

# 安装文档

请大家上课前一天在自己电脑完成相关系统安装，并可以打开 **Jupyter Notebook** 执行基本程序。有问题随时网上咨询，在上课期间没有时间安装软件。

## 1 引言

《Python 数据分析与机器学习前沿技术》课程采用 Anaconda 环境 + Jupyter Notebook 工具来进行讲解学习。

很多人想到学习 Python 就要去官网下载一个 Python，其实并不需要，因为 Anaconda 本身就自带了 Python 的解释器而且还自带了很多实用性工具，所以**只需要安装 Anaconda 环境就可以了**。

**注意：**相关软件可以自行在官网下载，也可以在我们提供的百度网盘下载

链接：<https://pan.baidu.com/s/1USYZHWyIZtMDbuS1m2A-xQ> 提取码：6zvz

## 2 电脑配置要求

为保证培训效果及教学质量请学员自带笔记本电脑，笔记本电脑配置要求如下：

操作系统：Windows、Linux 或 Mac OS X (**一定要 64 位版本的操作系统**)，

软件环境：Anaconda3-5.2.0 以上(**一定要选择 Anaconda3 系列**)，包含 Python3、Conda、pip、Jupyter Notebook 等软件，利用 conda 或 pip 安装 selenium、tensorflow、Keras、jieba、gensim、snownlp、wordcloud 等。

**CPU：**i5 以上，

**内存：**4G 以上，

**空余硬盘空间：**40G 以上，

**注意：**不同的操作系统安装方式略有不同，Windows 系统请见“**win 系统环境安装**”，Linux 或 Mac 系统请见“**mac 系统环境安装**”。

## 3 win 系统环境安装

### 3.1 Anaconda 安装

Anaconda 安装包可以在其官网下载到:

官网: <https://repo.anaconda.com/archive/>

清华镜像: <https://mirrors.tuna.tsinghua.edu.cn/anaconda/archive/>

进入后选择 Anaconda3-2019.03 这个版本(Python3.7), 3 的意思就是 Python3 以上的版本, x86\_64 就是 64 位的, 如果是 Win7 系统请选择 Anaconda3-5.2.0 这个版本 (Python3.6)。

#### Anaconda installer archive

Filename	Size	Last Modified	MD5
<a href="#">Anaconda2-2019.03-Linux-ppc64le.sh</a>	291.3M	2019-04-04 16:00:36	c65edf84f63c64a876aabc704a090b97
<a href="#">Anaconda2-2019.03-Linux-x86_64.sh</a>	629.5M	2019-04-04 16:00:35	dd87c316e211891df8889c52d9167a5d
<a href="#">Anaconda2-2019.03-MacOSX-x86_64.pkg</a>	624.3M	2019-04-04 16:01:08	f45d327c921ec856da31494fb907b75b
<a href="#">Anaconda2-2019.03-MacOSX-x86_64.sh</a>	530.2M	2019-04-04 16:00:34	fc7f811d92e39c17c20fac1f43200043
<a href="#">Anaconda2-2019.03-Windows-x86.exe</a>	492.5M	2019-04-04 16:00:43	4b055a00f4f99352bd29db7a4f691f6e
<a href="#">Anaconda2-2019.03-Windows-x86_64.exe</a>	586.9M	2019-04-04 16:00:53	042809940fb2f60d979eac02fc4e6c82
<a href="#">Anaconda3-2019.03-Linux-ppc64le.sh</a>	314.5M	2019-04-04 16:00:58	510c8d6f10f2ffad0b185adbbdddf7f9
<a href="#">Anaconda3-2019.03-Linux-x86_64.sh</a>	654.1M	2019-04-04 16:00:31	43caea3d726779843f130a7fb2d380a2
<a href="#">Anaconda3-2019.03-MacOSX-x86_64.pkg</a>	637.4M	2019-04-04 16:00:33	c0c6fbbeb5c781c510ba7ee44a8d8efcb
<a href="#">Anaconda3-2019.03-MacOSX-x86_64.sh</a>	541.6M	2019-04-04 16:00:27	46709a416be6934a7fd5d02b021d2687
<a href="#">Anaconda3-2019.03-Windows-x86.exe</a>	545.7M	2019-04-04 16:00:28	f1f636e5d34d129b6b996ff54f4a05b1
<a href="#">Anaconda3-2019.03-Windows-x86_64.exe</a>	661.7M	2019-04-04 16:00:30	bfb4da8555ef5b1baa064ef3f0c7b582

系统的位数, 可以查看:

<https://jingyan.baidu.com/article/27fa73265ed13046f8271f19.html>

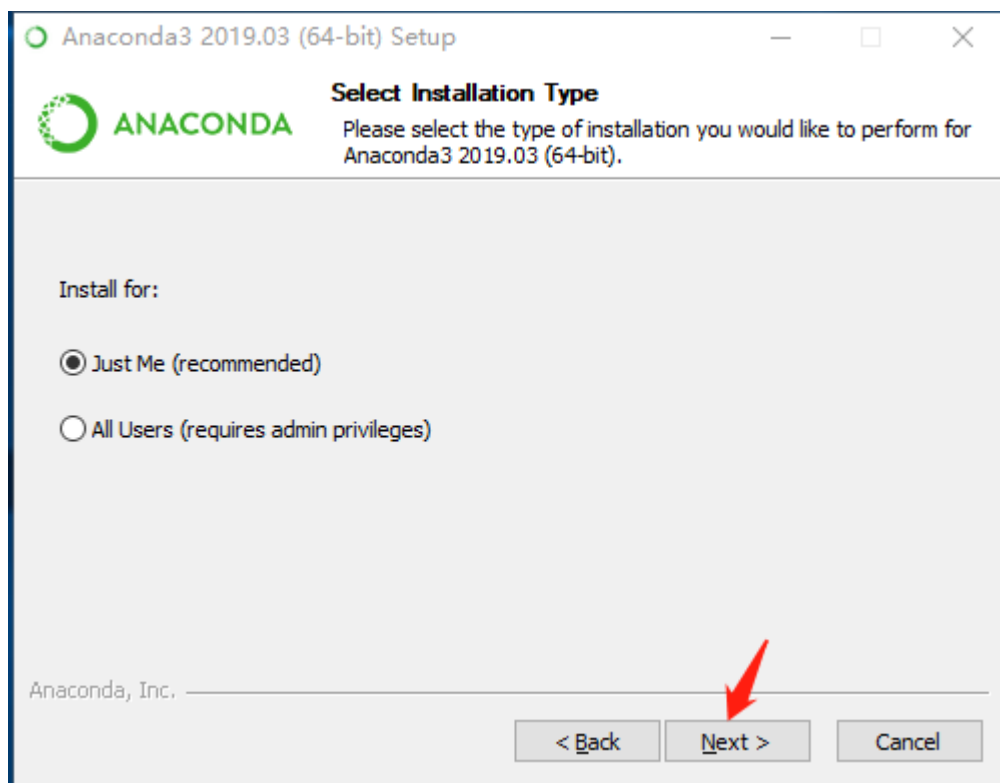
注意: win10 系统建议选择 **Anaconda3-2019.03** 安装, win7 系统必须选择 **Anaconda3-5.2.0** 安装。

#### 3.1.1 Win10 系统

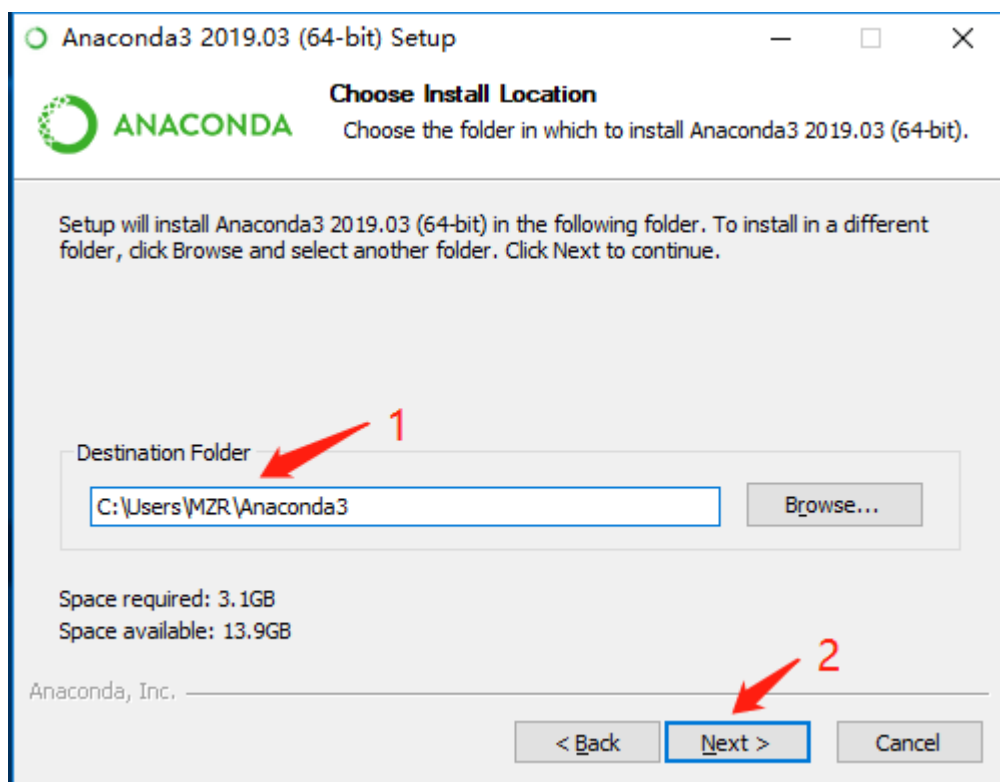
选择下载 Anaconda3-2019.03-Windows-x86\_64.exe, 下载完成后运行安装:



