

機器學習於材料資訊的應用

Machine Learning on Material Informatics

陳南佑(NAN-YOW CHEN)

nanyow@narlabs.org.tw

楊安正(AN-CHENG YANG)

acyang@narlabs.org.tw

Build your own Development Environment

Use python + pip + virtualenv

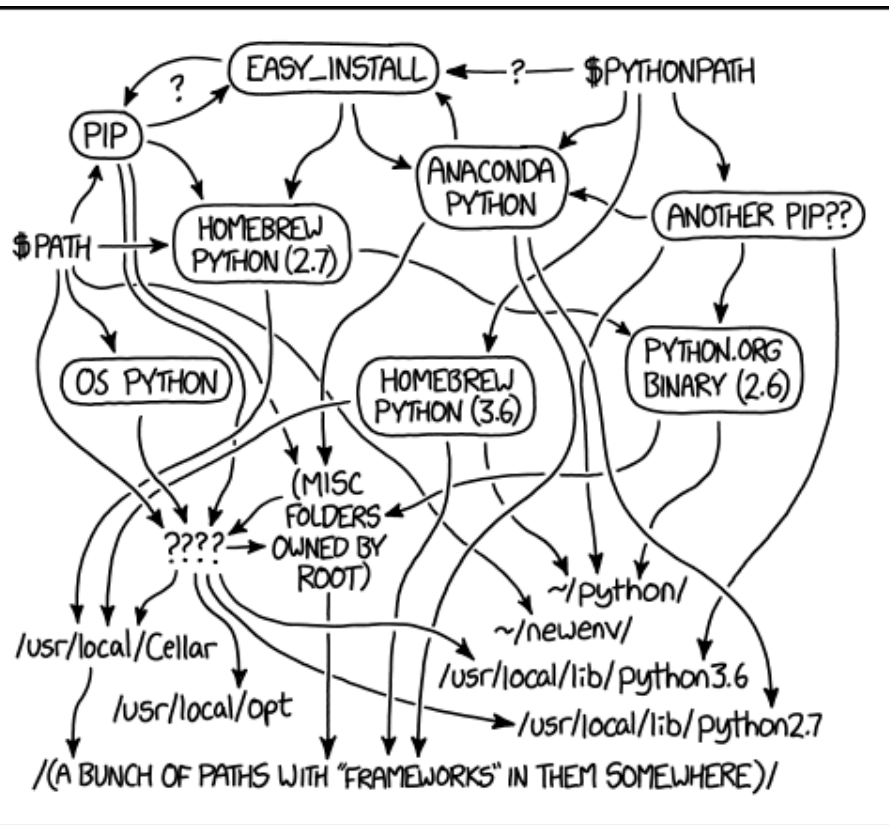
Windows

1. Download and install python.
2. Install virtualenv via pip
3. Activate virtualenv
4. Install package in virtualenv

Mac & Linux

Don't use system python
Don't use system python
Don't use system python

1. Download and compiler python.
2. Install virtualenv via pip
3. Activate virtualenv
4. Install package in virtualenv



MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED
THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

Build your own Development Environment

Use Miniconda/Anaconda

Conda is a package management system.

Miniconda/Anaconda is a distribution for python.

Anaconda is owned by Continuum Analytics™.

Packages in Conda are released by binary not source code. This means that if conda decide not to release certain package you will ...



ANACONDA®

Anaconda Navigator
File Help

ANACONDA NAVIGATOR

Sign in to Anaconda Cloud

Home

Environments

Learning

Community

Documentation

Developer Blog



Applications on scikit-learn

Channels

Refresh



jupyter

Notebook

5.7.4
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch



anypoint

1.1.1

Install



dioplas

0.4.1

Install



Glueviz

0.14.0

Multidimensional data visualization across files. Explore relationships within and among related datasets.

Install



hyperspyui

1.1.0



JupyterLab

0.35.3

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.



Orange 3

3.20.1

Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows



pysplot-gui

1.2.3

Home

Environments

Learning

Community

Documentation

Developer Blog



Search Environments

base (root)

LFPy

TWN-tools

idp

lfpv

mdanalysis

miniconda3

mpp

my-rdkit-env



Create



Clone



Import



Remove

Selected

Channels

Update index...

