## Python的基础应用

### 1.1、PyCharm或VSCode的安装

PyCharm和VSCode都是跨平台的，可以在Windows、Linux、Mac上安装。

### 1.2、Python国内镜像库的使用

最常用的镜像，就是阿里和清华的镜像。镜像的地址在网上都可以找到。

可以大大的加快包的安装。

常用的命令：pip install 包名 -i 镜像链接

### 1.3、Cuda的安装

确保后续安装的Pytorch能使用Cuda进行训练和推理。Cuda的训练速度平均是cpu的50~100倍。

### 1.4、Pytorch的安装和应用

包含cpu版本和gpu（cuda）版本的pytorch的安装。

Cuda版本的Pytorch的安装尤为重要的，因为这个涉及训练和推理的速度。一般而言gpu版本的速度是cpu版本的100倍左右。

### 1.5、Python程序运行计时程序

至少实现一种Python的计时程序，精度到毫秒。

## 自动化编程工具的使用

（至少选择一种使用）

### 2.1、PyCharm+JetBrain\_AI\_assistant

### 2.2、PyCharm+文心快码

### 2.3、VSCode+FittenCode

### 2.4、VSCode+Cline

### 2.5、Qoder

### 2.6、Trace

### 2.7、Cursor

## 三、爬虫技术（数据准备）

（学习为主，注意抓取频率不要太高）

（主要结合“证券分析大模型”项目进行）

### 3.1、Mysql的安装

Mysql建议安装8.0版本、同时安装Navicat方便数据的增删改查。

### 3.2、爬虫技术基础

#### 3.2.1、requests的应用

写一个抓取百度图片，比如桥梁的200张照片的程序。

#### 3.2.2、scrapy框架

采用scrapy框架写一个抓取车票信息的小程序。

### 3.3、证券信息爬取

#### 3.3.1、交易信息爬取（结构化数据）

推荐网站：同花顺、东方财富、雪球、网易财经、新浪财经。

重点要抓取几种种数据，一种是日交易数据、一种是60秒一次的实时数据。

还有龙虎榜的数据，以及资金流入和流出的数据。

爬取的数据，可以先用Excel暂存。

#### 3.3.2、论坛信息的抓取（非结构化数据）

东方财富的论坛。

方便后续做舆情和情感倾向性的分析。

#### 3.3.3、上市公司财报的抓取（非结构化数据）

上市公司的财报抓取后，方便后续做长文本的分析。后续可以用智谱和kimi（月之暗面）作长文本分析。

### 3.4、消防数据爬取

#### 3.4.1、消防法规爬取（非结构化数据）

用于构建消防知识库（RAG）

### 3.5、数据存储

#### 3.5.1、结构化数据的存储

将爬虫爬取的数据插入到Mysql数据库中待后续分析用。

#### 3.5.2、非结构化数据的存储

非结构化数据存放到向量数据库当中。

## 四、深度学习的基础应用

### 4.1、全连接神经网络

用Python编写一个BP神经网络（全连接）实现Minst数据集的手写数字识别。

### 4.2、CNN神经网络

采用AlexNet编写一个神经网络实现对MINST的手写数字识别。

采用RestNet编写一个神经网络实现对猫狗数据集的识别。

### 4.3、RNN神经网络

编写一个的基于基于WikiText数据集的RNN语言建模实例。

### 4.4、Transformer

#### 4.4.1、情感分类

可以利用Transformer的知识，手搓一个NLP神经网络出来，用于情感分类，采用数据集IMDB。

#### 4.4.2、文本生成

基于PyTorch和HuggingFace Transformers库的Transformer实例程序，使用WikiText-2数据集实现文本生成任务。

#### 4.4.3、Vision Transformer(Vit)

使用Vit模型对图片进行分类（可使用HuggingFace Transformers库）。