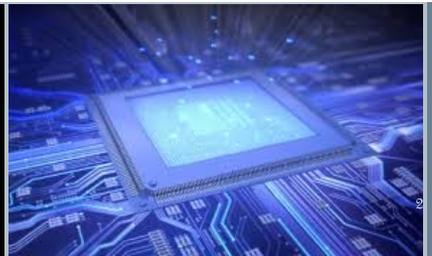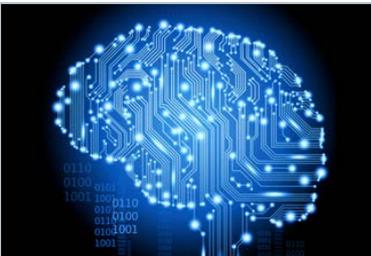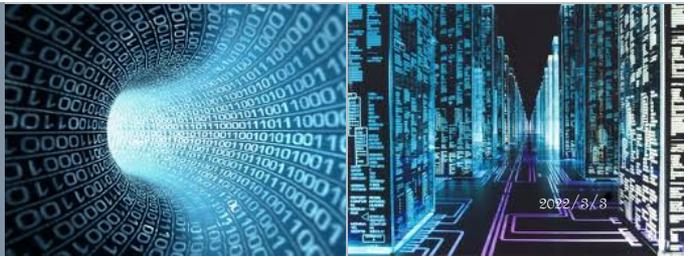# Computer programming and Data Science

William Hsu
Advanced Computation Laboratory
Department of Computer Science and Engineering
Department of Environmental Biology and Fisheries Science
National Taiwan Ocean University

Anaconda

Anaconda 整合環境

Anaconda

› Anaconda中文是森蚺，一種非常肥大的蟒蛇。
  – 簡單來說你可以把Anaconda當作是Python的懶人包。
  – 除了Python本身(python2, **3**) 還包含了Python常用的資料分析、機器學習、視覺化的套件。

› https://www.anaconda.com/

Anaconda

Packaged included in Anaconda 4+, or get with "`conda install PACKAGENAME`"

1. **NumPy** | numpy.org
   N-dimensional array for numerical computation

2. **SciPy** | scipy.org
   Collection of numerical algorithms and toolboxes, including signal processing and optimization

3. **MatPlotLib** | matplotlib.org
   Plotting library for Python

4. **Pandas** | pandas.pydata.org
   Powerful Python data analysis toolkit

5. **Seaborn** | stanford.edu/~mwaskom/software/seaborn/
   Statistical data visualization

6. **Bokeh** | bokeh.pydata.org
   Interactive web visualization library

7. **SciKit-Learn** | scikit-learn.org/stable
   Python modules for machine learning and data mining

8. **NLTK** | nltk.org
   Natural language toolkit

9. **Notebook** | jupyter.org
   Web-based interactive computational environment combines code execution, rich text, mathematics, plots and rich media

10. **R essentials** | conda.pydata.org/docs/r-with-conda.html
    R with 80+ of the most used R packages for data science
    "`conda install -c r r-essentials`"

Anaconda

› 當中比較重要的:

 – **Numpy**: Python做多維陣列（矩陣）運算時的必備套件，比起Python內建的list，Numpy的array有極快的運算速度優勢

 – **Pandas**: 有了Pandas可以讓Python很容易做到幾乎所有Excel的功能了，像是樞紐分析表、小記、欄位加總、篩選

 – **Matplotlib**: 基本的視覺化工具，可以畫長條圖、折線圖等等…

 – **SciKit-Learn**: Python 關於機器學習的model基本上都在這個套件，像是SVM, Random Forest…

Anaconda

› **Notebook(Jupyter notebook)**: 一個輕量級web-base
寫Python的工具，在資料分析這個領域很熱門。
  – 功能沒有比Pycharm, Spyder這些專業的IDE強大，但只要
    code小於500行、用Jupyter寫非常方便。
  – Jupyter也開始慢慢支援一些Multi cursor的功能了，可以讓
    你一次改許多的變數名稱。

Anaconda

› 優點：
  – 省時：一鍵安裝完90%一般人這一輩子會用到的Python套件，剩下的再用pip install個別去安裝即可

› 缺點：
  – 占空間：包含了一堆用不到的Python的套件(可安裝另一種miniconda)

› 安裝完Anaconda之後會獲得一個叫conda的指令可以用，可以用conda管理不同Python的版本以及套件。
  – 但是我們偏愛用pip來管理套件。

Anaconda 安裝

› Anaconda 下載連結 (要注意作業系統):
  https://www.anaconda.com/download/

The open-source Anaconda Distribution is the easiest way to perform Python/R data science and machine learning on Linux, Windows, and Mac OS X. With over 11 million users worldwide, it is the industry standard for developing, testing, and training on a single machine, enabling *individual data scientists* to:

- Quickly download 1,500+ Python/R data science packages
- Manage libraries, dependencies, and environments with Conda
- Develop and train machine learning and deep learning models with scikit-learn, TensorFlow, and Theano
- Analyze data with scalability and performance with Dask, NumPy, pandas, and Numba
- Visualize results with Matplotlib, Bokeh, Datashader, and Holoviews
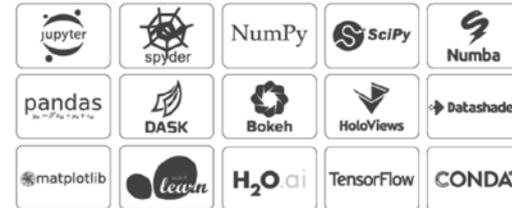
我們用這系統

Windows | macOS | Linux

Anaconda 2018.12 for Windows Installer

我們用這版本

Python 3.7 version

Download

64-Bit Graphical Installer (614.3 MB)
32-Bit Graphical Installer (509.7 MB)

Python 2.7 version

Download

64-Bit Graphical Installer (560.6 MB)
32-Bit Graphical Installer (458.6 MB)

Anaconda 安裝

下一步

你沒有選擇

Anaconda 安裝

Anaconda3 2018.12 (64-bit) Setup — □ ×

Select Installation Type

ANACONDA

Please select the type of installation you would like to perform for Anaconda3 2018.12 (64-bit).

Install for:

○ Just Me (recommended)

◉ All Users (requires admin privileges)

Anaconda, Inc. ——————————————

[< Back] [🛡Next >] [Cancel]

Anaconda3 2018.12 (64-bit) Setup — □ ×

Choose Install Location

ANACONDA

Choose the folder in which to install Anaconda3 2018.12 (64-bit).

Setup will install Anaconda3 2018.12 (64-bit) in the following folder. To install in a different folder, click Browse and select another folder. Click Next to continue.

Destination Folder

[C:\ProgramData\Anaconda3] [Browse...]

Space required: 2.8GB
Space available: 308.7GB

Anaconda, Inc. ——————————————

[< Back] [🛡Next >] [Cancel]

沒差、反正你的機器只有你一人用

下一步

Anaconda 安裝

據說勾起來比較好
反安裝(?)

VSCode 還不錯用

Anaconda 安裝

› 接著安裝完打開Command line輸入python

成功了

```
C:\Users\wwyhs>python
Python 3.7.1 (default, Dec 10 2018, 22:54:23) [MSC v.1915 64 bit (AMD64)] :: Anaconda, Inc. on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> quit
Use quit() or Ctrl-Z plus Return to exit
>>> ^Z

C:\Users\wwyhs>
```

使用 Anaconda
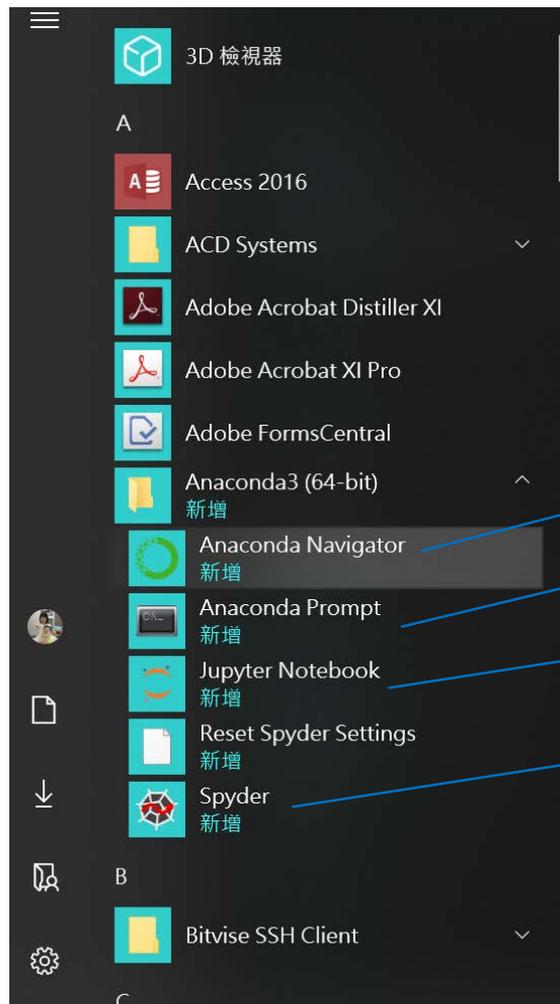
從這邊開始進入整合環境

命令提示列, 套件管理

IPython 筆記本

高級python發展環境

Anaconda 環境

使用Jupyter Notebook

會連到本地機器的python直譯核心

開啟舊檔案

可以直接開起ipynb檔案

互動式python環境

創立新檔案

New->選擇Python 3 檔案

命令模式(Command Mode)

編輯模式(Edit Mode)

```
In [ ]: |
```

儲存及載入(Saving/Loading)

建立一個新的notebook

複製目前notebook成一個新的
notebook

儲存目前notebook並記錄
checkpoint

下載notebook並存成各類型文件
- IPython notebook (.ipynb)
- Python (.py)
- HTML (.html)
- Markdown (.md)
- reST (.rst)
- LaTeX (.tex)
- PDF (.pdf)

File	Edit	View

New Notebook ▶
Open...

