AI활용 표현과 문제해결

[Practice Settings & Numpy]

허세훈

충남대학교 지능소프트웨어연구실 2021.03.08



Sehun Hu (허세훈)

E-mail: sehunhu5247@gmail.com

Laboratory: Intelligent Software

Office: W5602

실습 수업 진행방향

* 기본적으로 Python과 선형대수를 배우고 왔음을 가정

매주 실시간 강의

YouTube 링크를 통해 실습 수업시간에 진행 → **출석은 성적에 반영**

YouTube 댓글을 통해 출석 확인 수업 시작 시점부터 **1시간 이내**에 실시간 채팅에 **자신의 학번과 이름**을 입력

이론 시간의 내용 보완

이론 시간에 배웠던 내용들을 실제로 구현해보는 시간으로 활용 예정

과제 x 기말고사 반영 o 매주 실습에 따른 과제는 없으나, 기말고사 때 실습 시간에 배운 내용들이 나올 수 있음

이번 주차 실습 내용

Pycharm 설치

Anaconda

Jupyter Notebook

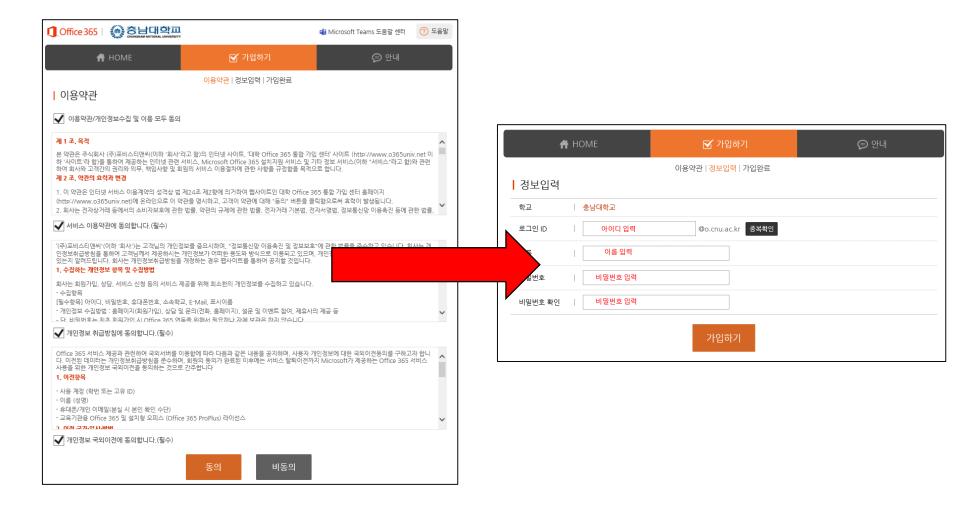
Numpy

- 학생 계정 생성 소개 Pycharm 설치 Professional 설 가상환 치

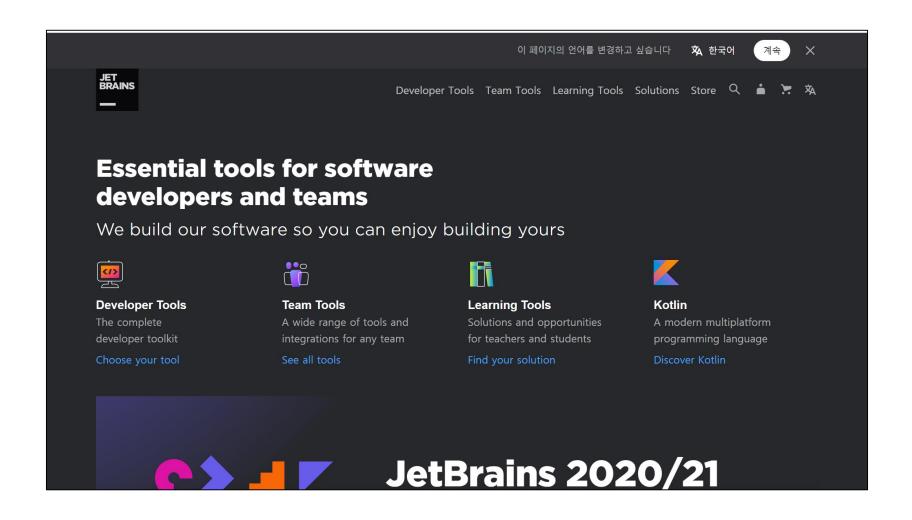
 - 가상환경

- 소개
- 실행 및 예시
- 소개
- 설치
- 배열/연산

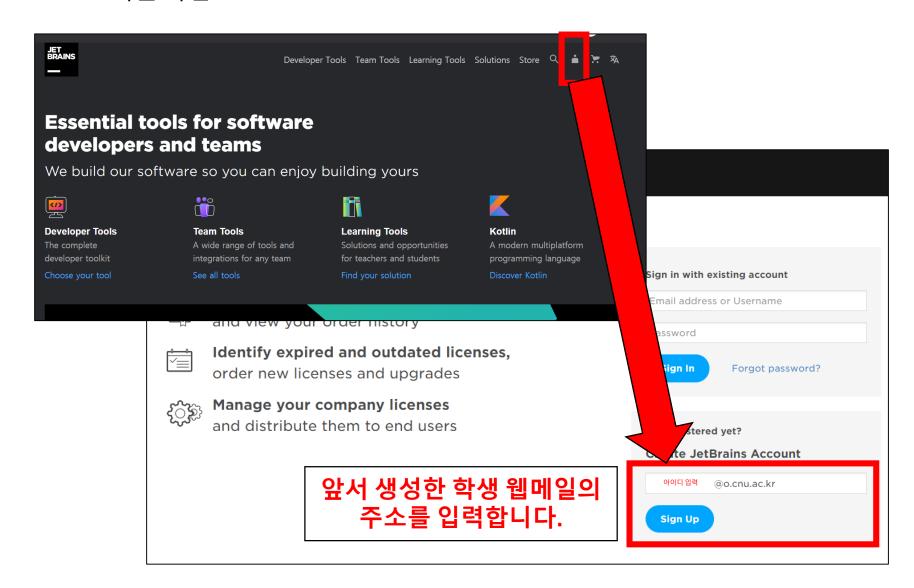
1. 학교 포털의 "학생 웹메일(Office 365)" 메뉴를 통해 계정 만들기



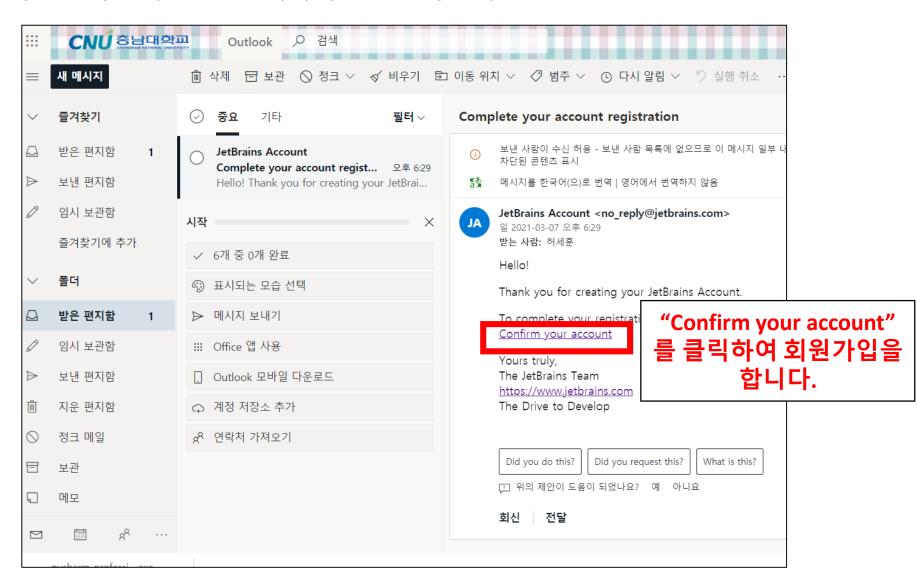
2. JetBrains(<u>https://www.jetbrains.com/</u>) 홈페이지에 접속