Search Packages

Name

Description

Version

matplotlib

Publication quality figures in python

3.2.0

Create new environment

Name:

mpp

自己取名字

Location: D:\anaconda3\envs\mpp

Packages:

☒ Python

3.8

選3.x都可以

☐ R

r

Cancel

Create

按這邊

6 packages available 6 packages selected

Apply

Clear



Sign in to Anaconda Cloud

Home

Environments

Learning

Community

Documentation

Developer Blog



Search Environments



base (root)

LFPy

TWN-tools

idp

lfp

mdanalysis

miniconda3

mpp



my-rdkit-env



Create



Clone



Import



Remove

自行切換filter來
方便查看

搜尋以下幾個套件

Selected



Channels

Update index...

Search Packages



Name



T

Description

Version



jupyterlab



Jupyterlab

2.0.1



matplotlib



Publication quality figures in python

3.2.0



numpy



Array processing for numbers, strings, records, and objects.

1.9.3



pandas



High-performance, easy-to-use data structures and data analysis tools.

1.0.1



pymatgen



2020.3.2



scikit-learn



A set of python modules for machine learning and data mining

0.22.2.p...



scipy



Scientific library for python

1.4.1



spyder



The scientific python development environment

4.0.1

8 packages available 8 packages selected

Apply

Clear

Home

Environments

Learning

Community

Documentation

Developer Blog



Applications on mpp Channels

Refresh



JupyterLab

2.0.1

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch



Jupyter Notebook

6.0.3

Web-based, interactive computing and reproducible computing. Edit and run notebooks, and create reusable docs while describing the data analysis.

Launch



Qt Console

4.7.1

PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.

Launch



Spyder

4.0.1

Scientific PYTHON Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection Features

Launch



anypytools

1.4.1

Install



CMD.exe Prompt

0.1.1

Run a cmd.exe terminal with your current environment from Navigator activated

Install



dioplas

0.4.1

Install



fsleyes

0.32.3

Install



Glueviz

0.15.2

Multidimensional data visualization across files. Explore relationships within and among related datasets.

Install



gpi

1.2.5

Install



hyperspyui

1.1.1

Install



labelme

3.21.1

Install



mercurial-app

5.2



Orange 3

3.24.1

Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.



Powershell Prompt

0.0.1

Run a Powershell terminal with your current environment from Navigator activated



psypilot-gui

1.2.4



RStudio

1.1.456

A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.



stratiz

0.1.3

點這開啟
jupyterlab

JupyterLab

localhost:8888/lab

File Edit View Run Kernel Tabs Settings Help

+

+

+

+

Name

Last Modified

3D Objects

24 days ago

adtemp

4 years ago

CMakeBuilds

2 years ago

Contacts

24 days ago

Desktop

6 days ago

Documents

3 days ago

Downloads

8 hours ago

Dropbox

13 days ago

Evernote

4 years ago

Favorites

8 days ago

GitBook

4 years ago

Google 雲端硬碟

4 days ago

Links

24 days ago

Music

24 days ago

OneAngstrom

4 months ago

OneDrive

9 days ago

ownCloud

3 years ago

Pictures

23 days ago

Roaming

5 years ago

Saved Games

24 days ago

Searches

24 days ago

SecurityScans

3 years ago

source

2 years ago

ssh

3 years ago

Tracing

4 years ago

Videos

24 days ago

1203087.pub

4 years ago

AzureStorageEmulatorDb42_log.ldf

3 years ago

AzureStorageEmulatorDb42.mdf

3 years ago

AzureStorageEmulatorDb45_log.ldf

3 years ago

AzureStorageEmulatorDb45.mdf

3 years ago

first.ppk

a year ago

firstkey.pem

a year ago

Y: idp.yml

4 months ago

Y: lfov.vaml

2 months ago

Launcher

Notebook

Python 3

Console

Python 3

Other

Terminal

Text File

Markdown File

Show Contextual Help

點這裡新增notebook

Build your own Development Environment

Use pip on CoLab

Google Colab (Colaboratory)是Google提供的雲端Jupyter Notebook開發環境。

最大賣點是提供了**免費**的GPU(型號為Tesla P100 GPU)和TPU。

每次開啟有**12小時**的連續使用時間限制，**12小時**過後虛擬機需要重新開啟才能繼續運行，也就是說不能在Colab上運行一個需要跑超過**12小時**的程式。

要有 google drive 的帳號，並且會操作 Jupyter Notebook。



Welcome To Colaboratory - Co

colab.research.google.com/notebooks/intro.ipynb#scrollTo=GJBs_fIRovLc

colab

Welcome To Colaboratory

File Edit View Insert Runtime Tools Help Unsaved changes since 3:20 AM

Connect Editing

New notebook

Open notebook... Ctrl+O

Upload notebook...

