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。**

# 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

# 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**”。

# win系统环境安装

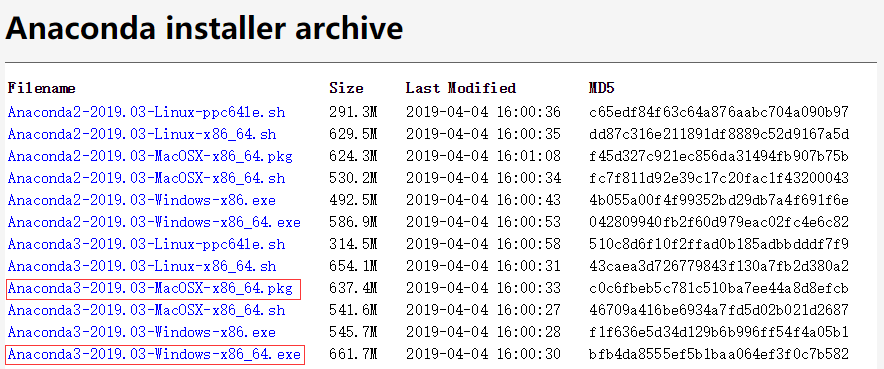
## Anaconda安装

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）。

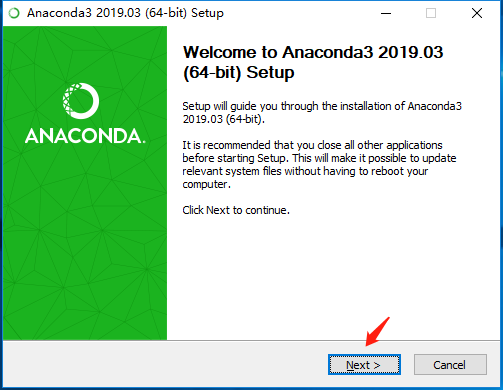


系统的位数，可以查看：<https://jingyan.baidu.com/article/27fa73265ed13046f8271f19.html>

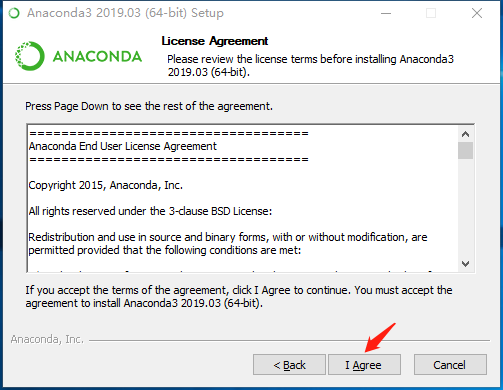
注意：win10系统建议选择**Anaconda3-2019.03**安装，win7系统必须选择**Anaconda3-5.2.0**安装。

### Win10系统

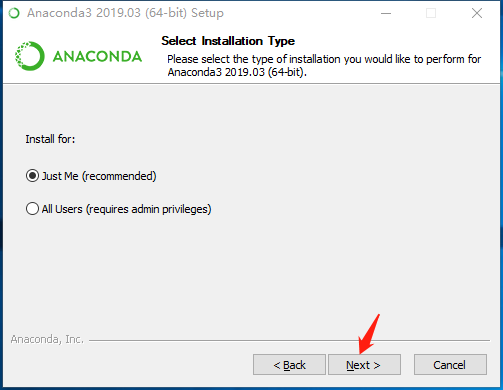
选择下载Anaconda3-2019.03-Windows-x86\_64.exe，下载完成后运行安装：



