LEARNING SPOONS

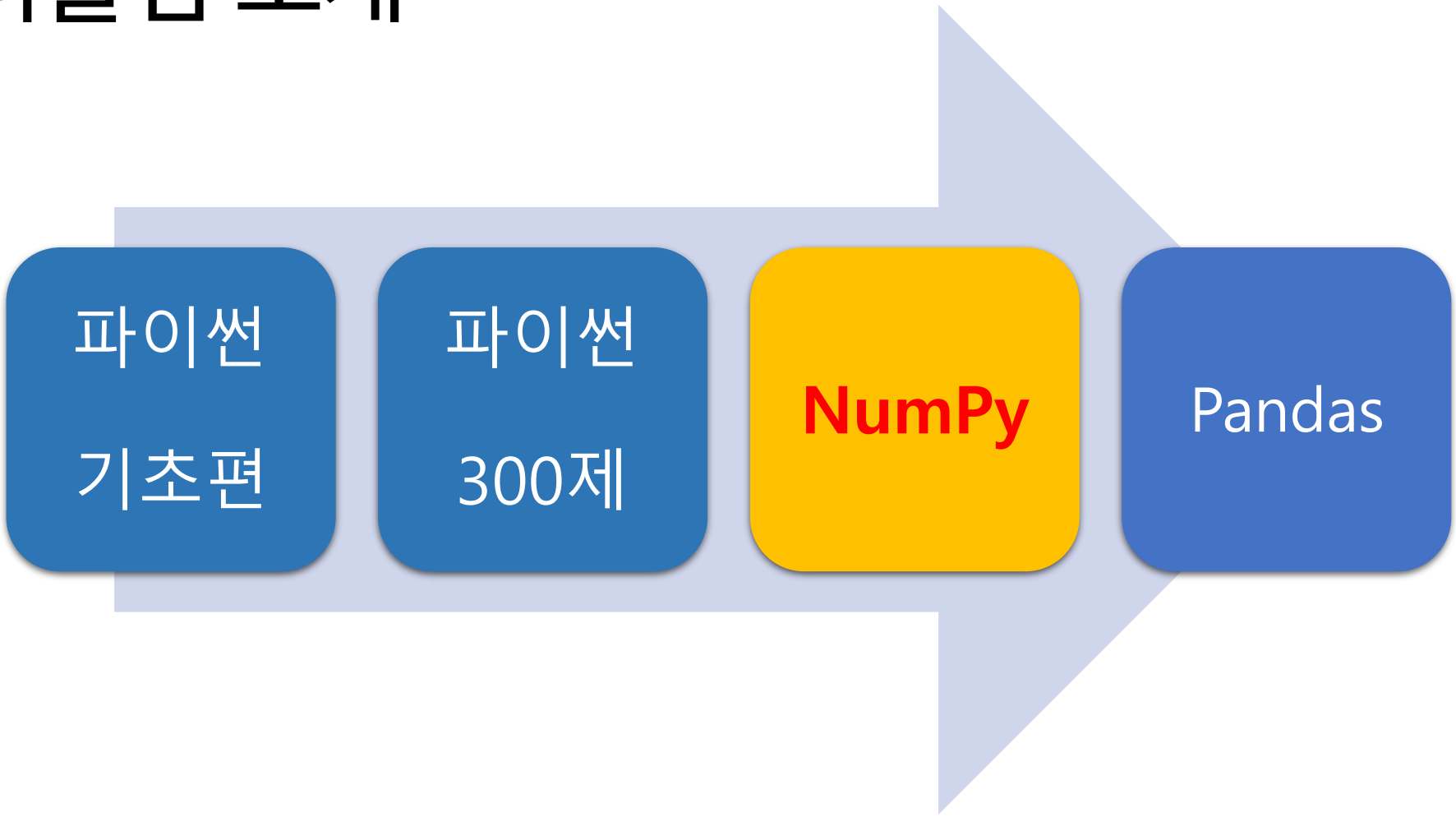
LEARNING SPOONS ONLINE

Introduction

데이터 분석을 위한 NumPy

LEARNING SPOONS

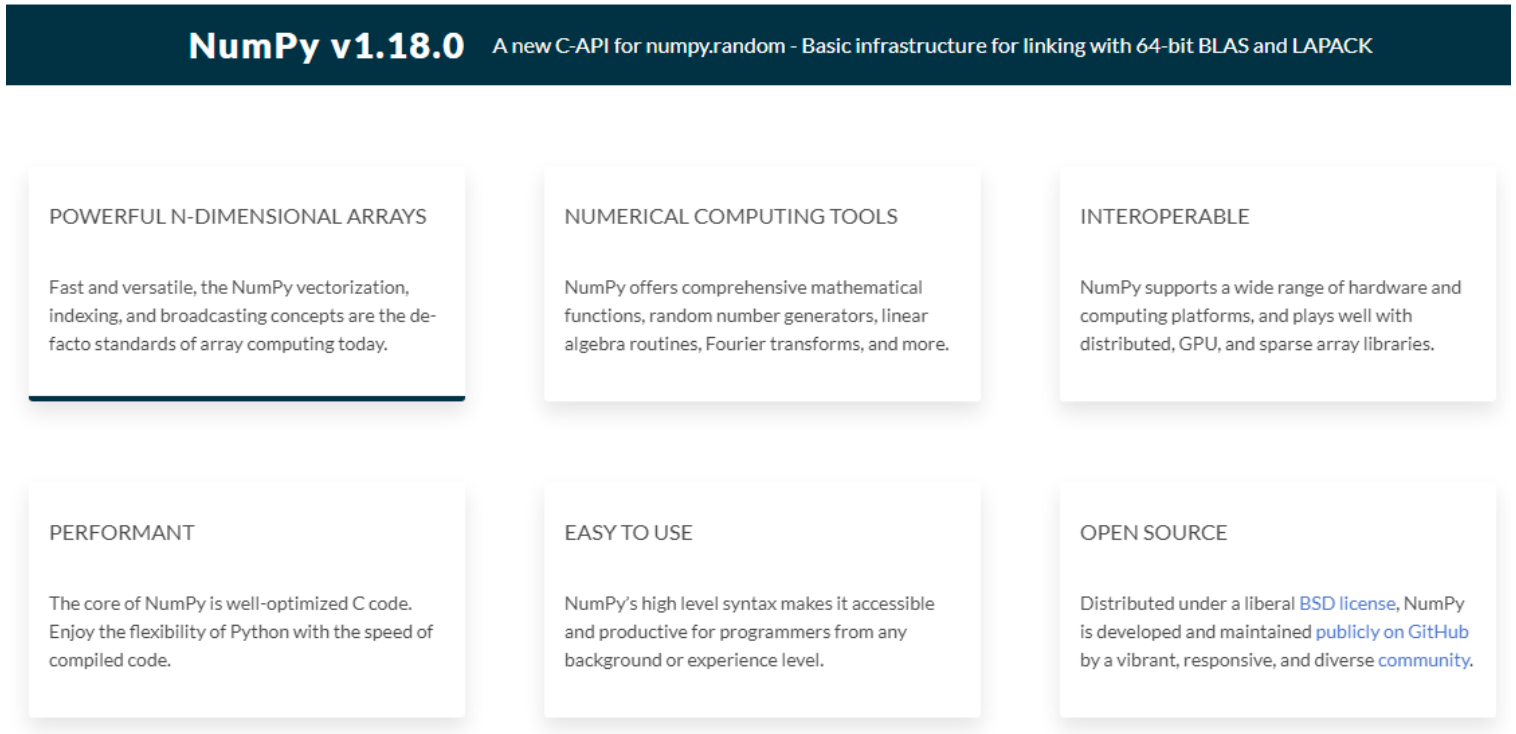
전체 커리큘럼 소개



LEARNING SPOONS

넘파이 NumPy

- 수학 및 과학 연산을 위한 파이썬 패키지

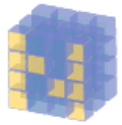


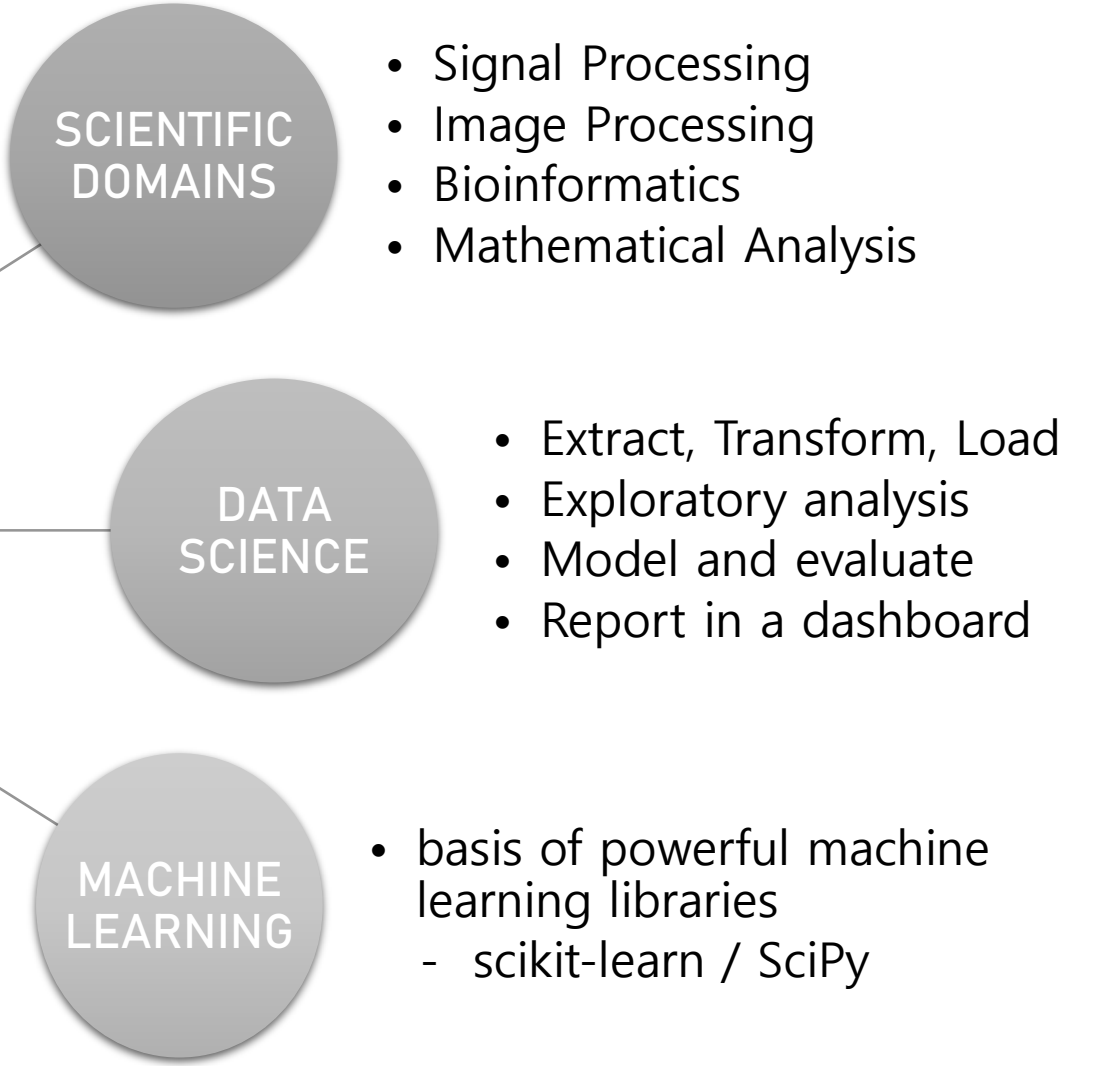
그림출처 : https://numpy.org/

LEARNING SPOONS

ECOSYSTEM

- 고수준의 라이브러리
  - 사용자에 친화적이므로 가져다 사용하기 쉬움

NumPy 



LEARNING SPOONS

LEARNING SPOONS ONLINE

Numpy와 ndarray

데이터 분석을 위한 Numpy

LEARNING SPOONS

Numpy 사용을 위한 import

- numpy 모듈을 np 이름으로 임포트

```
import numpy as np
```

LEARNING SPOONS

ndarray 클래스

- 넘파이에서 제공하는 자료구조
  - 파이썬의 기본 자료구조 리스트를 업그레이드한 클래스
  - 리스트의 기능 + 추가 기능

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

```
[1 2 3 4]
<class 'numpy.ndarray'>