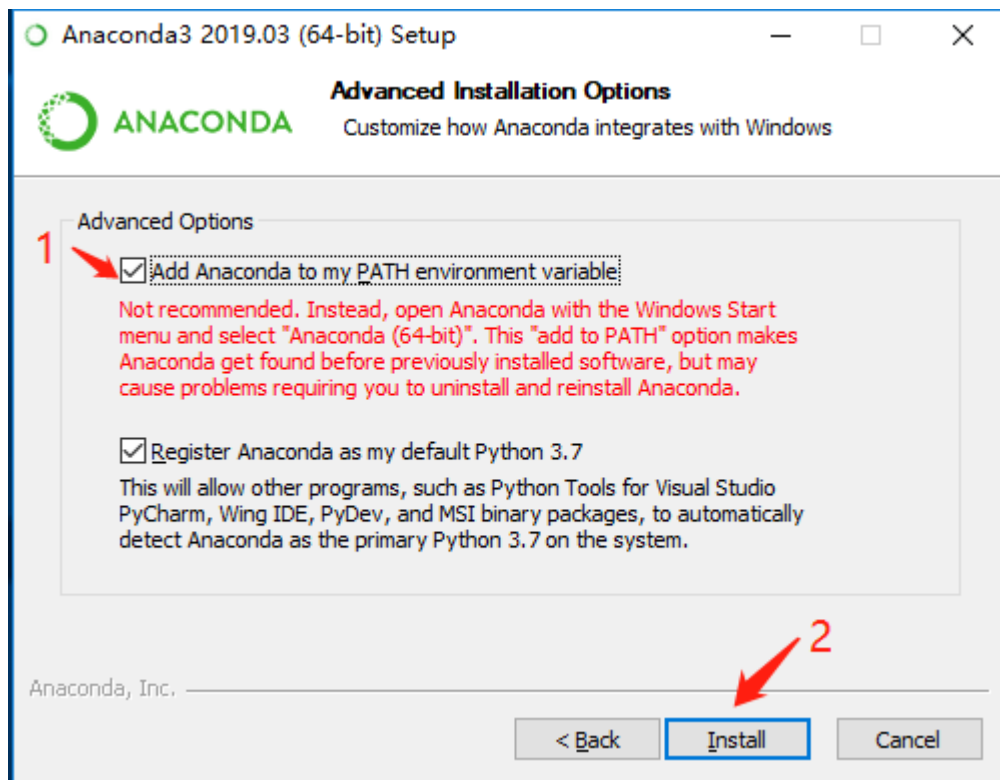
选择为所用用户安装或只是为自己安装。（直接下一步过去就行）



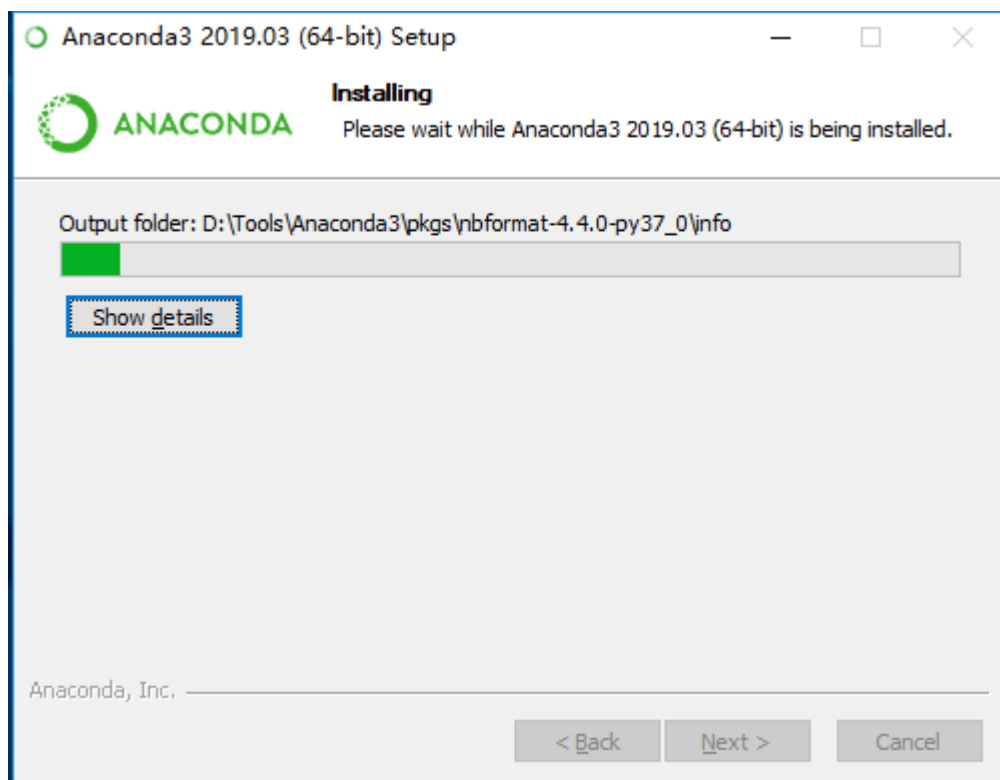
选择安装路径，默认安装在 C 盘就好（要保证有 10GB 以上的可用空间）



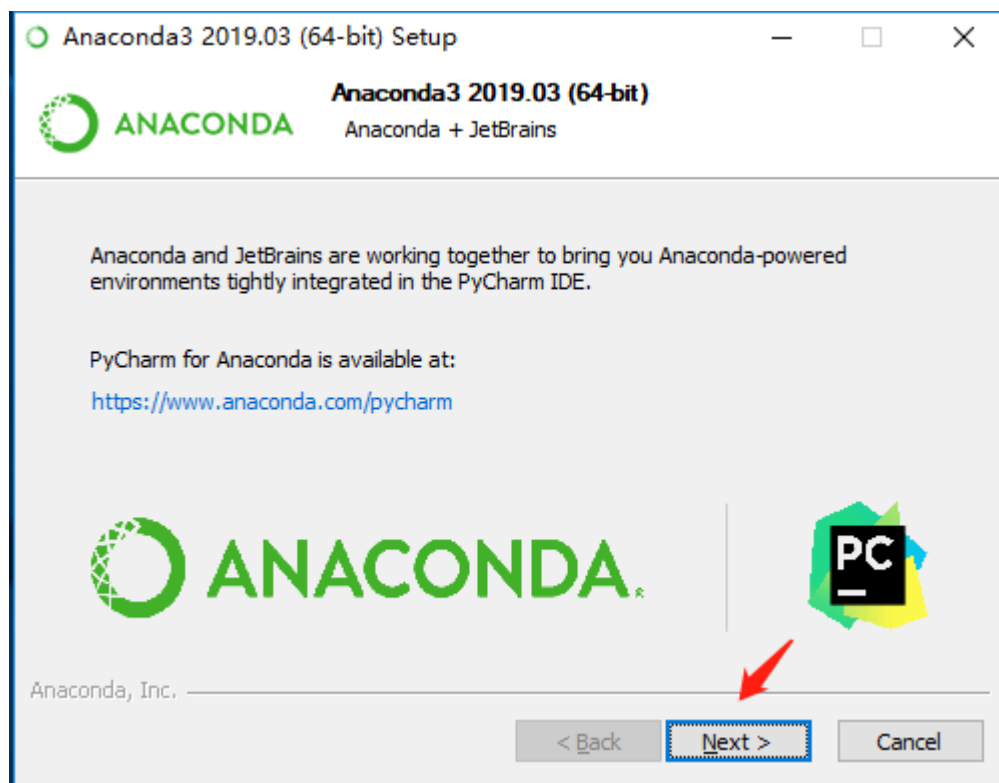
这里建议将第一个勾选（将 **Anaconda** 添加到环境变量中，不勾选的话后期要自己添加到环境变量）