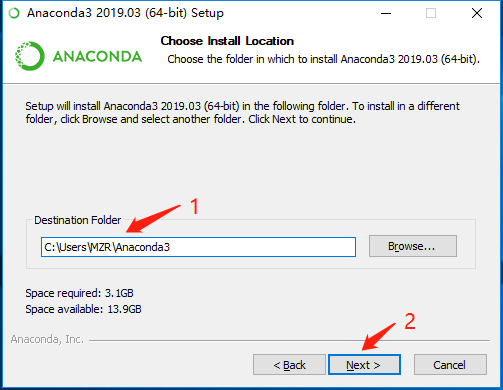
同意用户协议



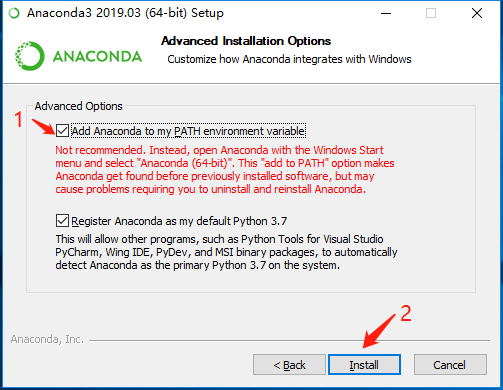
选择为所用用户安装或只是为自己安装。（直接下一步过去就行）



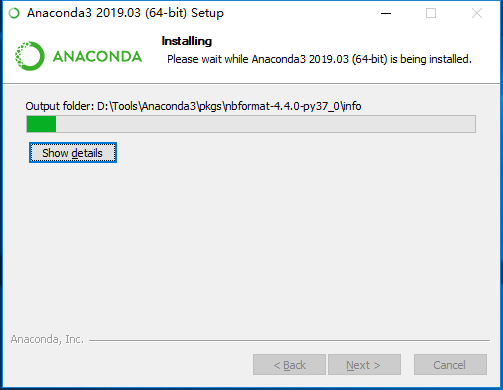
选择安装路径，默认安装在C盘就好（要保证有10GB以上的可用空间）



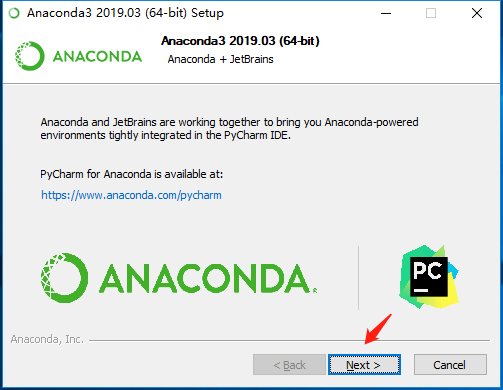
**这里建议将第一个勾选（将Anaconda添加到环境变量中，不勾选的话后期要自己添加到环境变量）**



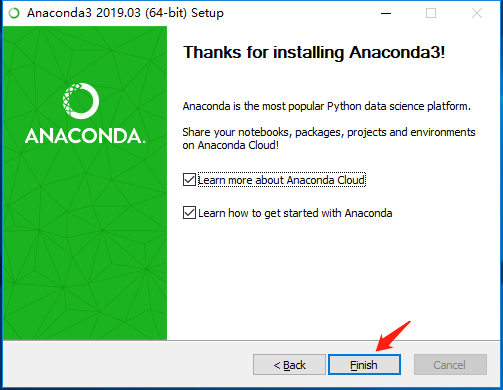
开始安装



安装完成，这里推荐安装PyCharm，直接点Next跳过。



安装完成

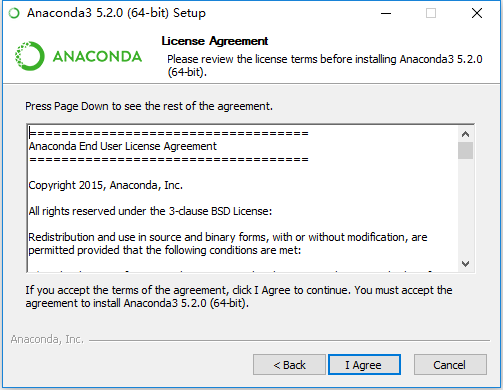


### Win7系统

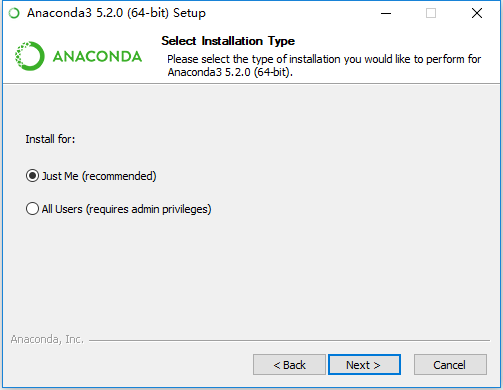
选择下载Anaconda3-5.2.0-Windows-x86\_64.exe，下载完成后运行安装：



