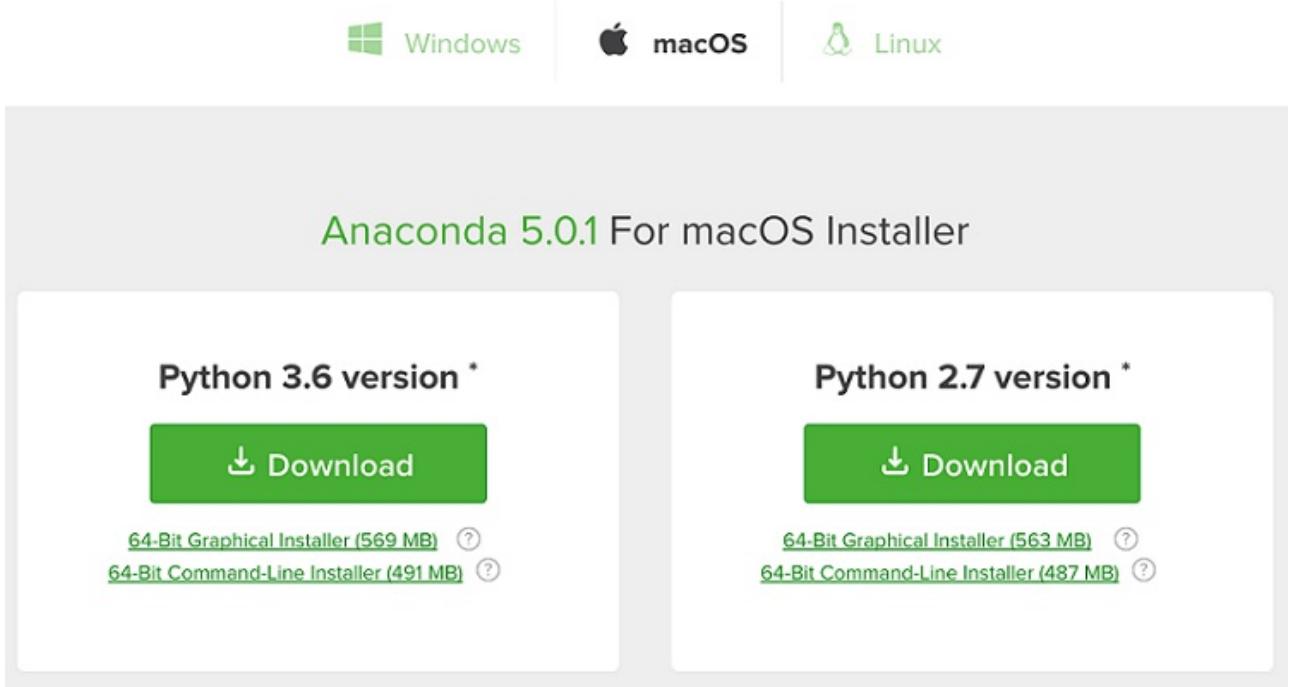


# 搭建数据分析环境： Anaconda、Jupyter Notebook

Anaconda 是一个针对数据分析领域的 Python 发行版本，它提供了包管理（packages）工具和虚拟环境（environment）管理，`conda` 命令可用于安装、卸载、更新包、创建不同版本的 Python 独立环境，可用于替换 `pip` 和 `virtualenv` 这两个工具。此外，Anaconda 自带了很多数据科学的依赖包以及 Jupyter Notebook 等工具。

## Anaconda 下载安装

可直接从 Anaconda [官方网站](#) (<https://www.anaconda.com/download/>) 进行下载，选择 Python3.6 的版本，因为 Python2.7 即将被废弃，下载后根据提示安装即可