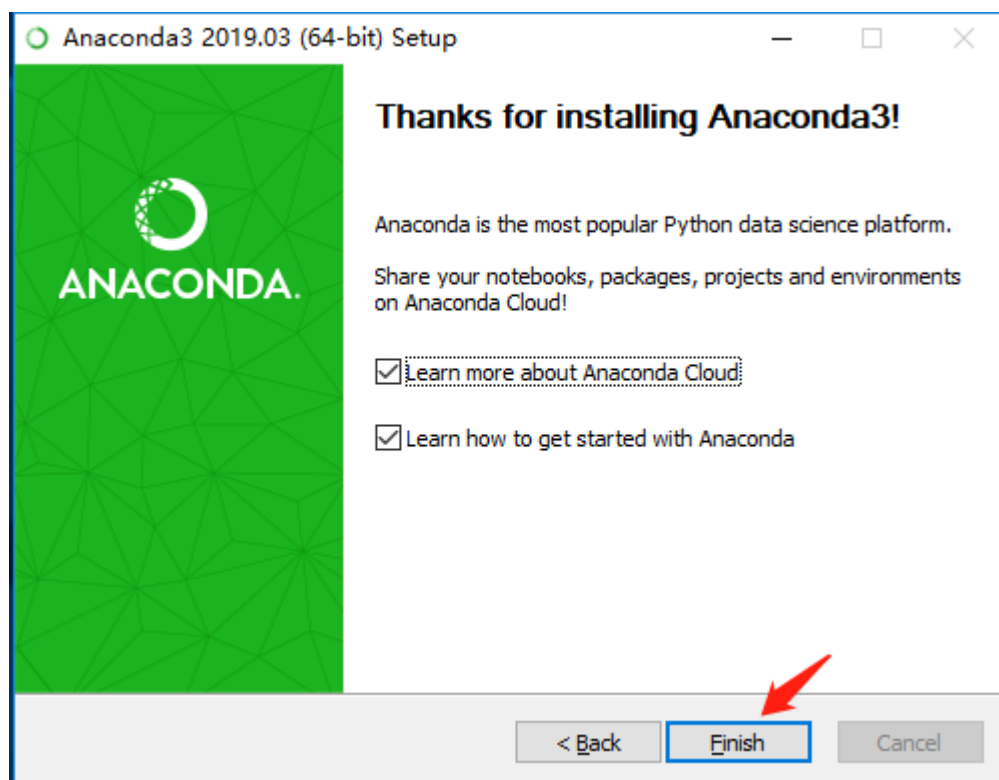
开始安装



安装完成，这里推荐安装 PyCharm，直接点 Next 跳过。



安装完成

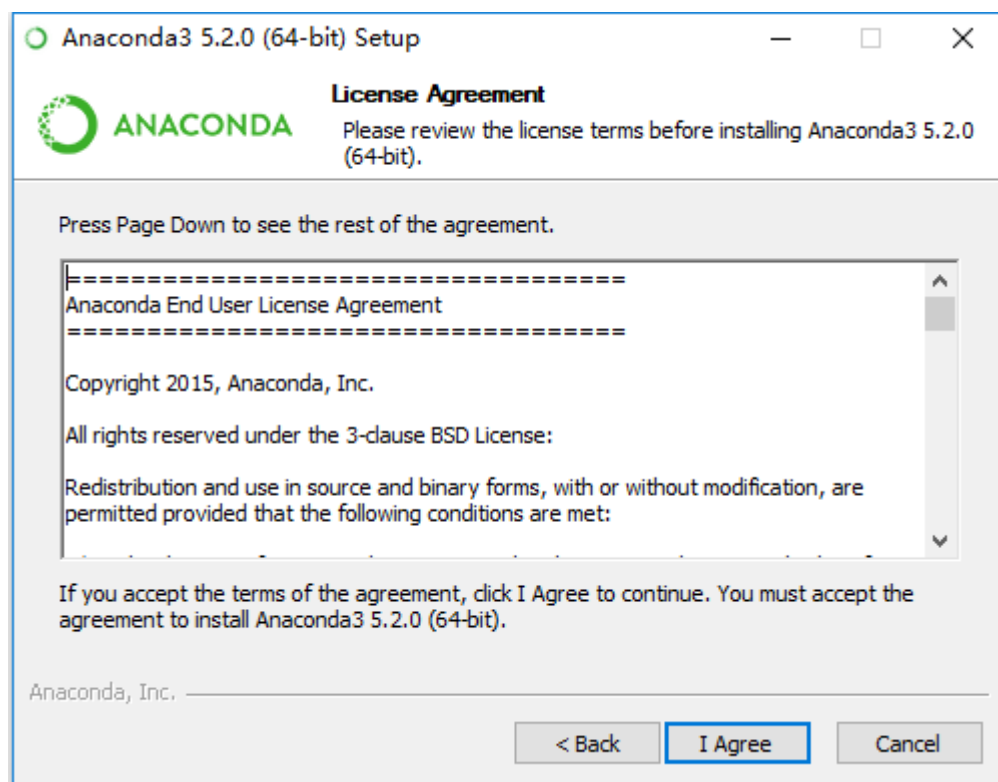


### 3.1.2 Win7 系统

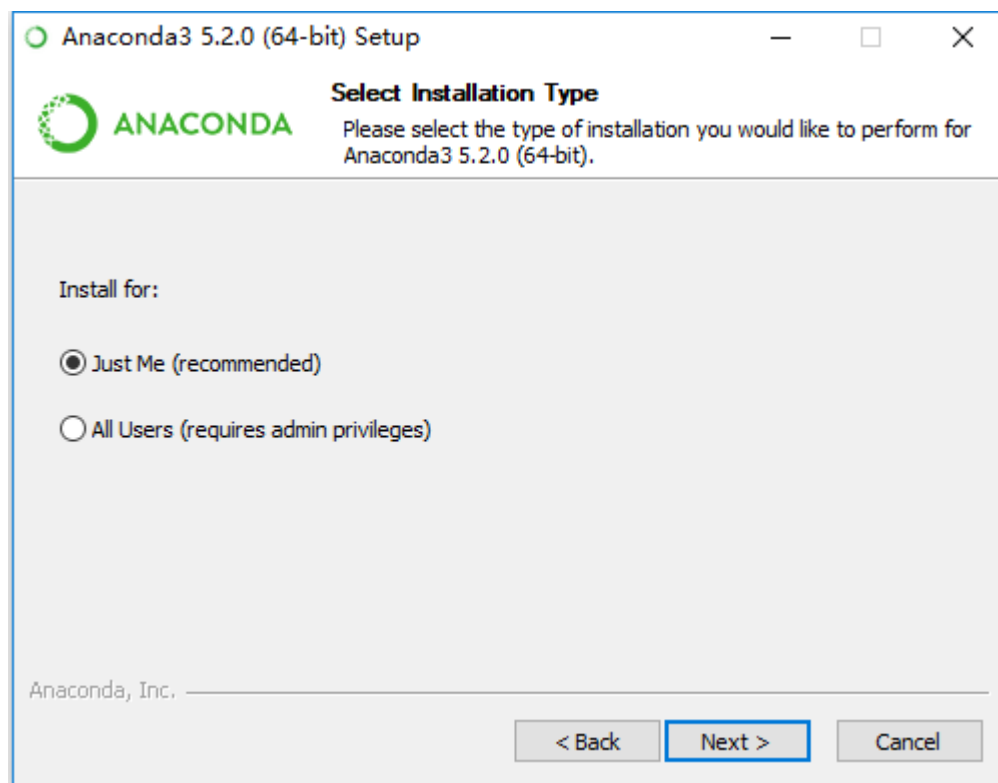
选择下载 Anaconda3-5.2.0-Windows-x86\_64.exe，下载完成后运行安装：