Rename...

Move to trash

Save a copy in Drive...

Save a copy as a GitHub Gist...

Save a copy in GitHub...

Save Ctrl+S

Save and pin revision Ctrl+M S

Revision history

Download .ipynb

Download .py

Update Drive preview

Print Ctrl+P

Code + Text

Copy to Drive

Getting started

The document you are reading is not a static web page, but an interactive environment called a **Colab notebook** that lets you write and execute code.

For example, here is a **code cell** with a short Python script that computes a value, stores it in a variable, and prints the result:

```
1 seconds_in_a_day = 24 * 60 * 60
2 seconds_in_a_day
```

86400

To execute the code in the above cell, select it with a click and then either press the play button to the left of the code, or use the keyboard shortcut "Command/Ctrl+Enter". To edit the code, just click the cell and start editing.

Variables that you define in one cell can later be used in other cells:

```
1 seconds_in_a_week = 7 * seconds_in_a_day
2 seconds_in_a_week
```

604800

Colab notebooks allow you to combine **executable code** and **rich text** in a single document, along with **images**, **HTML**, **LaTeX** and more. When you create your own Colab notebooks, they are stored in your Google Drive account. You can easily share your Colab notebooks with co-workers or friends, allowing them to comment on your notebooks or even edit them. To learn more, see [Overview of Colab](#). To create a new Colab notebook you can use the File menu above, or use the following link: [create a new Colab notebook](#).

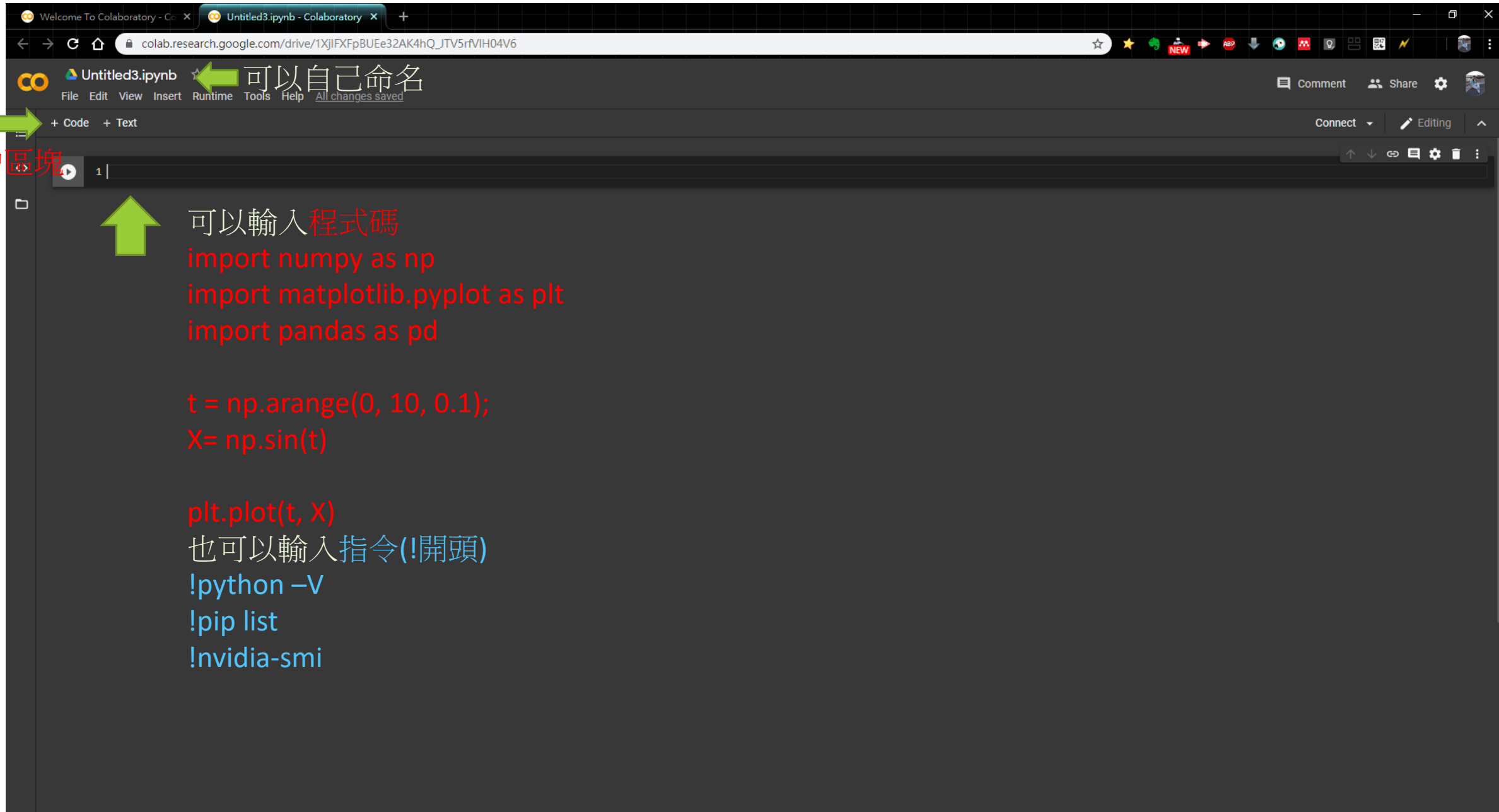
Colab notebooks are Jupyter notebooks that are hosted by Colab. To learn more about the Jupyter project, see [jupyter.org](#).