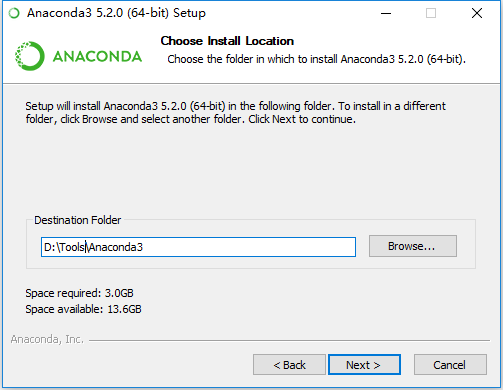
同意用户协议



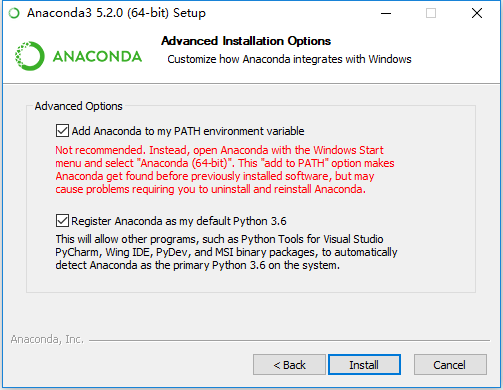
选择为所用用户安装或只是为自己安装。（直接下一步过去就行）



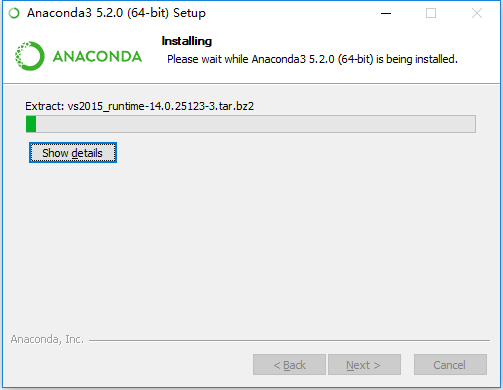
选择安装路径，默认安装在C盘就好（要保证有10GB以上的可用空间）



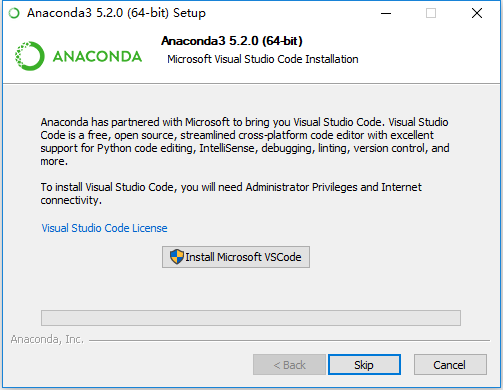
**这里建议将第一个勾选（将Anaconda添加到环境变量中，不勾选的话后期要自己添加到环境变量）**



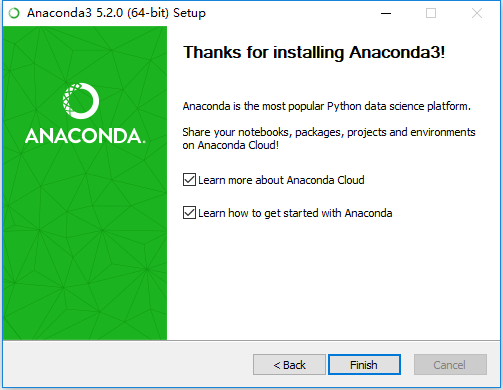
开始安装



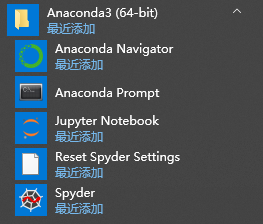
安装完成，这里提示是否安装微软的VSCode 直接Skip跳过。



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，，这三项，表明安装成功。

## 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

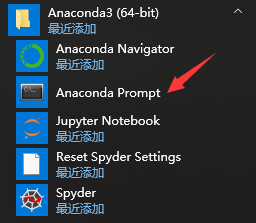
第六天：tensorflow、Keras

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



### 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。**

### 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 snownlp

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。

### 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

## 浏览器驱动下载

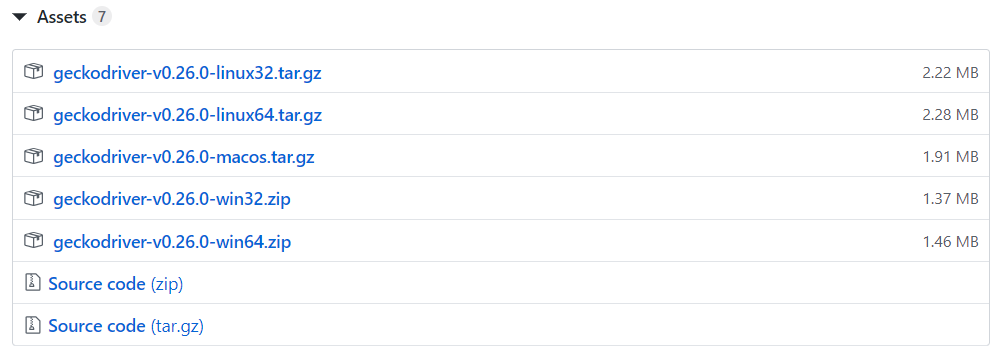
使用Selenium包爬取数据时，需要安装浏览器驱动。

建议使用Firefox或Chrome浏览器，二者选一种即可。

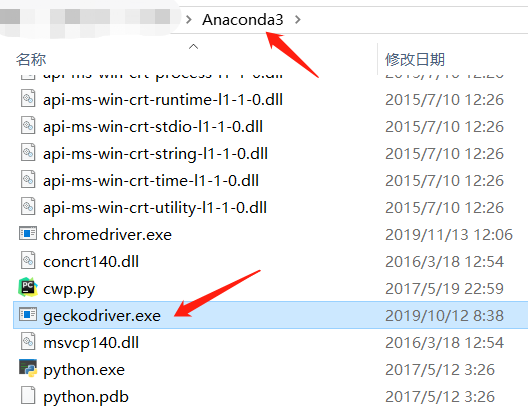
### Firefox浏览器

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：



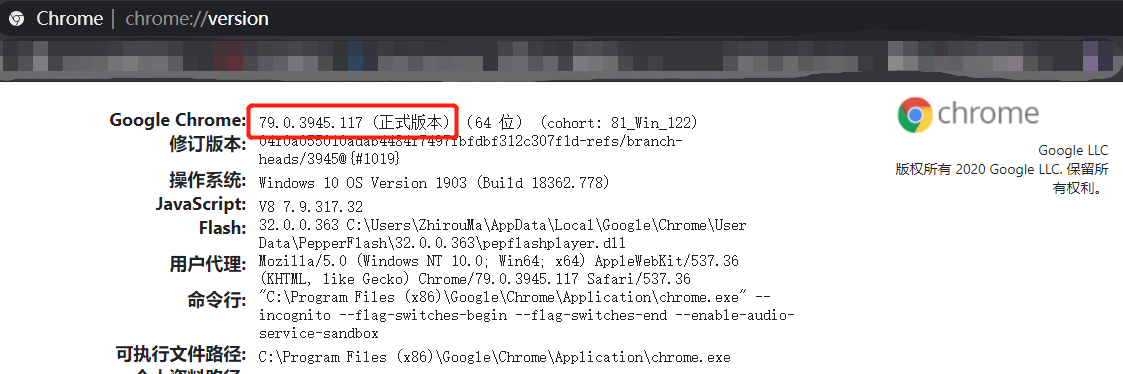
以win10 64位系统为例，下载geckodriver-v0.26.0-win64.zip，解压将geckodriver.exe放到anaconda安装目录下（与python.exe同目录）即可。