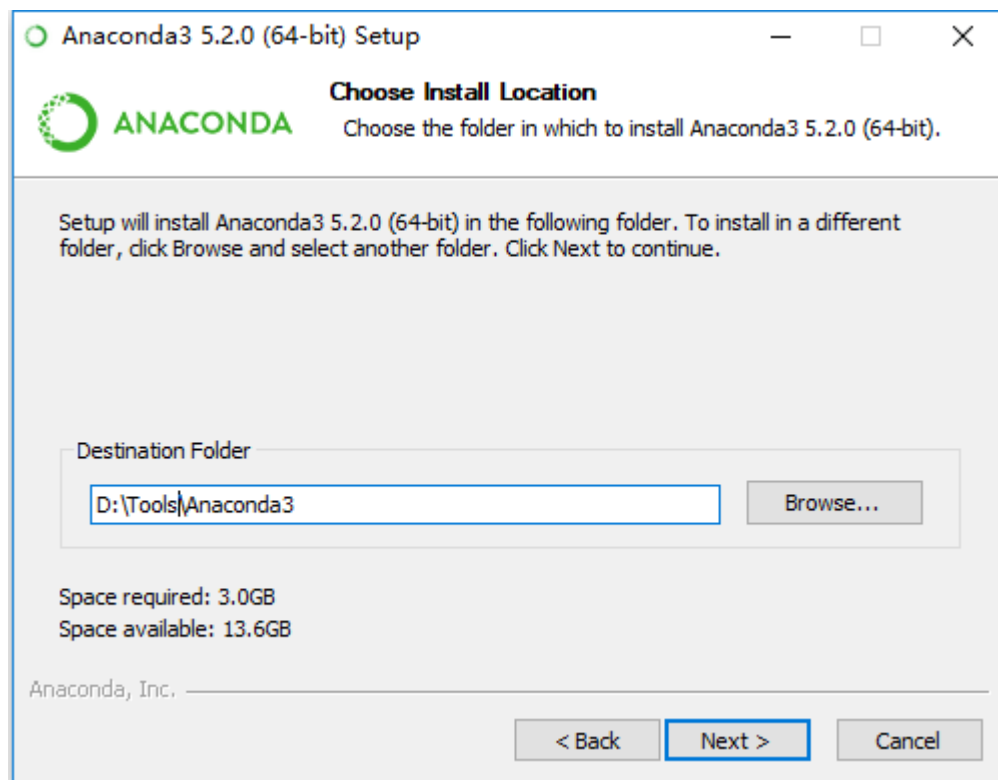
同意用户协议



选择为所用用户安装或只是为自己安装。（直接下一步过去就行）

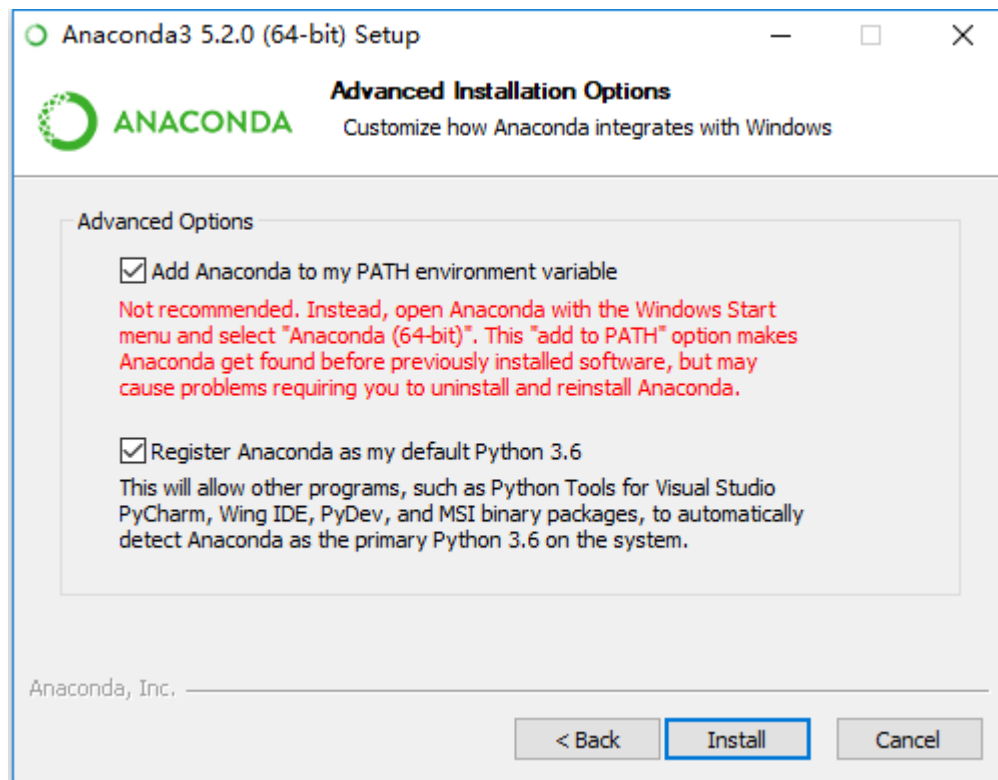


选择安装路径，默认安装在 C 盘就好（要保证有 10GB 以上的可用空间）

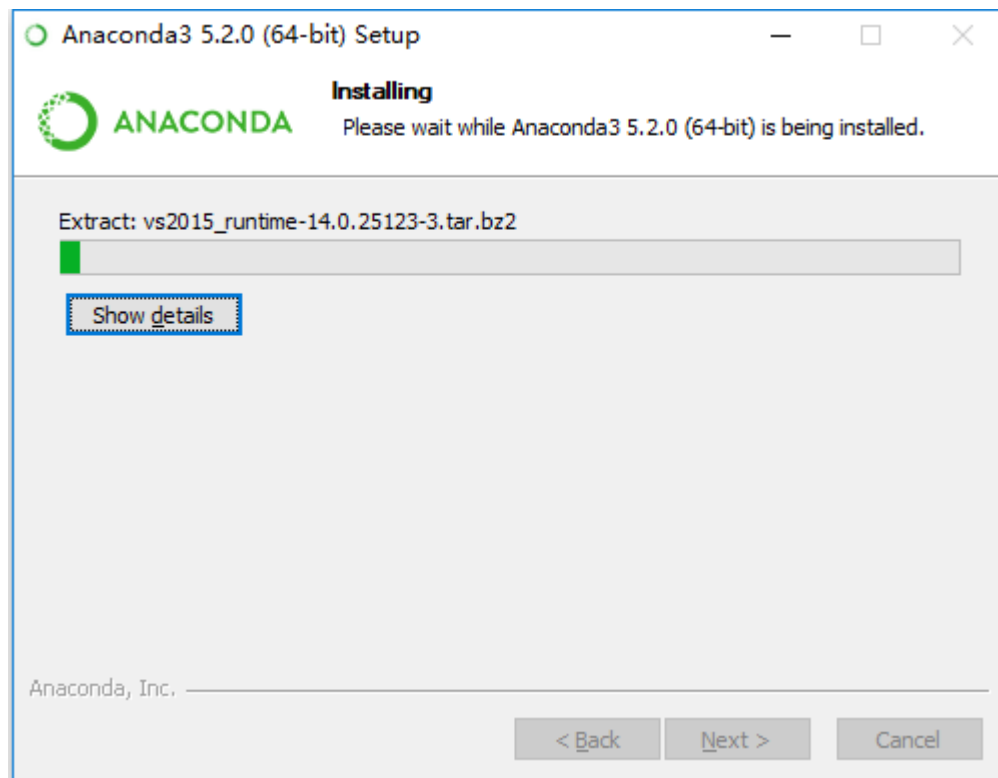




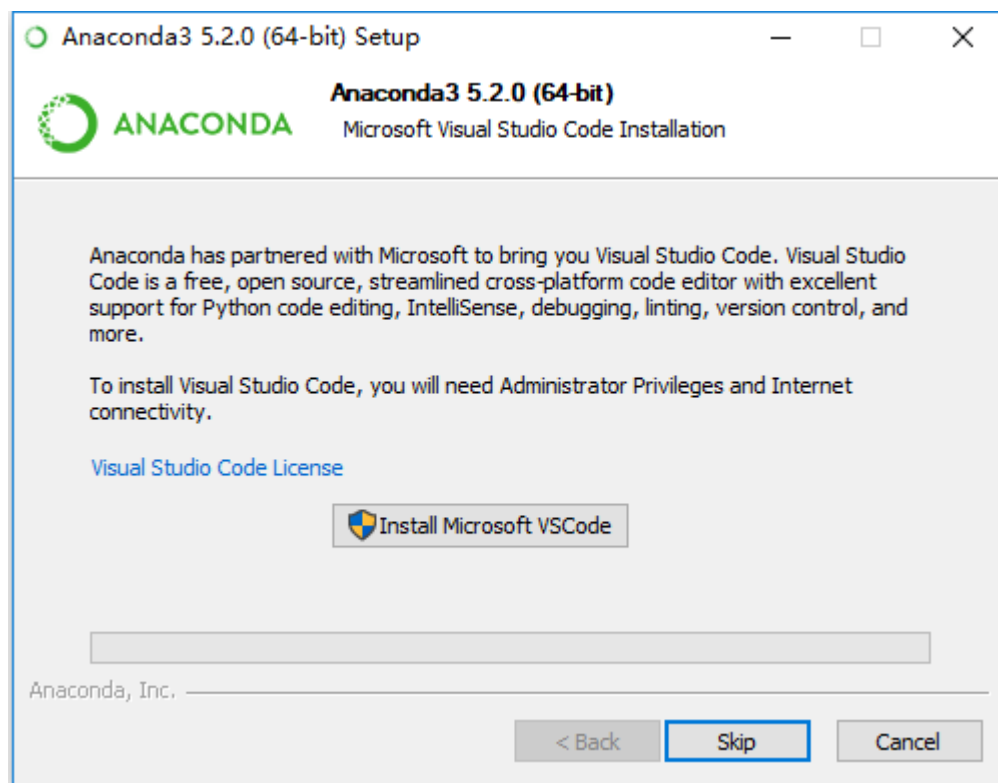
这里建议将第一个勾选（将 **Anaconda** 添加到环境变量中，不勾选的话后期要自己添加到环境变量）