Make a Copy...
Rename...
Save and Checkpoint

Revert to Checkpoint ▶

Print Preview
Download as ▶

Trusted Notebook

Close and Halt

開啟一個已經存在的
notebook

notebook重新命名

將notebook還原至之前
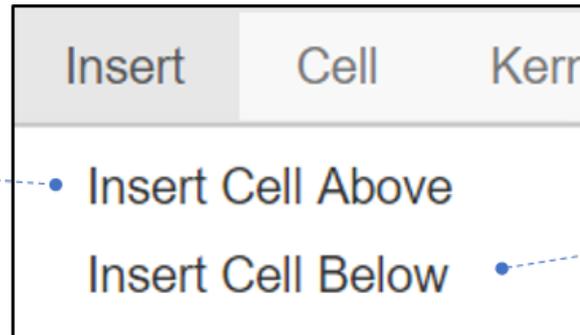checkpoint

預覽列印

關閉notebook並停止執行
script

程式碼及文本撰寫功能

› 程式碼和文本是由3種基本cells類型所包裝起來:
  Markdown cells、Code cells及Raw NBConvert cells。

將Cells剪下至剪貼簿	Cut Cells · · · 將Cells複製至剪貼簿
在目前Cell位置上方貼上複製的Cells	Copy Cells · · · Paste Cells Above · · · 在目前Cell位置下方貼上複製的Cells
	Paste Cells Below · · ·
將目前Cell置換為複製的Cells	Paste Cells & Replace · · · 刪除目前的Cells
	Delete Cells · · ·
恢復刪除的Cells	Undo Delete Cells · · · 在目前游標位置將Cell分開
	Split Cell · · ·
將目前Cell和上方Cell做合併	Merge Cell Above · · · 將目前Cell和下方Cell做合併
	Merge Cell Below · · ·
將目前的Cell往上移動	Move Cell Up · · · 將目前的Cell往下移動
	Move Cell Down · · ·
編輯notebook的metadata	Edit Notebook Metadata · · · 在選定的Cells當中尋找並替換
	Find and Replace · · ·
移除Cell的附件	Cut Cell Attachments · · · 複製目前Cell的附件
	Copy Cell Attachments · · ·
貼上目前Cell的附件	Paste Cell Attachments · · · 在選定的Cells當中插入圖片
	Insert Image · · ·

Edit   View   Insert

Insert

在目前Cell上方
插入一個新Cell

在目前Cell下方
插入一個新Cell

Executing

執行選定的Cell(s)

執行目前Cell(s)並且在下方插入一個新的Cell

執行目前Cell上方所有的Cells

更改目前Cell的類型
- Code
- Markdown
- Raw NBConvert

對所有Cell輸出做下列動作
- Toggle(隱藏/顯示切換)
- Toggle Scrolling(上下切換)
- Clear(清除)

Cell	Kernel	Widgets

- Run Cells
- Run Cells and Select Below
- Run Cells and Insert Below

Run All

- Run All Above
- Run All Below

- Cell Type ▶

- Current Outputs ▶
- All Output ▶

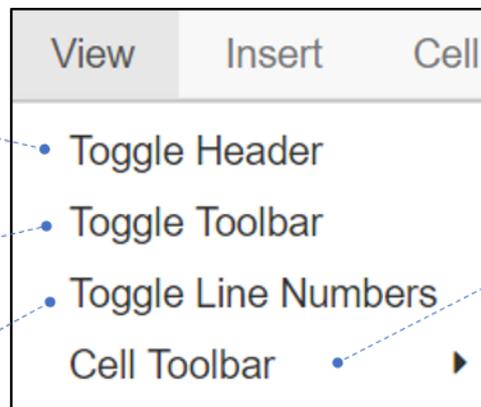執行目前Cell(s)並且移動至下方Cell

執行全部的Cells

執行目前Cell下方所有的Cells

對目前Cell輸出做下列動作
- Toggle(隱藏/顯示切換)
- Toggle Scrolling(上下切換)
- Clear(清除)

View

切換(隱藏/顯示) Jupyter notebook的logo和檔名

切換(隱藏/顯示) Jupyter notebook的工具列

在Cell裡顯示行號

View	Insert	Cell

Toggle Header

Toggle Toolbar

Toggle Line Numbers

Cell Toolbar ▶

切換Cell動作顯示
- None
- Edit Metadata
- Raw Cell Format
- Slideshow
- Attachments
- Tags

使用指南(Help)

UI瀏覽

Help

User Interface Tour — 內建快捷鍵列表

Keyboard Shortcuts

編輯內建快捷鍵 — Edit Keyboard Shortcuts

Notebook Help — Notebook使用指南

在notebook中可用的markdown敘述 — Markdown

Python — Python使用指南

IPython使用指南 — IPython

NumPy — NumPy使用指南

SciPy使用指南 — SciPy

Matplotlib — Matplotlib使用指南

SymPy使用指南 — SymPy

pandas — pandas使用指南

關於Jupyter Notebook — About

小工具(Widgets)