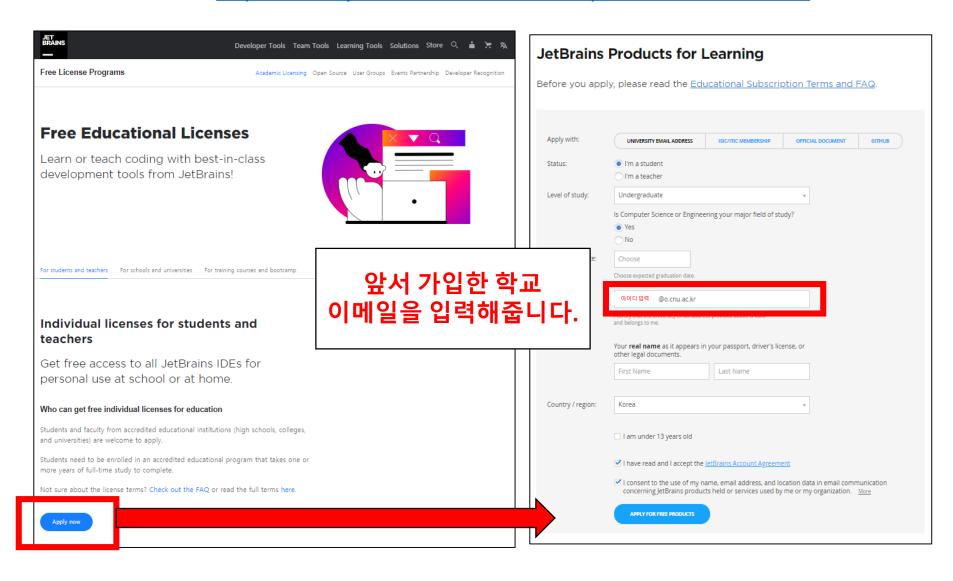
3. JetBrains 회원 가입



4. 학교 웹메일 계정으로 접속하여 발송된 메일 확인

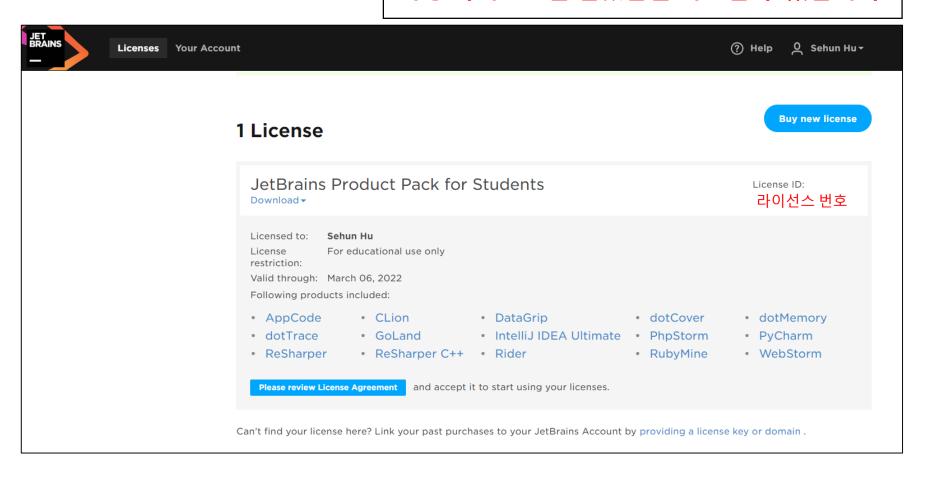


5. 학생 인증 받기: https://www.jetbrains.com/community/education/#students 에 접속

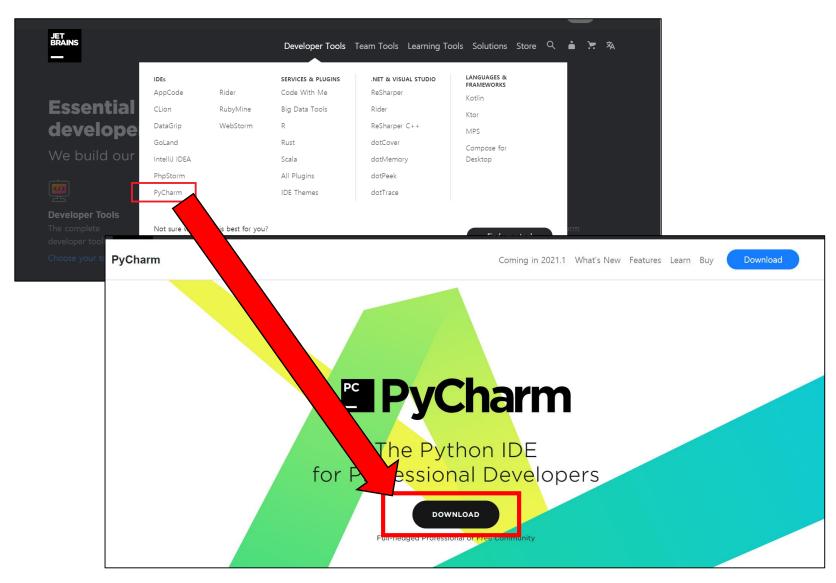


6. 학생 License 인증 확인

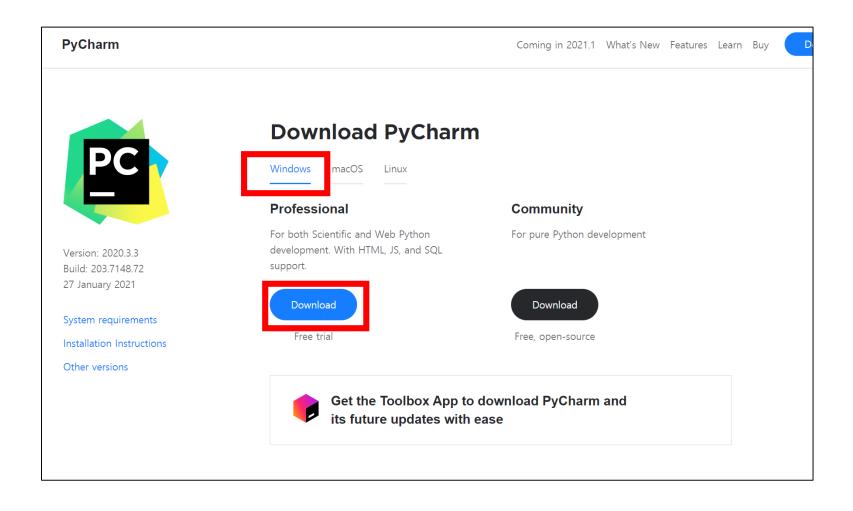
