

# Python Learning Resource

---

## 基础python学习

---

学好基础语法，data type

菜鸟教程（中文）：<https://www.runoob.com/python3/python3-tutorial.html>

w3schools (English): <https://www.w3schools.com/python/>

## Numpy

---

接下来要讲到的Pandas library是基于numpy的，语法有很多重合的地方，所以numpy非常值得学。

Numpy可以说是学python必学的一个package/library。主要用于各种数学计算和矩阵运算，具有非常高的速度。

Python是一种相对比较好写却相对慢的语言（已经很快了）。其中loop（迭代）是一个非常耗时间的操作。当数据量特别大的时候会特别慢。数学中，特别是线性代数的矩阵运算如果用纯python来写，就会用到很多这样的loop，非常耗时。Numpy是通过底层优化避免了loop的缺陷，加速数学计算。

所以说只要用通过python学数据科学，numpy是绝对避不开的。

Numpy有很多function/函数。有很多很复杂，光看并不一定能看懂，所以并不需要单纯的去学它。

如何学：在学pandas的时候，或者任何其他你需要用到数学和矩阵运算的时候，在网上搜numpy对应的用法。

你可能觉得矩阵运算，线性代数离你很远，跟数据处理关系不大，但是实际上数据就是一个一个table，每个table其实就是一个矩阵。每一个对table的处理都算是矩阵运算，大规模的改动，特别是跟数学有关的矩阵运算都应该尽量用numpy来完成。

官网教程（英文）：<https://numpy.org/devdocs/user/quickstart.html>

菜鸟教程（中文）：<https://www.runoob.com/numpy/numpy-tutorial.html>

搜到一个numpy function后可以看用法，官方文档里都会包含使用例子。

## python数据处理：Pandas（很有用，但是需要先学会python基础）

---

Pandas是一个python数据处理的package/library，非常强，感觉很类似R，跟SQL功能差不多，但本质上的目的是不同的。SQL语法能干的事它差不多都可以。SQL主要是作为一个真正的数据库来用的（高性能，高速度，大容量），主要用于存储信息，这些都不是pandas的工作，只是SQL的语法最终达到的数据处理目的和功能跟pandas是差不多的。

比如你想要从SQL数据库中提取一些你想要的data，并且写入一个文件供python进行下一步的分析。你有2种方法。

1. 用SQL语法可以选中一些table中符合你需求的信息，把选中的信息存为一个文件（e.g. csv）。用python读文件，进行下一步处理。

2. 可以把SQL数据库导出成文件，用python读取文件然后通过pandas处理和选中你想要的data。得到数据后进行下一步处理。

区别就在于选择信息是在哪里完成的（SQL or Pandas），没有孰优孰劣，只有哪种更方便更合适。

Pandas在我看来应该更为灵活，因为pandas是用python写的，做过处理的数据也可以随时用python做其他的处理，包括文件读写（file input/output）甚至是统计的分析（regression）和深度学习。而SQL并没有这么灵活，需要导出文件，再从python读入进行处理分析。

比如pandas可以在选中一些信息后，通过python对这些信息做一些比较复杂的处理，例如代入一些复杂的函数。SQL中也有类似的函数支持，但是应该是没有python强大，也没有直接在python中处理那么方便的。在处理data后把得到的data进行下一步处理。在SQL中，这样的工作是相对麻烦一点的。

所以说python非常重要，pandas也非常值得学习。

官网（英文）：<https://pandas.pydata.org>

官网（中文）：<https://www.pypandas.cn>

官网教程和文档（英文）：[https://pandas.pydata.org/docs/getting\\_started/index.html#getting-started](https://pandas.pydata.org/docs/getting_started/index.html#getting-started)

最好有一点python基础和对pandas的了解之后看（pandas常用知识点总结）：<https://zhuanlan.zhihu.com/p/99889912>

## SQL Learning Resource

SQL是一种服务器引擎。可以理解它是一个软件（引擎），装在一台服务器（电脑上），这个电脑的硬盘存了大量的数据（数据库），别的服务器（电脑）对这个数据库请求数据。但是不可能每次都要所有的数据，只要需要的一部分，所以SQL的语法就是SQL引擎用来从海量数据中找到你想要的数据的方法。你写出SQL query（请求：你想要什么信息），SQL引擎执行query，以极快的速度找到并返回你想要到信息。

菜鸟教程（中文）：<https://www.runoob.com/mysql/mysql-tutorial.html>

w3schools (English): <https://www.w3schools.com/sql/default.asp>

## 视频资源

个人经验而言，视频是我学习编程最好的方式，以上提到的网站一般是没有视频资源的时候学习，或者忘了一些细节，当作字典一样来查找想要的信息的。

一开始我建议你跟着视频学。如果一开始就看网站上的教程，可能不知道从哪开始或者看不懂。跟着视频开了头之后可以自己在网站上学。比如菜鸟教程上的教程都是比较简练和浓缩的。有一定基础后，看这个教程反而有可能可以学得比较快（如果看得懂的话）。看不懂还是去看视频一步步来。

YouTube是很好的资源，Python和SQL都有很多教程，质量非常好。但是如果在中国上不了YouTube，可以在Bilibili上找教程。Bilibili上有非常多非常好的教程，有些还是英文的。

具体看哪个视频没有特别的推荐，有的视频几个小时就讲完了，讲得很快但可能并不细致，有些视频讲几十个小时。你可以自己去搜视频，大概看一下喜欢哪个人的口音还有上课的节奏，然后坚持学。主要是看哪个适合你。如果是我学一个新的语言，因为我有类似的编程经验（逻辑上都是一样的），我就可以看速度比较快的教程节省时间。如果基础不好，跟不上的话就慢慢学，看哪个更合适。

基础python，SQL的教程有特别的，numpy，pandas的教程少一些，但是也有很多。在学好基础python后，numpy和pandas可以稍微看看视频然后自己去读网站上的教程反而可能会更快一些。

## 工具

---

学python的话推荐用pycharm，pycharm community版就够了。

但是VSCode是另一个我个人最喜欢的编辑器，支持几乎所有语言。

如果你学data science的话，推荐使用VSCode+Jupyter Notebook.

安装python的时候不用下载python原版，下载Anaconda，里面包含python和很多python的library。

Anaconda是专门给data science用的。

装好Anaconda后VSCode里就能打开Jupyter Notebook了。

Jupyter Notebook的好处是可以一段一段的run code，写一点就查一下当前的数据对不对，传统的python都是一次性跑完的。这样没什么不好的，反而方便使用debugger，但是对data science和machine learning来说所有人都是使用jupyter notebook的。

SQL的话用MySQL Workbench，若果你在电脑上装的引擎是MySQL的话。有很多不同的SQL，根据你选的教程来吧。

选一个好的教程后，第一步就是安装软件，安装的时候你可以问我。比如我见过菜鸟教程上安装软件搞得特别复杂，但是实际上没有特别复杂。