

# Introduction

## □ 자기 소개

- I am ... 학생? 직장인? 취준생?
- 머신러닝 관련하여 경험치는? (공부 수준은?)
- 이 스터디를 통해서 얻고 싶은 결과는? (왜 참여하게 되었는지)

## □ Ground Rule

- Slack !!!
- Camera !!

MODULABS 검색

# 04-1월-토-책거리-파이썬-머신러닝-완벽-가이드

+ 채팅피 추가

1 (10:30-10:50) 첫 모임 첫 시간에는 풀잎스쿨을 소개하는 OT가 진행됩니다.  
• 구글 MEET로 동시간대 모임 분들과 함께하는 자리니 모임 전 먼저 1월 18일 목요일 17:00부터 18:00까지 서서 OT에 참여해주세요.  
• 링크: <https://meet.google.com/htq-scov-txu>

2 (10:50-12:30) OT가 끝난 후 아래 모임 링크(ZOOM)으로 입장하셔서 모임을 이어가주세요!  
• 링크: <https://us02web.zoom.us/j/81650825482?pwd=dERob3UyUDlvdk1LaENkZ3ZqbFBjUT09>  
• 회의 ID: 816 5082 5482  
• 암호: 275743

그럼 토요일에 뵙겠습니다!  
메세지를 읽으신 후 ☒ 를 눌러주세요!

어제

이나진 오후 10:20  
#04-1월-토-책거리-파이썬-머신러닝-완벽-가이드에 참여했습니다.

오늘

박재규 오전 10:33  
@channel  
지금 시작했습니다!!! ^^  
(10:30-10:50) 첫 모임 첫 시간에는 풀잎스쿨을 소개하는 OT가 진행됩니다.  
• 구글 MEET로 동시간대 모임 분들과 함께하는 자리니 모임 전 먼저 아래 링크로 들어오셔서 OT에 참여해주세요.  
• 링크: <https://meet.google.com/htq-scov-txu>

meet.google.com  
Meet  
Real-time meetings by Google. Using your browser, share your video, desktop, and presentations with teammates and customers.

B I

#04-1월-토-책거리-파이썬-머신러닝-완벽-가이드에 메시지 보내기

+ Aa

# 파이썬 머신러닝 완벽 가이드

개정2판

간혹한 개념 예제와 함께  
기초부터 심화까지  
최신 기법까지 배우기

DS 데이터 사이언스 시리즈 081

권철민 지음



이제 나도 전문가

한빛미디어

# 혼자 공부하는 데이터 분석

데이터 분석을 어떻게 시작할지 막막할 때  
실생활 데이터로 기초를 제대로 다지고 싶을 때  
데이터에서 의미 있는 결과를 얻고 싶을 때



with 파이썬  
박해선 지음

★★★★★

[신] 혼공분석

혼자 공부하며 복잡한 문제를 해결하는 모습을 일컫는 신조어  
[유의어] 혼공족, 혼공파, 혼공머신

구글 코랩으로  
환경 설정 없이  
실습 가능

유튜브 강의  
발책  
용이 노트

TALK  
저자와 함께하는  
오픈채팅  
...  
<http://bit.ly/tensor-chat>

한빛미디어  
HANBIT MEDIA, INC.

## □ 스터디 방식

- 돌아가면서 요약 **강의** 진행
- 책 요약 + 보충 학습 + 실습

## □ 일정 조율

- 양해의 말씀 ...
- 설날에는 ?

No	Date	Agenda
1	01-20	Introduction + 01. 파이썬 기반의 머신러닝과 생태계 이해
2	01-27	02. 사이킷런으로 시작하는 머신러닝
3	(02-03)	03. 평가
4	(02-10)	04. 분류 #1
5	02-17	04. 분류 #2
6	02-24	05. 회귀 #1
7	03-02	05. 회귀 #2 + 06. 차원축소
8	03-09	07. 군집화
9	03-16	08. 텍스트 분석 #1
10	03-23	08. 텍스트 분석 #2
11	03-30	09. 추천 시스템
12	(04-06)	10. 시각화
13	(04-13)	