JetBrains에 로그인하면 다음과 같이 학생 라이선스를 받았음을 확인할 수 있습니다.



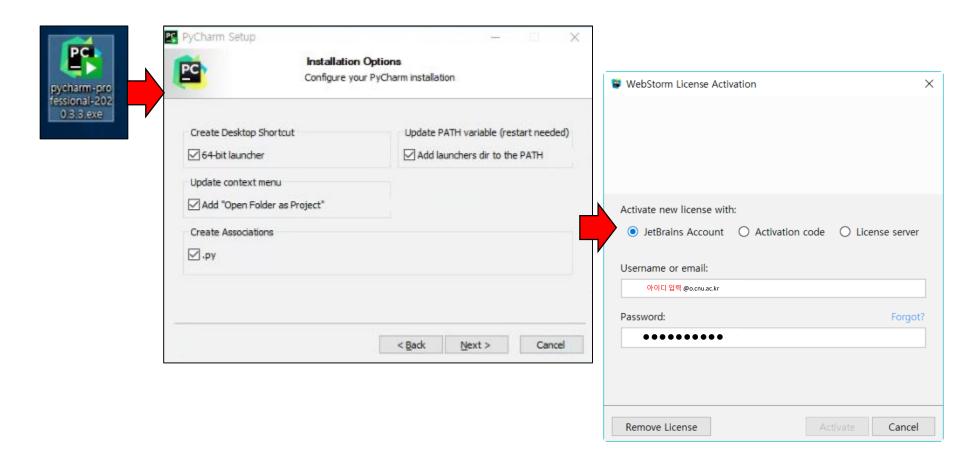
7. Developer Tools → PyCharm 클릭 → DOWNLOAD 버튼 클릭



8. 자신의 운영체제 선택 -> Professional 버전 Download



9. 설치 및 라이선스 입력





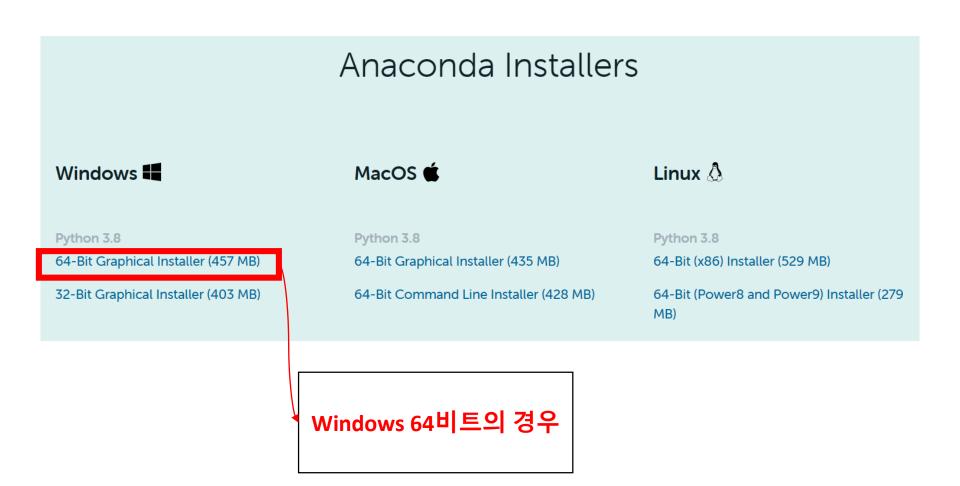
수많은 데이터 과학 패키지 제공 Numpy, Pandas, Scikit-learn과 같은 데이터 과학을 위한 패키지들을 제공