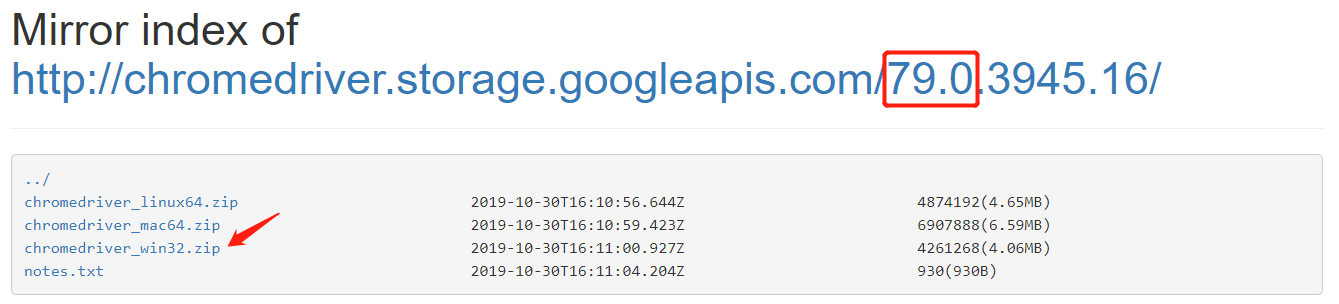
### Chrome浏览器

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

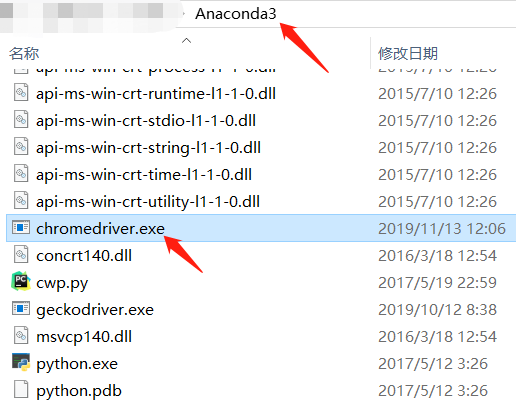
首先，需要查看chrome浏览器的版本号，在浏览器地址栏输入<chrome://version/>，79.0.3945.117为浏览器版本号，如下图所示：



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：

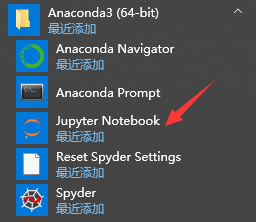


最后，下载chromedriver\_win32.zip，解压将chromedriver.exe放到anaconda安装目录下（与python.exe同目录）即可。