## 01

파이썬 기반의 머신러닝과  
생태계 이해

오늘은 여기까지~ ^^

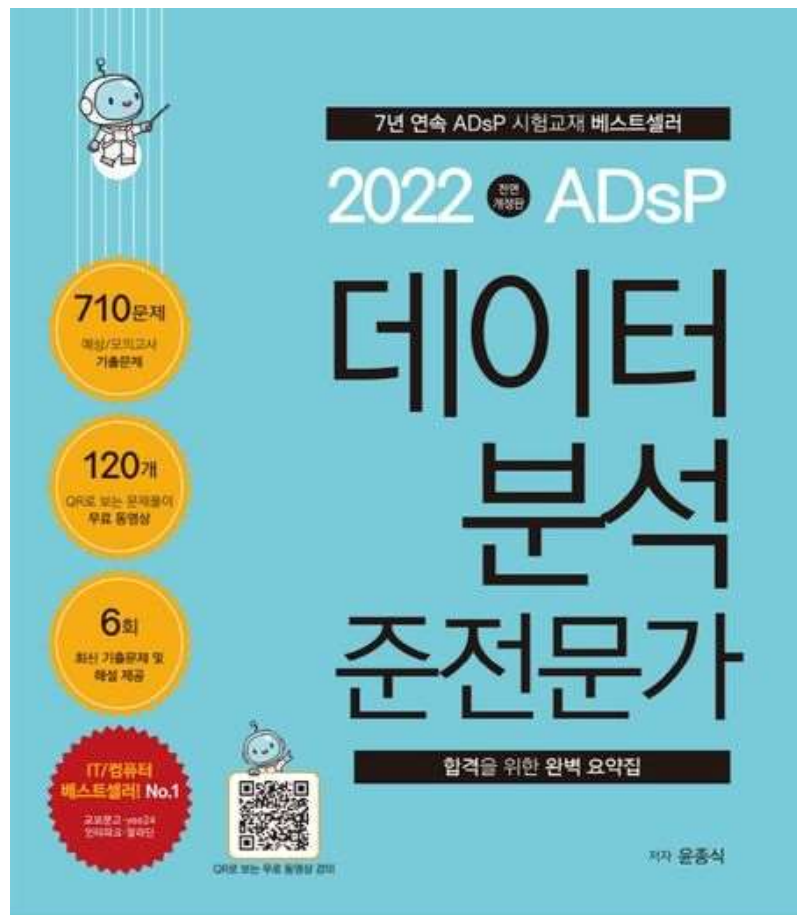
01. 머신러닝의 개념	1
머신러닝의 분류	2
데이터 전쟁	3
파이썬과 R 기반의 머신러닝 비교	4
02. 파이썬 머신러닝 생태계를 구성하는 주요 패키지	6
파이썬 머신러닝을 위한 S/W 설치	7
03. 넘파이	13
넘파이 ndarray 개요	14
ndarray의 데이터 타입	16
ndarray를 편리하게 생성하기 - arange, zeros, ones	18
ndarray의 차원과 크기를 변경하는 reshape( )	20
넘파이의 ndarray의 데이터 세트 선택하기 - 인덱싱(indexing)	23
행렬의 정렬 - sort( )와 argsort( )	33
선형대수 연산 - 행렬 내적과 전치 행렬 구하기	37
04. 데이터 핸들링 - 판다스	39
판다스 시작 - 파일을 DataFrame으로 로딩, 기본 API	40
DataFrame과 리스트, 딕셔너리, 넘파이 ndarray 상호 변환	49
DataFrame의 칼럼 데이터 세트 생성과 수정	53
DataFrame 데이터 삭제	55
Index 객체	58
데이터 셀렉션 및 필터링	62
정렬, Aggregation 함수, GroupBy 적용	73
결손 데이터 처리하기	77
apply lambda 식으로 데이터 가공	80
05. 정리	84

