

2019/02/20

# Weka安裝及補充教學



授課老師：林熙禎 教授

助教：

◦ 麥嘉芳 [may09061122@g.ncu.edu.tw](mailto:may09061122@g.ncu.edu.tw)

◦ 林佳蒼 [georgelin@g.ncu.edu.tw](mailto:georgelin@g.ncu.edu.tw)

# Outline

---

1. Weka 介紹與安裝說明
2. Weka - view data set
3. Weka - explorer
4. Weka 補充教學

ECT  
電子商務技術

# Weka介紹與安裝說明

# Weka介紹

---

- 紐西蘭Waikato大學所開發
- 包含常用的資料探勘的演算法及方法
- 對資料集可以進行前處理、分類、分群以及關聯規則，並可以比較不同演算法的處理結果
- Weka是以java為基礎，使用者可以用Java來撰寫演算法

# 安裝Windows

- 安裝連結：

<http://www.cs.waikato.ac.nz/ml/weka/downloading.html>

請安裝weka 3.8

- 依照作業系統及本機中是否包含jre選擇安裝檔類型

## Windows

Click **here** to download a self-extracting executable for 64-bit Windows that includes Oracle's 64-bit Java VM 1.8

(weka-3-8-2jre-x64.exe; 265.4 MB)

Click **here** to download a self-extracting executable for 64-bit Windows without a Java VM

(weka-3-8-2-x64.exe; 50.8 MB)

Click **here** to download a self-extracting executable for 32-bit Windows that includes Oracle's 32-bit Java VM 1.8

(weka-3-8-2jre.exe; 257.2 MB)

Click **here** to download a self-extracting executable for 32-bit Windows without a Java VM

(weka-3-8-2.exe; 50.8 MB)

These executables will install Weka in your Program Menu. Download the version without the Java VM if you already have Java 1.7 (or later) on your system.

# 安裝Mac

---

- 安裝連結：  
<http://www.cs.waikato.ac.nz/ml/weka/downloading.html>  
請安裝weka 3.8
- 依照作業系統及本機中是否包含jre選擇安裝檔類型

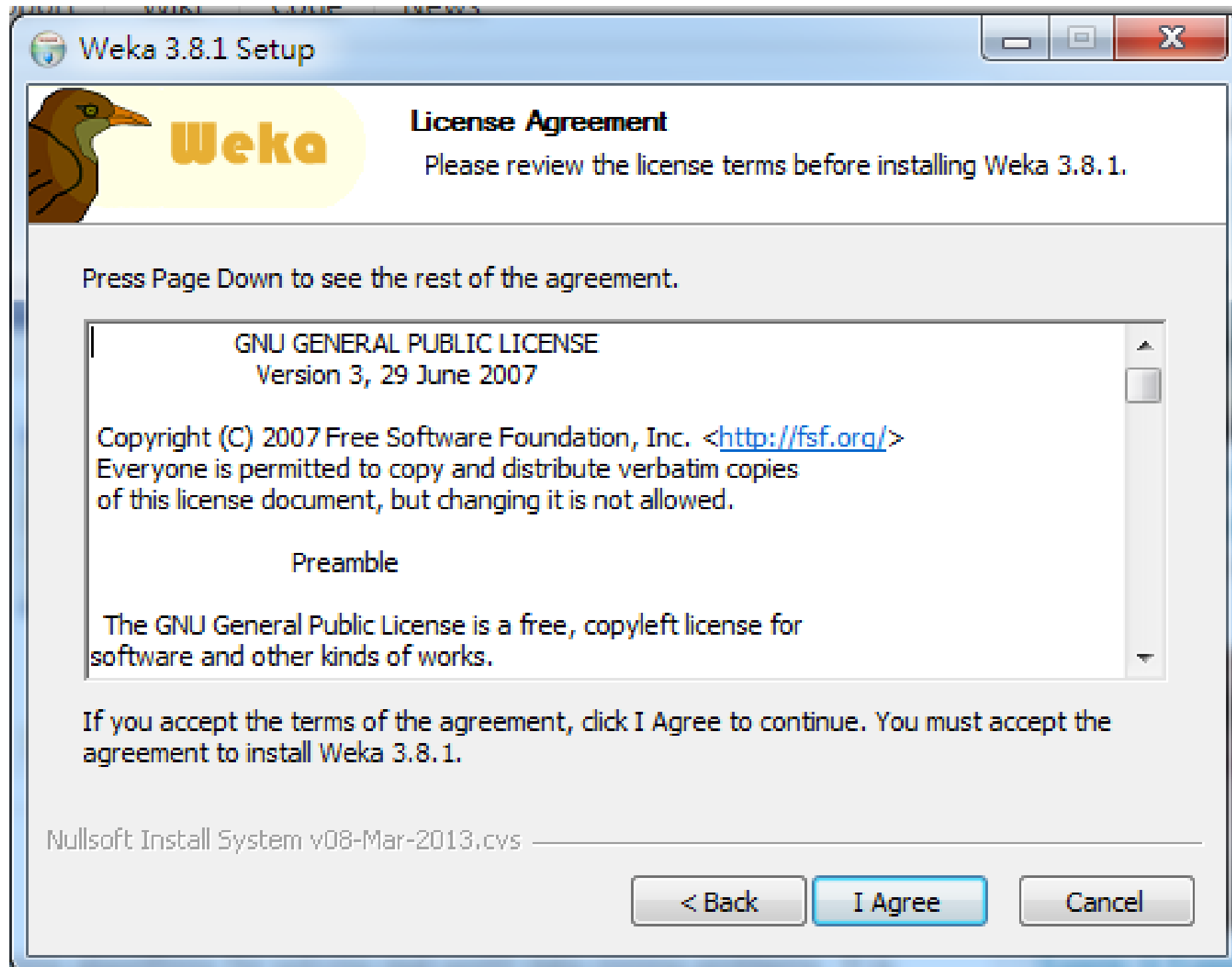
## Mac OS X

Click **here** to download a disk image for OS X that contains a Mac application including Oracle's Java 1.8 JVM (weka-3-8-2-oracle-jvm.dmg; 124.2 MB)