패키지 및 종속성 관리

Conda 가상환경을 통해, 프로젝트들 간의 패키지 및 종속성을 구분해서 관리 가능

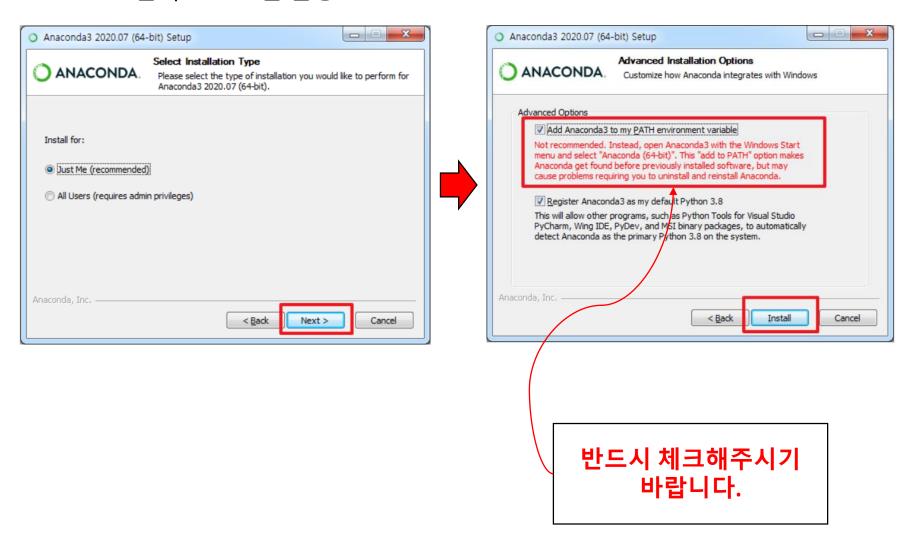
Anaconda | 설치

1. 아나콘다 설치 페이지(https://www.anaconda.com/products/individual#download-section)로 이동 → 자신에 맞는 Installer 파일 다운로드



Anaconda | 설치

2. Anaconda 설치 프로그램 실행



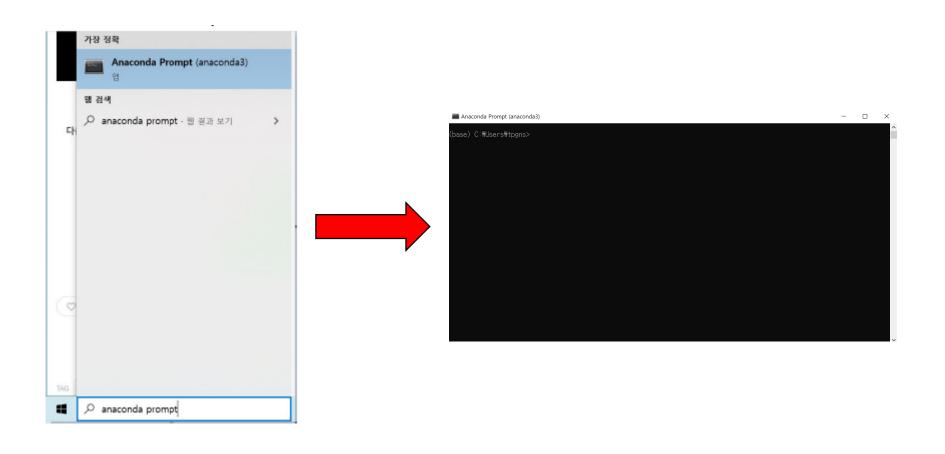
"독립적인 작업환경에서 작업을 할 수 있게 도와줌"

프로젝트간의 패키지 버전 문제 프로젝트마다 패키지의 버전들이 달라서 충돌이 발생하는 경우가 많다.

가상 환경을 통해 독자적으로 패키지 관리 Anaconda 가상 환경을 사용하면, 다른 프로젝트와 겹치지 않도록 할 수 있다.

Anaconda | 가상환경 생성 방법

1. Anaconda Prompt 실행 : Windows키를 누른 후, "anaconda prompt"라고 입력 후 실행



Anaconda | 가상환경 생성 방법

2. 가상환경 생성 명령어 입력

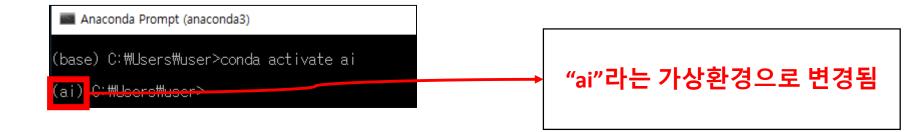
conda create -n {가상환경이름} python={Python 버전}



Anaconda | 가상환경 활성화/비활성화

1. 가상환경 활성화

conda activate {가상환경 이름}



2. 가상환경 비활성화

conda deactivate



Anaconda | 가상환경 관련 명령어들

	명령어 형태	예시
가상환경 생성	conda create –n {가상환경이름} python={python 버전}	conda create –n AI python=3.8
가상환경 목록 확인	conda env list	conda env list
가상환경 활성화	conda activate {가상환경 이름}	conda activate AI
가상환경 비활성화	conda deactivate {가상환경이 름}	conda deactivate Al
가상환경 삭제	conda env remove –n {가상환경 이름}	conda env remove –n Al
가상환경 내에서 패키지 설치	conda install {패키지 이름}	conda install numpy