01

# 파이썬 기반의 머신러닝과 생태계 이해

Machine Learning  
vs.  
Data Science

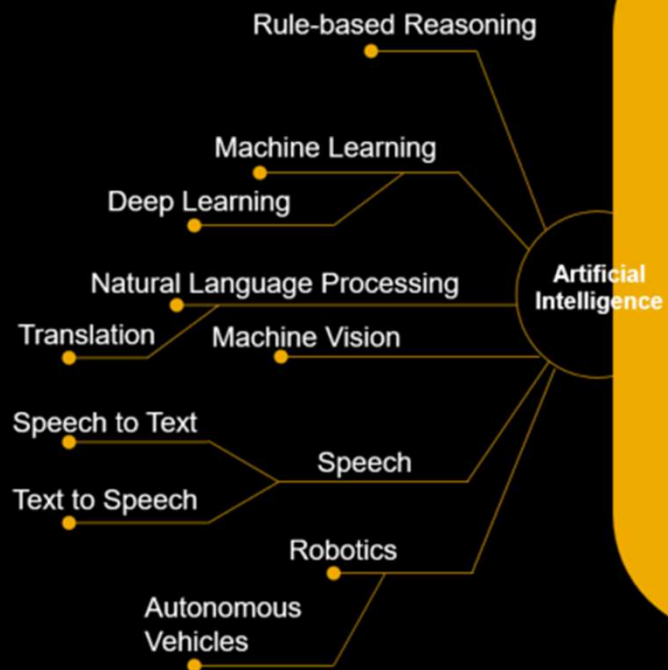




Korea Satisfaction Consumer Index  
2021 한국소비자만족지수 1위

DATA 교육부 선정  
EDU 빅데이터 전문교육기관

## AI Areas of Research



**Artificial Intelligence (AI)**

**Machine Learning (ML)**

**Deep Learning (DL)**

**Data Science**

## □ 머신러닝의 분류

지도 학습  
(Supervised Learning)

- 분류
- 회귀
- 추천 시스템
- 시각/음성 감지/인지
- 텍스트 분석, NLP

비지도 학습  
(Un-supervised Learning)

- 클러스터링
- 차원 축소
- 강화 학습

## ❑ Anaconda

Pricing

# Individuals

**FREE**

INCLUDES:

- Basic workspace for the student or practitioner
- 5GB for cloud-hosted notebooks
- Thousands of curated packages
- On-demand training courses
- Publish a single data app
- Unlimited notebook sharing

**Start Coding Now >**

### a. FREE PLANS.


Anaconda allows You to use the Free Offerings (as defined hereinafter), Trial Offerings (as defined hereinafter), Pre-Release Offerings (as defined hereinafter), and Scholarships (as defined hereinafter) (each, a "Free Plan"), without charge, as set forth in this Section 1.1(a). Your use of the Free Plan is restricted to Internal Business Purposes. If You receive a Free Plan to the Cloud Offerings, Anaconda grants You a non-transferable, non-exclusive, revocable, limited license to use and access the Cloud Offerings in strict accordance with these TOS. We reserve the right, in our absolute discretion, to withdraw or to modify your Free Plan access to the Cloud Offerings at any time without prior notice and with no liability.

i. FREE OFFERINGS. Anaconda maintains certain Cloud Offerings, including Anaconda Open Source that are generally made available to Community Users free of charge (the "Free Offerings") for their Internal Business Use. The Free Offerings are made available to You, and Community Users, at the **Free Subscription** level strictly for internal: (i) Personal Use, (ii) Educational Use, (iii) Open-Source Use, and/or (iv) Small Business Use.

(a) Your use of Anaconda Open Source is governed by the Anaconda Open-Source Terms, which are incorporated herein by reference.

(b) You may not use Free Offerings for commercial purposes, including but not limited to external business use, use in organizations over two hundred (200) employees (unless such use is for an Educational Purpose), third-party access to the Cloud Offerings, or Content redistribution or mirroring (each, a "Commercial Purpose"). Using the Free Offerings for a Commercial Purpose requires a Paid Plan with Anaconda.


<https://legal.anaconda.com/policies/en/?name=terms-of-service>



**pwang99** • 3 yr. ago

### Anaconda CEO - Peter Wang

Miniconda + conda forge is not affected by our commercial terms of service change. Please see the FAQ here: <https://www.anaconda.com/blog/anaconda-commercial-edition-faq>

 1   Share ...

<https://www.reddit.com/r/Python/comments/iqsk3y/comment/g4xuabr/>

## ❑ Miniconda + conda-forge

- <https://docs.conda.io/projects/miniconda/en/latest/>
- <https://conda-forge.org/>
- <https://www.whatwant.com/entry/miniconda>