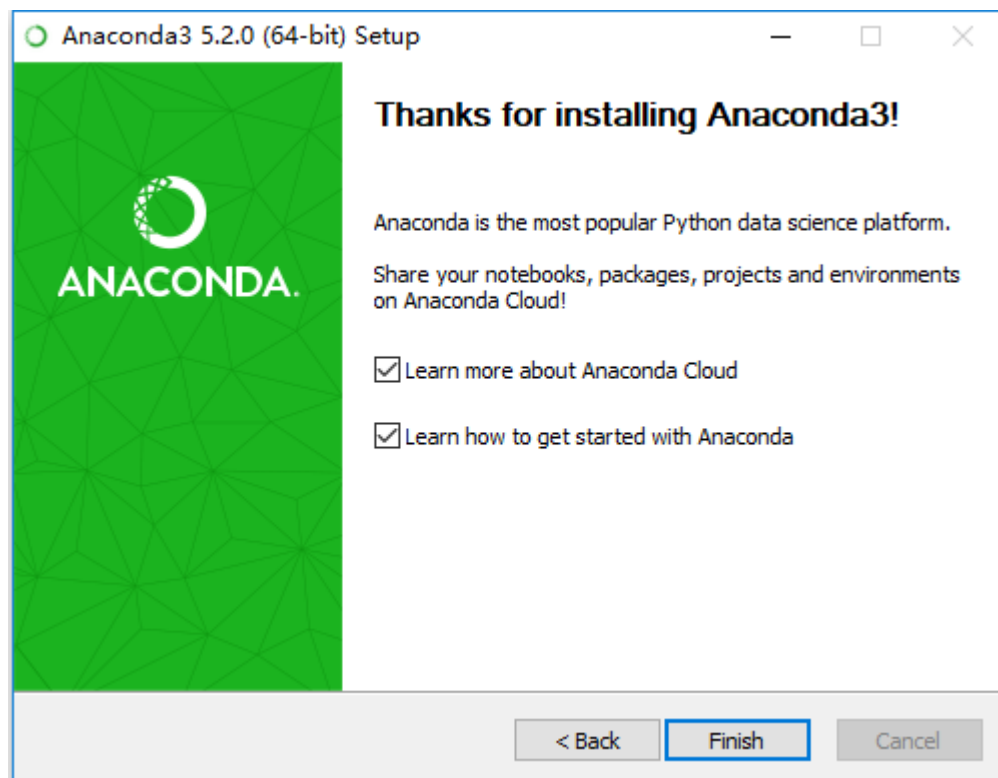
开始安装



安装完成，这里提示是否安装微软的 VSCode 直接 Skip 跳过。



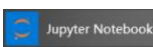


安装完成



安装完成后在开始菜单中就会有 anaconda 程序，如下图所示



不同版本安装后的菜单可能略有差别，但菜单中包含  Anaconda Navigator， Anaconda Prompt， Jupyter Notebook 这三项，表明安装成功。

## 3.2 工具包安装

本次课程需要的包如下：

第一天：numpy

第二天：pandas、matplotlib、seaborn

第三天：requests、beautifulsoup4、selenium

第四、五天：scikit-learn

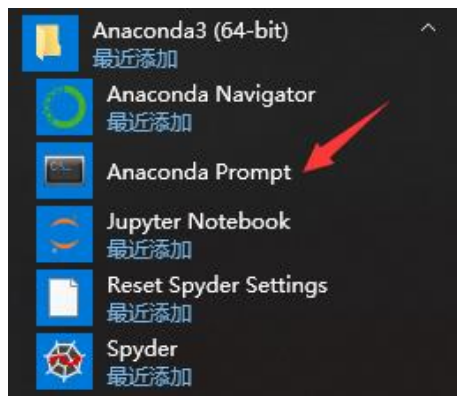
第六天：tensorflow、Keras

第七天：jieba、gensim、snownlp、wordcloud

其中，numpy、pandas、matplotlib、seaborn、requests、beautifulsoup4、scikit-learn 为 anaconda 自带的，只需安装其他包（标黄部分）。

具体安装命令如下，逐条在 Anaconda Prompt 命令行下执行。

建议在 Anaconda Prompt 命令行下安装第三方工具包，如下图



### 3.2.1 Win10 系统

```
pip install -i https://pypi.douban.com/simple selenium

pip install -i https://pypi.douban.com/simple tensorflow==1.15.0

pip install -i https://pypi.douban.com/simple Keras==2.3.1

pip install -i https://pypi.douban.com/simple jieba

pip install -i https://pypi.douban.com/simple gensim

pip install -i https://pypi.douban.com/simple snownlp

pip install -i https://pypi.douban.com/simple wordcloud
```

如果安装速度慢，可以通过设置镜像服务器安装。

格式：pip install -i 镜像服务器地址 包名

例如：pip install -i <https://pypi.douban.com/simple/> jieba

常用的镜像服务器地址

<https://pypi.douban.com/simple/> 豆瓣

<https://mirrors.aliyun.com/pypi/simple/> 阿里

<https://pypi.mirrors.ustc.edu.cn/simple/> 中国科学技术大学

<https://pypi.tuna.tsinghua.edu.cn/simple> 清华

镜像使用帮助详见：<https://mirrors.tuna.tsinghua.edu.cn/help/anaconda/>

**注意：安装 tensorflow、Keras 时，系统不同需要指定特定版本号。**

### 3.2.2 Win7 系统

```
pip install -i https://pypi.douban.com/simple selenium

pip install -i https://pypi.douban.com/simple tensorflow==1.13.1

pip install -i https://pypi.douban.com/simple Keras==2.2.4
```

```
pip install -i https://pypi.douban.com/simple jieba

pip install -i https://pypi.douban.com/simple gensim

pip install -i https://pypi.douban.com/simple snowlp

pip install -i https://pypi.douban.com/simple wordcloud

pip install -i https://pypi.douban.com/simple numpy==1.16.2
```

其中，win7 下 numpy 版本不能高于 1.16.2。

注意：包的安装过程需要联网，请检查网络环境并保持链接。

### 3.2.3 安装命令介绍

安装命令优先使用 `pip install`，如果不行再使用 `conda install`

例如安装 **selenium** 包，输入如下命令

```
pip install selenium
```

或者

```
conda install selenium
```

如果安装不成功，那么怎么卸载一个包呢，还以 **selenium** 包为例

```
pip uninstall selenium
```

或者

```
conda remove selenium
```

就行啦。

要查看当前环境中安装了哪些包可以用

```
pip list
```

或者

```
conda list
```

### 3.3 浏览器驱动下载

使用 Selenium 包爬取数据时，需要安装浏览器驱动。







建议使用 Firefox 或 Chrome 浏览器，二者选一种即可。

#### 3.3.1 Firefox 浏览器

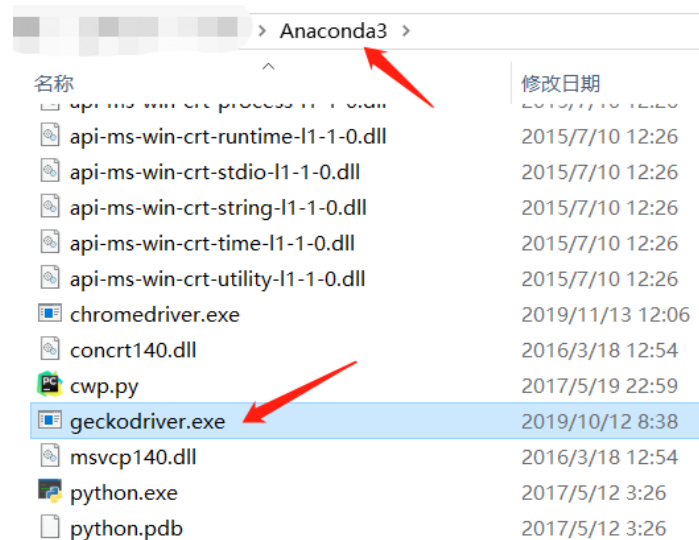
驱动地址：<https://github.com/mozilla/geckodriver/releases>