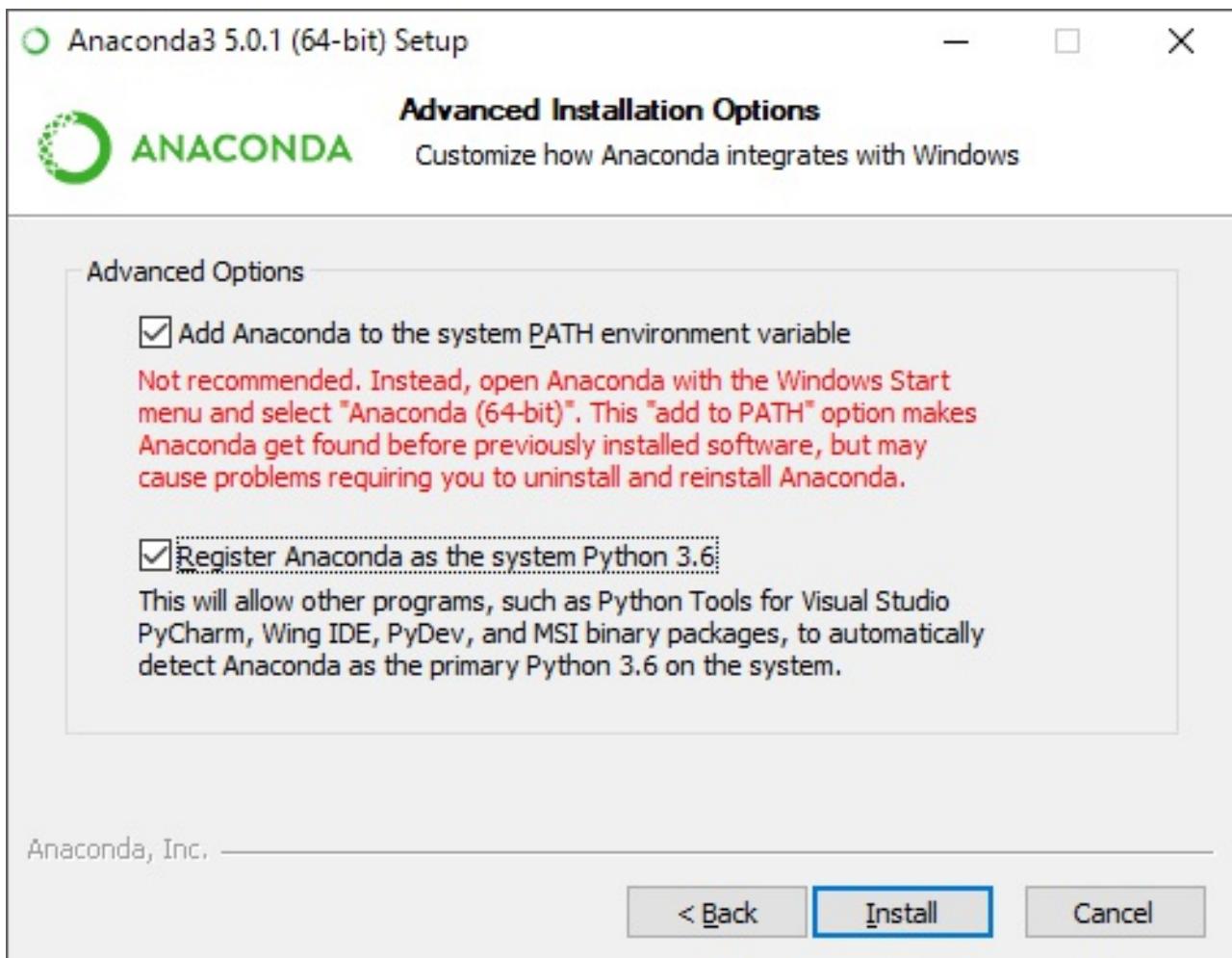
macOS/Linux 安装完成之后会自动把 Anaconda 添加到 PATH 环境变量（在 `~/.bash_profile` 文件中可以看到），如果你的终端默认 SHELL 不是 bash 的话（用 `echo $SHELL` 查看默认 shell 是啥），加了系统也找不到 `conda` 命令，比如我的 mac 默认 shell 是 zsh，需要把下面这行添加到 `~/.zshrc` 文件中

```
# added by Anaconda3 5.0.1 installer
export PATH="/Users/你的用户名/anaconda3/bin:$PATH"
```

再检查 `conda` 命令是否能用

```
conda -V
conda 4.3.30
```

Windows 平台安装的时候请自动勾选加入 PATH 路径，如果安装的时候没有勾选，要手动找到 Anaconda 的安装路径加入到 PATH 变量中，否则一样找不到 `conda` 命令。



为了使用 conda 安装包的过程中加快速度，可以把镜像地址修改为国内清华大学的镜像：编辑 ~/.condrc，（Windows 是在 C:\Users\你的用户名.condrc，如果没有该文件就创建一个），添加内容：

```
channels:
  -
    https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkgs/free/
  -
    https://mirrors.tuna.tsinghua.edu.cn/anaconda/cloud/conda-forge/
  -
    https://mirrors.tuna.tsinghua.edu.cn/anaconda/cloud/msys2/
ssl_verify: false
show_channel_urls: true
```

## 常用 conda 命令

### 包管理

```
# 查看帮助
conda -h
# 查看conda版本
conda --version
# 安装 matplotlib
conda install matplotlib
# 查看已安装的包
conda list
# 包更新
conda update matplotlib
# 删除包
conda remove matplotlib
```

### 环境管理

```
# 基于python3.6版本创建一个名字为test的python独立环境
conda create --name test python=3.6
# 激活此环境
activate test
source activate test # linux/mac
# 退出当前环境
deactivate test
# 删除该环境
conda remove -n test --all
# 或者
conda env remove -n test

# 查看所有安装的python环境
conda info -e
test *
D:\Programs\Anaconda3\envs\test
root D:\Programs\Anaconda3 (安
装 conda 默认生成的)
```

## 其他命令

```
# 更新conda本身
conda update conda
# 更新anaconda 应用
conda update anaconda
# 更新python, 假设当前python环境是3.6.1, 而最新版本是
3.6.2, 那么就会升级到3.6.2
conda update python
```