# Weka安裝步驟(1/8)

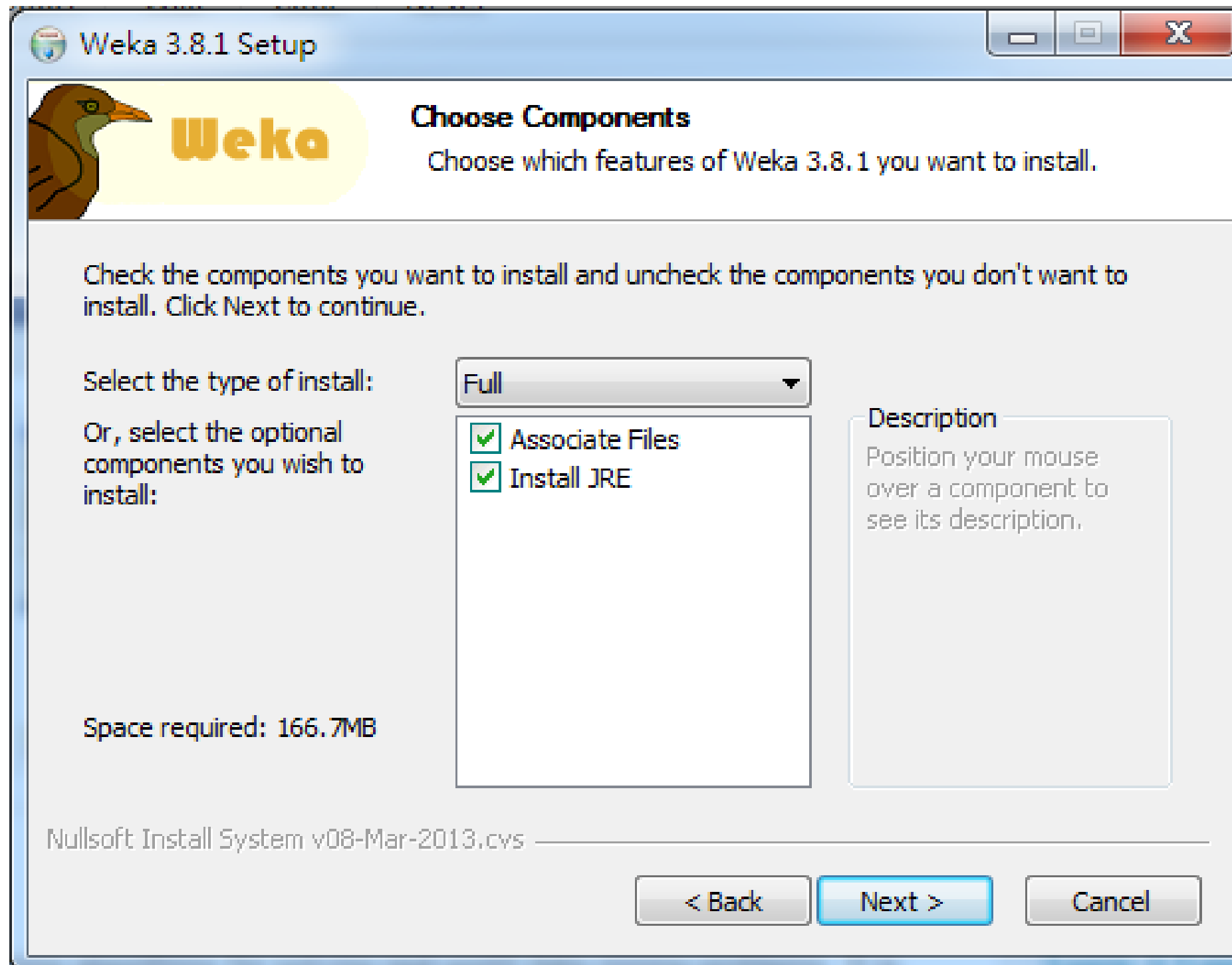


# Weka安裝步驟(2/8)

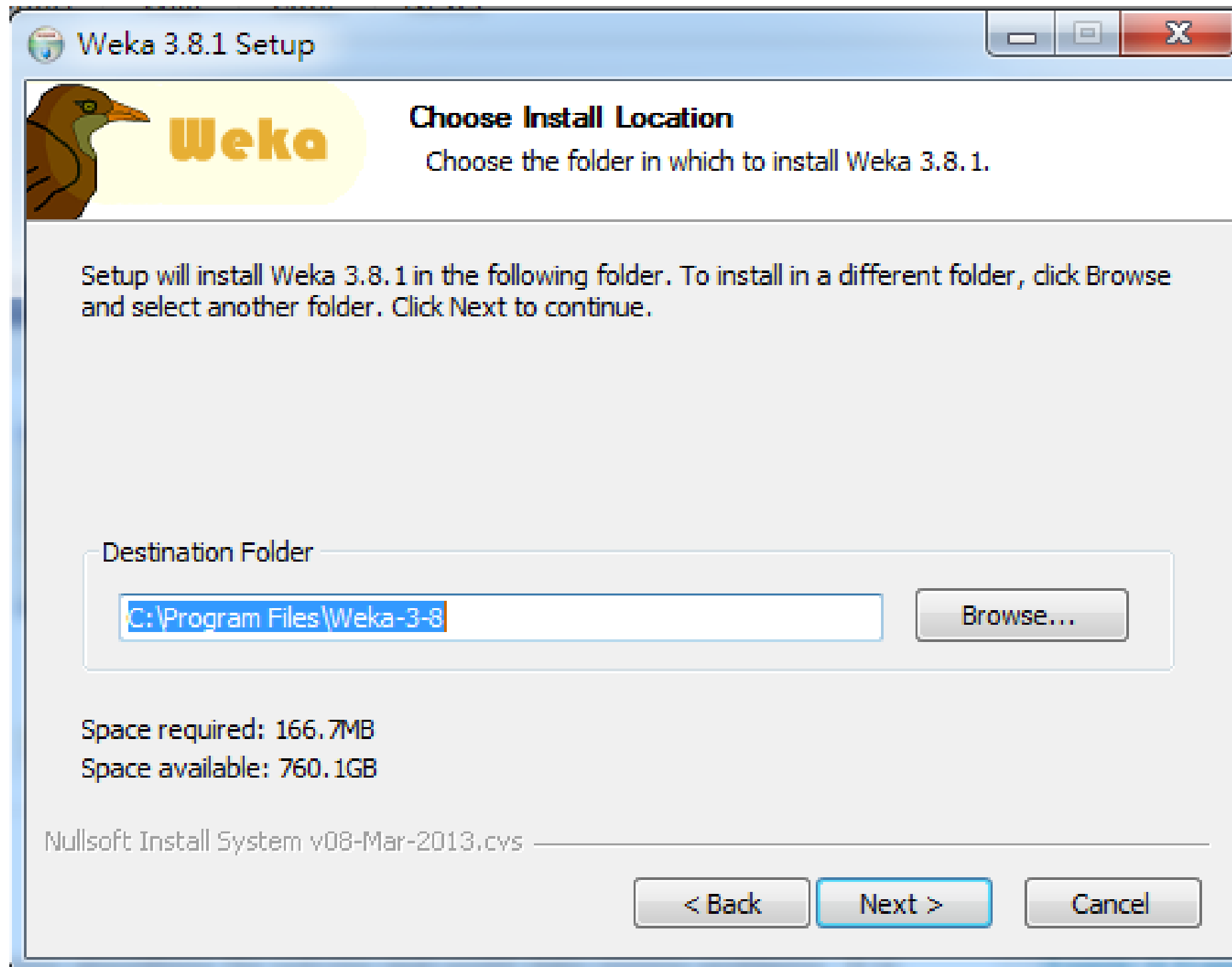




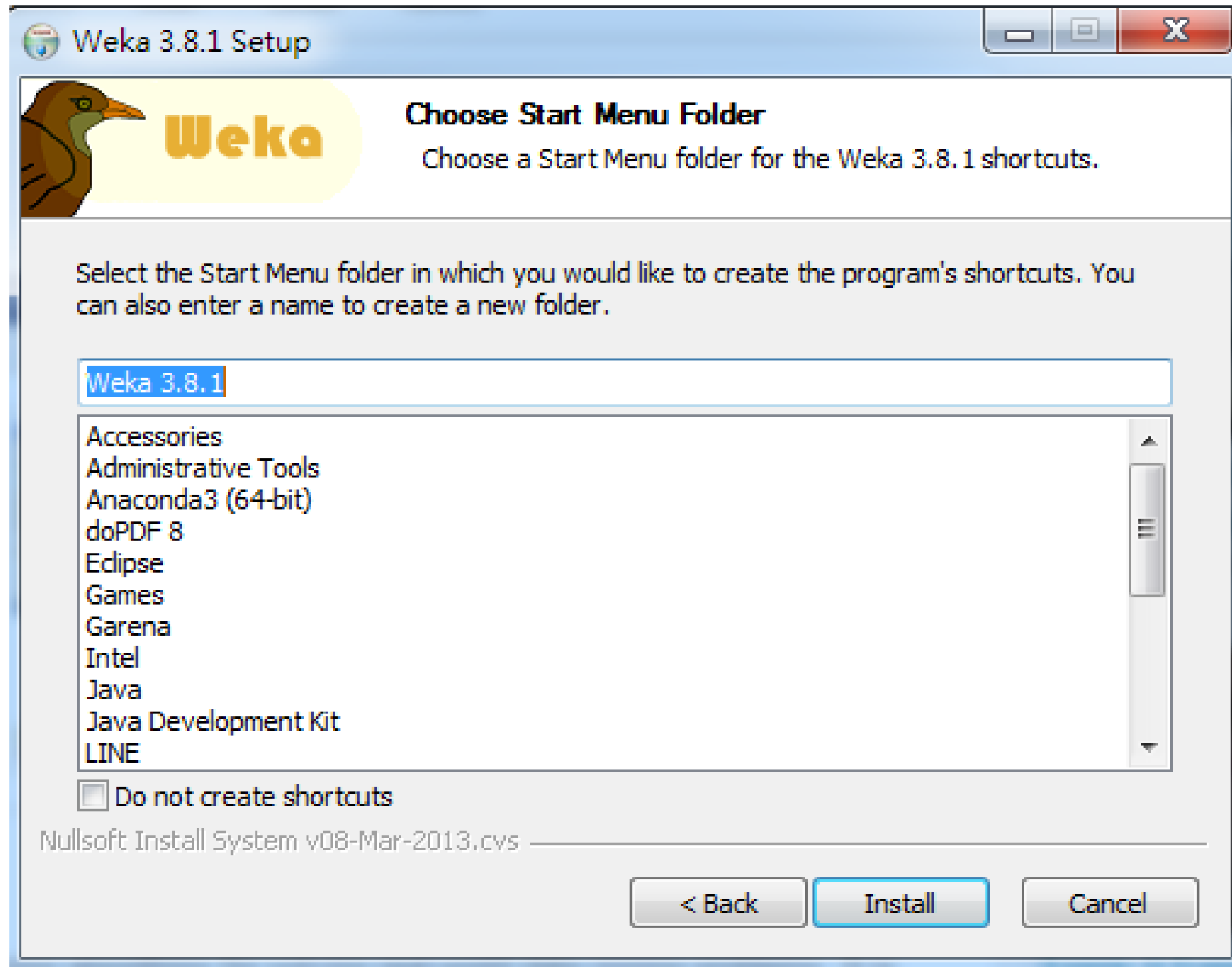
# Weka安裝步驟(3/8)



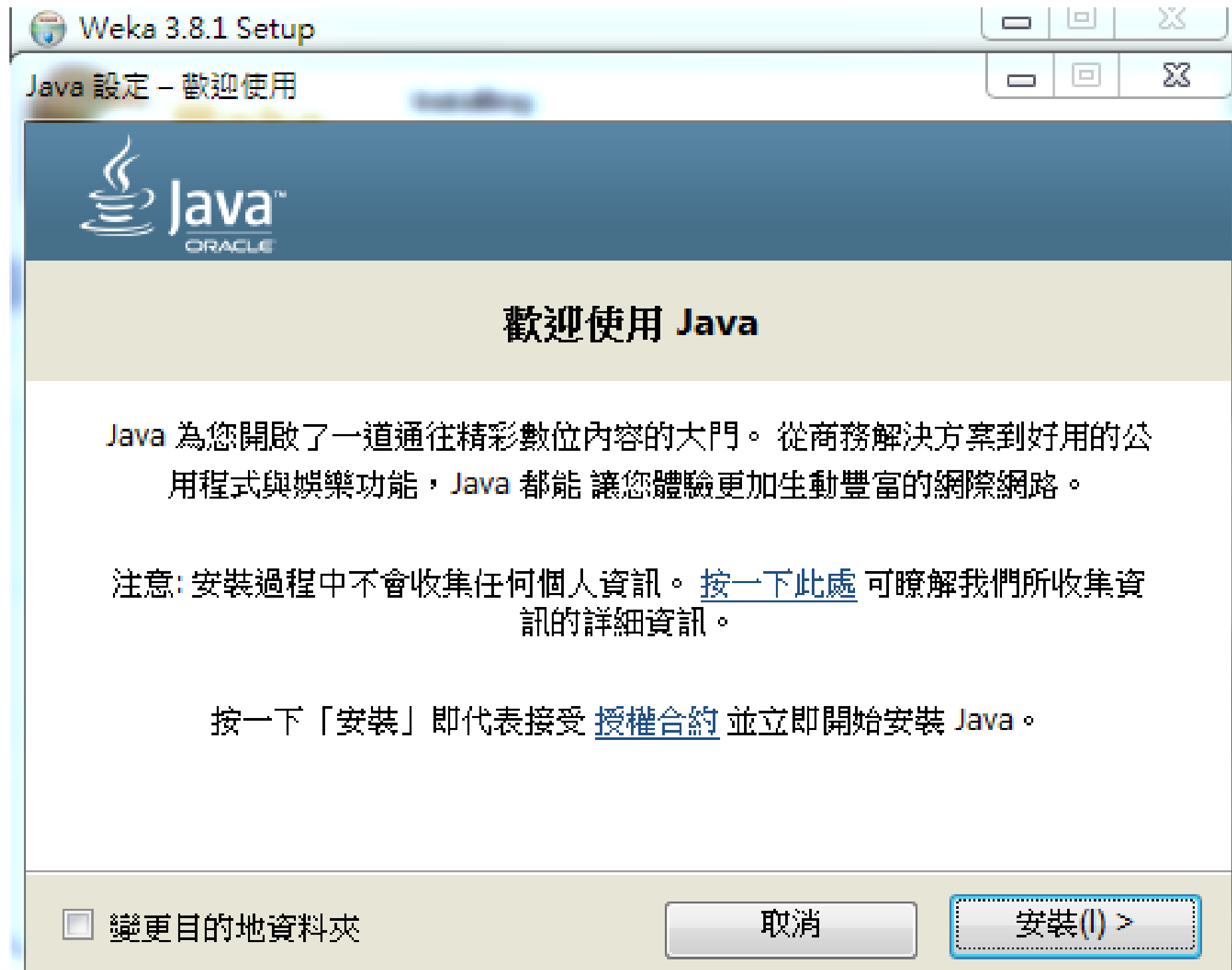
# Weka安裝步驟(4/8)



# Weka安裝步驟(5/8)



# Weka安裝步驟(6/8)



# Weka安裝步驟(7/8)



# Weka安裝步驟(8/8)



---

Weka - view data set

# Weka 預設資料集

電腦 > 本機磁碟(C:) > Program Files > Weka-3-8

The image displays two side-by-side Windows Explorer windows. The left window shows the directory structure of the Weka-3-8 installation, with the 'data' folder highlighted. The right window shows the contents of the 'data' folder, which contains various ARFF data files.

**Left Window: Weka-3-8 Directory Structure**

名稱	修改日期	類型	大小
changelogs	2017/2/13 上午 1...	檔案資料夾	
<b>data</b>	2017/2/13 上午 1...	檔案資料夾	
doc	2017/2/13 上午 1...	檔案資料夾	
COPYING	2016/12/19 上午 ...	檔案	35 KB
documentation.css	2016/12/19 上午 ...	Cascading Style ...	1 KB
documentation.html	2016/12/19 上午 ...	Chrome HTML D...	2 KB
README	2016/12/19 上午 ...	檔案	16 KB
remoteExperimentServer.jar	2016/12/19 上午 ...	Executable Jar File	43 KB
RunWeka.bat	2016/12/19 上午 ...	Windows 批次檔案	1 KB
RunWeka.class	2016/12/19 上午 ...	CLASS 檔案	5 KB
RunWeka.ini	2016/12/19 上午 ...	組態設定	3 KB
uninstall.exe	2017/2/13 上午 1...	應用程式	56 KB
Weka 3.8 (with console)	2017/2/13 上午 1...	捷徑	2 KB
Weka 3.8	2017/2/13 上午 1...	捷徑	2 KB
weka.gif	2016/12/19 上午 ...	GIF 影像	30 KB
weka.ico	2016/12/19 上午 ...	圖示	351 KB
weka.jar	2016/12/19 上午 ...	Executable Jar File	10,767 KB
wekaexamples.zip	2016/12/19 上午 ...	WinRAR ZIP 壓縮檔	14,413 KB
WekaManual.pdf	2016/12/19 上午 ...	Adobe Acrobat ...	6,467 KB
weka-src.jar	2016/12/19 上午 ...	Executable Jar File	10,605 KB