进入网页后，下拉找见驱动列表，选择相应的版本下载，如下图所示：

▼ Assets 7

 <a href="#">geckodriver-v0.26.0-linux32.tar.gz</a>	2.22 MB
 <a href="#">geckodriver-v0.26.0-linux64.tar.gz</a>	2.28 MB
 <a href="#">geckodriver-v0.26.0-macos.tar.gz</a>	1.91 MB
 <a href="#">geckodriver-v0.26.0-win32.zip</a>	1.37 MB
 <a href="#">geckodriver-v0.26.0-win64.zip</a>	1.46 MB
 <a href="#">Source code (zip)</a>	
 <a href="#">Source code (tar.gz)</a>	

以 win10 64 位系统为例，下载 `geckodriver-v0.26.0-win64.zip`，解压将 `geckodriver.exe` 放到 `anaconda` 安装目录下（与 `python.exe` 同目录）即可。



### 3.3.2 Chrome 浏览器

驱动地址 <http://npm.taobao.org/mirrors/chromedriver/>

首先，需要查看 chrome 浏览器的版本号，在浏览器地址栏输入 <chrome://version/>，79.0.3945.117 为浏览器版本号，如下图所示：

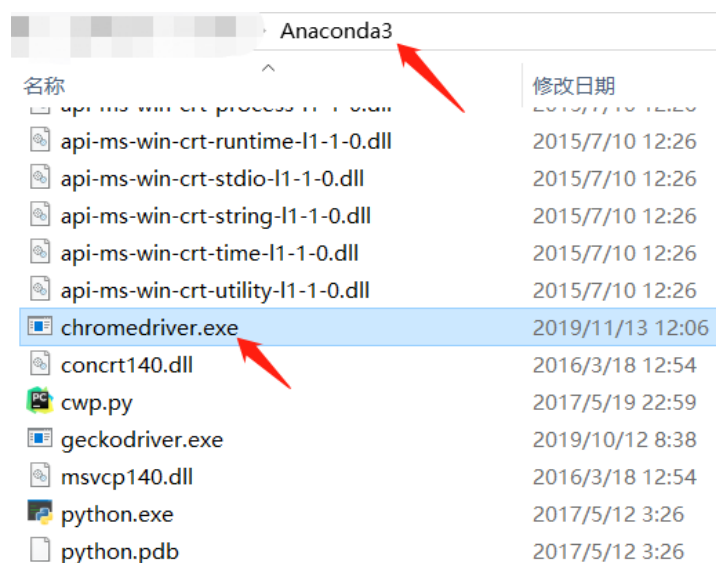


接着，win10 64 位系统为例，进入驱动页面选择大版本号相同（前两位相同）的驱动文件下载，如下图：

Mirror index of  
<http://chromedriver.storage.googleapis.com/79.0.3945.16/>

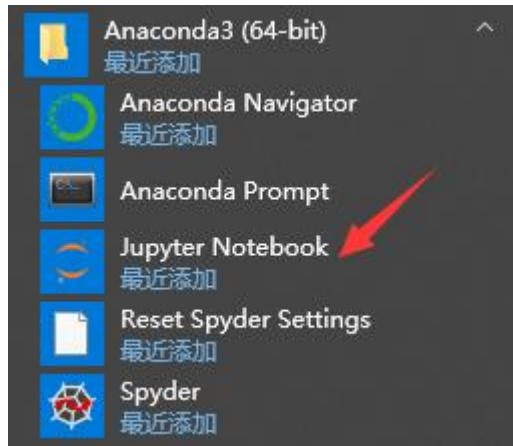
../	2019-10-30T16:10:56.644Z	4874192(4.65MB)
chromedriver_linux64.zip	2019-10-30T16:10:59.423Z	6907888(6.59MB)
chromedriver_mac64.zip	2019-10-30T16:11:00.927Z	4261268(4.06MB)
chromedriver_win32.zip	2019-10-30T16:11:04.204Z	930(930B)
notes.txt		

最后，下载 chromedriver\_win32.zip，解压将 chromedriver.exe 放到 anaconda 安装目录下（与 python.exe 同目录）即可。