安装完 Anaconda 之后, Jupyter Notebook 也装好了。

Jupyter Notebook 是一个强大的数据分析工具, 你可以在上面写代码、运行代码、写文档、列方程式、做数据可视化展示。正如其

名，它就像一个草稿本可以在上面随意地涂写改改画画，画错了还可以擦除重做。

## 启动jupyter

在命令行直接输入：

```
jupyter notebook
```

Jupyter 启动成功后，在浏览器中会自动打开 notebook 的主界面，新建一个notebook时要点击右上角的 New，选择 Python3，这里的 Python3 就是 jupyter 的内核，是安装 Anaconda 的时候的名字为root的默认 python 环境。



新建了 notebook 之后你就可以在单元格里面写代码或者写 markdown 文档，或者基于用 matplotlib 制图。

The screenshot shows a Jupyter Notebook interface. At the top, there's a toolbar with various icons: a file icon, a plus sign for new cells, a close icon, a refresh icon, up and down arrow icons, a play/pause icon, a stop/cancel icon, and a clear cell icon. To the right of the toolbar are buttons for 'Code' and '代码/文档'. Below the toolbar, there are two red annotations: one pointing to the '新增单元格' (New Cell) button with the text '新增单元格', and another pointing to the play/pause icon with the text '运行' (Run). The main area shows an input cell labeled 'In [2]:' containing the Python code: 'for i in range(5): print(i)'. To the right of this cell, the word '单元格' (Cell) is written in red. Below the code, the output cell shows the numbers 0, 1, 2, 3, and 4, each on a new line. At the bottom, there's an empty input cell labeled 'In [ ]:'.

## 补充

如何查看 jupyter 使用了哪些 kernel

```
~ jupyter kernelspec list
```

Available kernels:

weixin

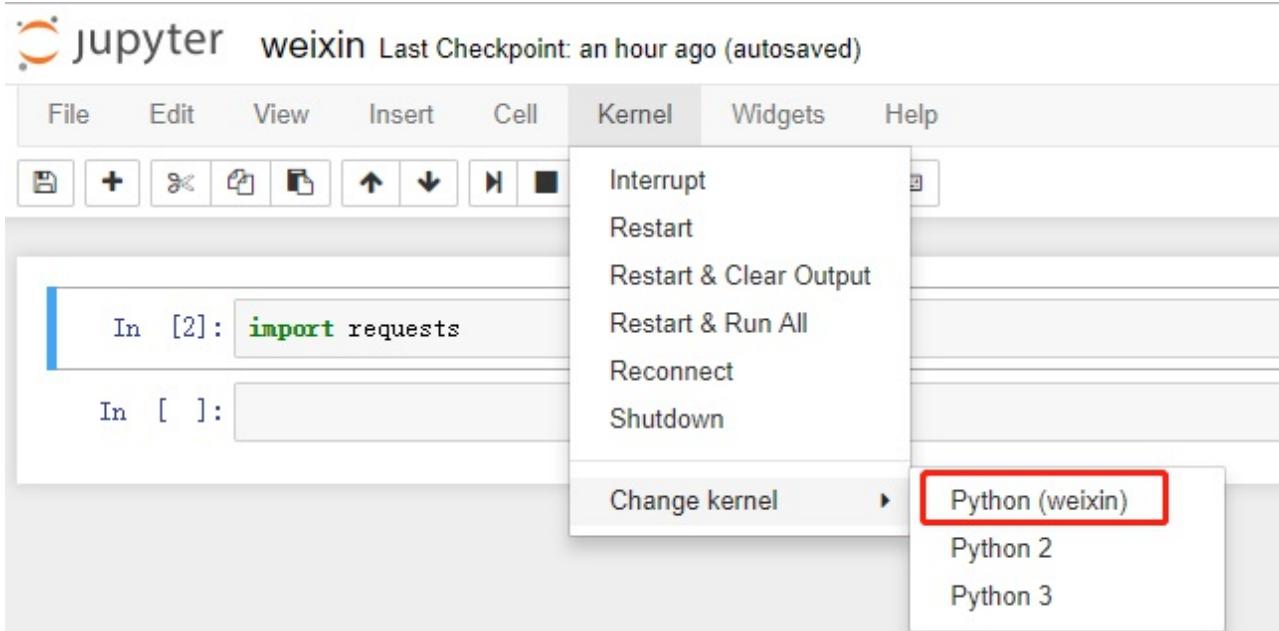
```
/Users/xxx/Library/Jupyter/kernels/weixin
```

python3

```
/Users/xxx/anaconda3/share/jupyter/kernels/python3
```

如何新增 kernel

```
# 创建python环境
conda create -n weixin python=3.6
# 激活
source activate weixin
# 加入到juypter
python -m ipykernel install --user --name weixin
--display-name "Python (weixin)"
```



新增了 kernel 之后，你可以在不同的 kernel 之间切换运行代码，本质上 kernel 还是 Python 的虚拟环境。

推荐一个Jupyter Notebook 的视频教程：[Jupyter Notebook Tutorial: Introduction, Setup, and Walkthrough](https://www.youtube.com/watch?v=HW29067qVWk) (<https://www.youtube.com/watch?v=HW29067qVWk>) (需要翻墙)