Jupyter Notebook | 소개

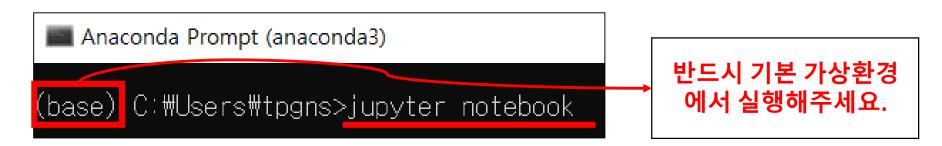


코드의 실행 결과를 즉각적으로 확인 가능 코드를 즉각적으로 실행함으로서 데이터를 분석하거나 시각화할 수 있음

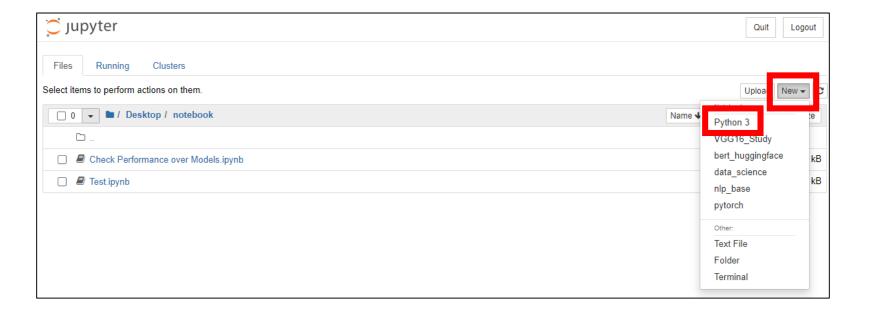
커뮤니케이션 또는 문서화 가능 Jupyter Notebook을 통해 생성되는 .ipynb 파일을 공유하거나 문서화하여 의사소통 가능

Jupyter Notebook | 실행

1. Anaconda Prompt에서 "jupyter notebook"이라는 명령어를 입력

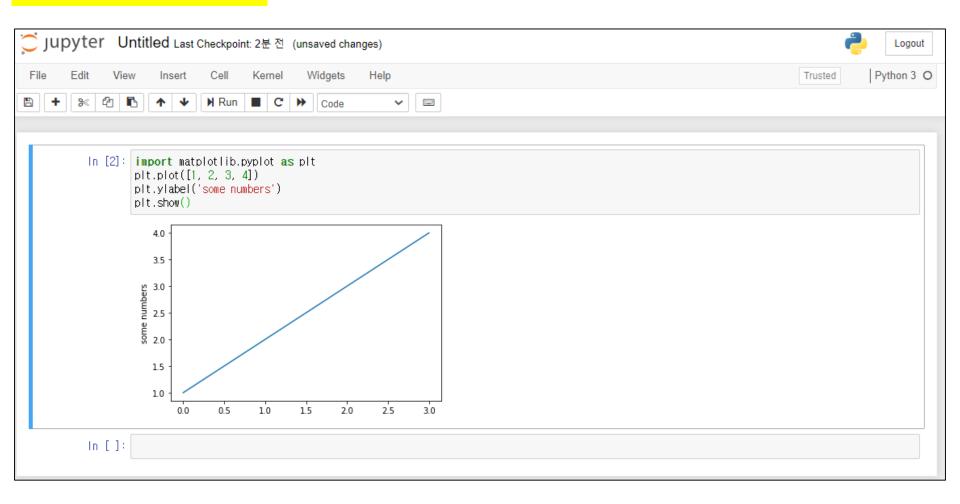


2. 쥬피터 노트북 파일 생성



Jupyter Notebook | 예시

나만의 코드 실행하기





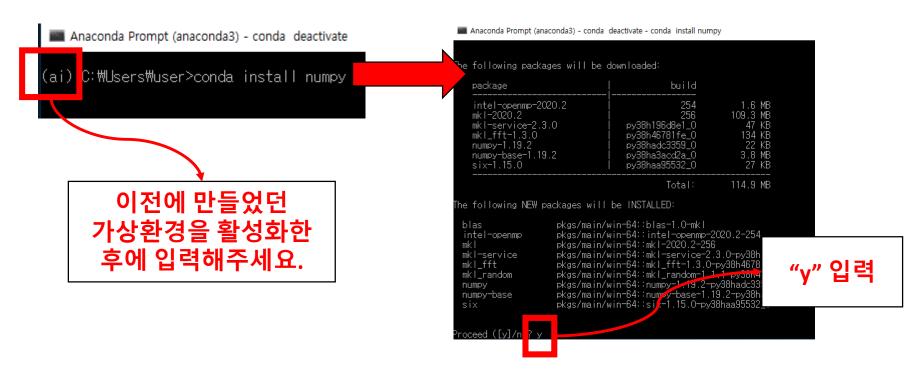
다차원 배열 자료형 및 연산 제공 Python에는 "리스트"라는 자료형은 있지만 "배열"은 없다. → Numpy는 배열 자료형을 제공