```

리스트를 ndarray로 업그레이드!

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

```
[[1 2]
 [3 4]]
<class 'numpy.ndarray'>
```

LEARNING SPOONS

이차원 데이터

- 가로축과 세로축이 있는 형태의 데이터

	고가	시가	저가	종가
2020/06/25	100	80	70	90
2020/06/26	120	110	100	110

```
arr = [ [100,  80,  70,  90],
        [120, 110, 100, 110]]
```

LEARNING SPOONS

zeros / ones 함수

- 정해진 값으로 채워진 ndarray를 생성
  - 파라미터에 데이터 수를 지정

```
a = np.zeros( 3 )
print(a)
```

```
[0. 0. 0.]
```

```
a = np.ones( 3 )
print(a)
```

```
[1. 1. 1.]
```

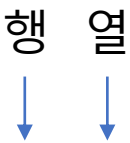
LEARNING SPOONS

zeros / ones 메서드

- 튜플을 전달하면 이차원 형태로 데이터를 구성
  - 튜플에 행과 열을 차례로 입력

```
a = np.zeros( (3, 3) )
print(a)
```

```
[[0. 0. 0.]
 [0. 0. 0.]
 [0. 0. 0.]]
```




```
a = np.ones( (3, 1) )
print(a)
```

```
[[1.]
 [1.]
 [1.]]
```