▼ Data science

With Colab you can harness the full power of popular Python libraries to analyze and visualize data. The code cell below uses **numpy** to generate some random data, and uses **matplotlib** to visualize it. To edit the code, just click the cell and start editing.

```
[ ] 1 import numpy as np
    2 from matplotlib import pyplot as plt
    3
    4 ys = 200 + np.random.randn(100)
```

新增區塊



The screenshot shows the Google Colaboratory web interface. At the top, there are two browser tabs: 'Welcome To Colaboratory - Co...' and 'Untitled3.ipynb - Colaboratory'. The address bar shows the URL 'colab.research.google.com/drive/1XjIFXfpBUeE32AK4hQ_JTV5rfVIH04V6'. The Colaboratory logo and 'Untitled3.ipynb' are in the top left. A green arrow points to the file name with the text '可以自己命名' (Can be named by yourself). Below the logo is a menu bar with 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. A green arrow points to the '+ Code' button with the text '新增區塊' (Add block). The main area is a code editor with a dark background. A green arrow points to the first line of code with the text '可以輸入程式碼' (Can input code). The code in the editor is as follows:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

t = np.arange(0, 10, 0.1);
X= np.sin(t)

plt.plot(t, X)
也可以輸入指令(!開頭)
!python -V
!pip list
!nvidia-smi
```

At the bottom right of the code editor, there are icons for undo, redo, link, comment, settings, and delete.

mpp_colab.ipynb - Colaboratory

colab.research.google.com/drive/1HThPx6suql8tQr1nQ7H-OASuHe3bdqmW#scrollTo=-4CUlwuPLyUR

CO mpp_colab.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings Profile

RAM Disk

Editing

+ Code + Text

<>

Python 3.6.9

1 !python -V

1 !nvidia-smi

NVIDIA-SMI has failed because it couldn't communicate with the NVIDIA driver. Make sure that the latest NVIDIA driver is installed and running.

mpp_colab.ipynb - Colaboratory

colab.research.google.com/drive/1HThPx6suql8tQr1nQ7H-OASuHe3bdqmW#scrollTo=-4CULwuPLYUR

mpp_colab.ipynb

FileEditViewInsertRuntimeToolsHelp

All changes saved

+ Code

[3]

⌂

▶

⌂

Undo insert cellCtrl+M Z

RedoCtrl+Shift+Y

Select all cellsCtrl+Shift+A

Cut selection

Copy selection

Paste

Delete selected cellsCtrl+M D

Find and replace...Ctrl+H

Find nextCtrl+G

Find previousCtrl+Shift+G

Notebook settings

Show/hide code

Clear all outputs

CommentShareSettingsProfile

RAMDisk

Editing

↑↓🔗💬⚙️🗑️⋮

ouldn't communicate with the NVIDIA driver. Make sure that the latest NVIDIA driver is installed and running.

mpp_colab.ipynb - Colaboratory

colab.research.google.com/drive/1HThPx6suql8tQr1nQ7H-OASuHe3bdqmW#scrollTo=-4CUlwuPlyUR

mpp_colab.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

RAM Disk

Editing

[3] 1 !python -V

Python 3.6.9

1 !nvidia-smi

NVIDIA-SMI has failed because it couldn't communicate with the NVIDIA driver. Make sure that the latest NVIDIA driver is installed and running.

Notebook settings

Runtime type

Python 3

Hardware accelerator

None

None GPU TPU

☐ Omit code cell output

aving this notebook

CANCEL

SAVE

mpp_colab.ipynb - Colaboratory

colab.research.google.com/drive/1HThPx6suql8tQr1nQ7H-OASuHe3bdqmW#scrollTo=-4CUIwuPLYUR

mpp_colab.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text

Comment Share

RAM Disk

Editing