儲存與Widget互動的Notebook

下載包含所有目前正在使用的
Widget序列化狀態的JSON文件

Widgets	Help

• Save Notebook Widget State
Clear Notebook Widget State

• Download Widget State
Embed Widgets

從Notebook中清除掉Metadata

嵌入目前的Widgets
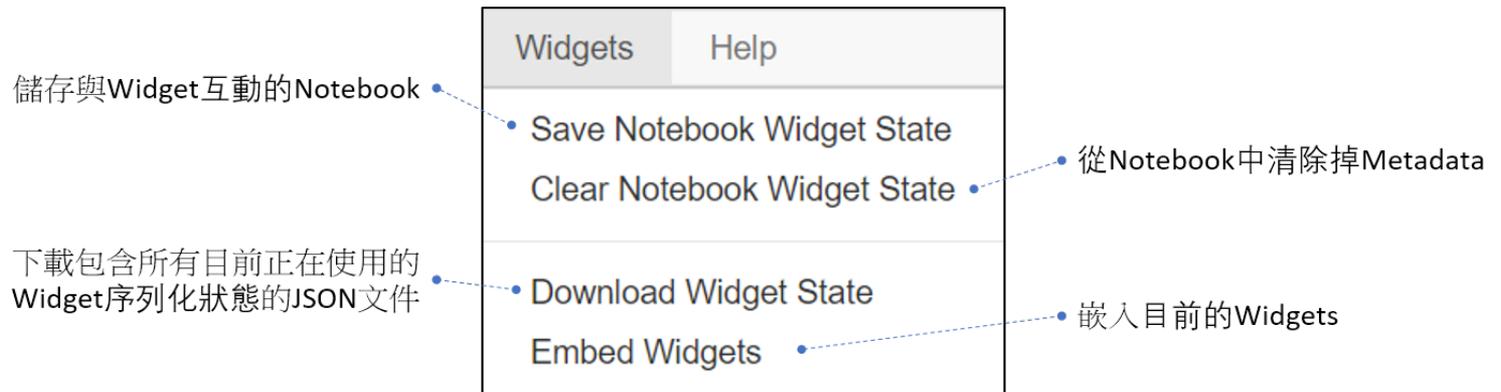
與不同的程式語言溝通

› Kernel主要提供與前端(front-end)介面間的計算及通信，其中有三個主要的Kernels如下

IPython

IRkernel

IJulia

與不同的程式語言溝通

› 安裝Jupyter Notebook時將會自動安裝IPython kernel。其它介面功能介紹如下圖

中斷kernel

重啟kernel

重啟kernel並清除所有輸出結果

重啟kernel並執行所有cells

重新連結kernel

關閉目前正在執行的kernel

切換到其它有安裝的kernel