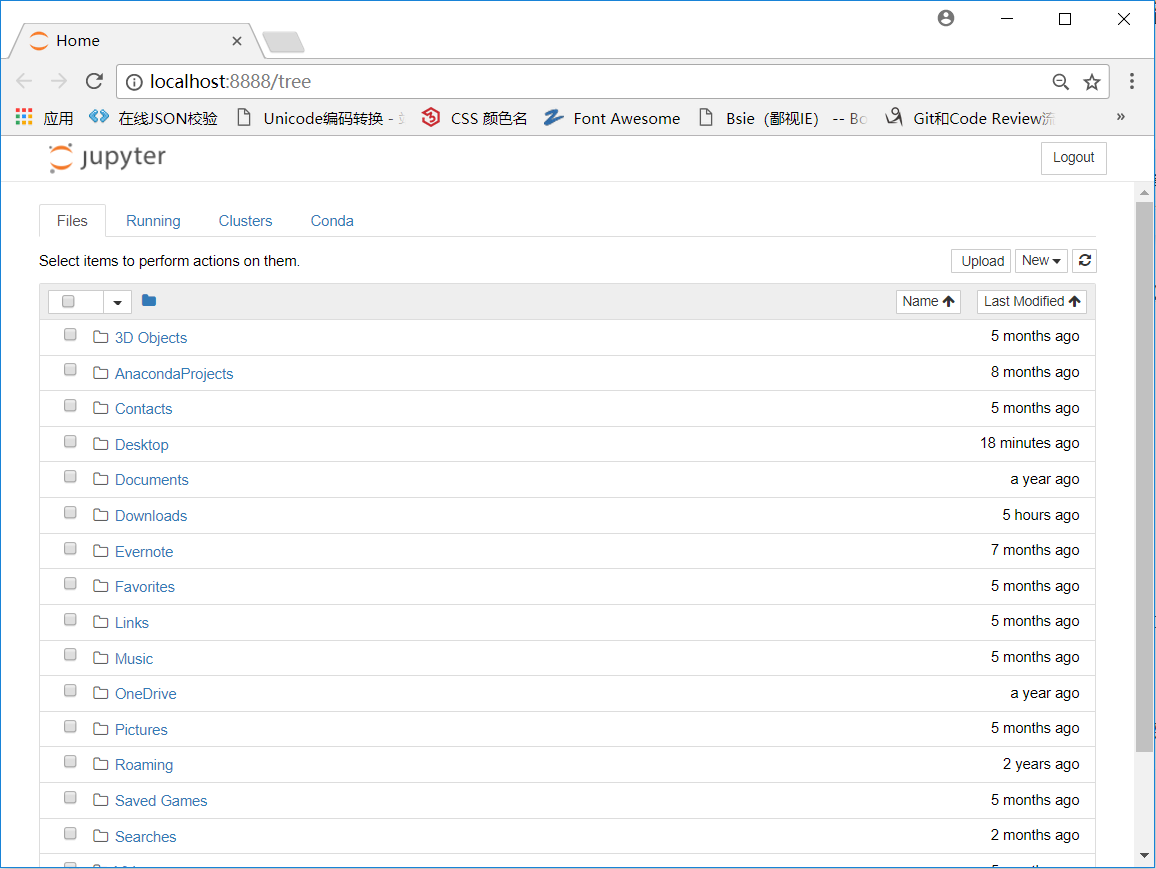


## Jupyter notebook

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。

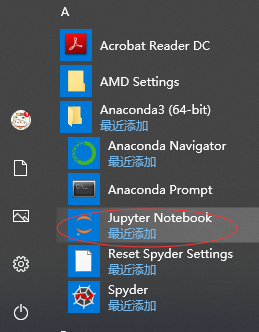
请大家确认一定要可以正常打开notebook。

## 检查包是否安装成功

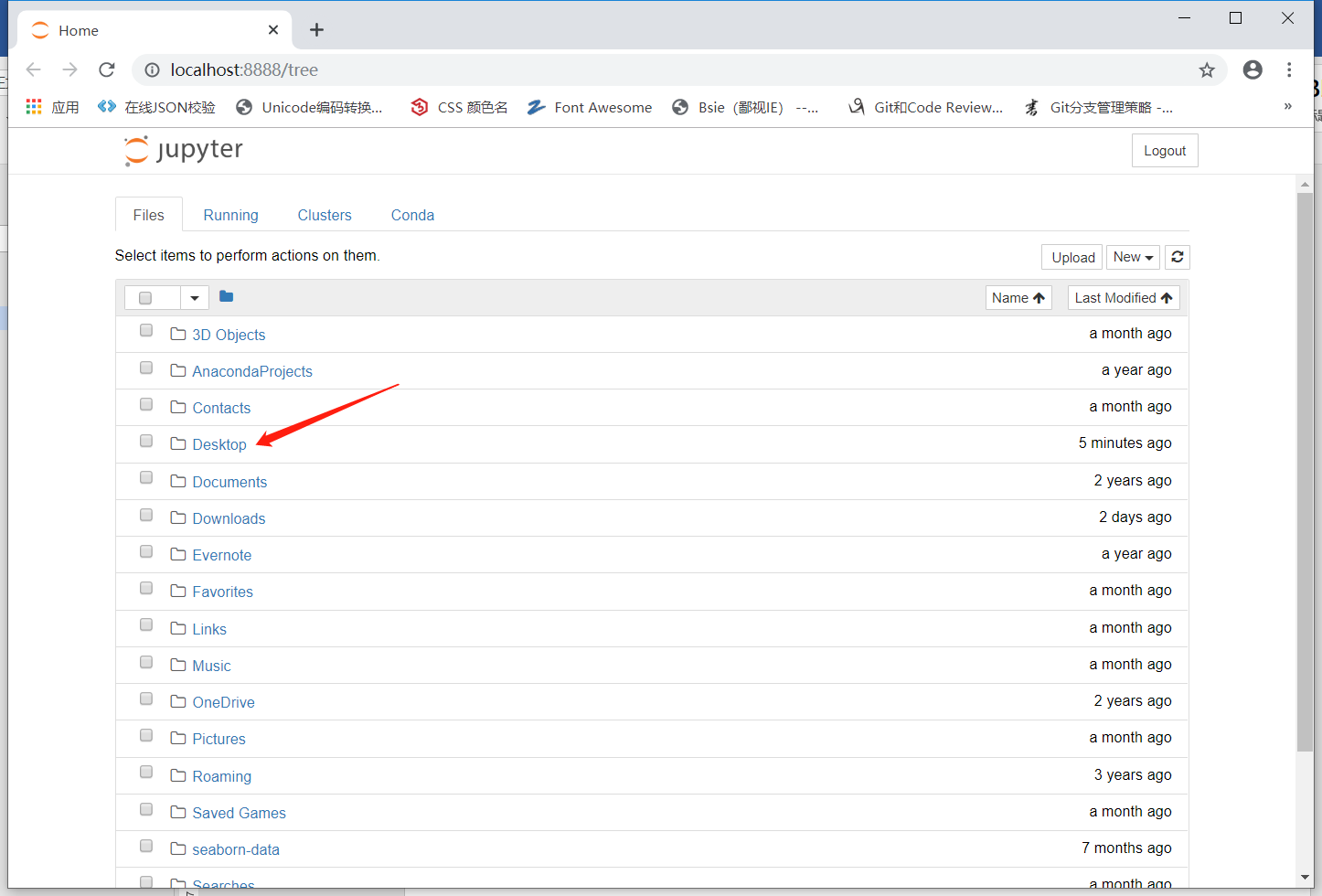
第一步：将 **Python课程-安装包检查.ipynb** 文件放到桌面



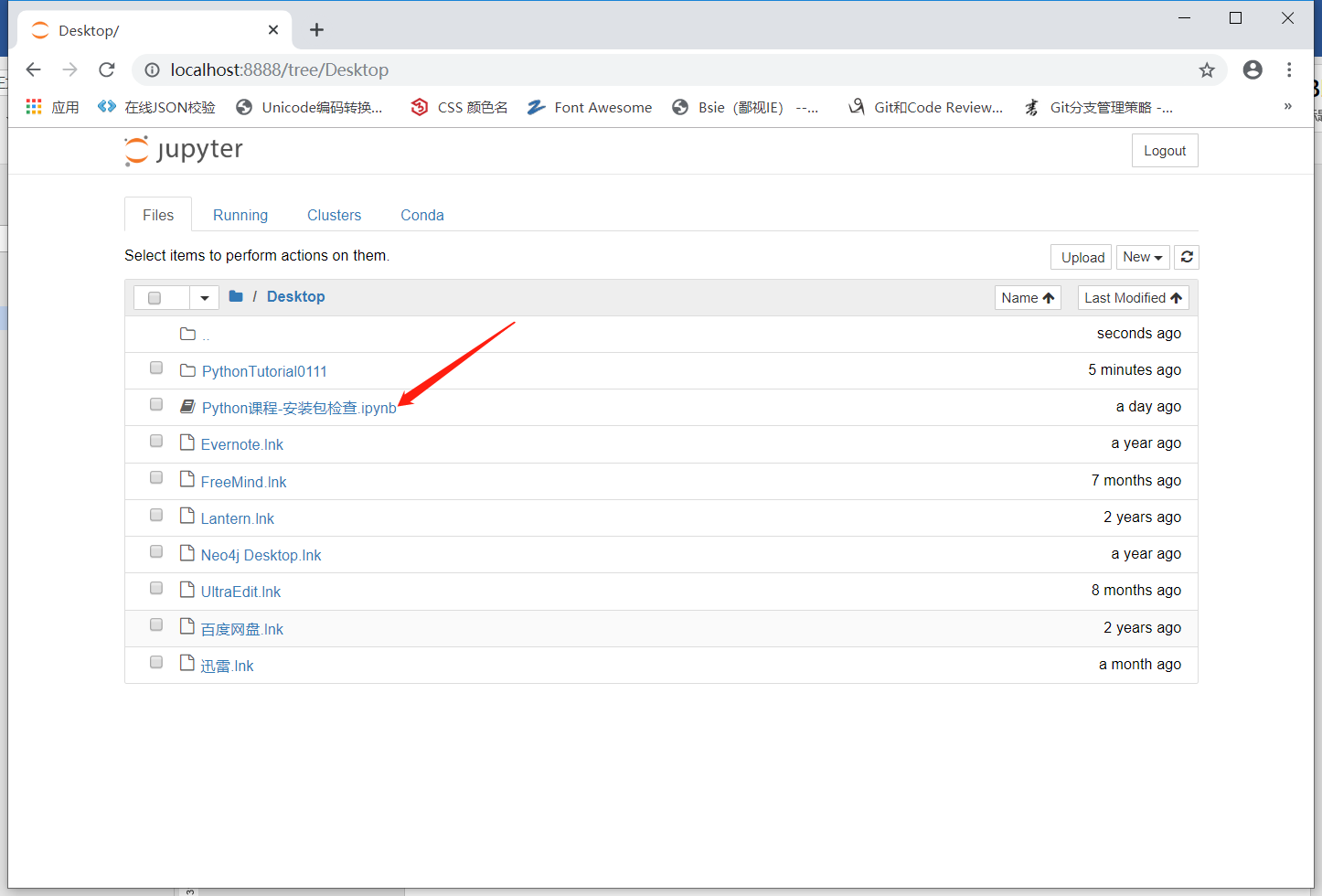
第二步：打开Jupyter Notebook



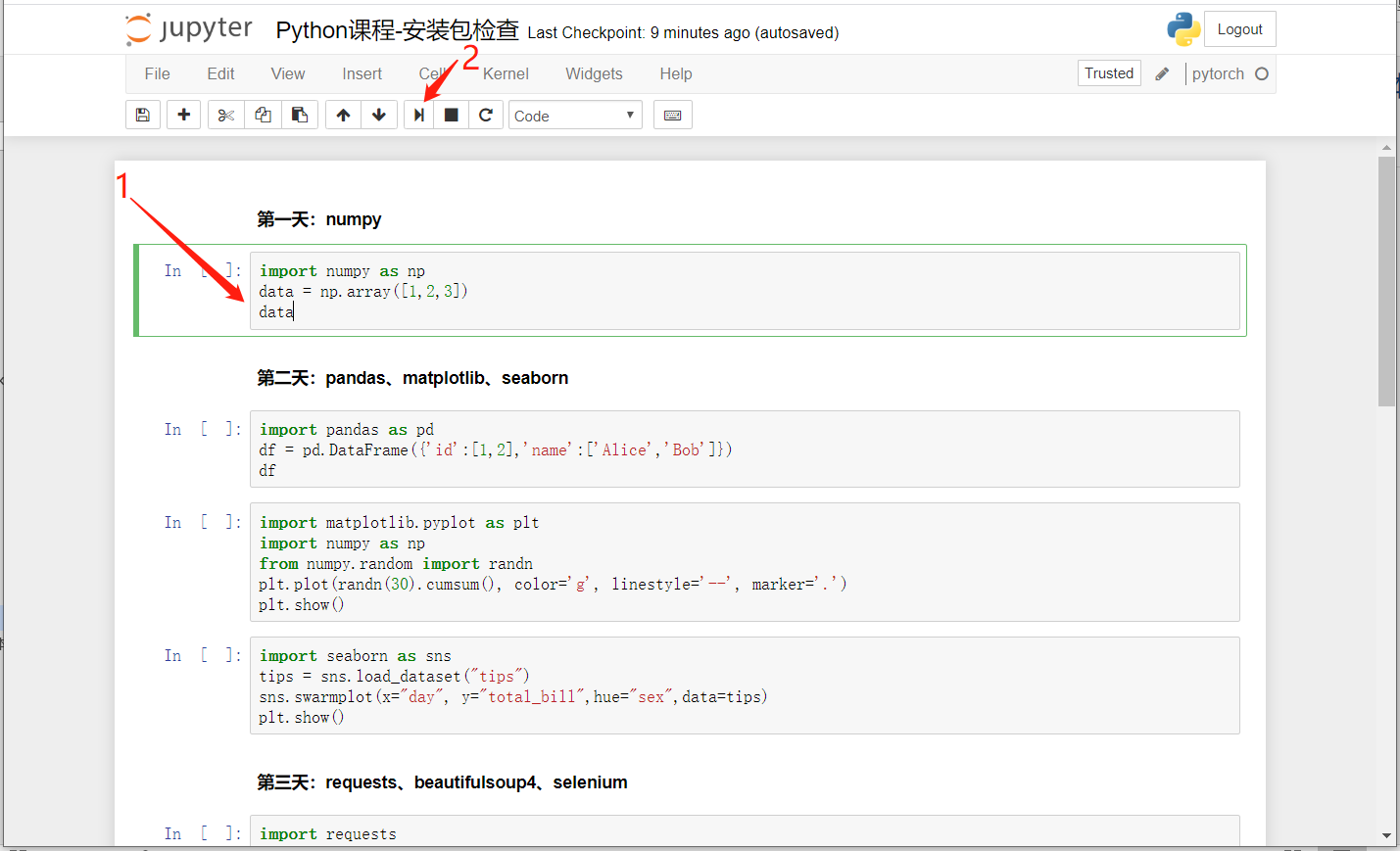