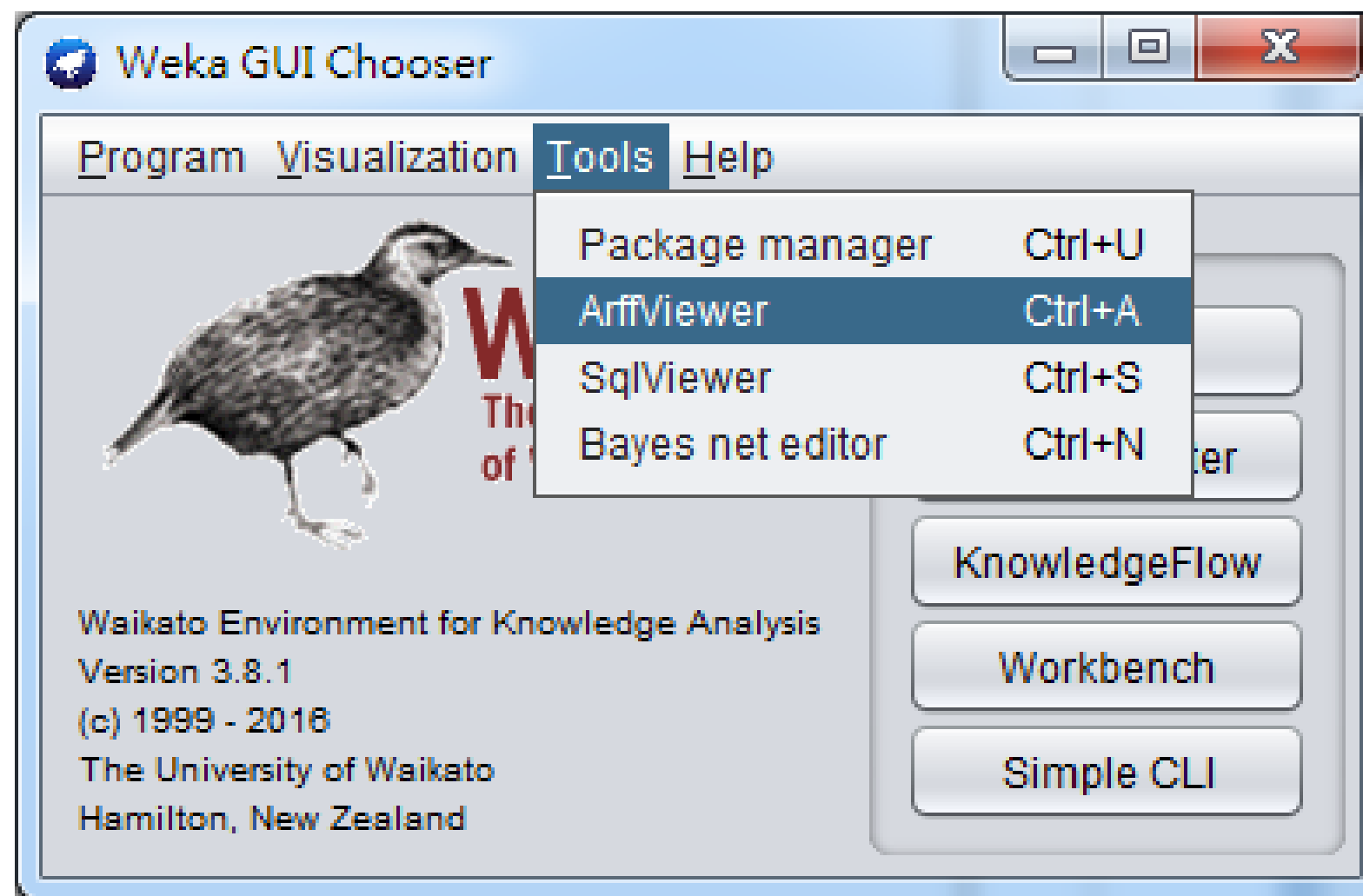
**Right Window: Weka-3-8\data Directory Contents**

名稱	修改日期	類型	大小
airline.arff	2016/12/19 上午 ...	ARFF Data File	3 KB
breast-cancer.arff	2016/12/19 上午 ...	ARFF Data File	29 KB
contact-lenses.arff	2016/12/19 上午 ...	ARFF Data File	3 KB
cpu.arff	2016/12/19 上午 ...	ARFF Data File	6 KB
cpu.with.vendor.arff	2016/12/19 上午 ...	ARFF Data File	7 KB
credit-g.arff	2016/12/19 上午 ...	ARFF Data File	159 KB
diabetes.arff	2016/12/19 上午 ...	ARFF Data File	37 KB
glass.arff	2016/12/19 上午 ...	ARFF Data File	18 KB
hypothyroid.arff	2016/12/19 上午 ...	ARFF Data File	304 KB
ionosphere.arff	2016/12/19 上午 ...	ARFF Data File	79 KB
iris.2D.arff	2016/12/19 上午 ...	ARFF Data File	4 KB
iris.arff	2016/12/19 上午 ...	ARFF Data File	8 KB
labor.arff	2016/12/19 上午 ...	ARFF Data File	9 KB
ReutersCorn-test.arff	2016/12/19 上午 ...	ARFF Data File	501 KB
ReutersCorn-train.arff	2016/12/19 上午 ...	ARFF Data File	1,198 KB
ReutersGrain-test.arff	2016/12/19 上午 ...	ARFF Data File	501 KB
ReutersGrain-train.arff	2016/12/19 上午 ...	ARFF Data File	1,198 KB
segment-challenge.arff	2016/12/19 上午 ...	ARFF Data File	196 KB
segment-test.arff	2016/12/19 上午 ...	ARFF Data File	108 KB
soybean.arff	2016/12/19 上午 ...	ARFF Data File	199 KB
supermarket.arff	2016/12/19 上午 ...	ARFF Data File	1,979 KB
unbalanced.arff	2016/12/19 上午 ...	ARFF Data File	182 KB
vote.arff	2016/12/19 上午 ...	ARFF Data File	40 KB
weather.nominal.arff	2016/12/19 上午 ...	ARFF Data File	1 KB
weather.numeric.arff	2016/12/19 上午 ...	ARFF Data File	1 KB



# Weka

## - view data set(1/2)



# Weka

## - view data set(2/2)

The screenshot displays the Weka ARFF-Viewer application. The main window has a menu bar with 'File', 'Edit', and 'View'. The 'File' menu is open, showing options like 'Open...', 'Save', 'Save as...', 'Close', 'Close all', 'Properties', and 'Exit'. A secondary window titled '開啟' (Open) is also visible, showing a list of ARFF files in the 'data' directory, with 'weather.nominal.arff' selected. Below the file list, there are fields for '檔案名稱(N):' (File name) and '檔案類型(T):' (File type), and buttons for '開啟' (Open) and '取消' (Cancel).

The 'weather.nominal.arff' window shows the following data:

Relation: weather.symbolic

No. 1: outlook 2: temperature 3: humidity 4: windy 5: play

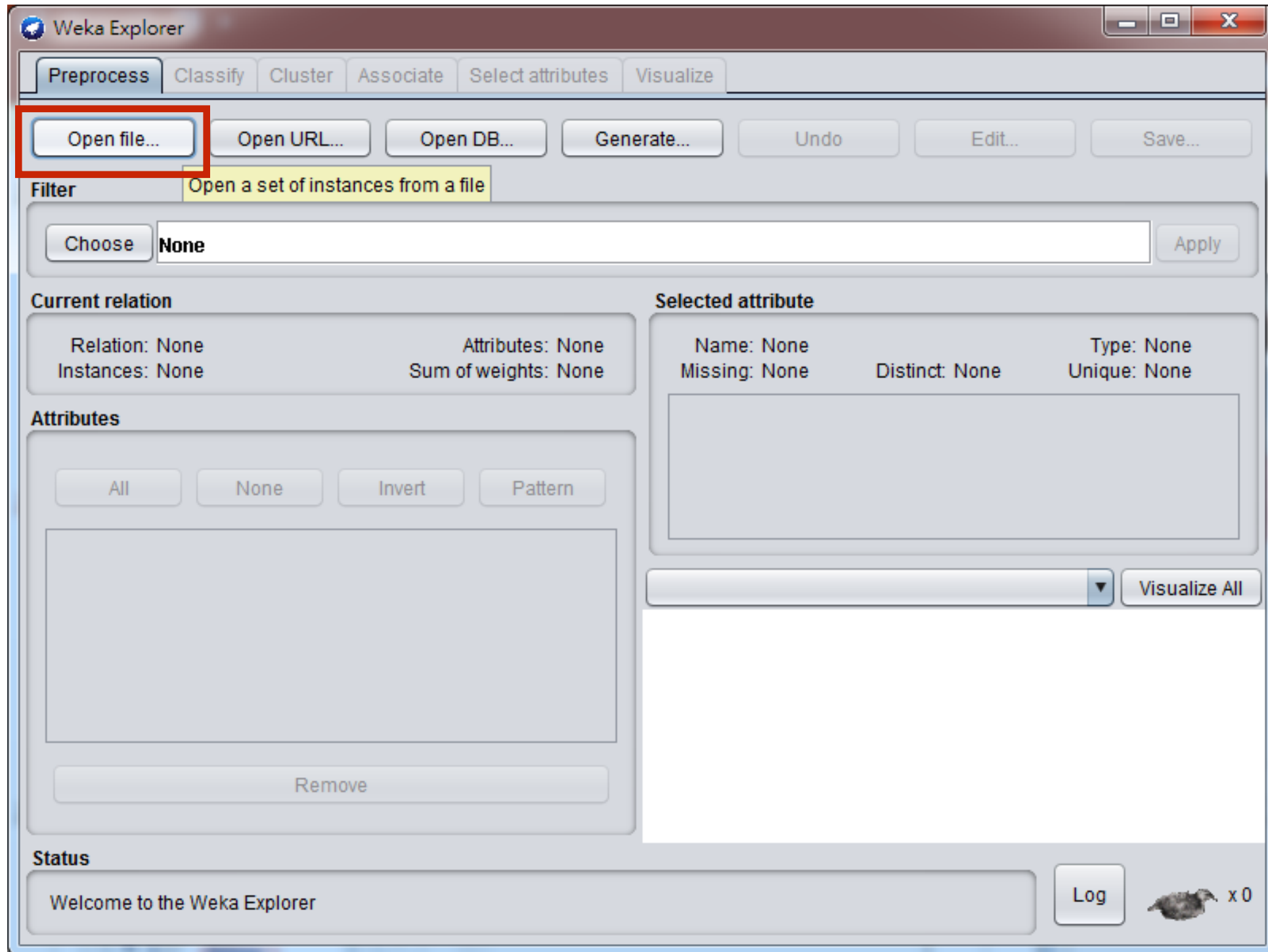
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