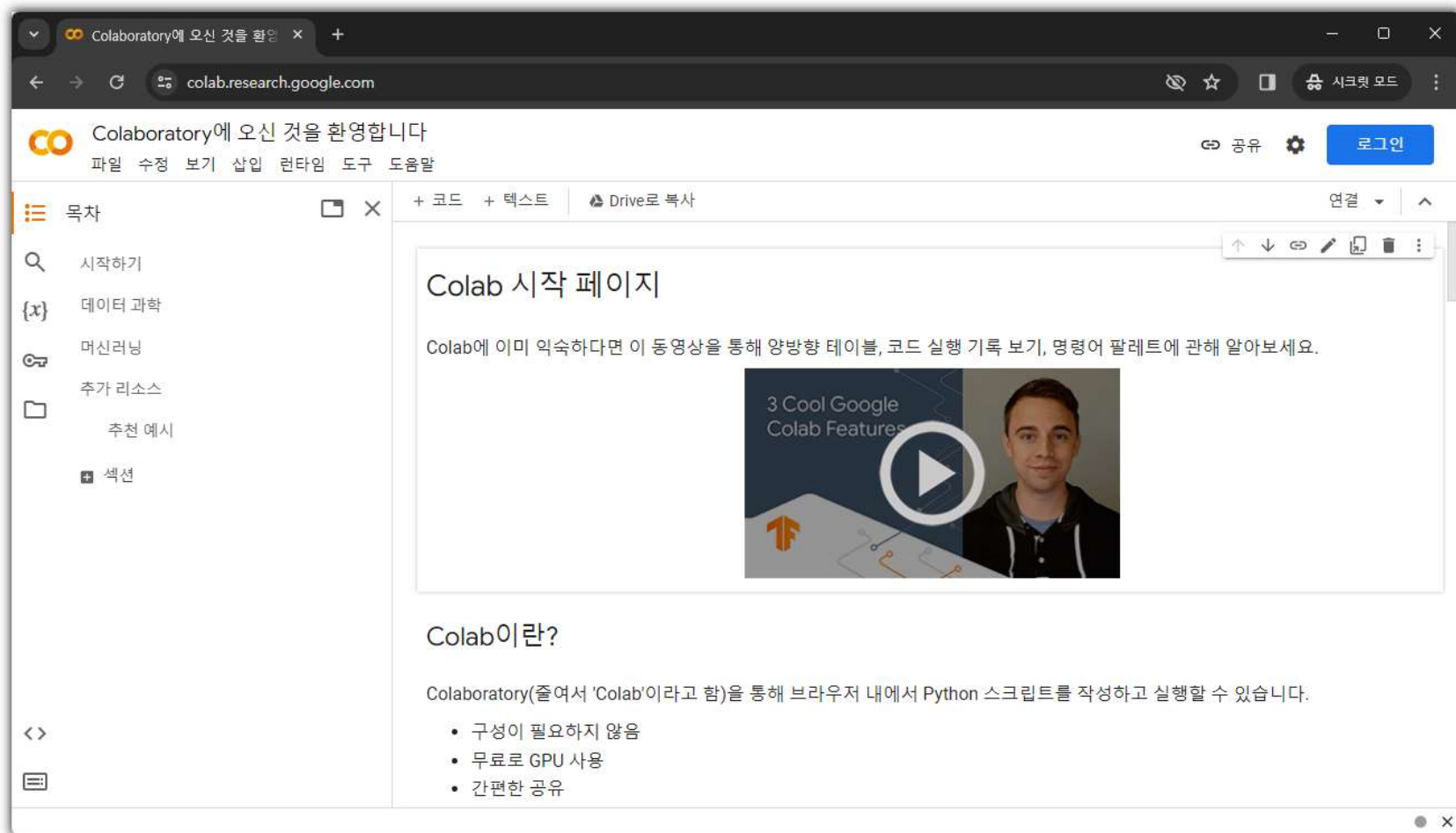
Latest - Conda 23.11.0 Python 3.11.5 released December 20, 2023