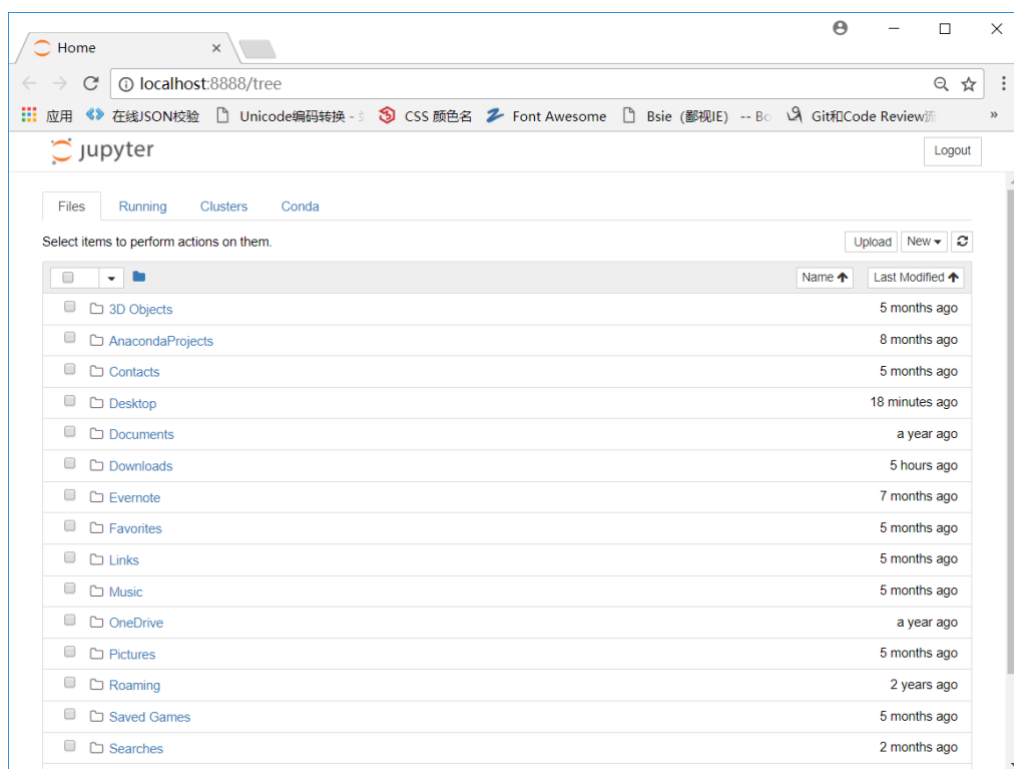


### 3.4 Jupyter notebook

安装 Jupyter 的最简单方法是使用 Anaconda。该发行版附带了 Jupyter notebook。你能够在默认环境下使用 notebook。



在如上图所示的菜单中打开 Jupyter Notebook，运行界面如下图所示



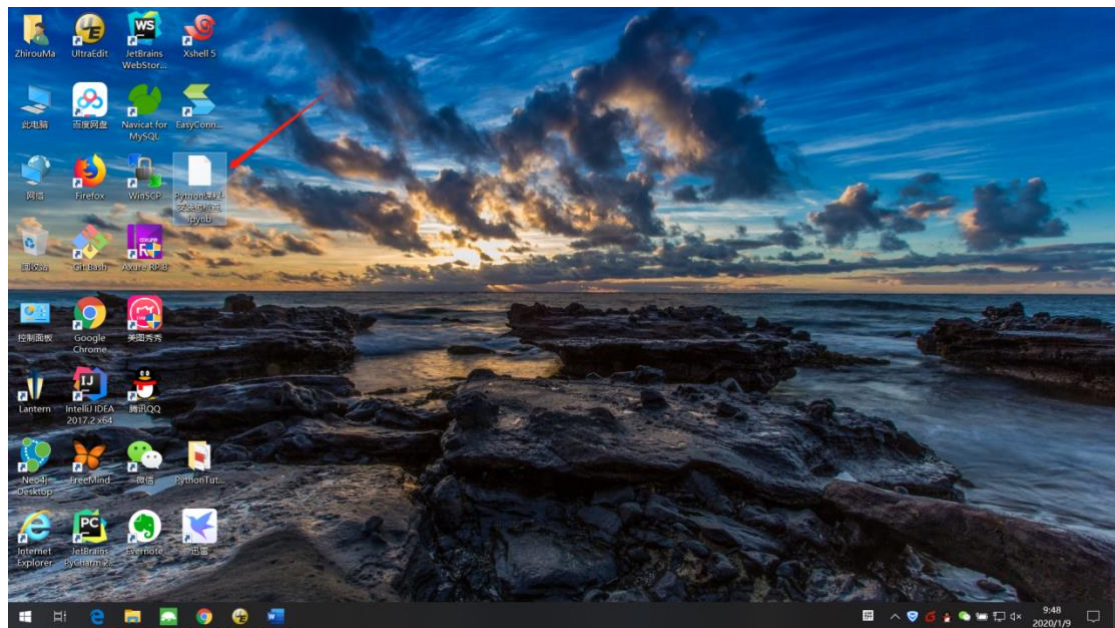
网页能启动成功，表明 Jupyter notebook 安装成功。

请大家确认一定要可以正常打开 notebook。

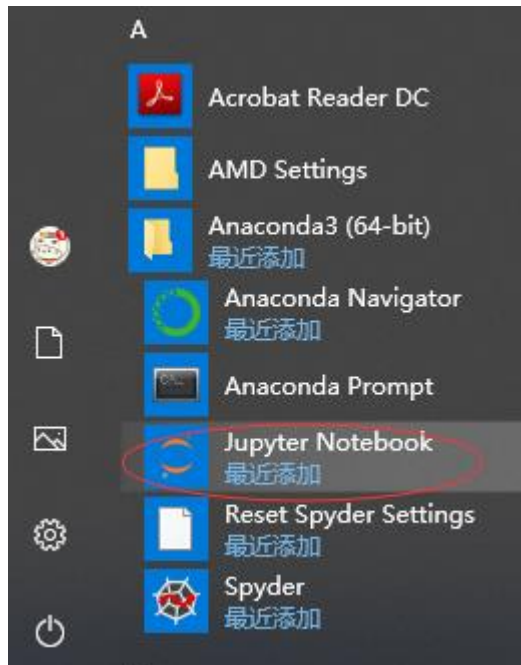


### 3.5 检查包是否安装成功

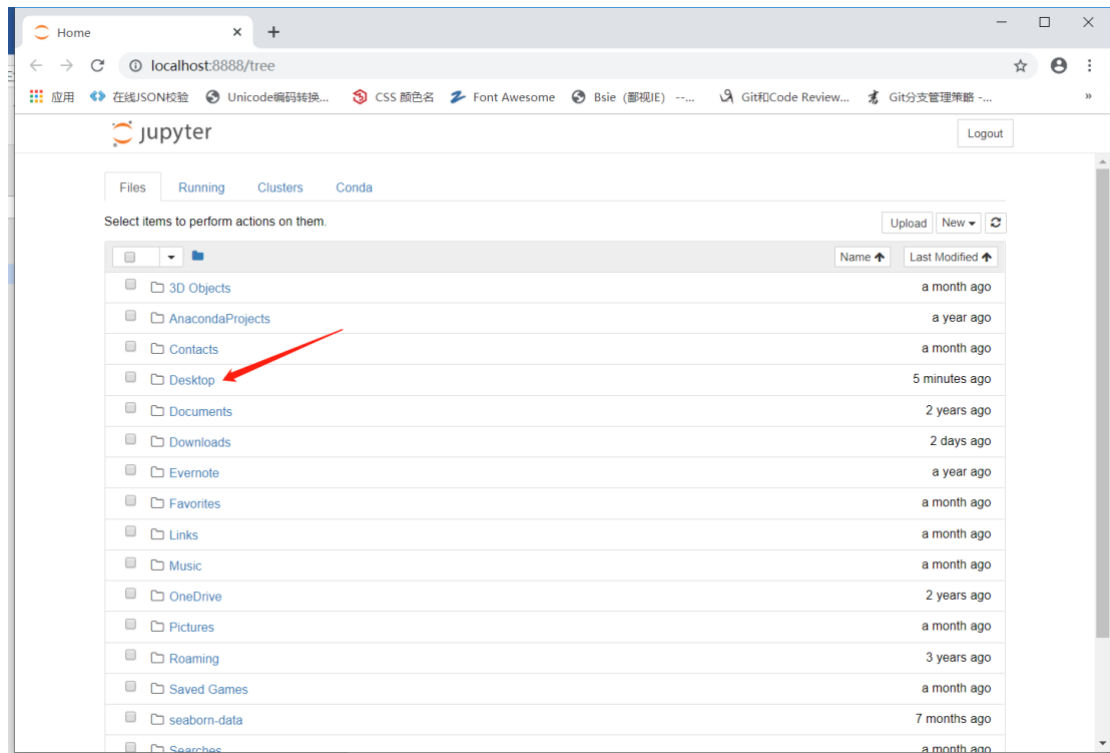
第一步：将 **Python 课程-安装包检查.ipynb** 文件放到桌面



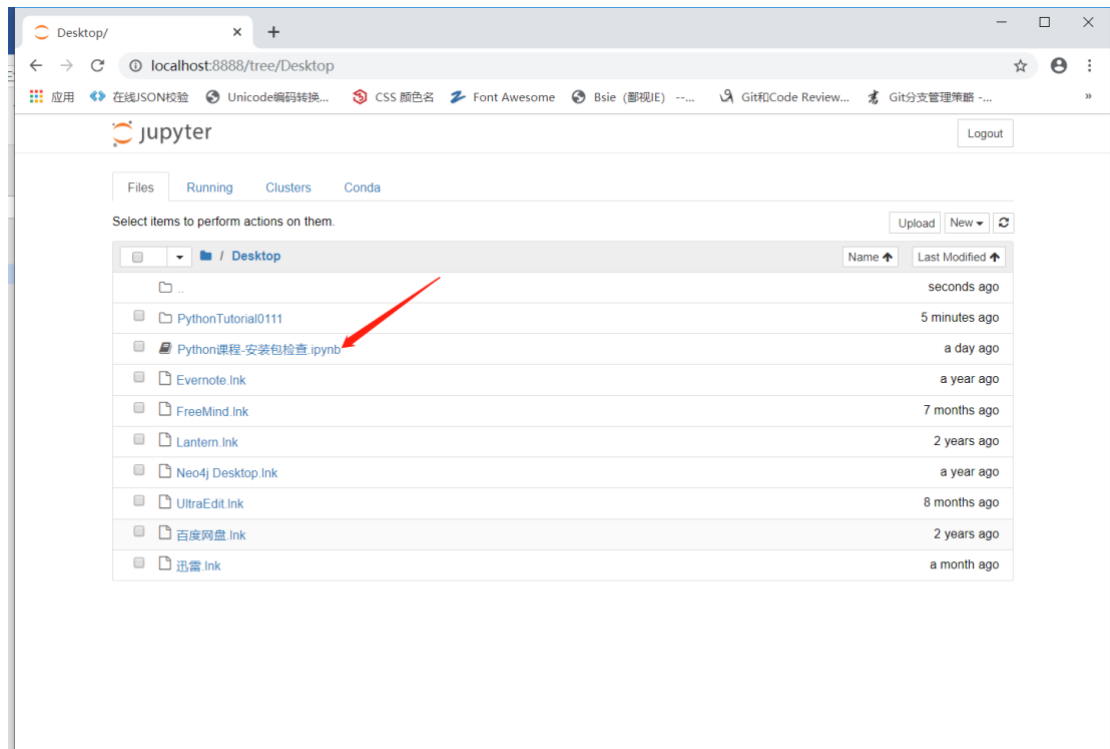
第二步：打开 Jupyter Notebook



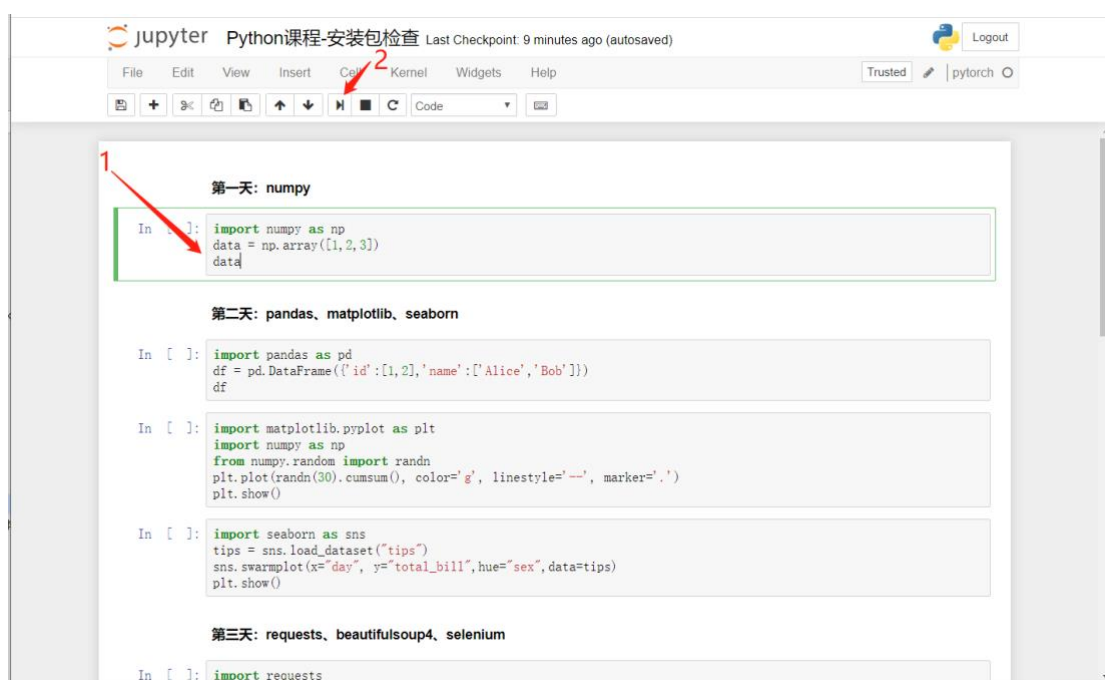
第三步：选择并打开 **Python 课程-安装包检查.ipynb** 文件



从文件夹【Desktop】进入，选择 **Python 课程-安装包检查.ipynb** 文件打开



第四步：选中待执行语句 cell，点击 run 按钮运行语句



图中 In[]表示程序的执行状态, In[\*]表示程序正在执行, In[1]表示程序执行成功, 中括号中的数字 1 表示执行次序; 执行成功, 如下图所示

### 3.5.1 第一天

第一天: numpy

```
In [1]: import numpy as np
data = np.array([1, 2, 3])
data
```

```
Out[1]: array([1, 2, 3])
```

## 3.5.2 第二天

第二天: pandas, matplotlib, seaborn

```
In [2]: import pandas as pd
df = pd.DataFrame({'id':[1,2], 'name':['Alice', 'Bob']})
df
```

```
Out[2]:
```

	id	name
0	1	Alice
1	2	Bob

```
In [3]: import matplotlib.pyplot as plt
import numpy as np
from numpy.random import randn
plt.plot(randn(30).cumsum(), color='g', linestyle='--', marker='.')
plt.show()
```

<Figure size 640x480 with 1 Axes>

```
In [4]: import seaborn as sns
x = np.arange(8)
y = np.array([1, 5, 3, 6, 2, 4, 5, 6])
df = pd.DataFrame({"x-axis": x, "y-axis": y})

sns.barplot("x-axis", "y-axis", palette="RdBu_r", data=df)
plt.xticks(rotation=90)
plt.show()
```



能画出图（不一定一样的图），即安装成功。

## 3.5.3 第三天

第三天: requests, beautifulsoup4, selenium

```
In [5]: import requests
response = requests.get("https://baidu.com")
response
```

```
Out[5]: <Response [200]>
```

```
In [6]: from bs4 import BeautifulSoup
html_doc = "<html><head><title>The Dormouse's story</title></head></html>"
soup = BeautifulSoup(html_doc)
```

```
Out[6]: <html><head><title>The Dormouse's story</title></head></html>
```

```
In [7]: from selenium import webdriver
browser = webdriver.Chrome() # 如果是火狐浏览器, 请将此语句换为 browser = webdriver.Firefox()
browser.quit()
```

运行后会打开一个浏览器，然后再关闭，即安装成功。

**注意：**如果是火狐浏览器，请语句换为 `browser = webdriver.Firefox()`

**常见错误：**浏览器驱动版本不对、位置不对，详见 3.3 节。

可以直接将 **mnist.npz** 和 **imdb.npz** 两个文件拷贝到用户目录下的 **.keras\datasets** 文件夹下

此电脑 > 系统 (C:) > 用户 > ZhirouMa > .keras > datasets				
名称	修改日期	类型	大小	
cifar-10-batches-py	2009/6/5 4:47	文件夹		
cifar-100-python	2010/2/20 8:16	文件夹		