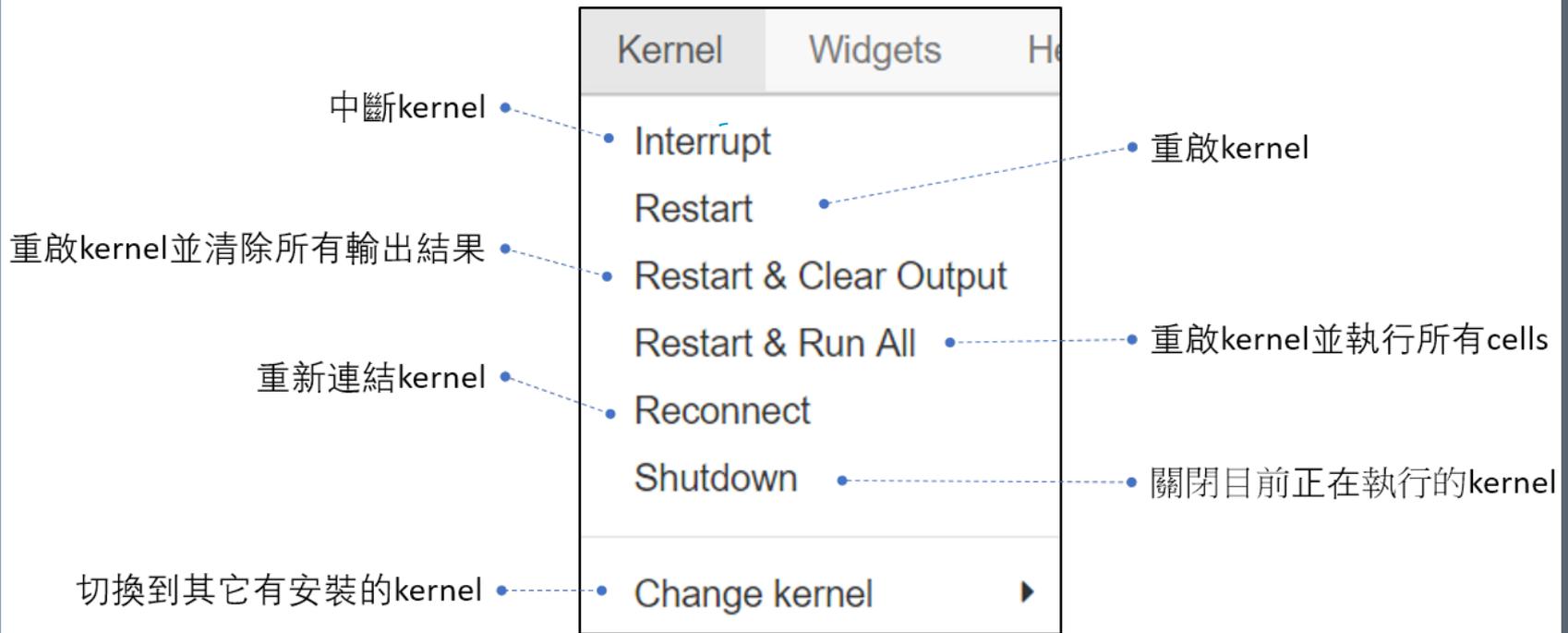
Jupyter notebook 常用快捷鍵

› 在cell旁邊為藍色時
  – 按下x：刪除當前選擇的cell
  – 按下a：在當前選擇的上方新增一個cell
  – 按下b：在當前選擇的下方新增一個cell
  – 按下Shift-Enter：執行當前的cell並且選到下一個cell
  – 按下Ctrl-Enter：執行當前cell
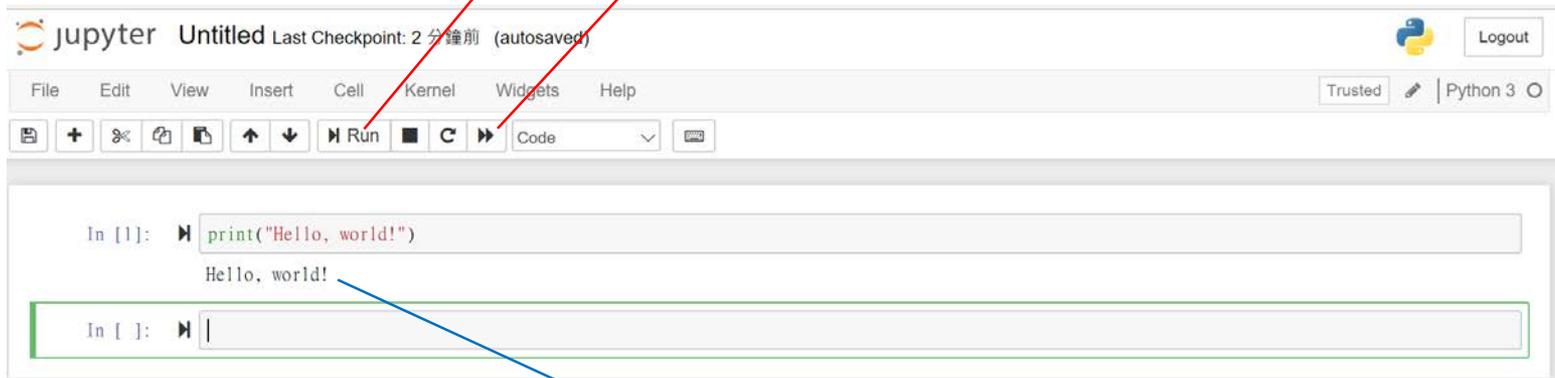 – 按下M：轉成markdown模式，可以看到紅色框框內容從code變成markdown

› 更多文件說明
  – https://jupyter.readthedocs.io/en/latest/

程式設計的第一個作業
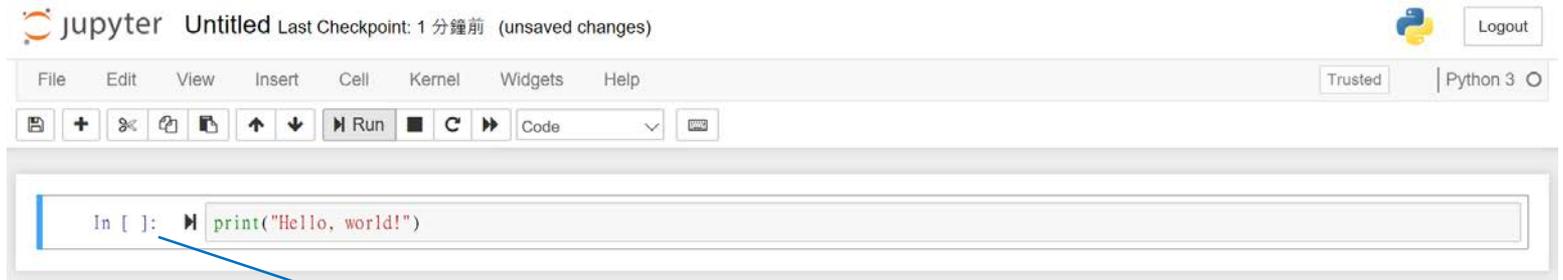
將程式碼打入In[*] 格子內

如何執行程式

按下 Run, 會執行目前格子內的程式

按下這個會整個文件重新執行

輸出結果就會在這邊

Visual studio code (VSCode)