# Weka - Explorer

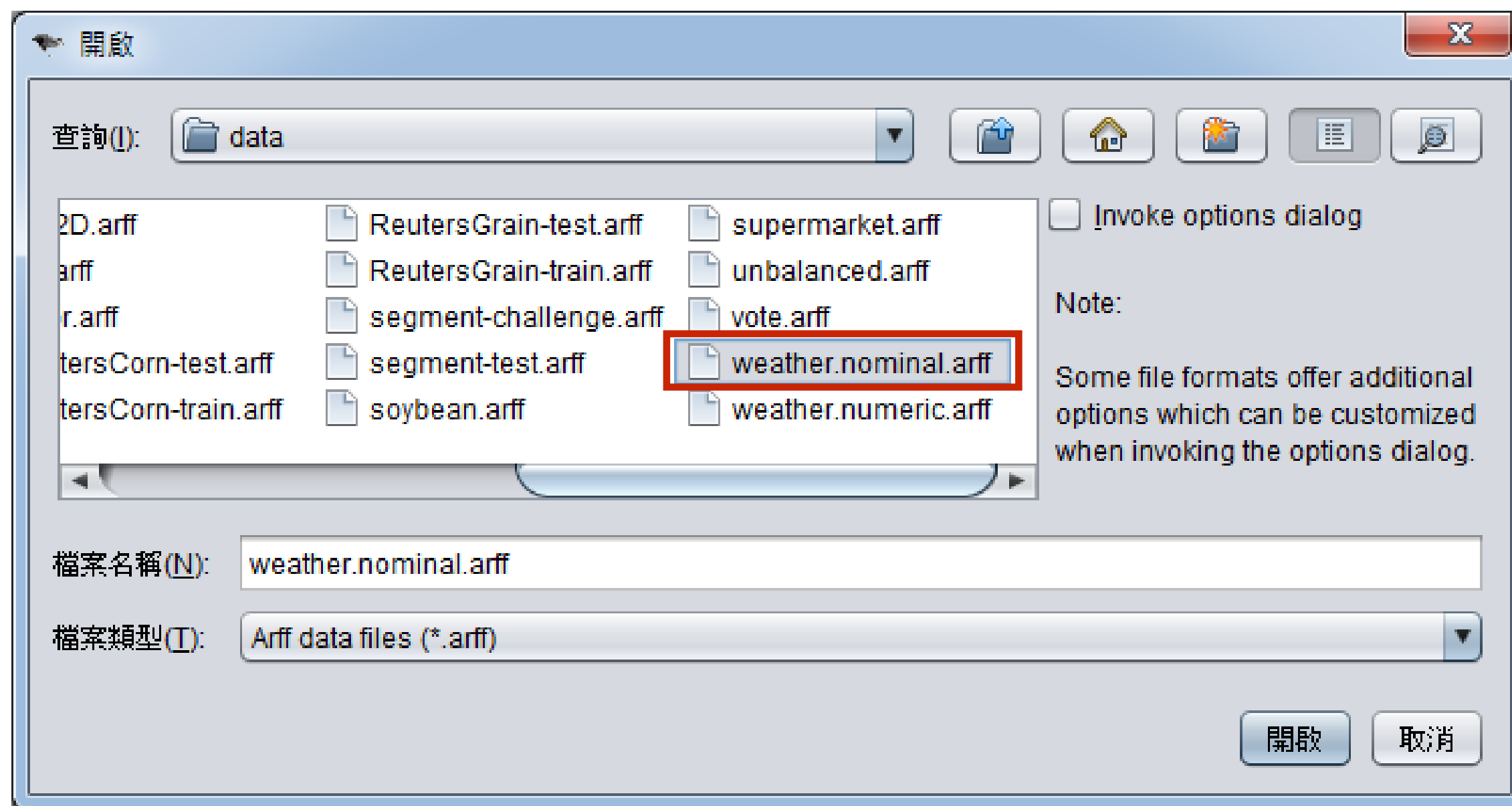
# Weka-explorer(1/5)



# Weka-explorer(2/5)



# Weka-explorer(3/5)



# Weka-explorer(4/5)

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter

Choose None Apply

Current relation

Relation: weather.symbolic  
Instances: 14  
Attributes: 5  
Sum of weights: 14

Attributes

All None Invert Pattern

No.	Name
1	outlook
2	temperature
3	humidity
4	windy
5	play

Remove

Selected attribute

Name: outlook  
Missing: 0 (0%)  
Distinct: 3  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

Class: play (Nom) Visualize All

Status

OK Log x 0

編輯檔案內容

Filter及參數

Data內容

# Weka-explorer(5/5)

- Preprocess：選擇和修改要處理的資料集
  - Classify：訓練和測試關於分類或回歸的方案
  - Cluster：從資料集當中學習分群
  - Associate：從資料集當中學習關聯規則
  - Select attributes：選擇資料集中最相關的欄位或屬性
  - Visualize：查看資料的互動式二維圖像
- 
- 支持的檔案格式有.arff, csv等，也支援URL或是DB的開啟方式並且支援數據轉換



# Weka - 補充教學

# Weka 補充教學

---

- 若發現 Weka3.8 有演算法缺漏問題 (找不到老師上課示範的演算法)，有兩種解決辦法：
  1. 直接刪掉，安裝 Weka3.6 版本，或同時安裝兩種版本
  2. 透過GUI的Package安裝來安裝演算法的Package

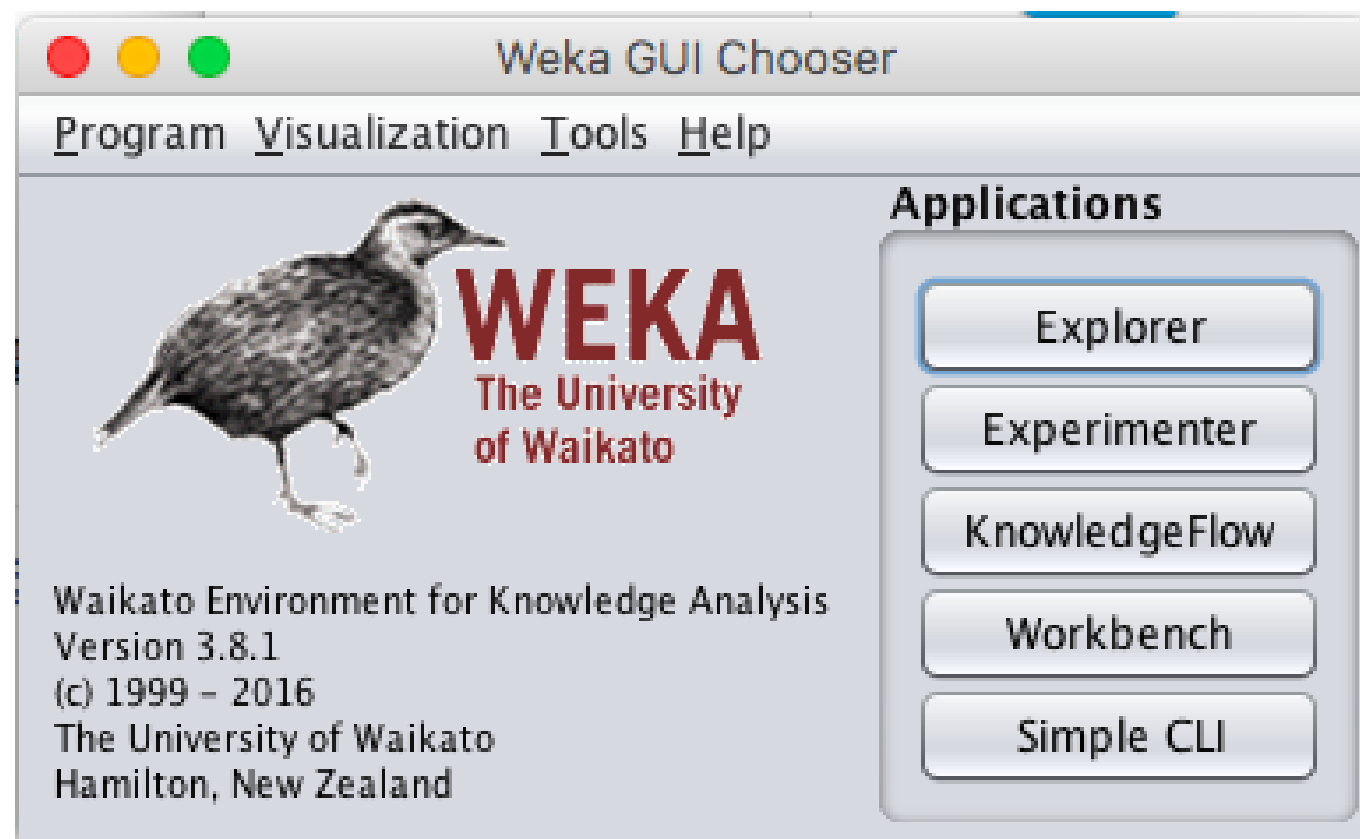
# 安裝 Weka3.6

---

- 直接使用跟熙禎老師相同的Weka 3.6 Version
  1. MacOS: <https://sourceforge.net/projects/weka/files/weka-3-6-osx/3.6.15/>
  2. Windows:  
<https://sourceforge.net/projects/weka/files/weka-3-6-windows-x64/3.6.15/>

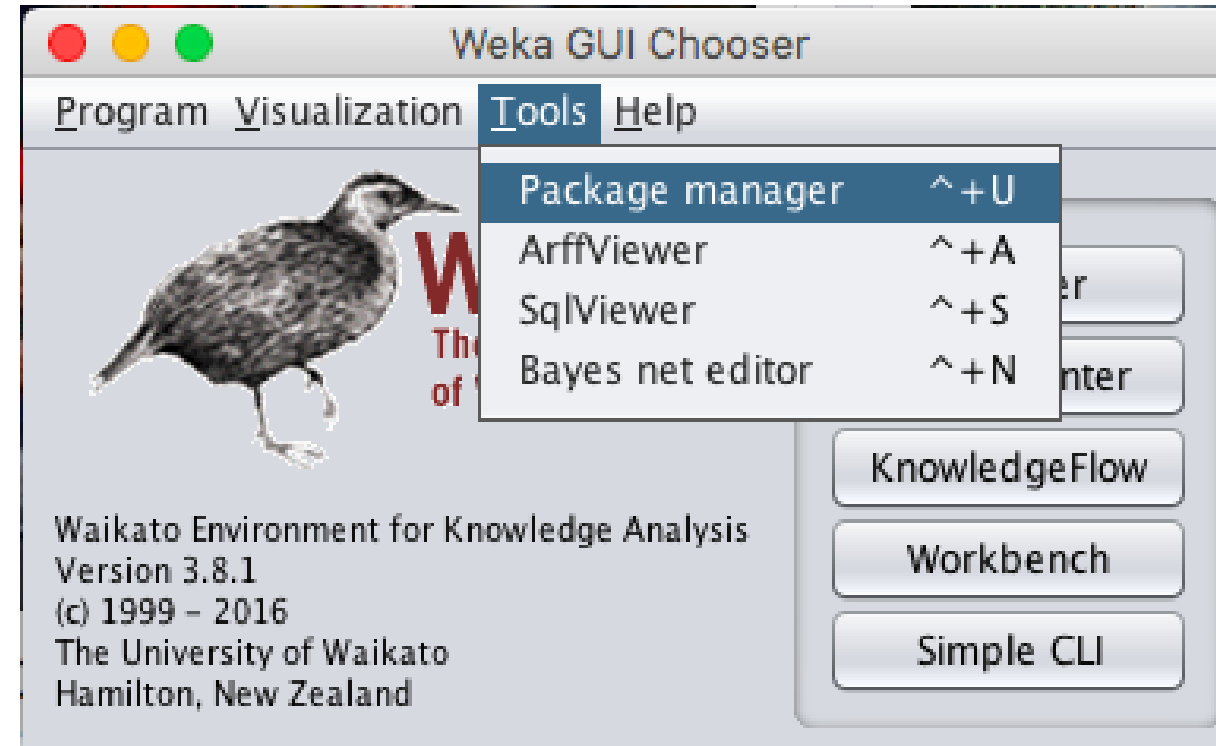
# Weka3.8安裝Package(1/4)