LEARNING SPOONS

arange 함수

- 주어진 구간을 ndarray로 반환하는 함수
  - 빌트인 함수인 range와 같은 방식으로 파라미터 전달
  - 파라미터로 시작, 끝, 증감폭을 순서대로 전달

```
a = np.arange( 5 )
print(a)
```

```
[0 1 2 3 4]
```

```
a = np.arange( 1, 5 )
print(a)
```

```
[1 2 3 4]
```

```
a = np.arange( 1, 5, 2 )
print(a)
```

```
[1 3]
```

LEARNING SPOONS

ndarray의 차원 변환

- ndarray의 shape을 변환하는 함수

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

```
[[1 2]
 [3 4]]
```

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

```
[[1]
 [2]
 [3]
 [4]]
```

LEARNING SPOONS

LEARNING SPOONS ONLINE

데이터 타입

데이터 분석을 위한 Numpy

numpy의 데이터 타입

- 수치 연산을 위해 세분화된 데이터 타입
  - 실수와 복소수를 구분
  - 8 / 16 / 32 / 64 비트 형 데이터 타입

데이터 타입	구분
bool	불린형
int8 / int16 / int32 / int64 uint8 / uint16 / uint32 / uint64	정수형
float8 / float16 / float32 / float64	부동소수형
complex64 / complex128	복소수형
string	문자형

LEARNING SPOONS

이진법 (1/3)

- 컴퓨터는 이진수로 모든 값을 기록함
  - 0과 1 두 개의 수를 사용

```
1 2 3 10
```

$$1 \times 10^2 + 2 \times 10^1 + 3 \times 10^0$$

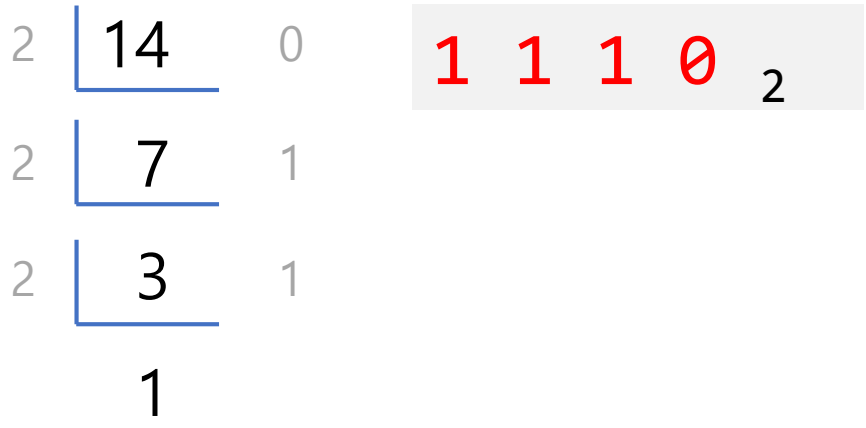
```
1 0 1 2
```

$$1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$$

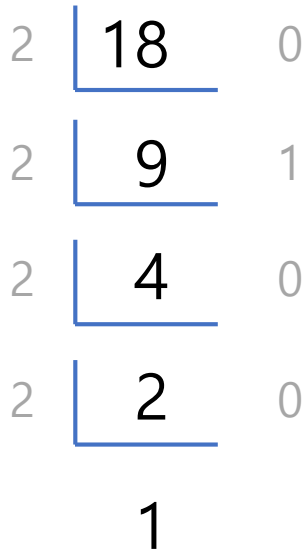
LEARNING SPOONS

이진법 (2/3)

- 십진수를 이진수로 변환



1110_2



10010_2

LEARNING SPOONS

이진법 (3/3)

- 네 자리 이진수로 표현할 수 있는 값의 범위

$$X\ X\ X\ X_{2}$$

  - 부호없는 숫자
    - 0000_2 ~ 1111_2 (0 ~ 15)의 16개의 숫자를 표현할 수 있음
    - 한 자리에 나올 수 있는 수가 2개 (0 또는 1) 이므로 2 x 2 x 2 x 2

  - 부호있는 숫자
    - 상위 비트를 부호 비트로 사용
    - 1111_2 ~ 0111_2 (-8 ~ 7)의 16개의 숫자를 표현할 수 있음

LEARNING SPOONS

넘파이의 다양한 속성

- 생성할 때 타입을 지정해 주거나 강제로 형 변환

```
a = np.zeros( (2, 3) )
print(a.dtype)
print(a.shape)
print(a.ndim)
print(a.itemsize)
```

```
float64
(2, 3)
2
8
```

```
a = np.array( [1,2,3,4] )
print(a.dtype)
print(a.shape)
print(a.ndim)
print(a.itemsize)
```

```
int32
(4,)
1
4
```

LEARNING SPOONS

넘파이의 형변환

- 생성할 때 타입을 지정해 주거나 강제로 형 변환

```
a = np.zeros( 3, dtype=np.in32 )
print(a)
print(a.dtype)
```

```
[0 0 0]
int32
```

```
a = np.zeros( 3 )
print(a.dtype)

a = np.int32(a)
print(a.dtype)

a = a.astype(np.int8)
print(a.dtype)
```

LEARNING SPOONS

LEARNING SPOONS ONLINE

Numpy와 함수

데이터 분석을 위한 Numpy

LEARNING SPOONS

임의의 수 생성

- randint() 함수는 임의의 정수를 생성
  - 숫자를 넣을 경우 0부터 x-1까지의 범위에서 생성

```
a = np.random.randint(4)
print(a)
```

```
3
```

```
a = np.random.randint(10, 20)
print(a)
```

```
15
```

실행할 때마다 다른 값이 생성

```
a = np.random.randint(10, 20, size=(2,2))
print(a)
```

```
[[15 18]
 [19 18]]
```

LEARNING SPOONS

ndarray의 sampling (1/2)

- 입력된 값에서 size개 샘플링 (복원 추출)

```
a = np.random.choice( np.arange(4), size=(1,4) )
print(a)
```

```
[[1 2 0 3]]
```

```
a = np.random.choice( 4, size=(1,4) )
print(a)
```

```
[[1 2 0 3]]
```

정수만 입력하면 arange() 함수를 사용하는 것과 같음

LEARNING SPOONS

ndarray의 sampling (2/2)

- 입력된 값에서 size개 샘플링 (비복원 추출)

```
a = np.random.choice(4, size=3, replace=False)
print(a)
```

```
[3 1 0]
```

LEARNING SPOONS

연습 문제 - 1

- 로또 번호 6개를 생성하는 "로또번호생성기" 이름의 함수를 작성하라
  - 중복된 번호를 허용
  - numpy의 randint 함수를 사용

LEARNING SPOONS

linspace 함수

- 특정 범위를 입력된 개수로 분할
  - 시작, 끝, 분할 개수를 차례로 입력

```
a = np.linspace(0, 10, 3)
print(a)
```

```
[ 0.  5. 10.]
```

LEARNING SPOONS

연습 문제 - 2

- 다음 정의역에 대해 $y = x^2$를 만족하는 y를 반복문을 사용하여 계산하라.
  - `x = np.linspace(0, 10, 100)`