```
[ ] 1 !python -V
```

Python 3.6.9

```
[ ] 1 !nvidia-smi
```

```
Tue Mar 10 02:10:25 2020
```

NVIDIA-SMI 440.59 Driver Version: 418.67 CUDA Version: 10.1									
GPU	Name	Persistence-M	Bus-Id	Disp.A	Volatile Uncorr. ECC	Fan	Temp	Perf	Pwr:Usage/Cap
0	Tesla P100-PCIE...	Off	00000000:00:04.0	Off	0	N/A	45C	P0	27W / 250W
					Memory-Usage	GPU-Util	Compute M.		
					0MiB / 16280MiB	0%	Default		

Processes:					GPU Memory
GPU	PID	Type	Process name		Usage
No running processes found					

P100 18NTD/gpu-hr
免費讓你用

mpp_colab.ipynb - Colaboratory

colab.research.google.com/drive/1HTHPx6suql8tQr1nQ7H-OASuHe3bdqmW#scrollTo=pRNsGSLONZww

mpp_colab.ipynb

File Edit View Insert Runtime Tools Help All changes saved

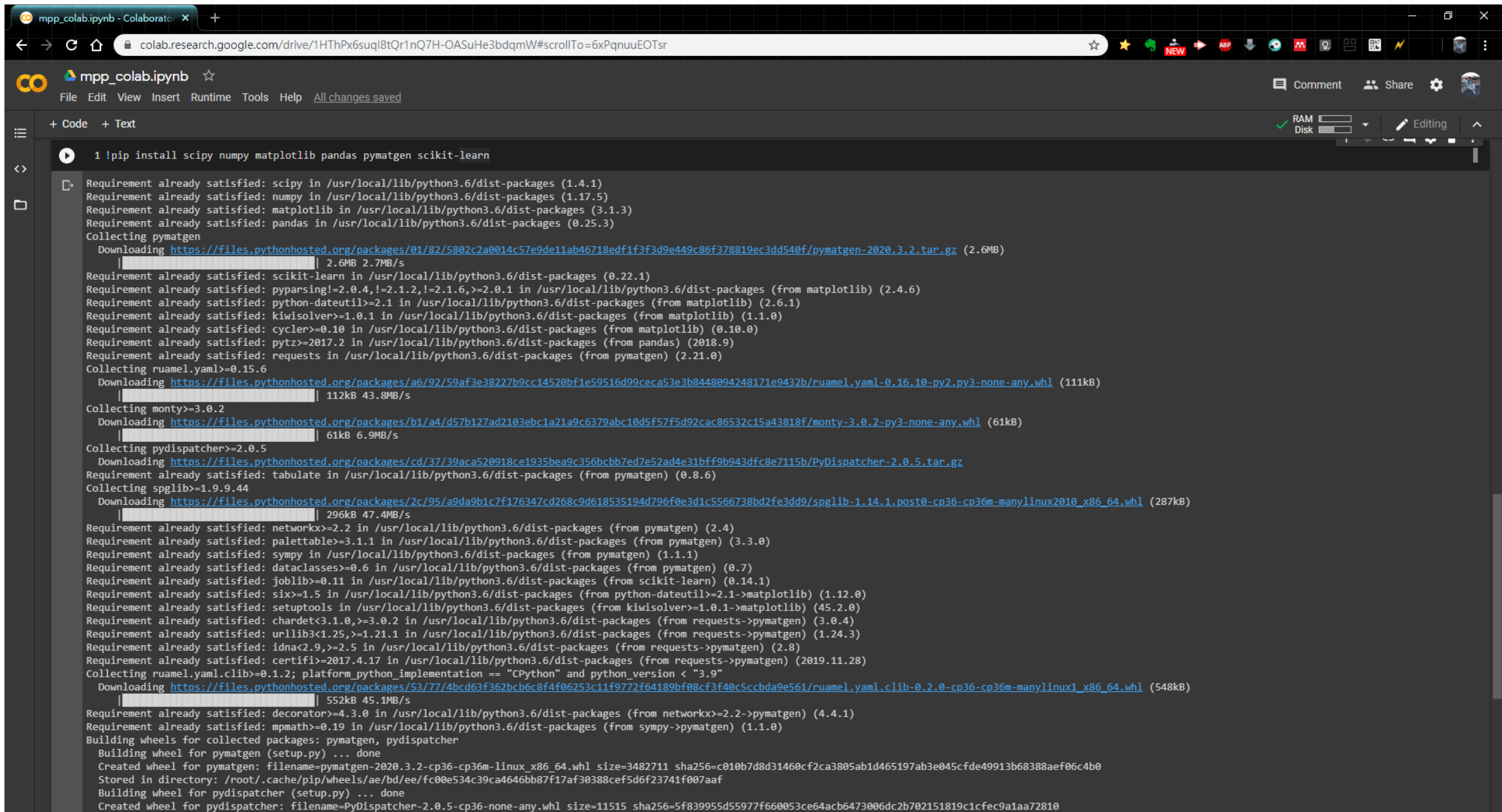
+ Code + Text

RAM Disk

Editing

1 !pip list

Package	Version
absl-py	0.9.0
alabaster	0.7.12
albumations	0.1.12
altair	4.0.1
asgiref	3.2.3
astor	0.8.1
astropy	4.0
atari-py	0.2.6
atomicwrites	1.3.0
attrs	19.3.0
audioread	2.1.8
autograd	1.3
Babel	2.8.0
backcall	0.1.0
backports.tempfile	1.0
backports.weakref	1.0.post1
beautifulsoup4	4.6.3
bleach	3.1.0
blis	0.2.4
bokeh	1.4.0