다양한 수학적인 연산 제공

수학적인 연산을 Function 형태로 제공

Numpy | 설치 및 사용

1. Anaconda Prompt에서 "conda install numpy"이라는 명령어를 입력



2. import를 통해 numpy를 사용

```
import numpy as np # numpy 라이브러리를 np라는 이름으로 import합니다.

arr = np.array([1, 2, 3, 4])
print(arr)
```

Numpy | 배열 생성 방법

(1) Python 리스트 → Numpy 배열

```
import numpy as np
list_python = [1, 2, 3, 4] # 파이썬 리스트를 만듭니다.
numpy_arr = np.array(list_python) # 파이썬 리스트를 numpy 배열로 만듭니다.
print(numpy_arr.shape) # (4, )
```

(2) Numpy에서 제공하는 함수를 통해 Numpy 배열 생성하기

```
import numpy as np

zero_arr = np.zeros((2, 2)) # 전부 0으로 채워진 (2, 2)짜리 배열 생성
one_arr = np.ones((2, 3)) # 전부 1로 채워진 (2, 3)짜리 배열 생성
five_arr = np.full((2, 3), 5) # 전부 5로 채워진 (2, 3)짜리 배열 생성
identity_arr = np.eye(3) # (3, 3)짜리 단위 행렬 생성
```

Numpy | 배열 슬라이싱

"지정한 범위가 아닌 원소들을 잘라냄"

```
import numpy as np
p_list = [[1, 2, 3], [4, 5, 6], [7, 8, 9]]
arr = np.array(p_list) # Python 리스트로부터 (3, 3)짜리 배열 생성
ex1 = arr[0:2, 0:2] # row와 column이 0~1까지인 원소들만 선택
print(ex1)
ex2 = arr[1:, 1:] # row와 column이 1~끝까지인 원소들만 선택
print(ex2)
ex3 = arr[-1, 0:] # row가 마지막이고 column이 0~끝까지인 원소들만 선택
print(ex3)
```

Numpy | 배열 인덱싱

"배열의 인덱스를 통해 원하는 원소들만 선택"

(1) 자신이 원하는 원소들을 Index로 접근해서 가져오기

```
import numpy as np
target_arr = np.array([11, 22, 33, 44, 55, 66, 77, 88, 99])
idx = np.array([0, 2, 4, 6, 8]) # 선택하고 싶은 Index들을 np.array 형태에 저장합니다.
print(target_arr[idx]) # [11, 33, 55, 77, 99]
```

[주의] Index를 위한 배열은 Numpy Array 형태이어야 함

(2) Boolean 값으로 인덱싱하기

```
import numpy as np

target_arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

bool_idx = (target_arr % 2 == 0) # 2로 나누어 떨어지는 지를 boolean 값으로 저장합니다.

print(bool_idx) # [True, False, True, False, True, False, True, False, True, False]

print(target_arr[bool_idx]) # [0, 2, 4, 6, 8]
```

[한줄로도 작성 가능] target arr[target arr % 2 == 0]

Numpy | 배열의 크기 및 자료형

(1) 배열 크기 확인

```
import numpy as np

a = np.array([1, 2, 3])
print(a.shape) # (3, )

b = np.array([[1, 2, 3], [4, 5, 6]])
print(b.shape) # (2, 3)

shape가 (3, )인 것과 (3, 1)은 완전히
다른 의미입니다.
(3, 1)은 np.array([[1], [2],[3]])을 의미
```

(2) 배열 자료형 확인

```
import numpy as np
a = np.array([1, 2, 3])
print(a.dtype) # dtype('int64')
b = np.array([1, 2, 3.])
print(b.dtype) # dtype('float64')
```

numpy 자료형

arr1이나 arr2는 int64라는 자료형을 갖는 것에 반해 arr3는 float64라는 자료형을 갖는다.

이는 arr3내부 데이터를 살펴보면 3.5라는 실수형 데이터를 갖기 때문임을 알 수 있다.

numpy에서 사용되는 자료형은 아래와 같다. (자료형 뒤에 붙는 숫자는 몇 비트 크기인지를 의미한다.

