Python Programming

Jaewoong Han

Machine Learning &
Control Systems Lab.

Basic

Install Necessary Programs

- Python
  - https://www.python.org/downloads/
  - Choose latest version or appropriate version. (Recommend 3.8 upper)

- Visual Studio Code (VSCode)
  - https://code.visualstudio.com/
  - It's most popular open-source code editor.

- Anaconda (not now…)
  - https://www.anaconda.com/
  - It is useful when you use virtual environment.

- CUDA, cuDNN for using GPU (recommend, but not required)
  - https://pytorch.org/get-started/locally/
  - Find appropriate CUDA, cuDNN version for your computer
  - https://en.wikipedia.org/wiki/CUDA
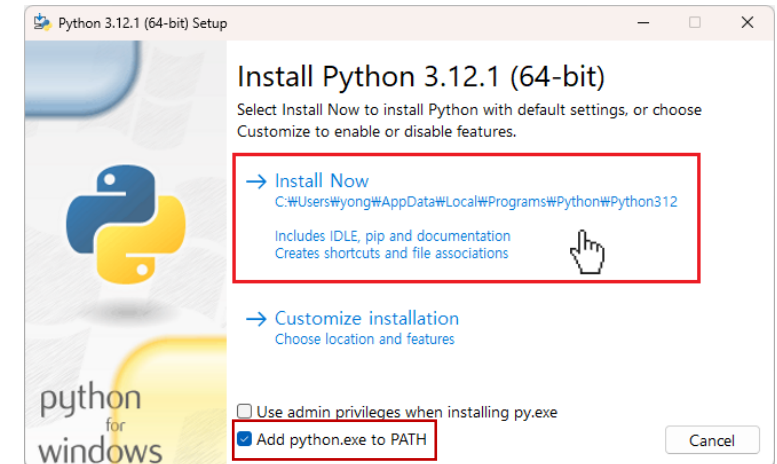
Python coding you should study on your own

- Data Structures
  - List, Tuples, Dictionaries, …
- Control Structures
  - if, elif, else
  - for, while
- Functions
  - Define functions, parameters, return, …
- Object-Oriented Programming (OOP)
  - class class_name(father_class_name):
    - super(class_name, self). __init__()
- Error Handling
  - try, except, finally
  - raise valuerror, …
- Modules and Packages
  - pip install ~
  - from ~ import ~

Basic

How to run a Python file using Command Prompt (cmd)
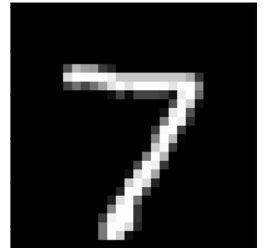

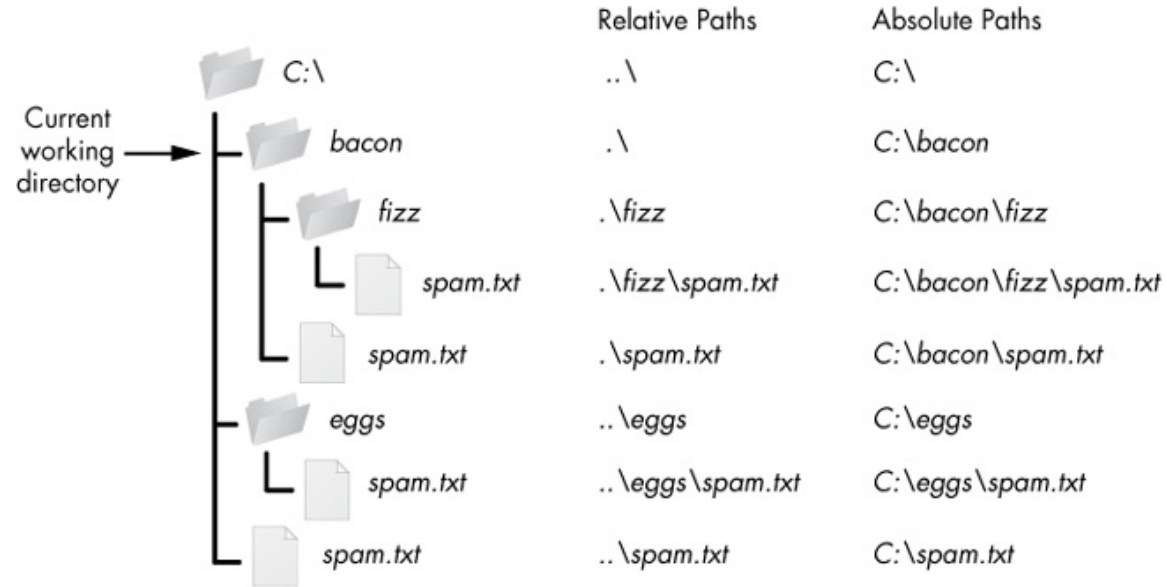Module Path (prefer)

File Path

001 CSV

Handling csv files

- Reading and writing csv files is one of the basic functions that we use a lot when dealing with data. In this course, we are going to read the MNIST dataset in the form of 1D vector and convert it into 2D vector form.

