; 利用 **Tuple**锁定坐标; 最后将所有信息归档进一个 **Dict** 档案中

Day4 (2.5)

1.基础学习

- 流程控制 + 异常捕获

- 学习 `if-elif-else` 分支判断、`for/while` 循环结构、`try-except` 异常捕获机制

- 推导式 + 函数式编程

- 学习列表推导式、匿名函数 `lambda`、高阶函数 `map` `filter`
 - 尝试用**一行代码**（推导式）生成 1 到 10 的平方数列表
 - 尝试用 `map` 结合 `lambda` **给全班同学的名字加上统一前缀**（例如将 `["张三", "李四"]` 变为 `["QG_张三", "QG_李四"]`）

2.练习

- 能源核心数据清洗

- 数据: `raw_data = ["85", "92", "ERROR", "105", "78", "WARNING", "99", "120"]`。（可以自己设计数据）
- 实现自动跳过非数字项、仅保留 ≥ 80 数值、归一化为 `0.xx-1.xx` 小数, 且根据结果是否 > 1.0 对应报「核心过载」或输出「运转正常」

思考题: 是先清洗再过滤, 还是先过滤再清洗? 如何利用 `try-except` 配合函数式编程?

Day5 (2.6)

1.基础学习

- 面向对象

- 类与实例化 类变量 实例变量及类方法 私有变量和私有方法 封装 继承 多态

- 学习查阅官方文档

- 看代码示例, 先跑通再配合文档内容理解
- 注意文档细枝末节, 查阅 Python 相关工具 / 库的官方文档时, 需重点关注 API / 函数 / 类的**隐含行为、默认参数、数据处理规则**等易被忽略的细节

2.练习（面向对象实操）

基于“封装 + 继承 + 多态”设计「图形计算器」：

- 定义基类 Shape（私有属性 **area**，封装计算面积的私有方法 `calc_area`）
- 子类 Circle / Rectangle 继承 Shape，重写面积计算逻辑（多态体现）
- 类变量记录所有图形的创建数量，实例变量存储各自尺寸
- 要求：通过实例调用公开方法获取面积，禁止直接访问私有属性
- 查 Python 官方文档：确认 `str.strip()` 的默认参数、隐含行为（比如是否仅删空格？）
- 总结一条结论，再写一段文档相关代码及运行结果

Day6 实战（2.7）

- 设计面向对象思想实现图书库存管理系统，落地类变量、封装、继承、异常处理等需求

基本库学习（2.8-2.10）

Day7（2.8）

1.基础学习

- numpy库学习
 - 基于列表构建矩阵、特殊矩阵构建、矩阵乘法、矩阵广播机制、矩阵转置、矩阵的逆、矩阵存取
- [numpy官方文档](#)

2.练习：

- numpy
 - 创建一个形状为 (3,4) 的二维数组，元素为 0-11 的连续整数
 - 查看数组的 `shape`、`dtype`、`ndim` 属性
 - 将数组变形为 (4,3)，并展平为一维数组
 - 文档查阅：搜索 NumPy 官方文档「`np.reshape`」，确认 `reshape` 中 `-1` 的用法（如 `arr.reshape(4,-1)` 是什么含义）

Day8（2.9）

1.基础学习

- pandas库学习
 - 初步学习读取数据 简单的数据清洗
- matplotlib库学习
 - 学会基础的绘图语法,比如散点图、曲线图、条形图、扇形图等等，不同绘图之间差异较大，不建议广度地学习，不需要学太深，需要时再学习查阅即可，熟能生巧

2.练习:

- pandas
 - 处理 CSV 数据, 以加州房价为例, 使用pandas处理
 - [加州房价数据集](#)
- matplotlib
 - 配合 Matplotlib 画图, 用 NumPy 生成 x 轴数据, 绘制正弦曲线

Day9 实战 (2.10)

- 泰坦尼克号数据集可视化
 - 分别设计对各指标的生还数(率)进行对比可视化
 - 分析可视化结果
- 参考资料
 - [Titanic数据集](#)
 - [泰坦尼克号数据可视化分析](#)