boto	2.49.0
boto3	1.11.15
botocore	1.14.15
Bottleneck	1.3.1
branca	0.3.1
bs4	0.0.1
bz2file	0.98
cachetools	3.1.1
certifi	2019.11.28
cffi	1.14.0
chainer	6.5.0
chardet	3.0.4
chart-studio	1.0.0
Click	7.0
cloudpickle	1.2.2
cmake	3.12.0
colorlover	0.3.0
community	1.0.0b1
contextlib2	0.5.5
convertdate	2.2.0
coverage	3.7.1
coveralls	0.5
crcmod	1.7
cufflinks	0.17.0
cupy-cuda101	6.5.0



The screenshot shows a Google Colab notebook interface. The browser address bar displays the URL: `colab.research.google.com/drive/1HTPx6suql8tQr1nQ7H-OASuHe3bdqmW#scrollTo=6xPqnuuEOTsr`. The notebook title is `mpp_colab.ipynb`. The code cell contains the command: `!pip install scipy numpy matplotlib pandas pymatgen scikit-learn`. The output shows the installation progress for each package, including requirements already satisfied and new packages being collected and downloaded. The packages installed are `scipy`, `numpy`, `matplotlib`, `pandas`, `pymatgen`, and `scikit-learn`. The output also shows the installation of `ruamel.yaml` and `pydispatcher` as dependencies. The final output shows the creation of wheels for the installed packages.