VSCode是什麼

> **Visual Studio Code**（簡稱**VS Code**）是一個由微軟開發的，同時支援Windows、Linux和macOS作業系統且開放原始碼的文字編輯器，它支援偵錯，並內建了Git 版本控制功能，同時也具有開發環境功能，例如程式碼補全（類似於 IntelliSense）、程式碼片段和程式碼重構等，該編輯器支援用戶自訂組態，例如改變主題顏色、鍵盤捷徑等各種屬性和參數，還在編輯器中內建了擴充程式管理的功能。
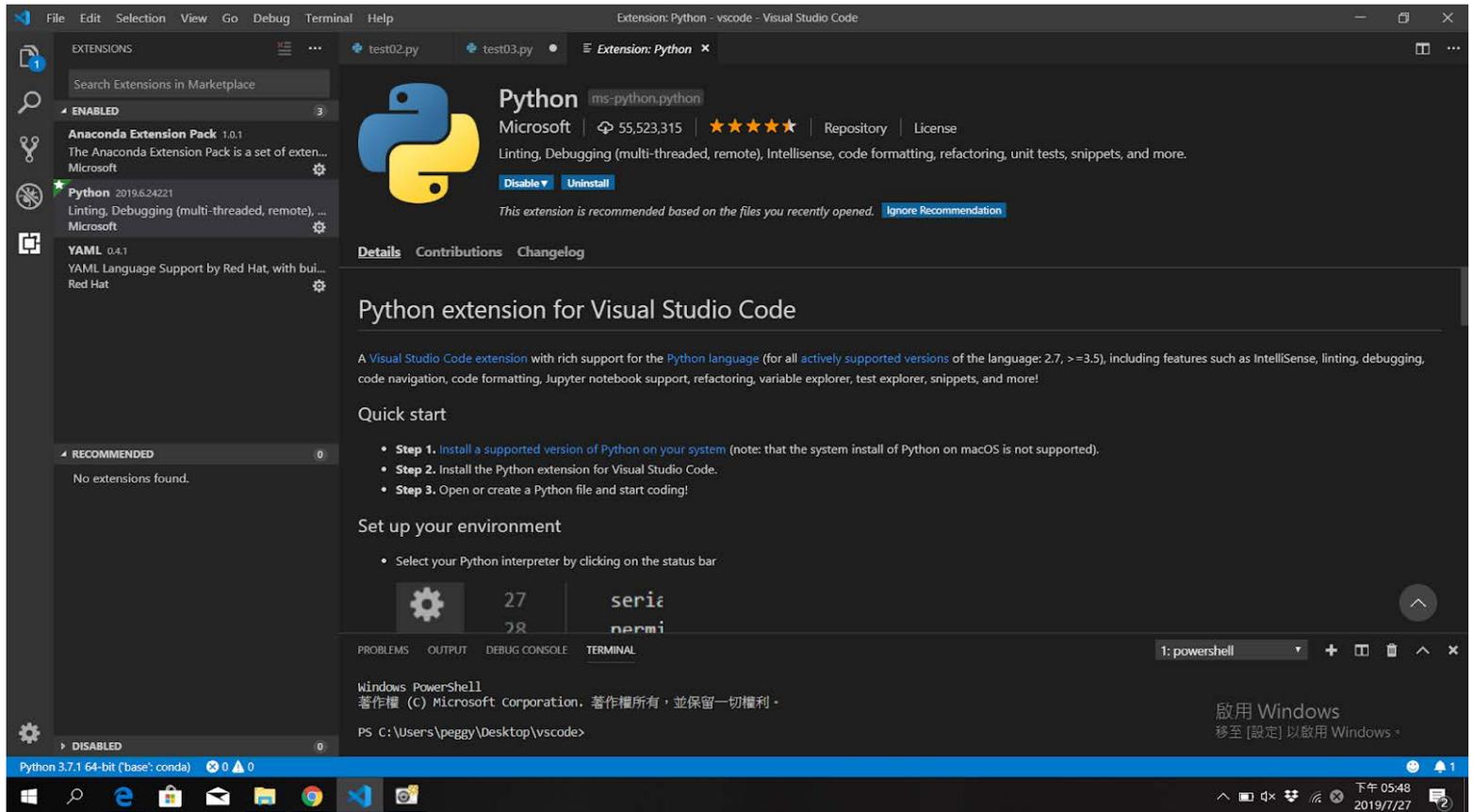
下載地點