线性代数与多元微分 (2.23-2.26)

参考资料

- [麻省理工学院—线性代数课 \(完整版72讲\)](#)

尝试去搞懂这些数学基础有什么作用

Day10 (2.23)

- 向量
 - 线性组合
 - 点积(内积)
- 矩阵乘法

Day11 (2.24)

- 矩阵转置以及对应的性质
- 矩阵的逆以及对应的性质
- 行列式

Day12-13 (2.25-2.26)

- 多元微分学
 - **偏导的概念** 微分 **方向导数和梯度** **多元函数极值与偏导的关系** 条件极值

视频快速过概念, 课本系统学推导, 用NumPy/linalg做小实验(如矩阵乘法、求逆、梯度计算)

- 复习熟练矩阵的内容

机器学习基础 (2.27-3.5)

参考资料

- [2025年公认最好的【吴恩达机器学习课程】](#)
- 因为原版是英文,中文字幕是机翻,所以有些概念可能会很抽象,有时间的话,建议看原汁原味的英文教材
- [李宏毅机器学习入门](#)

Day14 (2.27)

概念学习

1.基础学习

- **基础概念**
 - 什么是机器学习,了解机器学习与AI的关系
 - 什么是深度学习,了解一个基础的神经网络是如何构成的(输入层,隐藏层,输出层)
 - 了解数据集的划分——训练集、验证集、测试集,及其作用意义
- **神经网络**
 - 神经网络的架构是怎样的,其基本组成结构——**神经元**是什么
 - 网络的推理:什么是前向传播,了解神经网络是如何预测结果的
 - 网络的训练:什么是反向传播,了解神经网络是如何“学习”与“训练”的

Day15 (2.28)

更层次的了解,机器学习相关概念背后的原理与实现方法

1.基础学习

- **前向传播**
 - 即为网络的预测过程
 - 学习**激活函数**是什么,并了解主流的几个激活函数
 - 一个神经元是如何前向传播的
 - 由神经元组成的网络是如何前向传播的
- **反向传播**
 - 即为神经网络“学习”的过程
 - **损失函数**是什么
- **监督学习、无监督学习、半监督学习**

Day16 (3.1)

扩展内容与练习

1.基础学习

- 过拟合是什么，为什么会过拟合
- 数据归一化的方法及作用
- 正则化是什么
- 数据清洗
- 训练时的Batch Size，学习率

2.练习:

- 要求会推导计算一个神经网络前向传播的公式
- 在python中，使用之前学习过的numpy，分别用直接计算，和矩阵的方式，实现一个神经元
- 将上述神经元组成网络，实现前向传播

Day17 (3.2)

参考资料

- [5分钟深度学习 #01 梯度下降算法哔哩哔哩bilibili](#)
- [可能是全网最详细的线性回归原理讲解!!! 哔哩哔哩bilibili](#)
- 两个视频都要看完!不要只看5分钟那个

1.基础学习

- 多元线性回归梯度下降法求解
 - 梯度下降法
- 多元线性回归解析求解
 - 最小二乘法构建损失函数

2.练习

- 求矩阵与向量偏导并手推公式

Day18 实战 (3.3)

- 完成多元线性回归的类封装
 - 代码实现解析解求解和梯度下降法求解
 - **参考sklearn接口形式**不允许调用机器学习库，仅可使用numpy
 - 以波士顿房价数据集为例，划分训练集，测试集，并完成数据分析和预处理
 - 完成模型训练并予以评价分析

Day19-20 (3.4-3.5)

1.基础学习

- 学习K-means算法
 - 熟练掌握K-means算法的流程

2.练习

- 完成K-means的类封装
 - 参考sklearn接口形式不允许调用机器学习库，仅可使用numpy 以鸢尾花数据集为例，完成模型训练并予以评价分析结果

Day21 (3.6)

- 写一篇总结与体会
 - 收获
 - 寒假训练营遇到的问题以及解决方法
 - 遇到的好玩的事情分享

寒假训练营结束咯