```
1 !pip install scipy numpy matplotlib pandas pymatgen scikit-learn

Requirement already satisfied: scipy in /usr/local/lib/python3.6/dist-packages (1.4.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages (1.17.5)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.6/dist-packages (3.1.3)
Requirement already satisfied: pandas in /usr/local/lib/python3.6/dist-packages (0.25.3)
Collecting pymatgen
  Downloading https://files.pythonhosted.org/packages/01/82/5802c2a0014c57e9de11ab46718edf1f3f3d9e449c86f378819ec3dd540f/pymatgen-2020.3.2.tar.gz (2.6MB)
    | 2.6MB 2.7MB/s
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.6/dist-packages (0.22.1)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/python3.6/dist-packages (from matplotlib) (2.4.6)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.6/dist-packages (from matplotlib) (2.6.1)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.6/dist-packages (from matplotlib) (1.1.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.6/dist-packages (from matplotlib) (0.10.0)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.6/dist-packages (from pandas) (2018.9)
Requirement already satisfied: requests in /usr/local/lib/python3.6/dist-packages (from pymatgen) (2.21.0)
Collecting ruamel.yaml>=0.15.6
  Downloading https://files.pythonhosted.org/packages/a6/92/59af3e38227b9cc14520bf1e59516d99ceca53e3b8448094248171e9432b/ruamel.yaml-0.16.10-py2.py3-none-any.whl (111kB)
    | 112kB 43.8MB/s
Collecting monty>=3.0.2
  Downloading https://files.pythonhosted.org/packages/b1/a4/d57b127ad2103ebc1a21a9c6379abc10d5f57f5d92cac86532c15a43818f/monty-3.0.2-py3-none-any.whl (61kB)
    | 61kB 6.9MB/s
Collecting pydispatcher>=2.0.5
  Downloading https://files.pythonhosted.org/packages/cd/37/39aca520918ce1935bea9c356bcb7ed7e52ad4e31bfff9b943dfc8e7115b/PyDispatcher-2.0.5.tar.gz
Requirement already satisfied: tabulate in /usr/local/lib/python3.6/dist-packages (from pymatgen) (0.8.6)
Collecting spglib>=1.9.44
  Downloading https://files.pythonhosted.org/packages/2c/95/a9da9b1c7f176347cd268c9d618535194d796f0e3d1c5566738bd2fe3dd9/spglib-1.14.1.post0-cp36-cp36m-manylinux2010_x86_64.whl (287kB)
    | 296kB 47.4MB/s
Requirement already satisfied: networkx>=2.2 in /usr/local/lib/python3.6/dist-packages (from pymatgen) (2.4)
Requirement already satisfied: palettable>=3.1.1 in /usr/local/lib/python3.6/dist-packages (from pymatgen) (3.3.0)
Requirement already satisfied: sympy in /usr/local/lib/python3.6/dist-packages (from pymatgen) (1.1.1)
Requirement already satisfied: dataclasses>=0.6 in /usr/local/lib/python3.6/dist-packages (from pymatgen) (0.7)
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.6/dist-packages (from scikit-learn) (0.14.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.6/dist-packages (from python-dateutil>=2.1->matplotlib) (1.12.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.6/dist-packages (from kiwisolver>=1.0.1->matplotlib) (45.2.0)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /usr/local/lib/python3.6/dist-packages (from requests->pymatgen) (3.0.4)
Requirement already satisfied: urllib3<1.25,>=1.21.1 in /usr/local/lib/python3.6/dist-packages (from requests->pymatgen) (1.24.3)
Requirement already satisfied: idna<2.9,>=2.5 in /usr/local/lib/python3.6/dist-packages (from requests->pymatgen) (2.8)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.6/dist-packages (from requests->pymatgen) (2019.11.28)
Collecting ruamel.yaml.clib>=0.1.2; platform_python_implementation == "CPython" and python_version < "3.9"
  Downloading https://files.pythonhosted.org/packages/53/77/4bcd6f362bcb6c8f4f06253c11f9772f64189bf08cf3f40c5ccbda9e561/ruamel.yaml.clib-0.2.0-cp36-cp36m-manylinux1_x86_64.whl (548kB)
    | 552kB 45.1MB/s
Requirement already satisfied: decorator>=4.3.0 in /usr/local/lib/python3.6/dist-packages (from networkx>=2.2->pymatgen) (4.4.1)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.6/dist-packages (from sympy->pymatgen) (1.1.0)
Building wheels for collected packages: pymatgen, pydispatcher
  Building wheel for pymatgen (setup.py) ... done
  Created wheel for pymatgen: filename=pymatgen-2020.3.2-cp36-cp36m-linux_x86_64.whl size=3482711 sha256=c010b7d8d31460cf2ca3805ab1d465197ab3e045cfde49913b68388aef06c4b0
  Stored in directory: /root/.cache/pip/wheels/ae/bd/ee/fc00e534c39ca4646bb87f17af30388cef5d6f23741f007aaf
  Building wheel for pydispatcher (setup.py) ... done
  Created wheel for pydispatcher: filename=PyDispatcher-2.0.5-cp36-none-any.whl size=11515 sha256=5f839955d55977f660053ce64acb6473006dc2b702151819c1cfec9a1aa72810
```

`!pip install scipy numpy matplotlib pandas pymatgen scikit-learn`