第三步：选择并打开 **Python课程-安装包检查.ipynb** 文件



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

### 第一天



### 第二天



能画出图（不一定一样的图），即安装成功。

### 第三天

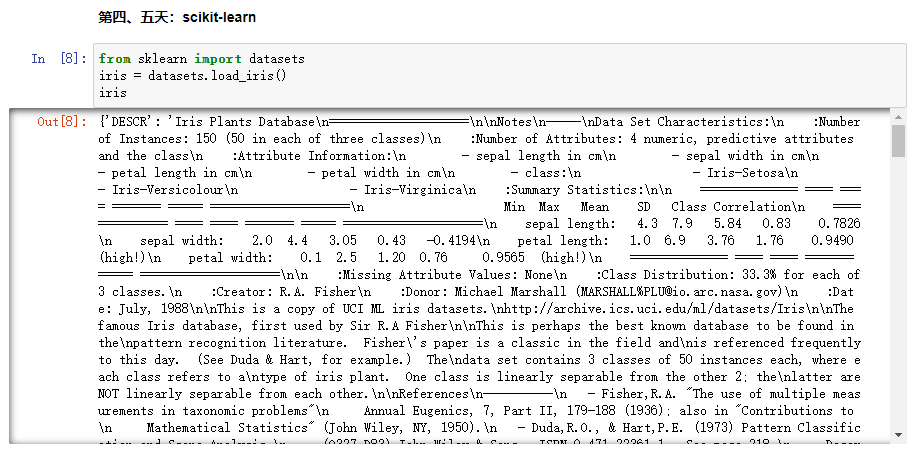


运行后会打开一个浏览器，然后再关闭，即安装成功。

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

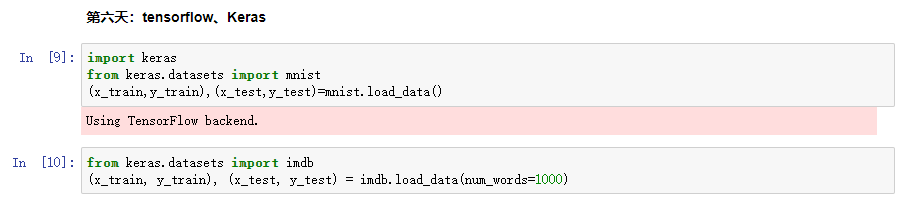
常见错误：浏览器驱动版本不对、位置不对，详见3.3节。

### 第四、五天



运行不报错，即安装成功。