LEARNING SPOONS

다양한 함수

- 공식 홈페이지 참고
  - https://numpy.org/doc/stable/reference/routines.html

- Array creation routines
  - Ones and zeros
  - From existing data
  - Creating record arrays (**numpy.rec**)
  - Creating character arrays (**numpy.char**)
  - Numerical ranges
  - Building matrices
  - The Matrix class
- Array manipulation routines
  - Basic operations
  - Changing array shape
  - Transpose-like operations
  - Changing number of dimensions
  - Changing kind of array
  - Joining arrays
  - Splitting arrays
  - Tiling arrays
  - Adding and removing elements
  - Rearranging elements
- Binary operations
  - Elementwise bit operations
  - Bit packing
  - Output formatting
- String operations
  - String operations
  - Comparison
  - String information
  - Convenience class
- C-Types Foreign Function Interface (**numpy.ctypeslib**)
- Datetime Support Functions
  - numpy.datetime_as_string
  - numpy.datetime_data
  - Business Day Functions
- Data type routines
  - numpy.can_cast
  - numpy.promote_types
  - numpy.min_scalar_type
  - numpy.result_type
  - numpy.common_type
  - numpy.obj2sctype
  - Creating data types
  - Data type information
  - Data type testing
  - Miscellaneous
- Optionally Scipy-accelerated routines (**numpy.dual**)
  - Linear algebra
  - FFT
  - Other

LEARNING SPOONS

LEARNING SPOONS ONLINE

인덱싱과 슬라이싱

데이터 분석을 위한 Numpy

LEARNING SPOONS

넘파이의 인덱싱 (1/3)

- 리스트와 동일한 인덱싱을 제공

```
a = np.arange(5)
print(a[0])
```

```
0
```

```
a = np.arange(5)
print(a[-1])
```

```
4
```

LEARNING SPOONS

넘파이의 인덱싱 (2/3)

- 행 단위로 우선 접근

```
a = np.arange(9).reshape(3,3)
print(a[0])
```

```
[0 1 2]
```

```
print(a[1])
```

```
[3 4 5]
```

0	1	2	← [0]
3	4	5	← [1]
6	7	8	← [2]

LEARNING SPOONS

넘파이의 인덱싱 (3/3)

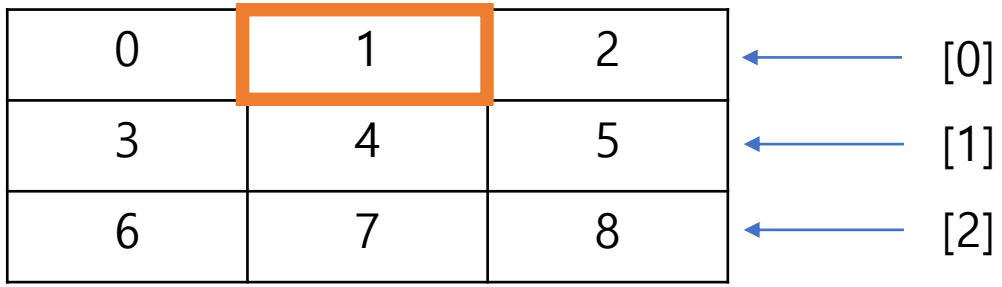
- 행 -> 열 순서로 접근

```
a = np.arange(9).reshape(3,3)
print(a[0][1])
```

```
1
```

```
print(a[0, 1])
```

```
1
```



LEARNING SPOONS

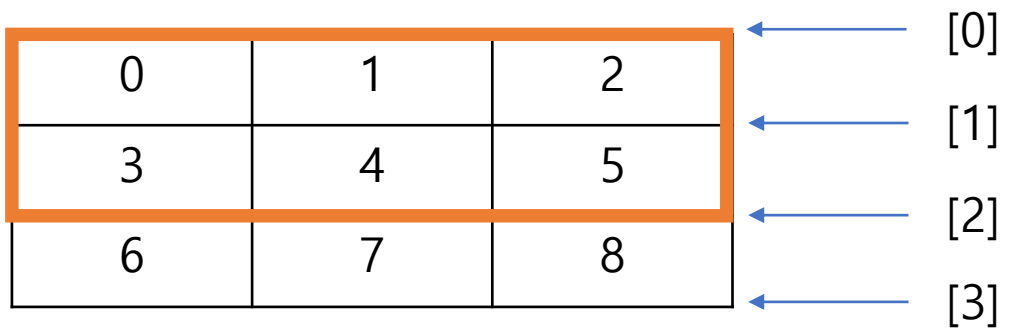
넘파이의 슬라이싱 (1/3)

- 행 -〉 열 순서로 접근

```
a = np.arange(9).reshape(3,3)
print(a[0:2])
```

```
[[0 1 2]
 [3 4 5]]
```

슬라이싱한 결과는 ndarray

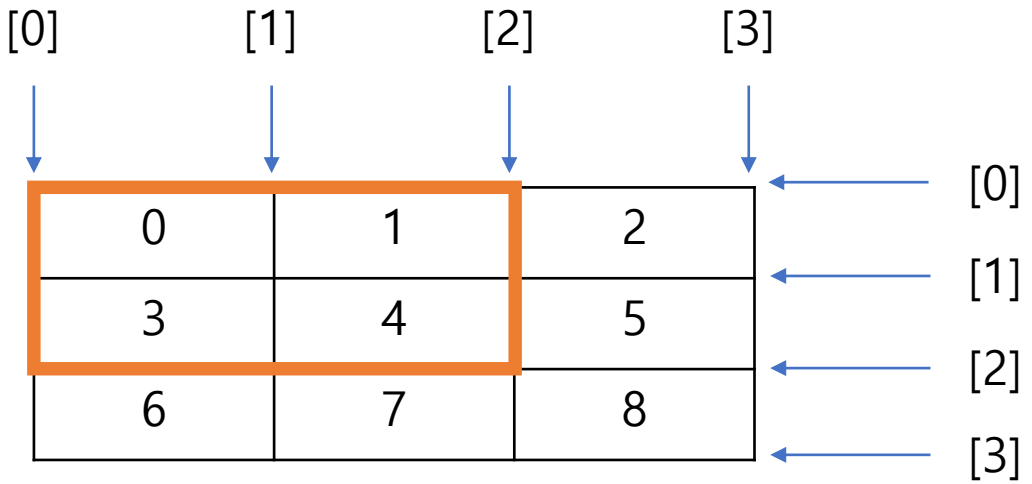


넘파이의 슬라이싱 (2/3)

- 행 -〉 열 순서로 접근

```
a = np.arange(9).reshape(3,3)
print(a[0:2, 0:2])
```

```
[[0 1]
 [3 4]]
```