- 打開 Weka3.8 的GUI介面



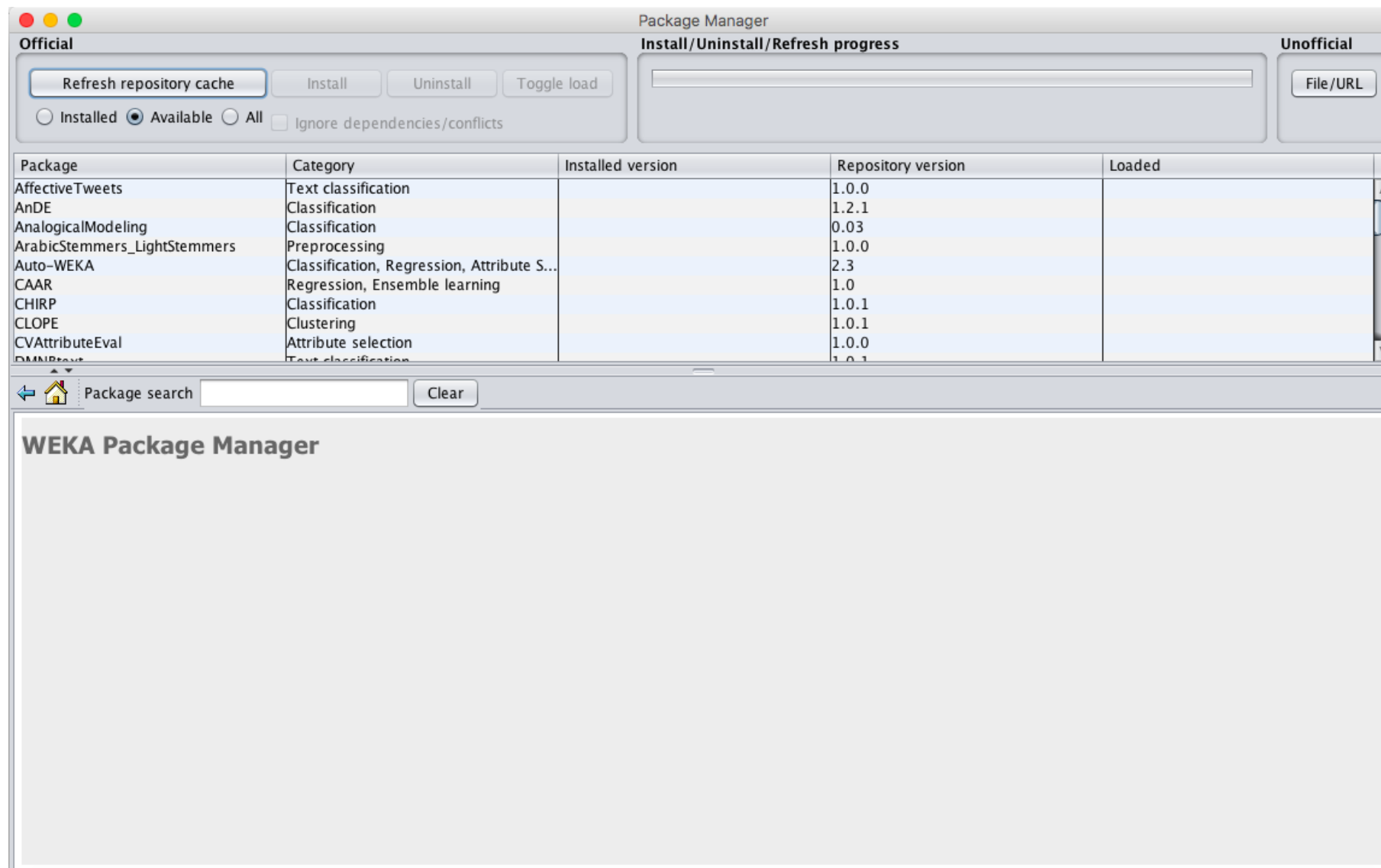
# Weka3.8安裝Package(2/4)

- 點選工具列的「Tools」的「Package manager」



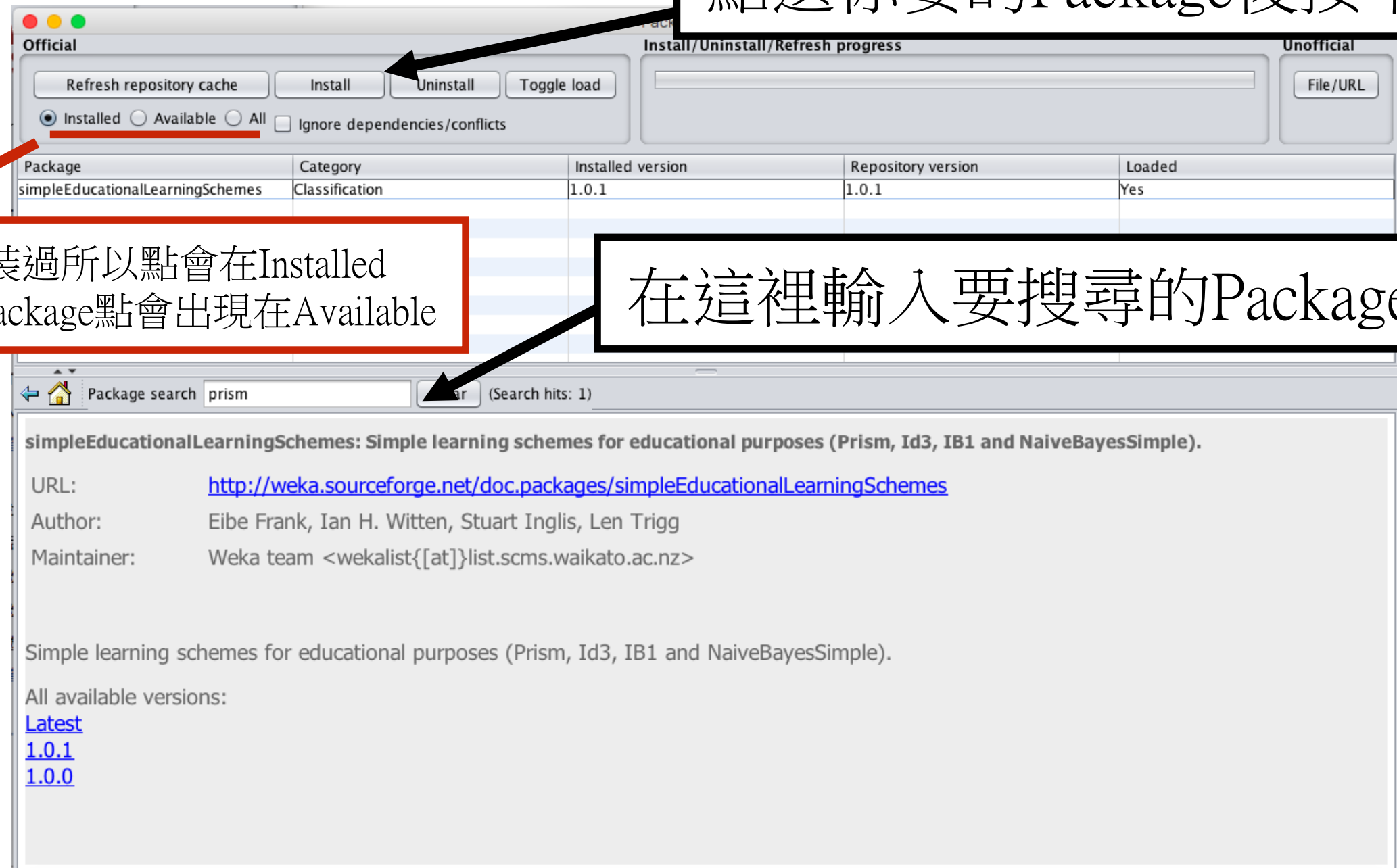
# Weka3.8安裝Package(3/4)

- 點選後如下圖



# Weka3.8安裝Package(4/4)

點選你要的Package後按下安裝



已經安裝過所以點會在Installed  
未安裝的Package點會出現在Available

在這裡輸入要搜尋的Package

2019/02/20

# Anaconda安裝與套件介紹





# Outline

---

1. Anaconda介紹與安裝說明
2. Jupyter Notebook介紹
3. Pandas套件
4. SciKit-Learn套件

ECT  
電子商務技術

# Anaconda介紹與安裝說明

# Anaconda介紹

---

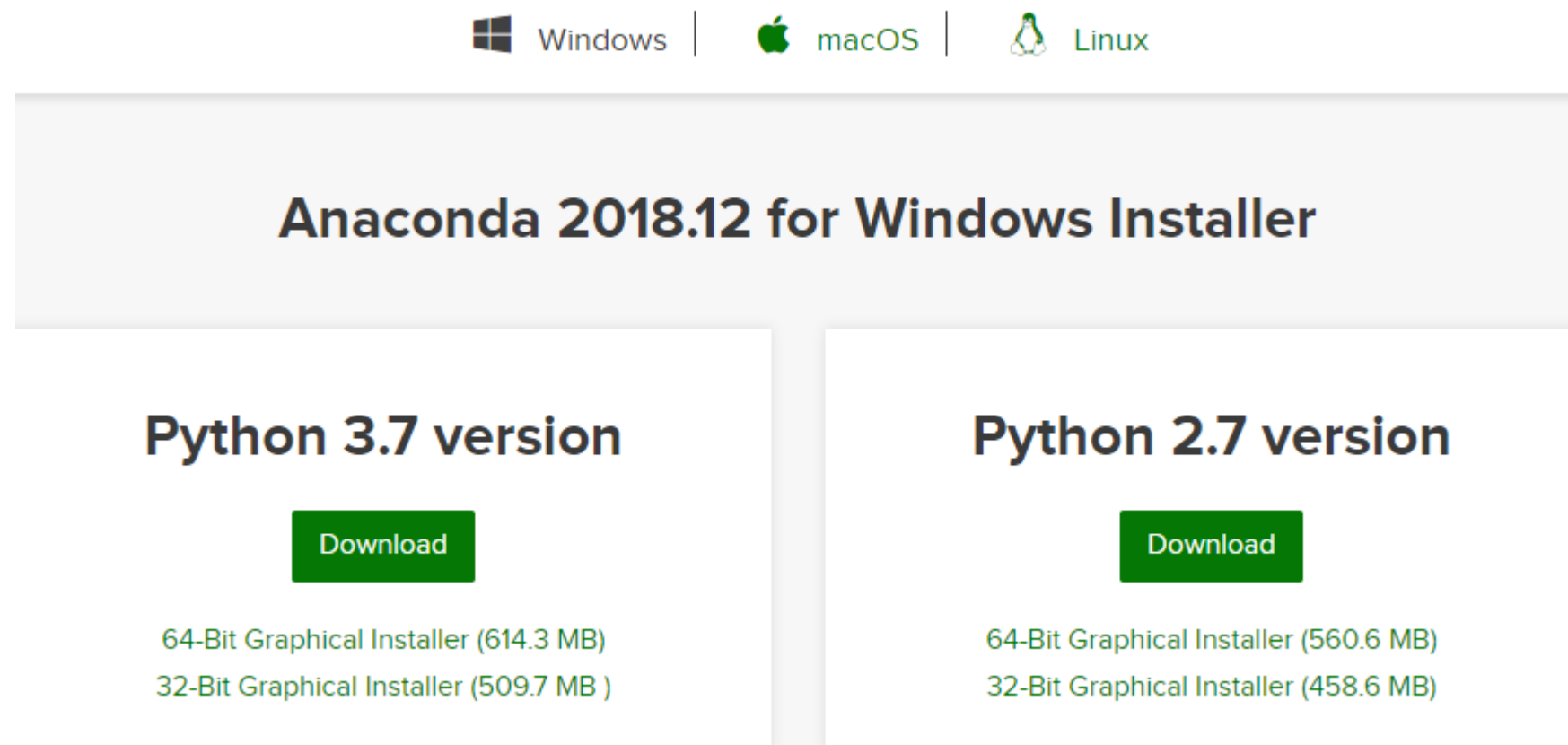
- Python的懶人包
- 安裝方便，適用於Python初學者
- 除了Python本身(python2, 3) 還包含了Python常用的資料分析、機器學習、視覺化的套件

