# Knowledge Discovery and Data Mining

## Lab 1 Introduction to Python, Anaconda Jupyter Environment

Xuan Song
Songx@sustech.edu.cn

# Python

- Python is an **interpreted**, **high-level** and **general-purpose** programming language.
- Created by Guido van Rossum and first released in 1991.
- Aims to help programmers write clear, logical code for small and large-scale projects.

# Why to Learn Python?

- Easy to learn
- Easy to read
- **Large standard library**

| | |
|---|---|
| Automation | Graphical user interfaces |
| Data analytics | Networking |
| Image processing | Test frameworks |
| Machine learning | Databases |
| Text processing | Mobile App |
| Multimedia | Web frameworks |

# Python Programming Examples

- Example 1

```
In [1]: print("hello world!")
```

```
hello world!
```

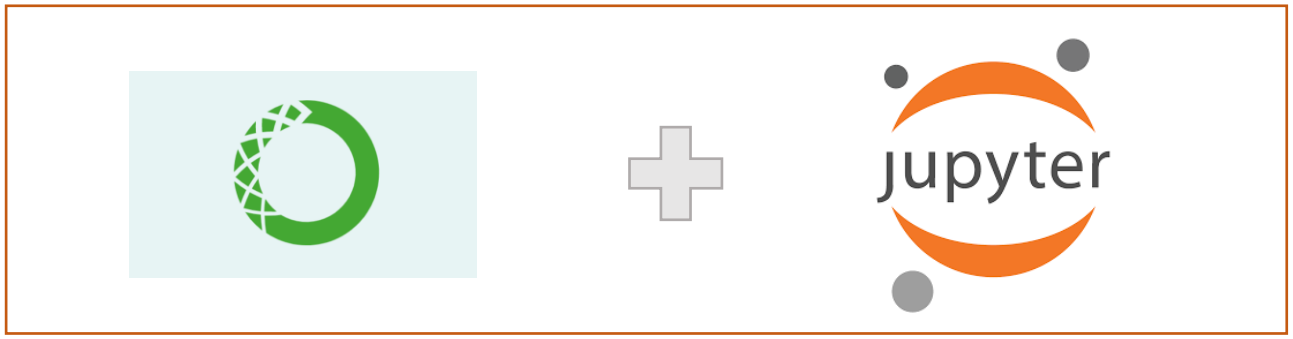
- Example 2

```
In [2]: import math
        print(math.sin(math.pi/2))
```

```
1.0
```

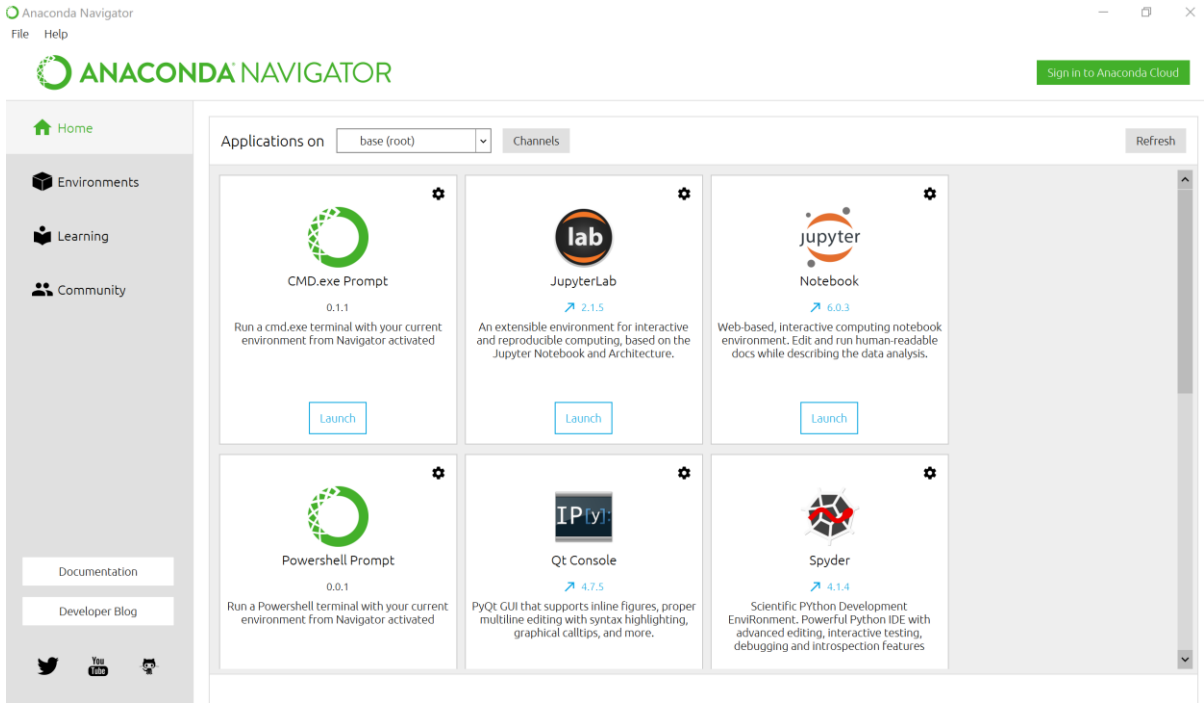
# Python Environment


Recommended



# Install Anaconda

- Installation of Anaconda

# Install Jupyter Notebook

- Installation of Jupyter notebook
  - Installing Jupyter using Anaconda and conda
  - Installing Jupyter with pip

If you have any problem to install Jupyter notebook, you can refer to the following websites:

*(1)* *https://jupyter.readthedocs.io/en/latest/install/notebook-classic.html*

*(2)* *https://www.jianshu.com/p/91365f343585*

# Try to Install Packages

- Install some packages
  - pandas
  - numpy
  - matplotlib
  - scikit-learn

# Try to Use Jupyter Notebook

- Implement the sample code mentioned in the previous slides.

```
In [1]: print("hello world!")

hello world!
```

```
In [2]: import math
        print(math.sin(math.pi/2))

1.0
```

# Exercise1

- Implement a function in Python that takes a collection of intervals as input and merges all overlapped intervals as output.

```
def Function(interval):
    '''
    write your code here
    '''

    return merged_interval
```

Example1:

Input: interval = [[1,3],[2,6],[8,10],[15,18]]

Output: [[1,6],[8,10],[15,18]]

Example2:

Input: interval = [[1,4],[4,5]]

Output: [[1,5]]

# Exercise2

- 1. Reading and writing TXT file in jupyter notebook.
- 2. Reading and writing CSV file in jupyter notebook.

**Hints:**

1. txt file:
https://www.geeksforgeeks.org/reading-writing-text-files-python/
https://pythonexamples.org/python-read-text-file/

2. csv file
https://realpython.com/python-csv/

# Other Resources

- Python:
  - https://www.w3schools.com/python/
  - https://www.runoob.com/python/python-tutorial.html

- Anaconda and Jupyter notebook:
  - https://www.anaconda.com/products/individual/get-started
  - https://blog.csdn.net/zaishuiyifangxym/article/details/83269834
  - https://mirrors.tuna.tsinghua.edu.cn/anaconda/archive/
  - https://juejin.im/post/6844903842497167374


# End of Lab1