LEARNING SPOONS

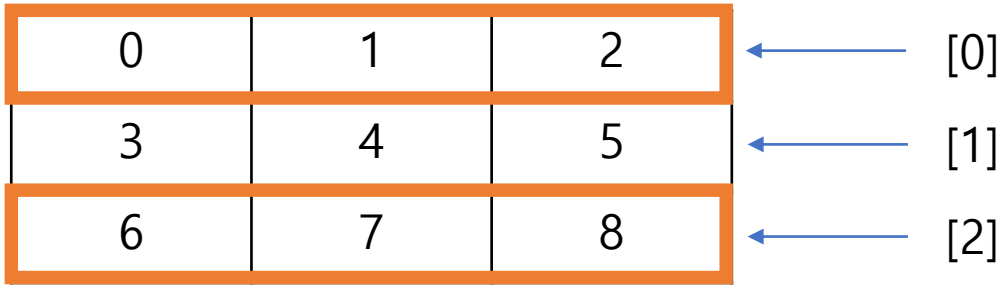
넘파이의 슬라이싱 (3/3)

- 비연속적인 행을 슬라이싱
  - 인덱스의 리스트로 표현

```
a = np.arange(9).reshape(3,3)
target = [0, 2]
print(a[target])
```

```
a = np.arange(9).reshape(3,3)
print(a[[0, 2]])
```

```
[[0 1 2]
 [6 7 8]]
```



LEARNING SPOONS

연습 문제 - 1

- 학생들의 점수를 ndarray로 표현하라.
  - 문자열을 제외하고 점수만 표현

	국어	영어	수학
중간	30	25	66
기말	23	44	23

LEARNING SPOONS

연습 문제 - 2

- 기말고사 점수의 합을 출력하라.
  - sum() 함수 사용

	국어	영어	수학
중간	30	25	66
기말	23	44	23

LEARNING SPOONS

연습 문제 - 3

- 영어 점수의 평균을 출력하라.
  - sum() 함수 사용

	국어	영어	수학
중간	30	25	66
기말	23	44	23

LEARNING SPOONS

LEARNING SPOONS ONLINE

Numpy와 수치연산

데이터 분석을 위한 Numpy

LEARNING SPOONS

Numpy의 사칙연산 (1/2)

- 브로드캐스팅 (Boradcasting)
  - 전체 데이터에 대해 적용 됨

```
a = np.arange(4)
print( a * 3 )
```

```
[ 0  3  6  9]
```

```
a = np.arange(4)
print( a + 4 )
```

```
[ 4  5  6  7]
```

```
a = np.arange(4)
print( a / 2 )
```

```
[0.   0.5  1.   1.5]
```

LEARNING SPOONS

Numpy의 사칙연산 (2/2)

- 브로드캐스팅 (Boradcasting)
  - ndarray와 ndarray의 연산도 가능

```
a = np.arange(4)
b = np.arange(4, 8)
print(a+b)
```

```
[ 4  6  8 10]
```

```
a = np.arange(4)
b = np.arange(5)
print( a - b )
```

```
ValueError: operands could not be broadcast together with shapes (4,) (5,)
```

LEARNING SPOONS

연습문제 - 1

- 중간고사 성적의 분산을 출력하라
  - 점수는 ndarray에 저장
  - $var(X) = E((X - \mu)^2)$

	국어	영어	수학
중간	30	25	66
기말	23	44	23

LEARNING SPOONS

연습문제 - 2

- 중간/기말고사 성적의 분산을 각각 출력하라
  - 점수는 ndarray에 저장
  - $\mathrm{var}(X) = E((X - \mu)^2)$

	국어	영어	수학
중간	30	25	66
기말	23	44	23

LEARNING SPOONS

연습문제 - 3

- 국어/영어/수학 과목별 분산을 출력하라
  - 점수는 ndarray에 저장
  - $var(X) = E((X - \mu)^2)$

	국어	영어	수학
중간	30	25	66
기말	23	44	23

LEARNING SPOONS

Numpy의 수치연산 함수

- Sums, products, differences
  - mean()
  - max()
  - min()
  - median()
  - prod()
  - cumprod()
  - cumsum()
  - np.abs()
  - np.square()

LEARNING SPOONS

LEARNING SPOONS ONLINE

Numpy와 비교연산

데이터 분석을 위한 Numpy

LEARNING SPOONS

ndarray의 비교 연산 (1/2)

- 크다/작다/같다/다르다 등의 비교 연산을 지원
  - 연산의 결과는 Boolean 데이터 타입

```
a = np.arange(5)
print( a > 2 )
```

```
[False False False  True  True]
```

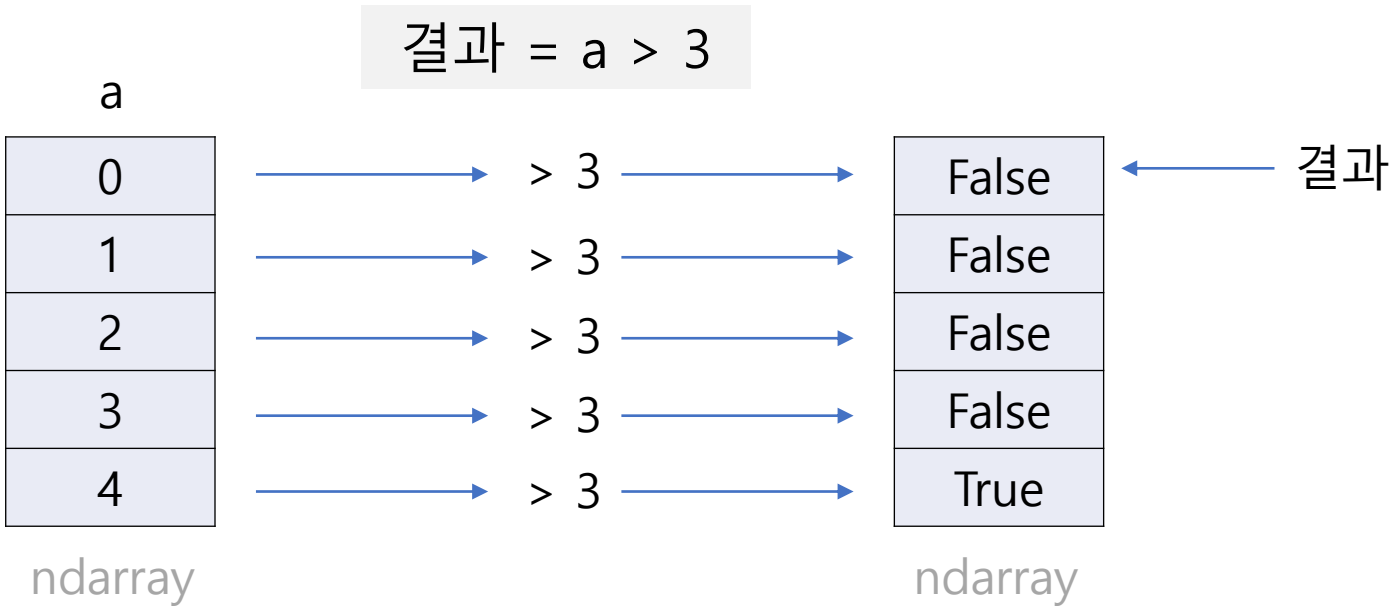
```
a = np.arange(5)
print( a != 2 )
```

```
[ True  True False  True  True]
```

LEARNING SPOONS

ndarray의 비교 연산 (2/2)

- 모든 데이터에 비교 연산이 적용
  - 연산의 결과는 ndarray



LEARNING SPOONS

연습문제 - 1

- 결과 ndarray에 저장되는 값을 채워 넣으세요.