# 安裝Windows

- 安裝連結：<https://www.anaconda.com/distribution/>

請安裝python3.7版

- 依照作業系統選擇安裝檔類型



Windows | macOS | Linux

## Anaconda 2018.12 for Windows Installer

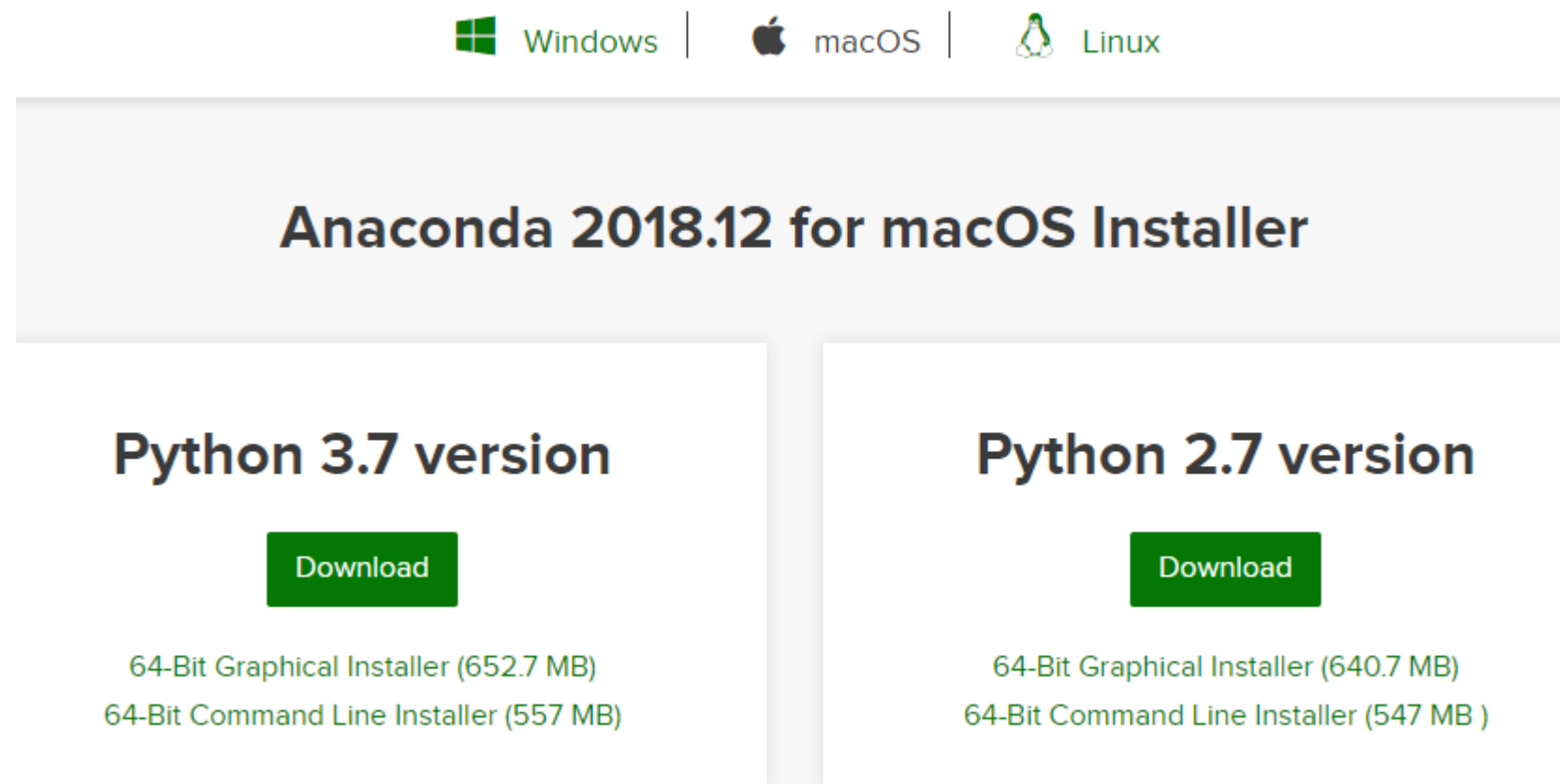
Python 3.7 version	Python 2.7 version
<a href="#">Download</a>	<a href="#">Download</a>
<a href="#">64-Bit Graphical Installer (614.3 MB)</a>	<a href="#">64-Bit Graphical Installer (560.6 MB)</a>
<a href="#">32-Bit Graphical Installer (509.7 MB)</a>	<a href="#">32-Bit Graphical Installer (458.6 MB)</a>

# 安裝Mac

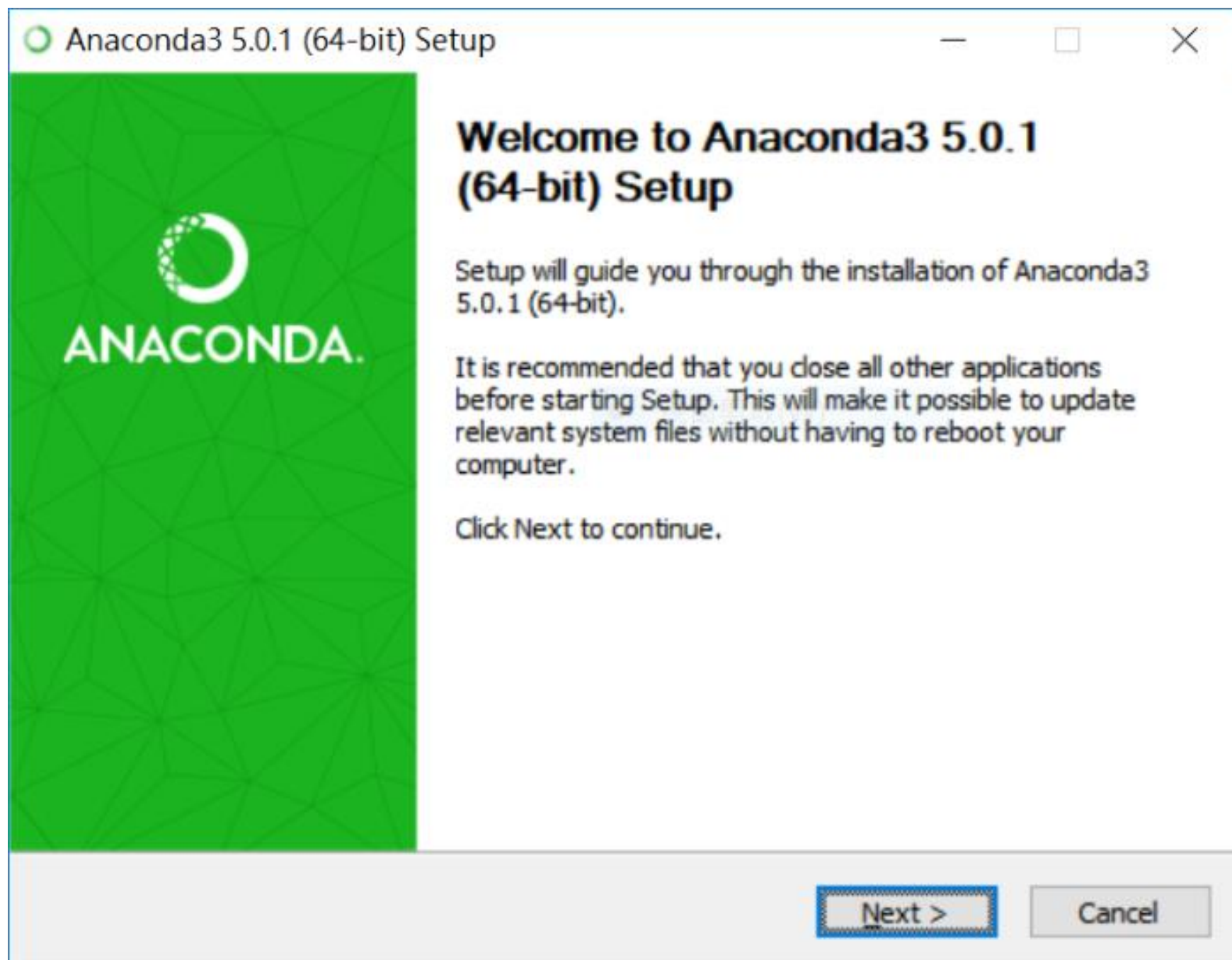
- 安裝連結：<https://www.anaconda.com/distribution/>