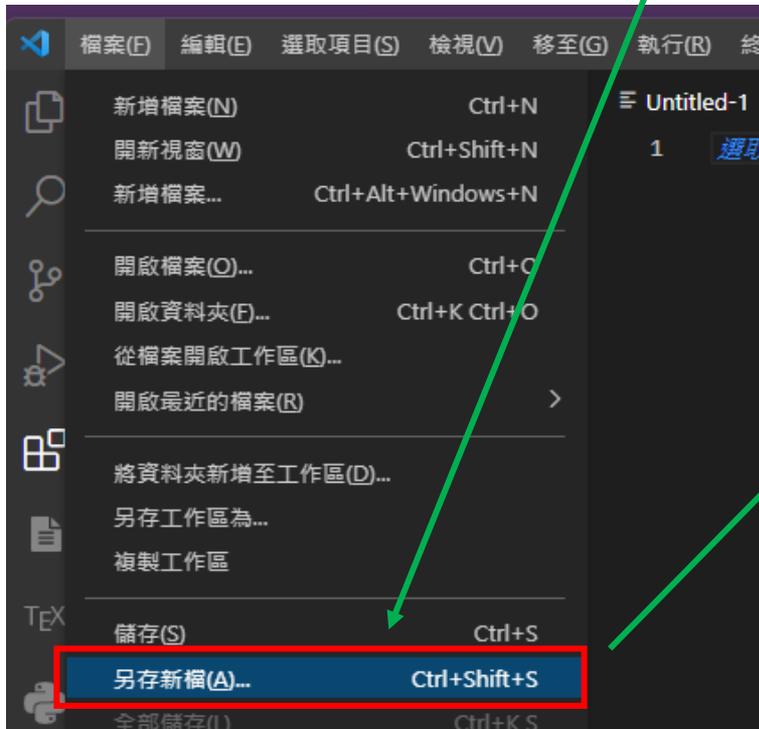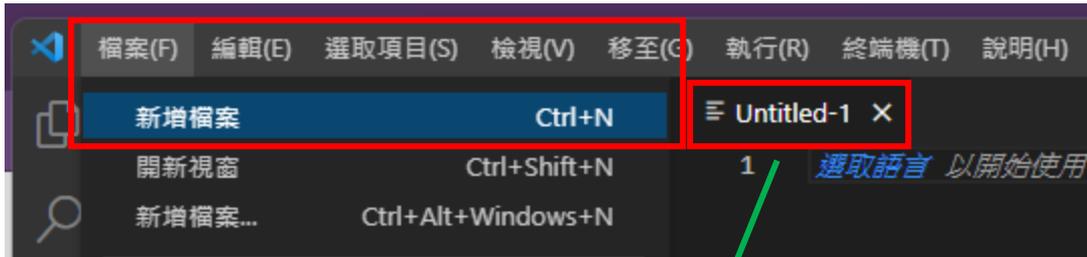> https://code.visualstudio.com/
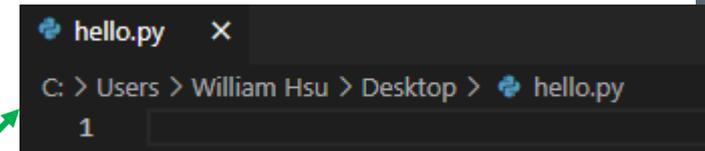
主畫面

設定VSCode

> › VSCode只是個編輯器，需要配合擴充套件才能處理各種語言

> › 安裝python擴充套件

第一個程式

存成某個檔名，副檔名用.py

執行他

› 把第一個程式打進去

› 未來記得隨時儲存自己的程式

電子化評測

電子化評測

› 本課程90% 作業與考試都由電腦自動化批改
  – 由你們上傳程式碼
  – 電腦會自動編譯然後將隱藏測資拿來驗證你的程式正確性
    › 有可能是設定好的, 也有可能是另一隻老師寫的程式產生的
  – 會根據各種判斷來檢測程式的正確性
    › 拿到滿分不一定是滿分 (因為通常所有可能性不可能測試到, 只是我們準備的測試你都通過了)
    › 拿不到滿分一定代表你踩了某個坑了

電子化評測

› http://judge.aclab.cool/

註冊一個帳號

Register

使用者名稱

電子郵件

密碼 (?)

密碼 2 (again, for confirmation)

時區 (select your closest major city)

Taipei ▼

或 pick from map, detect

Default language

Python 3.8.9 ▼

Affiliated organizations

By registering, you agree to our Terms & Conditions. **Register!**

註冊帳號

› 註冊完嘗試登入，助教會協助

› 登入後加入組織 "2022

查看習題

進入題組

當時間到時候就會出現 'Join'，即可加入

第0週練習

 資料和信息   統計数据   Rankings   Participation   MOSS   編輯   複製   **Spectate contest**

Starting in 1 天 12:14:24

4 天 **05:00** long starting on 二月 24, 2022, 12:00 CST

第0週上機練習.

- 建立帳號
- 上傳程式

❷ 問題

題目	分數	AC率	用戶
現在放棄	100	0.0%	0
創造三角形 (II)	100p	100.0%	1

💬 評論

There are no comments at the moment.

New comment

| B | *I* | x^2 | *f* |

閱讀題目與解題

上傳解答按這邊

現在放棄 🔒 2020程式設計與資料處理

當你不想上課時候, 應該怎麼做呢?

那就寫一個程式, 印出 現在放棄暑假就開始了!

輸入

沒有任何輸入資料

輸出

現在放棄暑假就開始了!

💬 評論

There are no comments at the moment.