Load files

- It is preferred to use relative paths in most cases
- Use 'os.path' library instead of using string path

001 CSV

Load files

```
PS C:\Users\            > cd "C:\Users\            \Documents\MLCS\인 턴 십 \MLCS-Internship-Program_revise\002 Neural Networks"
PS C:\Users\          \Documents\MLCS\인 턴 십 \MLCS-Internship-Program_revise\002 Neural Networks> python
Python 3.8.10 (tags/v3.8.10:3d8993a, May  3 2021, 11:48:03) [MSC v.1928 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import os, glob
>>> current_path = os.getcwd()
>>> current_path
'C:\\Users\\          \\Documents\\MLCS\\인 턴 십 \\MLCS-Internship-Program_revise\\002 Neural Networks'
>>> data_path = os.path.join('data', 'stl10_binary')
>>> data_path
'data\\stl10_binary'
>>> data = glob.glob(os.path.join(data_path, '*.bin'))
>>> data
['data\\stl10_binary\\test_X.bin', 'data\\stl10_binary\\test_y.bin', 'data\\stl10_binary\\train_X.bin', 'data\\stl10_bin
ary\\train_y.bin', 'data\\stl10_binary\\unlabeled_X.bin']
```

- 'os.path.join' creates path.
- 'glob.glob' returns the files in given pathB but note that the files are not sorted.

Jaewoong Han   7

Using Numpy Library

- Using numpy library to treat numerical data in matrix form will be a daily job when you become a data scientist.

- In this course, we are going to read 2D MNIST dataset (from the previous course) as numpy arrays.

- Subsequently, we will split the data into train, validation, and test dataset and save those files as .npz files.

Our current directory is
··· /MLCS-Internship-Program/001 Python Programming/002 numpy
We need to load train datasets in
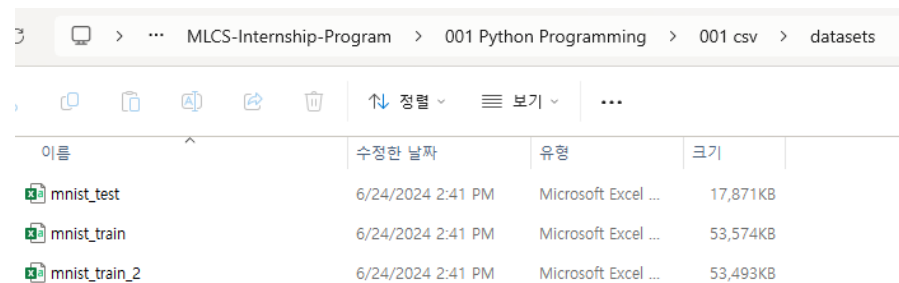··· /MLCS-Internship-Program/001 Python Programming/001 csv/ ~

Create an empty numpy array.

```python
import os, glob
import numpy as np

path_to_train_2d_datasets = os.path.join('..','001 csv','outputs','train','*.csv')
train_2d_files = glob.glob(path_to_train_2d_datasets)

# We already know that the width and height of the 2D MNIST dataset is 28.
train = np.empty((len(train_2d_files),28,28))

for data_idx, data_path in enumerate(train_2d_files):
    # Read the csv and replace the elements
    csv_data = # may need csv library to load the data
    for i in range(28):
        for j in range(28):
            train[data_idx,i,j] = csv_data[i][j]
```

MLCS-Internship-Program > 001 Python Programming > 001 csv > datasets

이름	수정한 날짜	유형	크기
mnist_test	6/24/2024 2:41 PM	Microsoft Excel ...	17,871KB
mnist_train	6/24/2024 2:41 PM	Microsoft Excel ...	53,574KB
mnist_train_2	6/24/2024 2:41 PM	Microsoft Excel ...	53,493KB

Question