cifar-10-batches-py.tar.gz	2019/5/30 18:51	WinRAR 压缩文件	166,503 KB	
cifar-100-python.tar.gz	2019/8/22 19:49	WinRAR 压缩文件	165,041 KB	
imdb.npz	2019/5/24 8:50	NPZ 文件	17,056 KB	
mnist.npz	2019/5/6 20:17	NPZ 文件	11,222 KB	

常见错误 2: Tensorflow 与 Keras 版本不对应。

一般情况下, Anaconda3-2019.03 (Python3.7) 版本下, Tensorflow 为 1.15.0、Keras 为 2.3.1; Anaconda3-5.2.0 (Python3.6) 版本下, Tensorflow 为 1.13.1、Keras 为 2.2.4。

Tensorflow 与 Keras 版本对应关系见下面网站 ,  
<https://docs.floydhub.com/guides/environments/>

Framework	Env name (–env parameter)	Description	Docker Image	Packages and Nvidia Settings
TensorFlow 1.14	tensorflow-1.14	TensorFlow 1.14.0 + Keras 2.2.5 on Python 3.6.	floydhub/tensorflow	TensorFlow-1.14
TensorFlow 1.13	tensorflow-1.13	TensorFlow 1.13.0 + Keras 2.2.4 on Python 3.6.	floydhub/tensorflow	TensorFlow-1.13
TensorFlow 1.12	tensorflow-1.12	TensorFlow 1.12.0 + Keras 2.2.4 on Python 3.6.	floydhub/tensorflow	TensorFlow-1.12
	tensorflow-1.12:py2	TensorFlow 1.12.0 + Keras 2.2.4 on Python 2.	floydhub/tensorflow	
TensorFlow 1.11	tensorflow-1.11	TensorFlow 1.11.0 + Keras 2.2.4 on Python 3.6.	floydhub/tensorflow	TensorFlow-1.11
	tensorflow-1.11:py2	TensorFlow 1.11.0 + Keras 2.2.4 on Python 2.	floydhub/tensorflow	
TensorFlow 1.10	tensorflow-1.10	TensorFlow 1.10.0 + Keras 2.2.0 on Python 3.6.	floydhub/tensorflow	TensorFlow-1.10
	tensorflow-1.10:py2	TensorFlow 1.10.0 + Keras 2.2.0 on Python 2.	floydhub/tensorflow	

### 3.5.6 第七天

第七天: jieba, gensim, snownlp, wordcloud

```
In [11]: import jieba
         from gensim import models
         from snownlp import SnowNLP
         import wordcloud

D:\Tools\Anaconda3\envs\pytorch\lib\site-packages\gensim\utils.py:862: UserWarning: detected Windows; aliasing c
hunkize to chunkize_serial
warnings.warn("detected Windows; aliasing chunkize to chunkize_serial")
```

运行不报错，即安装成功。

注意：粉色信息为提示信息或警告信息，不影响使用。

## 4 mac 系统环境安装

### 4.1 Anaconda 安装

Anaconda 安装包可以在以下两个地方下载到：

官网：<https://repo.anaconda.com/archive/>

清华镜像：<https://mirrors.tuna.tsinghua.edu.cn/anaconda/archive/>

下载适用的安装包，Anaconda3-5.2.0-MacOSX-x86\_64.pkg 或 Anaconda3-5.2.0-MacOSX-x86\_64.sh，进行安装

其他操作同 Windows 下类似，详见 3.2-3.5 节。

mac 系统下.keras\datasets 文件夹位置在~/.keras/datasets

## 5 其他工具

### 5.1 代码自动补全插件（选择安装）

jupyter notebook 代码自动补全插件安装，可以参考这个网页的步骤安装  
<https://blog.csdn.net/mengfei2656/article/details/89287140>

安装完以后，在 Notebook 写代码时，如果一个对象有函数可以调用，会自动提示函数名，还会提示变量名。

注意：不要在 360 浏览器上使用插件。

提示：由于防火墙或杀毒软件等原因，可能会安装不成功。该插件不是课程必须的软件包，根据个人情况选择安装。