- 부호가 있는 정수 int(8, 16, 32, 64)
- 부호가 없는 정수 uint(8,16,32,54)
- 실수 float(16, 32, 64, 128)
- 복소수 complex(64, 128, 256)
- 불리언 bool
- 문자열 string
- 파이썬 오프젝트 object
- 유니코드 unicode_

Numpy | 배열끼리의 연산

(1) 기본적으로는 크기가 서로 동일할 때 연산 가능 (2) 브로드캐스트라는 기능을 통해, 연산 가능

```
import numpy as np
a = np.array([1, 2, 3])
b = np.array([4, 5, 6])
c = a + b # 각 원소끼리 더하기
print(c)
c = a - b # 각 원소끼리 빼기
print(c)
c = a * b # 각 원소끼리 곱하기
                              행렬 곱
print(c)
c = a / b # 각 원소끼리 나누기
print(c)
```

```
import numpy as np
a = np.array([[1, 2, 3], [4, 5, 6]])
print(a.shape)
b = np.array([10, 11, 12])
print(b.shape)
print(a + b) # array([[11, 13, 15], [14, 16, 18]])
print(a * b) # array([[10, 22, 36], [40, 55, 72]])
print(a * 10) # array([[10, 20, 30], [40, 50, 60]])
print(a ** 2) # array([[ 1, 4, 9], [16, 25, 36]])
```

* 브로드캐스트 : 서로 크기가 다른 배열끼리 연산이 가능하도록 하는 기능

Numpy | 수학 연산 함수들

(1) 하나의 배열/수에 적용되는 함수들

함수	설명	예시
abs()	절대값 계산	np.abs(arr1)
sqrt()	제곱근 계산	np.sqrt(arr1)
square()	제곱 계산	np.square(arr1)
exp()	e의 지수로 삼은 값 계산	np.exp(arr1)
log()	로그 취하기	np.log(arr1)
sign()	부호 계산	np.sign(arr1)
ceil()	소수 첫 번째 자리에서 올림	np.ceil(arr1)
isnan()	NaN 값인 경우, True, 아닌 경우 False 반환	np.isnan(arr1)
sin()	사인 함수	np.sin(arr1)
cos()	코사인 함수	np.cos(arr1)

Numpy | 수학 연산 함수들

(2) 두 개의 배열에 적용되는 함수들

함수	설명	예시
add()	동일한 위치의 성분끼리 더하기	np.add(arr1, arr2)
subtract()	동일한 위치의 성분끼리 빼기	np.subtract(arr1, arr2)
multiply()	동일한 위치의 성분끼리 곱하기	np.multiply(arr1, arr2)
divide()	동일한 위치의 성분끼리 나누기	np.divide(arr1, arr2)
maximum()	동일한 위치의 성분끼리 비교하여 최대값 계산하기	np.maximum(arr1, arr2)
minimum()	동일한 위치의 성분끼리 비교하여 최소값 계산하기	np.minimum(arr1, arr2)
dot()	행렬 곱	np.dot(arr1, arr2)

Numpy | 수학 연산 함수들

(3) 통계 관련 함수들

함수	설명	예시
sum()	행렬 내의 성분 합 계산	 np.sum(arr1) → 전체 성분의 합 np.sum(arr1, axis=0) → 행 간의 합 np.sum(arr1, axis=1) → 열 간의 합
mean()	행렬 내의 성분 평균 계산	 np.mean(arr1) → 전체 성분의 평균 np.mean(arr1, axis=0) → 행 간의 평균 np.mean(arr1, axis=1) → 열 간의 평균
std()	행렬 내의 표준편차 계산	np.std(arr1)
var()	행렬 내의 분산 계산	np.var(arr1)
argmax()	행렬 내의 최대값이 위치한 인덱스 반환	np.argmax(arr1)np.argmax(arr1, axis=0)
argmin()	행렬 내의 최소값이 위치한 인덱스 반환	np.argmin(arr1)np.argmin(arr1, axis=0)

Numpy | 정렬 관련 함수들

- sort(): 정렬된 값을 반환
- argsort(): 정렬된 값의 인덱스를 반환

```
import numpy as np
a = np.array([2, 5, 3, 6, 8, 1])
b = np.sort(a) # 정체 성분에 대해서 오름차순으로 정렬
print(b)
c = np.sort(a)[::-1] # 전체 성분에 대해서 내림차순으로 정렬
print(c)
d = np.argsort(a) # 오름차순 순으로 인덱스를 반환
print(d)
```

Numpy | 난수 생성

- Numpy의 random 모듈 사용
- rand() 함수에 크기를 입력해준다.

```
import numpy as np
a = np.random.rand(5)
print(a)
b = np.random.rand(2, 3)
print(b)
```