결과 = 시가 > 종가

시가	종가	결과
92600	96600	
92400	91400	
92100	93100	
94300	96300	
92300	90300	

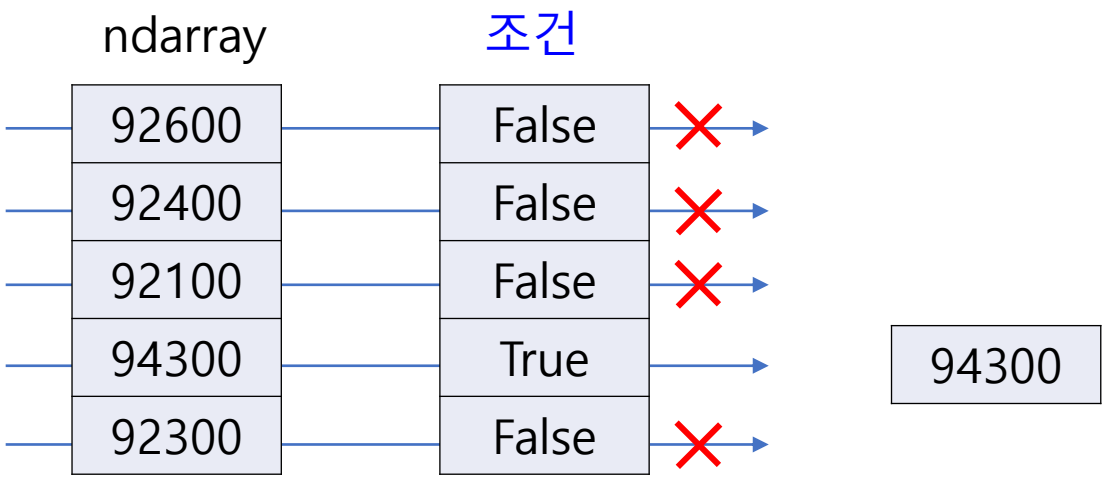
ndarray

LEARNING SPOONS

ndarray와 조건문

- 조건을 사용해서 원본 값을 필터링!
  - 연산의 결과는 ndarray

```
>> 조건 = ndarray > 93000

>> ndarray [조건]
94300
```



줄여서 한번에 써도 됨

```
>> ndarray [ndarray > 93000]
94300
```

LEARNING SPOONS

코드를 한국말로 해석해 봅시다.

```
>> ndarray [          ]
```

ndarray에서 값을 가져와라

```
>> ndarray [     0    ]
```

ndarray 에서 0번째 값을 가져와라

```
>> ndarray [    조건   ]
```

ndarray에서 조건을 만족하는 값을 가져와라

```
>> ndarray [ndarray>100]
```

ndarray 에서 100보다 큰 값을 가져와라

LEARNING SPOONS

연습문제 - 1

- LG전자의 종가 데이터가 ndarray 형태로 저장되어 있다. LG 전자의 주가가 85000원 보다 작을 때의 가격을 출력하라.

```
lge = np.array([93000, 82400, 99100, 81000, 72300])
```

```
[82400 81000 72300]
```

LEARNING SPOONS

연습문제 - 2

- LG전자의 종가 데이터가 ndarray 형태로 저장되어 있다. LG 전자의 주가가 85000원 미만으로 떨어진 횟수는?

```
lge = np.array([93000, 82400, 99100, 81000, 72300])
```

3

LEARNING SPOONS

연습문제 - 3

- 고가와 저가의 차이가 100 이상인 날의 고가를 출력하라.

```
저가 = np.array([10, 200, 200, 400, 600])
고가 = np.array([100, 300, 400, 500, 600])
```

```
[300 400 500]
```

LEARNING SPOONS

연습문제 - 4

- 종가가 80000원 이상 90000원 미만인 값을 출력하라

```
종가 = np.array([93000, 82400, 99100, 81000, 72300])
```

```
[82400 81000]
```

LEARNING SPOONS

LEARNING SPOONS ONLINE

Numpy와 대입연산

데이터 분석을 위한 Numpy

LEARNING SPOONS

1차원 ndarray (1/2)

- ndarray에 대입 연산은 전체 데이터에 적용
  - 어디에 값을 채울지 명확히 지정해야 함
  - 인덱싱으로 특정 위치의 값을 변경

```
a = np.arange(5)
a[ 2 ] = 0
print(a)
```

```
[0 1 0 3 4]
```

```
a = np.arange(5)
a[ -1 ] = 0
print(a)
```

```
[0 1 2 3 0]
```

```
a = np.arange(5)
a = 0
print(a)
```

```
0
```

LEARNING SPOONS

1차원 ndarray (2/2)

- ndarray에 대입 연산은 전체 데이터에 적용
  - 슬라이싱으로 다수의 값을 한 번에 변경

```
a = np.arange(5)
a[ : ] = 0
print(a)
```

```
[0 0 0 0 0]
```

```
a = np.arange(5)
a[ : 3 ] = 0
print(a)
```

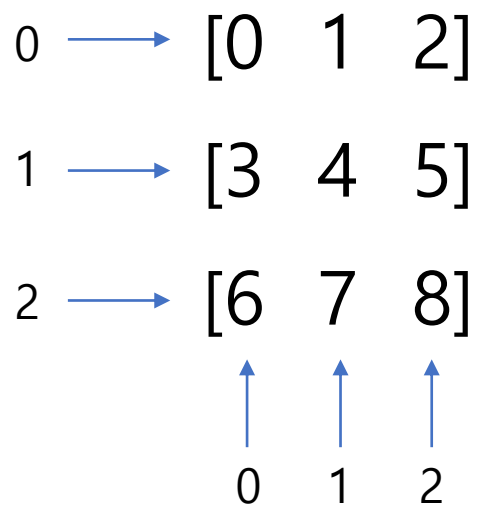
```
[0 0 0 3 4]
```

```
a = np.arange(5)
a[ [1, 3] ] = 0
print(a)
```

```
[0 0 2 0 4]
```

LEARNING SPOONS

2차원 ndarray (1/2)



- ndarray에 대입 연산은 전체 데이터에 적용
  - 2차원 ndarray를 인덱싱을 한 결과 1차원 ndarray
  - 1차원 ndarray에 대입 연산을 적용하면 전체 데이터가 변경

```
a = np.arange(9).reshape(3, 3)
a[ 0 ] = 0
print(a)
```