### 第六天

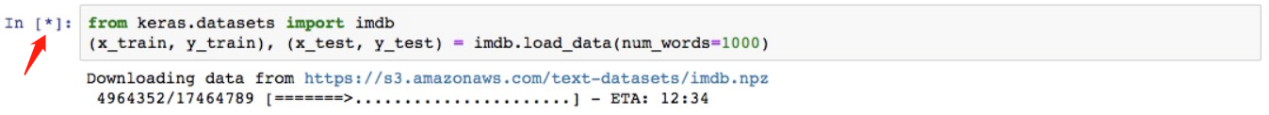


运行不报错，即安装成功。

如果**In[ ]**中出现**数字**表示程序执行完成，如下图所示：

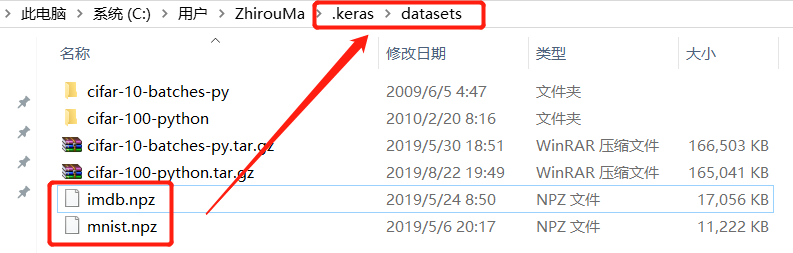


如果**In[ ]**中出现**星号**表示程序正在执行，如下图所示：



常见错误1：由于网络原因，下载时会出现超时中断。

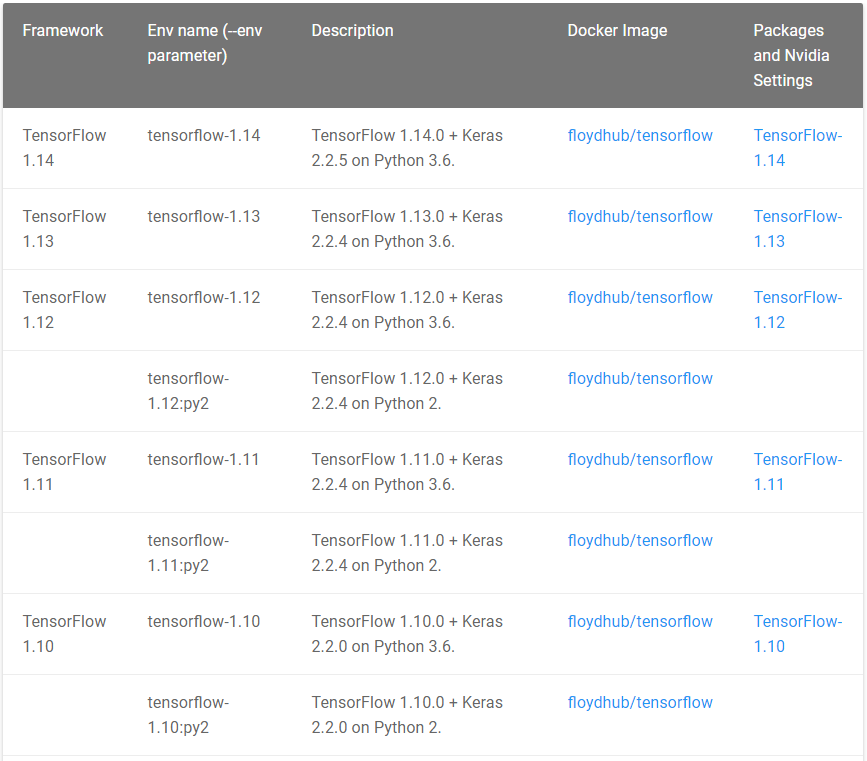
可以直接将**mnist.npz**和**imdb.npz**两个文件拷贝到用户目录下的.keras\datasets文件夹下



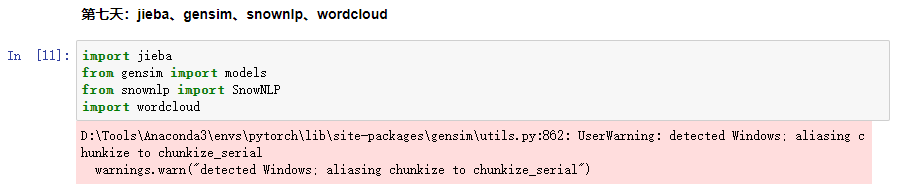
常见错误2：Tensorflow与Keras版本不对应。

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/>



### 第七天



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。

# mac系统环境安装

## 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**

# 其他工具

## 代码自动补全插件（选择安装）

jupter 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的原因，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。