請安裝python3.7版

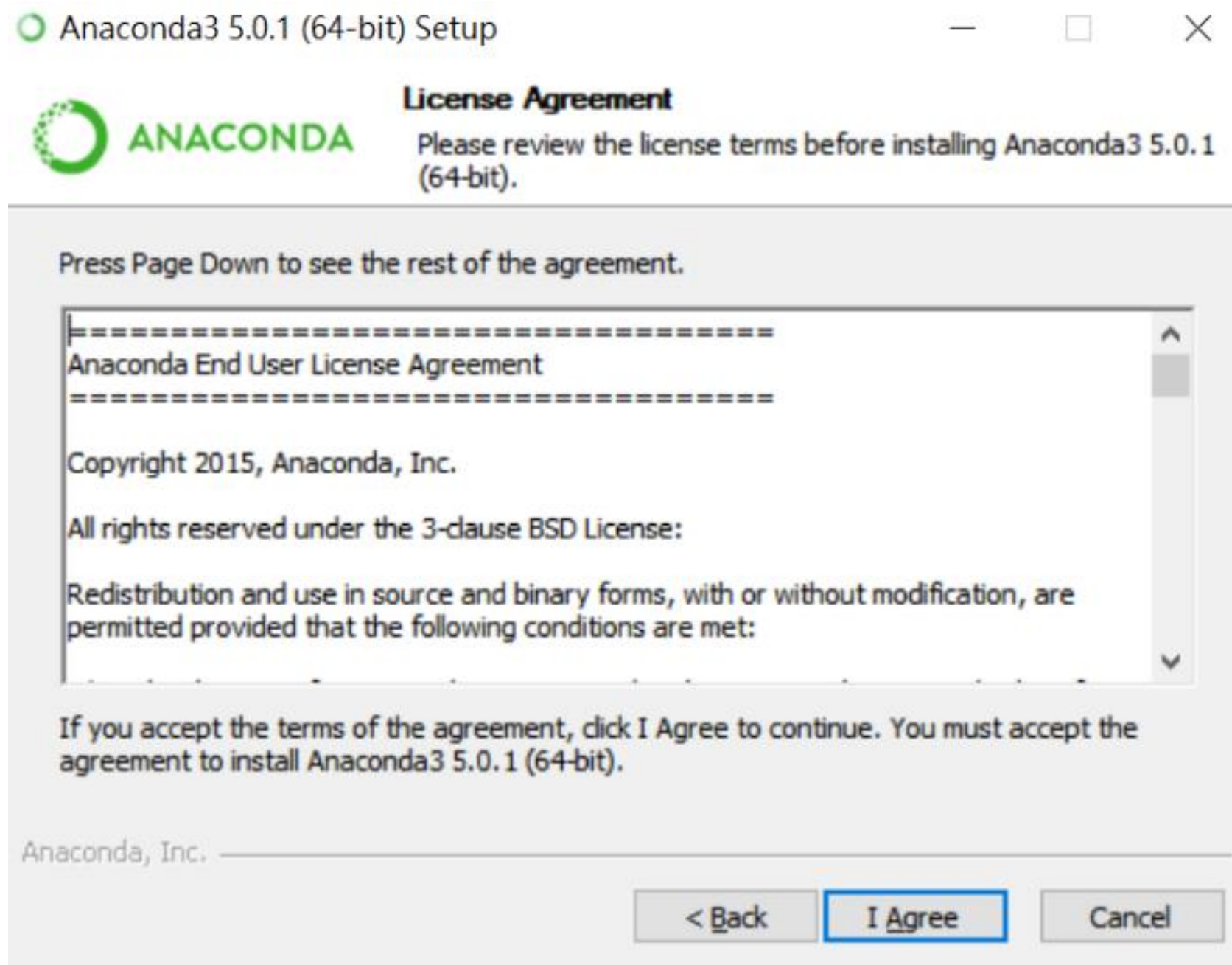
- 依照作業系統選擇安裝檔類型



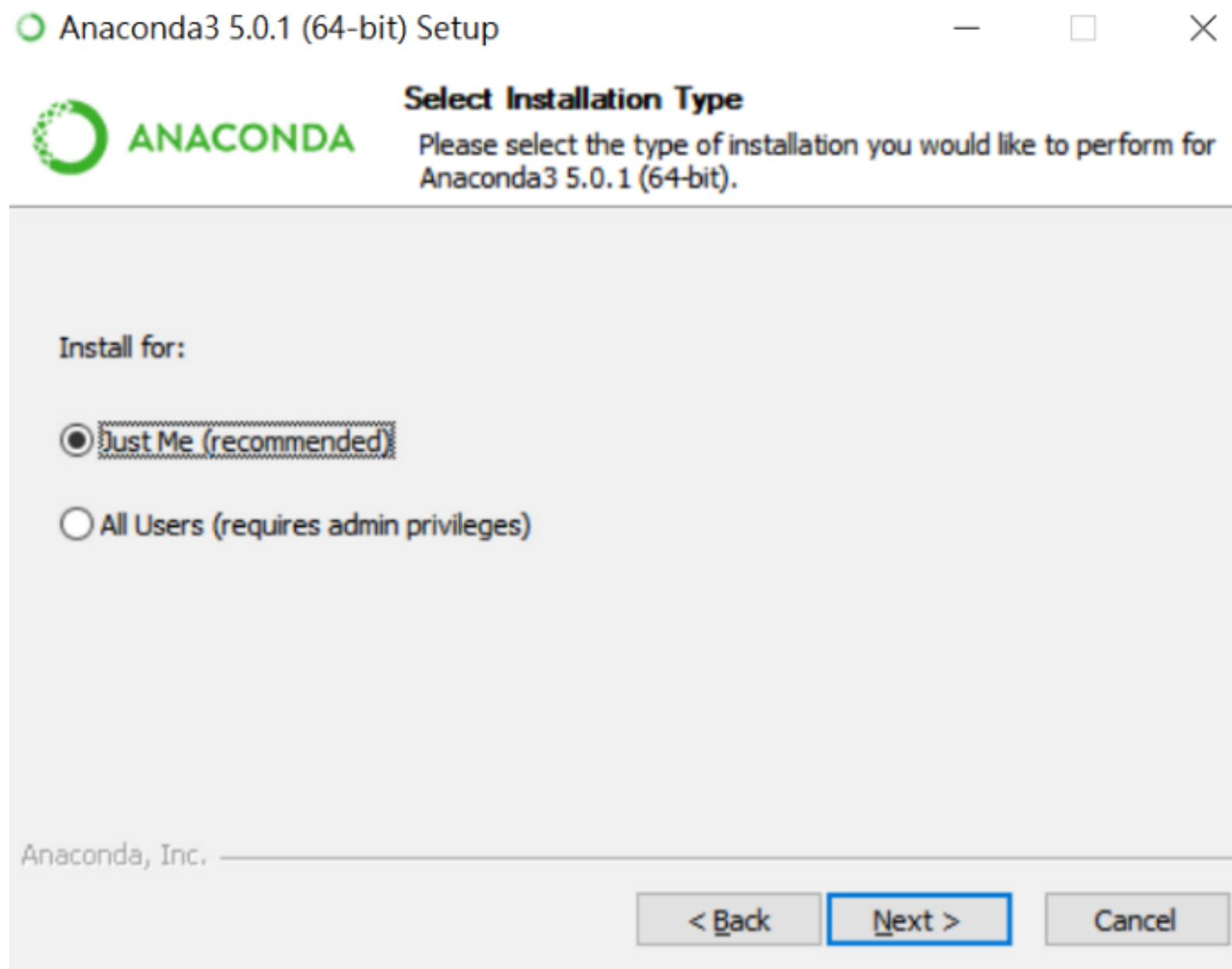
# Anaconda安裝步驟(1/6) <sup>ECT</sup>電子商務技術