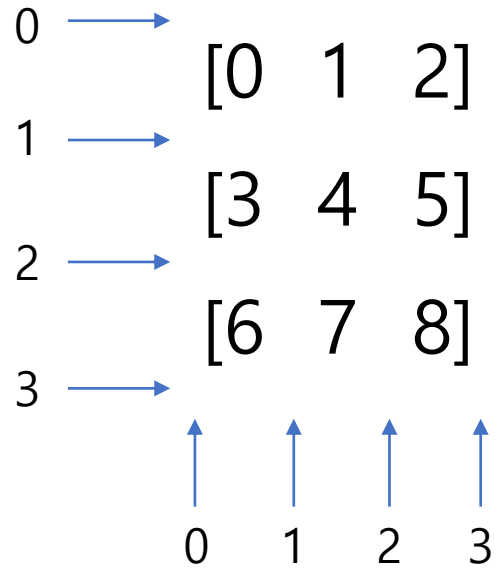
```
[[0 0 0]
 [3 4 5]
 [6 7 8]]
```

```
a = np.arange(9).reshape(3, 3)
a[ 1, 1 ] = 0    # a[1][1]
print(a)
```

```
[[0 1 2]
 [3 0 5]
 [6 7 8]]
```

LEARNING SPOONS

2차원 ndarray (2/2)



- ndarray에 대입 연산은 전체 데이터에 적용
  - 2차원 ndarray를 슬라이싱한 결과 2차원 ndarray
  - 2차원 ndarray에 대입 연산을 하면 선택된 모든 데이터가 변경

```
a = np.arange(9).reshape(3, 3)
a[ [0, 2] ] = 0
print(a)
```

```
[[0 0 0]
 [3 4 5]
 [0 0 0]]
```

```
a = np.arange(9).reshape(3, 3)
a[ 0:2, 0:2 ] = 0
print(a)
```

```
[[0 0 2]
 [0 0 5]
 [6 7 8]]
```

LEARNING SPOONS

조건문과 ndarray (1/2)

- 조건 연산으로 슬라이싱한 뒤 대입 연산을 적용
  - 조건 연산의 결과는 True, False가 저장된 ndarray

```
a = np.arange(5)
a[ a > 2 ] = 0
print(a)
```

```
[0 1 2 0 0]
```

```
a = np.arange(9).reshape(3, 3)
a[ a > 4 ] = 10
print(a)
```

```
[[ 0  1  2]
 [ 3  4 10]
 [10 10 10]]
```

LEARNING SPOONS

조건문과 ndarray (2/2)

- np.where 함수
  - 조건을 충족하거나 그렇지 않은 경우, 각각에 적용할 연산을 한 번에 지정
  - np.where(조건, 참일때 실행, 거짓일때 실행)
  - 연산을 적용한 결과를 ndarray로 반환

```
a = np.arange(9).reshape(3, 3)
a = np.where(a > 4, 10, a)
print(a)
```

a가 4보다 크면 10을 대입하고,
그렇지 않으면 a를 그대로 사용

```
[[ 0  1  2]
 [ 3  4 10]
 [10 10 10]]
```

LEARNING SPOONS

LEARNING SPOONS ONLINE

이미지로 연습하는 Numpy

데이터 분석을 위한 Numpy

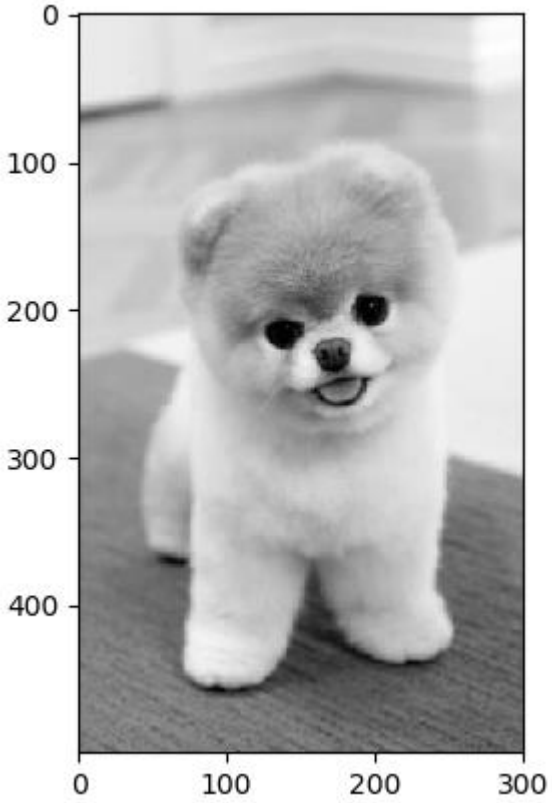
LEARNING SPOONS

이미지 읽고 출력하기

- matplotlib를 사용한 이미지 입출력

```
from matplotlib import image
import matplotlib.pyplot as plt

img = image.imread('gray.jpg')
img = img[:, :, 0].copy()

print(img.dtype)
print(img.shape)

plt.imshow(img, cmap='gray')
plt.show()
```

```
uint8
(500, 301)
```