New comment

B I x² f 🌐 👤 66 ¹⁰¹ 🖼 📋 📋 ≡ ≡

📄 View as PDF

Submit solution

全部的提交紀錄
Best submissions

Manage tickets
Edit problem
Edit test data
Manage submissions
Clone problem

✔ 分數: 100
⏱ 時間限制: 1.0s
☰ 記憶體限制: 129M

📝 作者:
wwyhsu

❯ 題目類型

✔ 允許的語言
C, C++, Python, ~~ZHG~~

🖥 裁判服務器:
DeepSea, FarSea,
NearSea, OverSea,
RoughSea, SmoothSea,
UnderSea

Report an issue

題目說明，包含故事，目的，以及範例輸入資料和輸出資料供參考測試

電子評測的格式要求很嚴謹，要一模一樣才會通過

上傳題解

Submit to 現在放棄
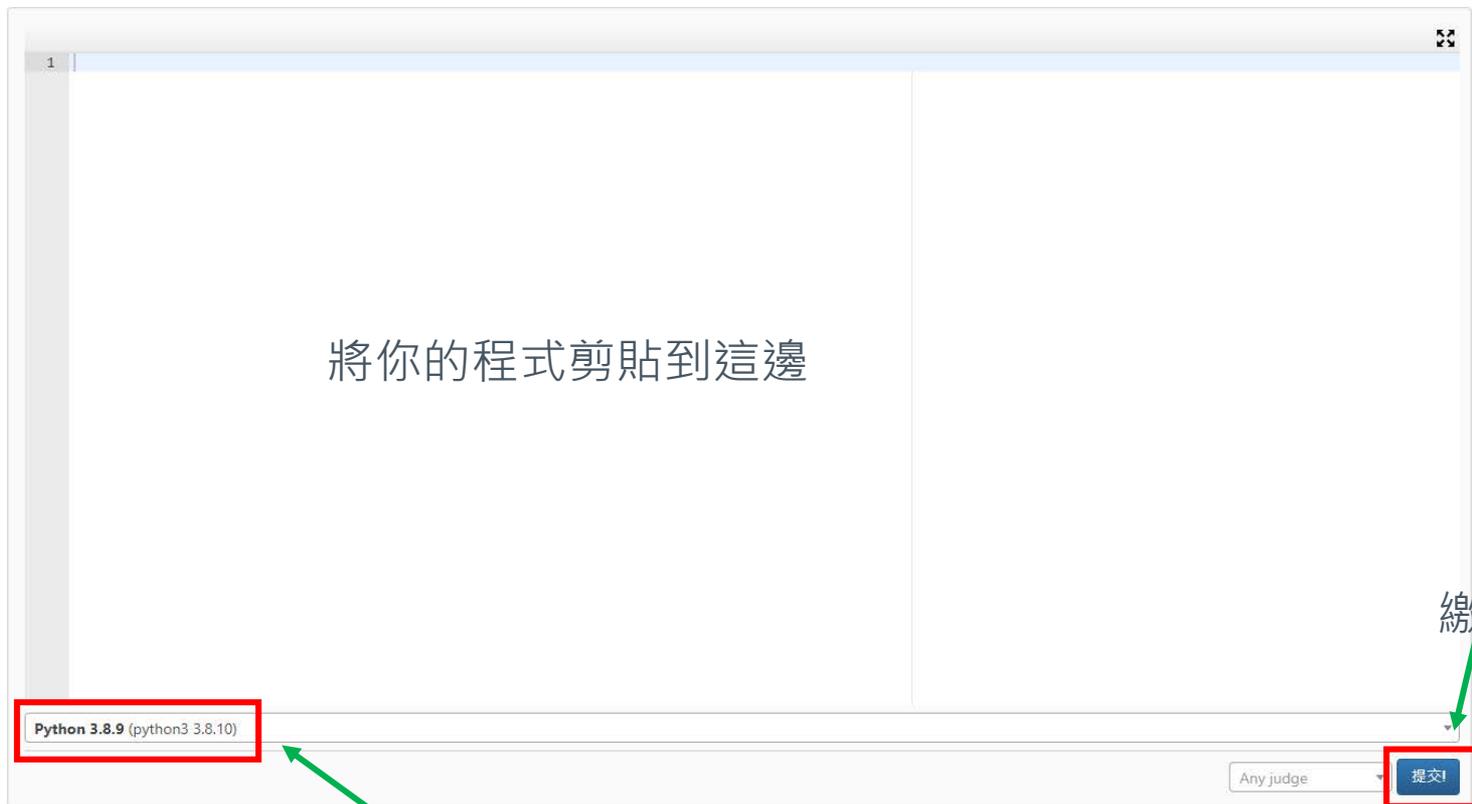
將你的程式剪貼到這邊

Python 3.8.9 (python3 3.8.10)

Any judge

提交!

繳交

記得選對語言

觀察評測結果

Submission of 現在放棄 by wwyhsu

View source
重新提交
重新評分

Execution Results

✔

Test case #1: AC [0.043s,9.14 MB] (100/100)

Resources: 0.043s, 9.14 MB
Maximum runtime on single test case: 0.043s
Final score: 100/100 (100.0/100 points)

› 大家想看到的就是AC
  – AC: 正確
  – WA: 答案錯誤
  – IR: 語法有誤
    › 程式執行碰到錯誤
  – TLE: 程式執行超過時間限制
  – MLE:程式執行超過記憶體限制
  – CE: 編譯/直譯錯誤