# Anaconda安裝步驟(2/6) ECT 電子商務技術

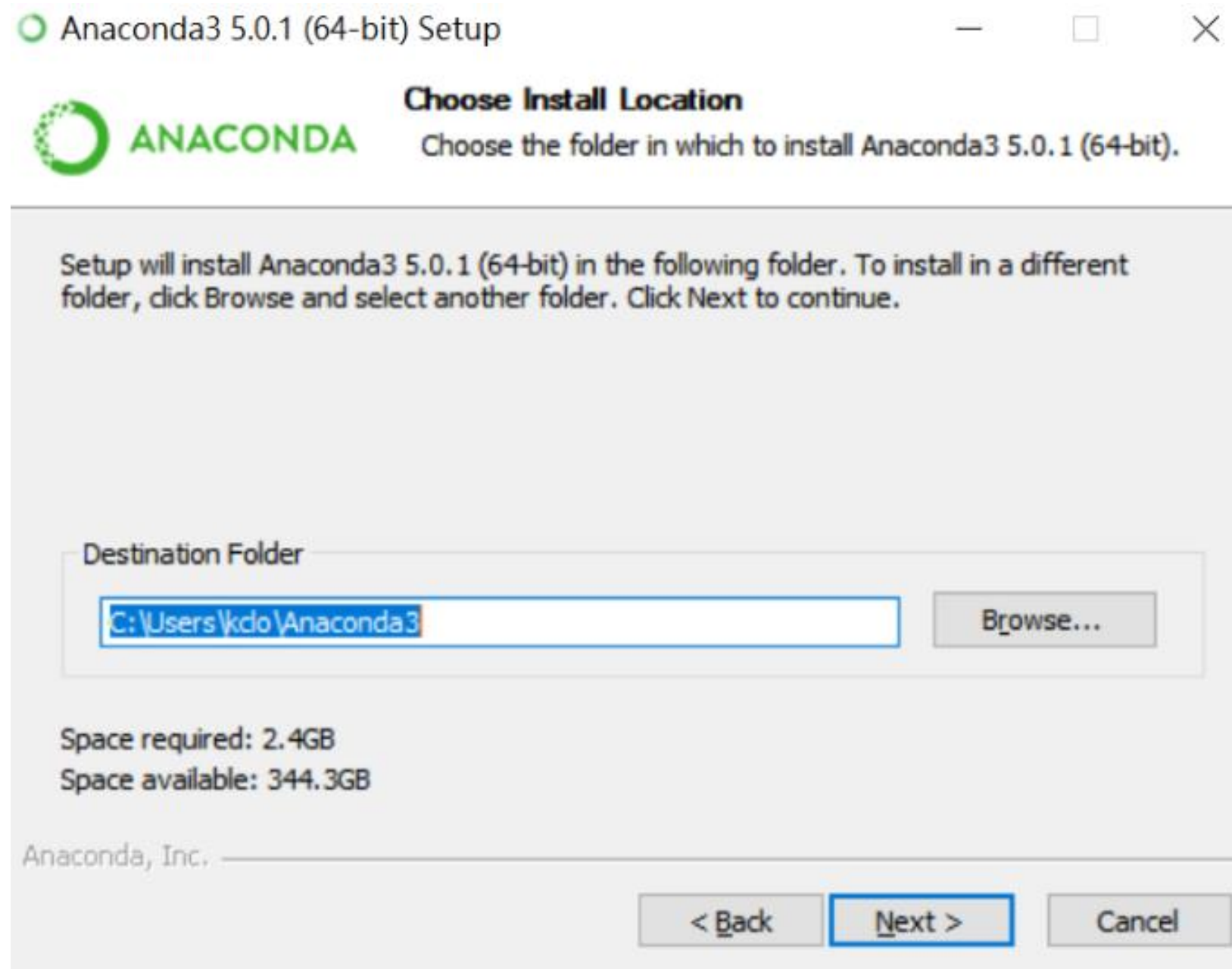


# Anaconda安裝步驟(3/6) <sup>ECT</sup>電子商務技術

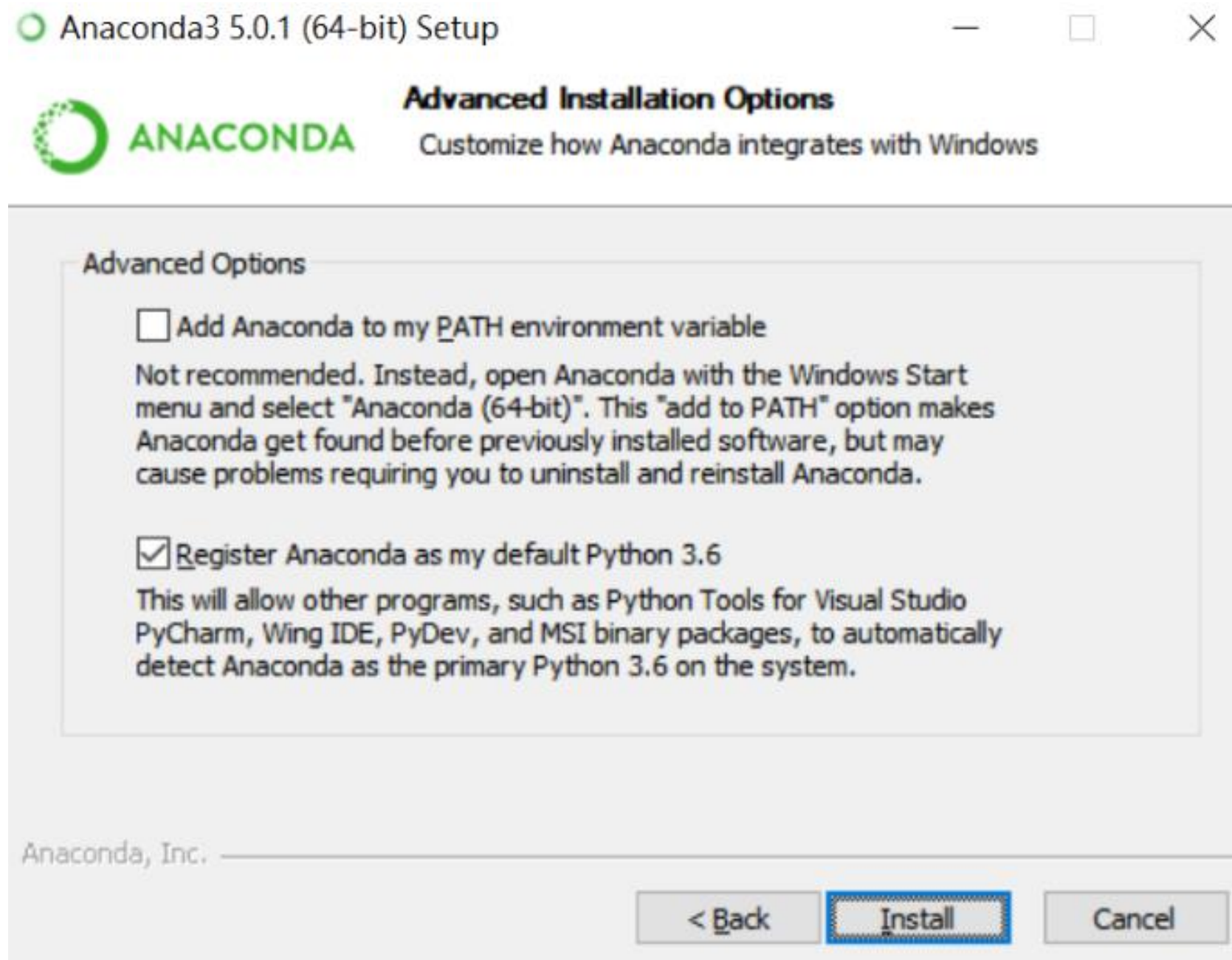




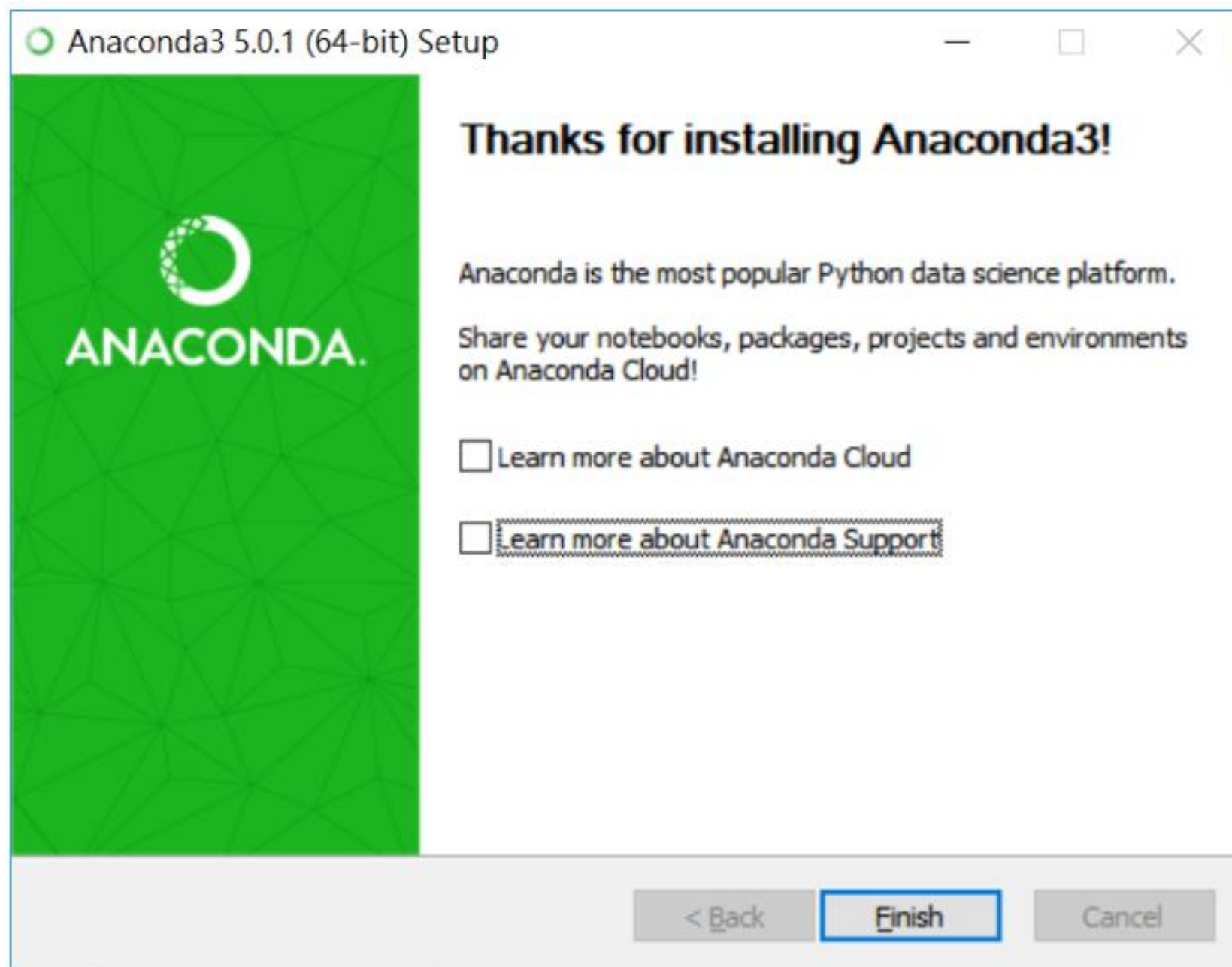
# Anaconda安裝步驟(4/6) <sup>ECT</sup>電子商務技術



# Anaconda安裝步驟(5/6) ECT 電子商務技術

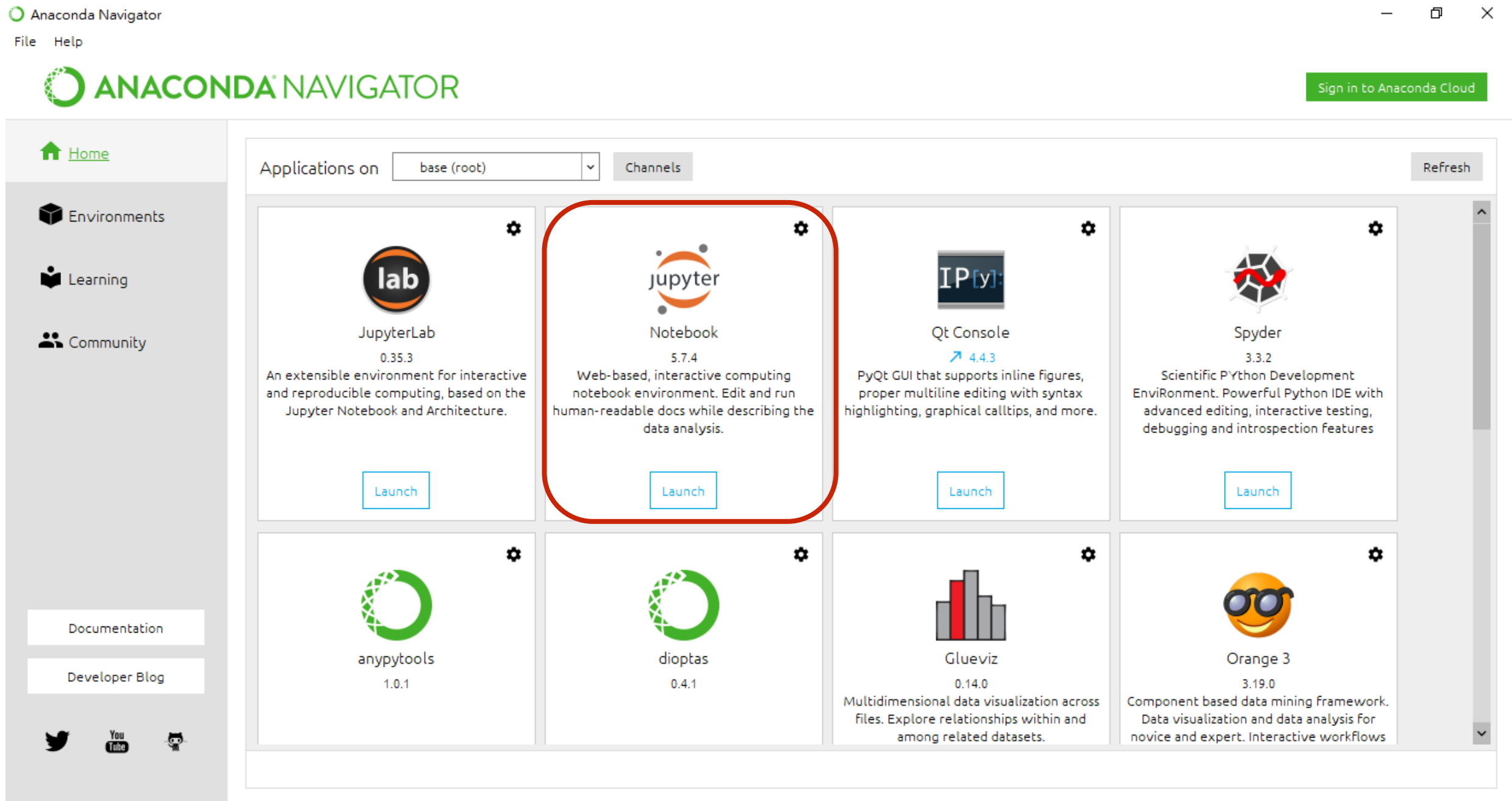


# Anaconda安裝步驟(6/6) <sup>ECT</sup>電子商務技術



# Anaconda安裝完成介面

ECT  
電子商務技術



ECT  
電子商務技術

# Jupyter Notebook介紹

# Jupyter Notebook介紹

---

- 介於編輯器(例如Atom)及IDE(Spider、PyCharm、Vim)之間的應用環境
- 在編寫程式時具有直譯的特性，達到高互動執行結果
- 很容易呈現資料視覺化的結果

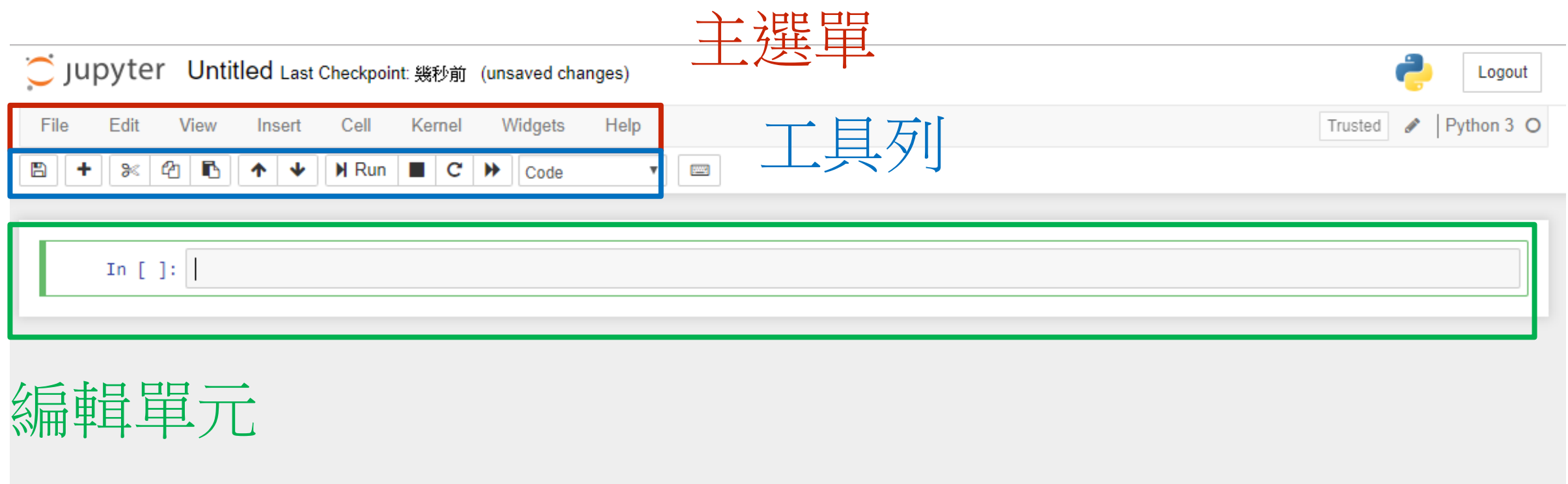
# Jupyter Notebook

ECT  
電子商務技術



# Jupyter Notebook

ECT  
電子商務技術

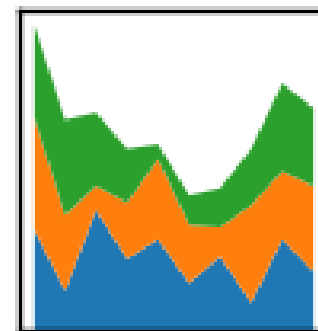
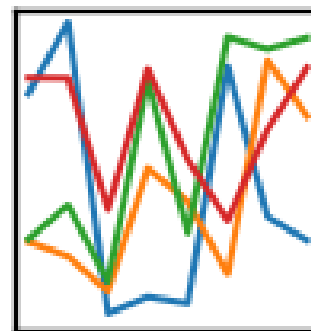
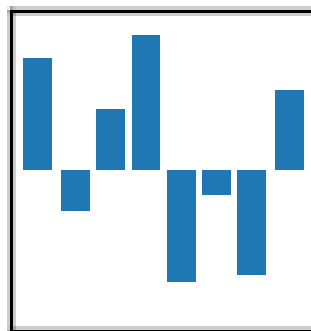




# Pandas套件

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



# Pandas介紹

---

- Pandas可以讓Python很容易做到幾乎所有Excel的功能，像是樞紐分析表、小記、欄位加總、篩選
- 便於進行**資料前處理**
- 兩種主要資料結構
  - Series：用來處理時間序列相關的資料
  - DataFrame：用來處理結構化(Table like)的資料

ECT  
電子商務技術

# SciKit-Learn套件



Machine Learning with Scikit-Learn

# SciKit-Learn介紹

---

- 基於python的機器學習模型套件
- 提供大量常見的機器學習演算法和許多實用的資料集合
- 6大類功能：
  - 分類(Classification)
  - 回歸(Regression)
  - 分群(Clustering)
  - 降低維度(Dimensionality reduction)
  - 模型選擇(Model selection)
  - 資料前處理(Preprocess)