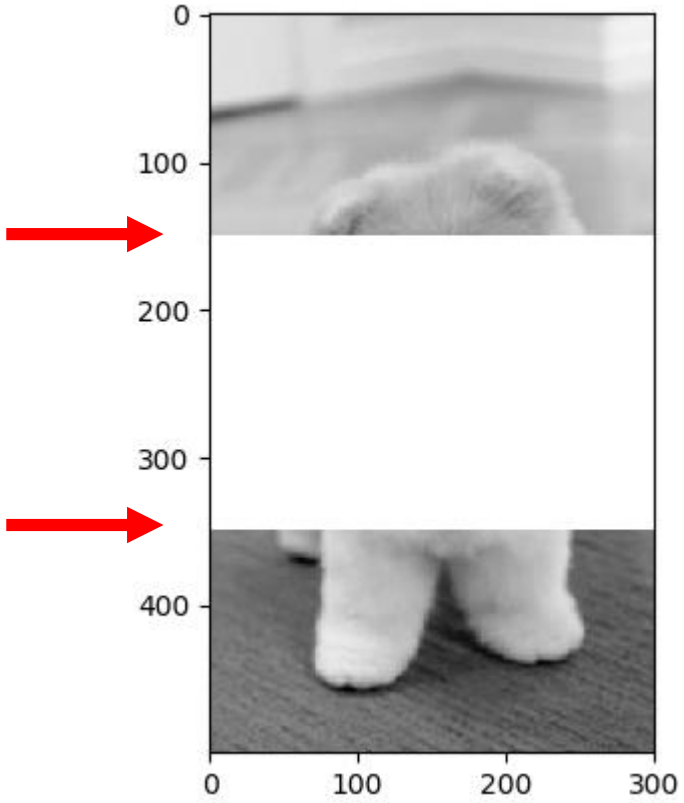


LEARNING SPOONS

이미지 수정

- 이미지 기본 속성 이해하기
  - uint8 영상은 최대 255의 값을 갖음
  - 0은 검정 255는 흰색을 의미

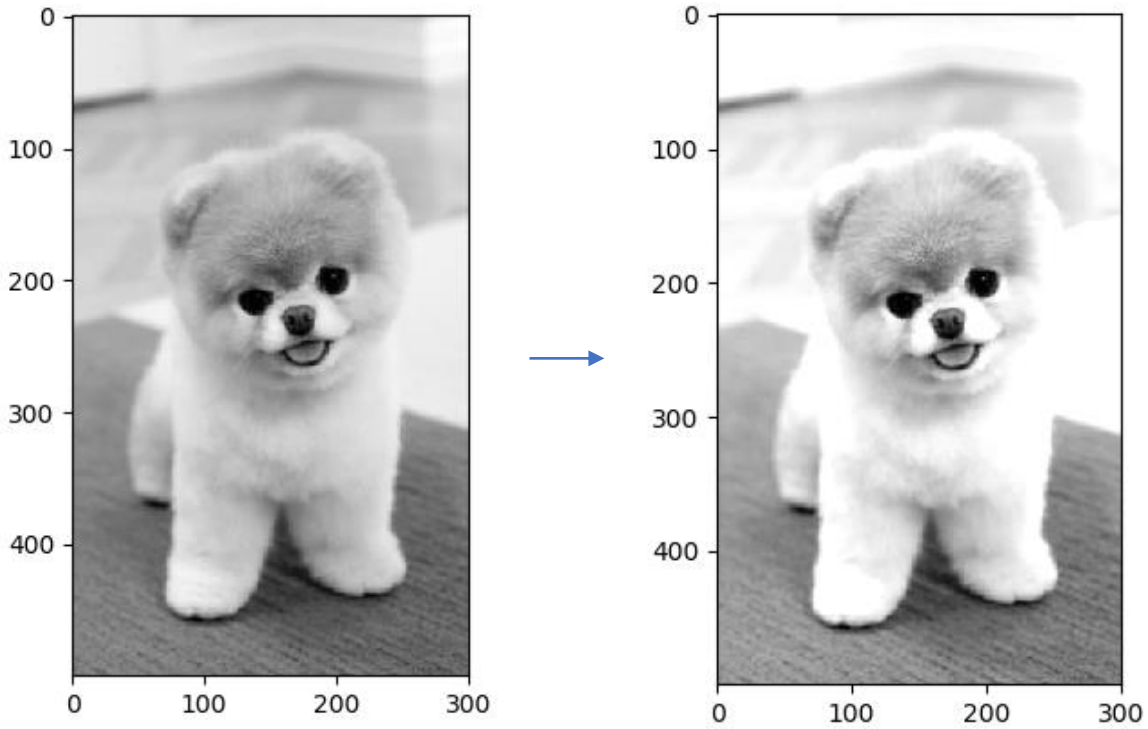
```
from matplotlib import image
import matplotlib.pyplot as plt

img = image.imread('gray.jpg')
img = img[:, :, 0].copy()

img[ 150 : 350 ] = 255

plt.imshow(img, cmap='gray')
plt.show()
```



LEARNING SPOONS

연습 문제 - 1

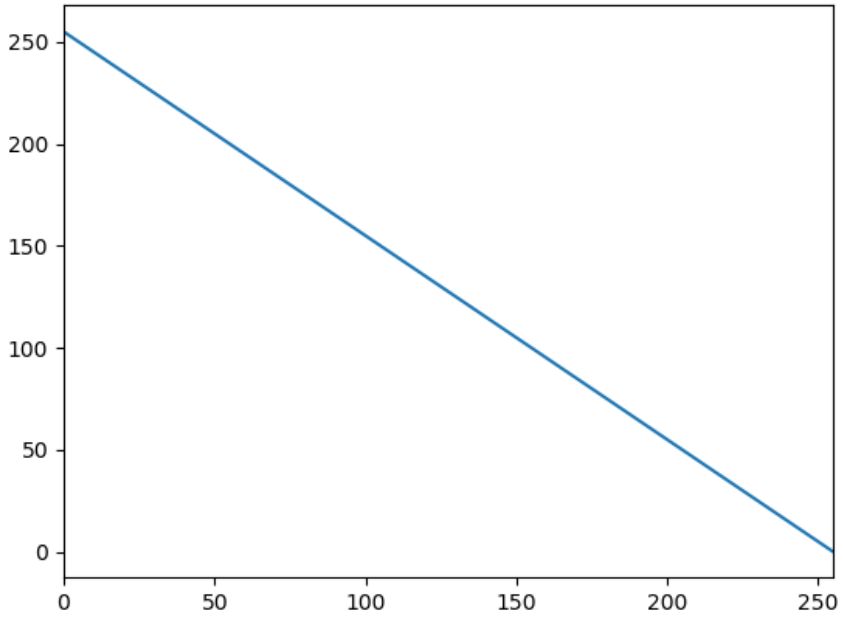
- 예외 처리를 포함하여 영상을 50 만큼 밝게 표현하라.
  - 참고로 205보다 큰 경우 255로 지정해야 올바르게 출력 됨.

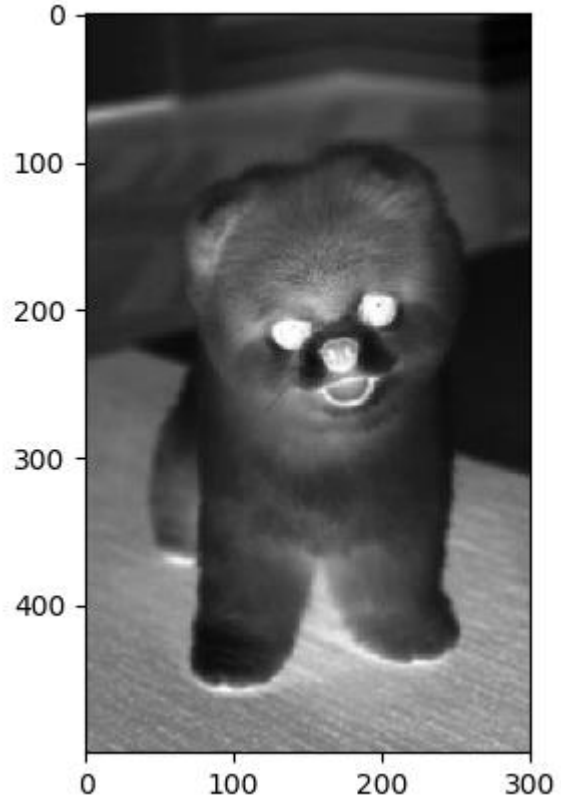


LEARNING SPOONS

연습 문제 - 2

- 반전영상을 생성하라.
  - 다음 그래프($y = 255 - x$)에 대응되도록 색상을 변경
  - 예를 들어 0을 255로, 255를 0으로 변환





LEARNING SPOONS

연습 문제 - 3

- 1) 영상을 세로축 절반 크기로 축소하라.
- 2) 영상을 가로/세로 절반 크기로 축소하라.
  - 힌트 : 슬라이싱 증감폭

LEARNING SPOONS

연습 문제 - 4

- 좌측 상단에 로고 파일을 합성하라.
  - 로고 파일은 "logo.jpg"
  - 로고 파일의 크기를 축소해야 함

```
bg = image.imread('gray.jpg')
bg = bg[:, :, 0].copy()

logo = image.imread('logo.jpg')
logo = logo[:, :, 0].copy()




plt.imshow(bg, cmap='gray')
plt.show()
```