Platform	Name	SHA256 hash
Windows	Miniconda3 Windows 64-bit	c9b32faa9262828702334b16bcb5b53556e630d54e5127f5c36c7ba7ed43179a
macOS	Miniconda3 macOS Intel x86 64-bit bash	2b7f9e46308c28c26dd83abad3e72121ef63916eaf17b63723b5a1f728dc3032
	Miniconda3 macOS Intel x86 64-bit pkg	74ab9e8c3e9b3c2fc7c44d710ed9bad19085d951d819c1284a46eeb0bdf2578
	Miniconda3 macOS Apple M1 64-bit bash	5694c382e6056d62ed874f22692224c4f53bca22e8135b6f069111e081be07aa
	Miniconda3 macOS Apple M1 64-bit pkg	912c0b58e800f26e08d515526a8d3455755e83963b40e78597176540ea2401ca
Linux	Miniconda3 Linux 64-bit	c9ae82568e9665b1105117b4b1e499607d2a920f0aea6f94410e417a0eff1b9c
	Miniconda3 Linux-aarch64 64-bit	dec447fb99dbd0fc5004481ec9bf8c04f9ba28b35a9292afe49ecfe400237f
	Miniconda3 Linux-s390x 64-bit	53a9e9eb97cd6e318f4f184add069436e1a46124cf864bf2d7bd67043e50e471

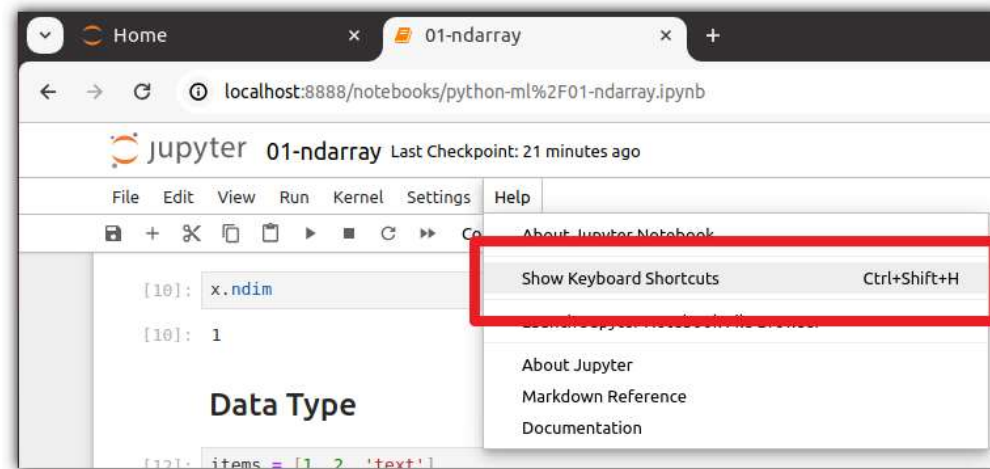
```
conda config --add channels conda-forge
conda config --set channel_priority strict
conda install <package-name>
```

## □ Colab

- <https://colab.research.google.com/>



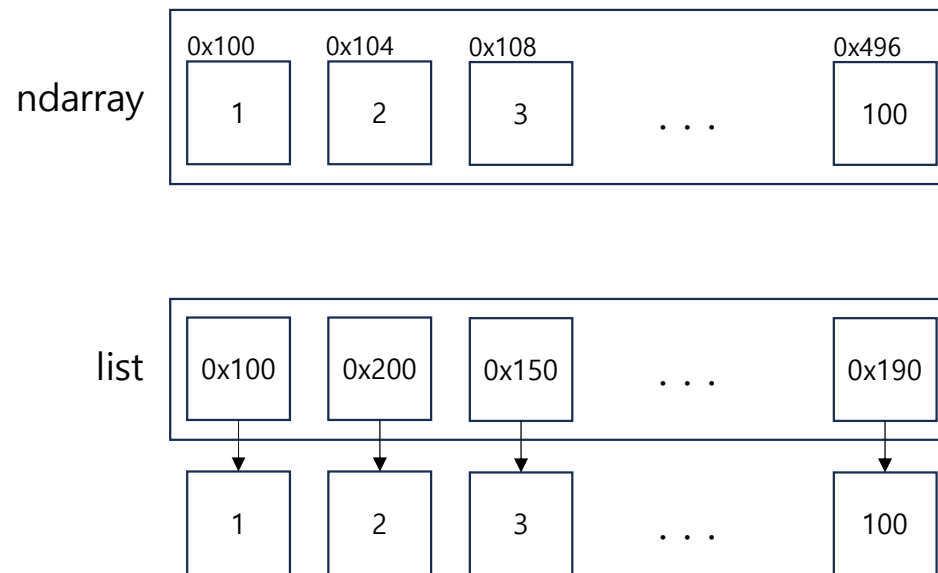
## □ Jupyter Notebook



## □ Numpy 모듈

- Why Numpy's ndarray ?

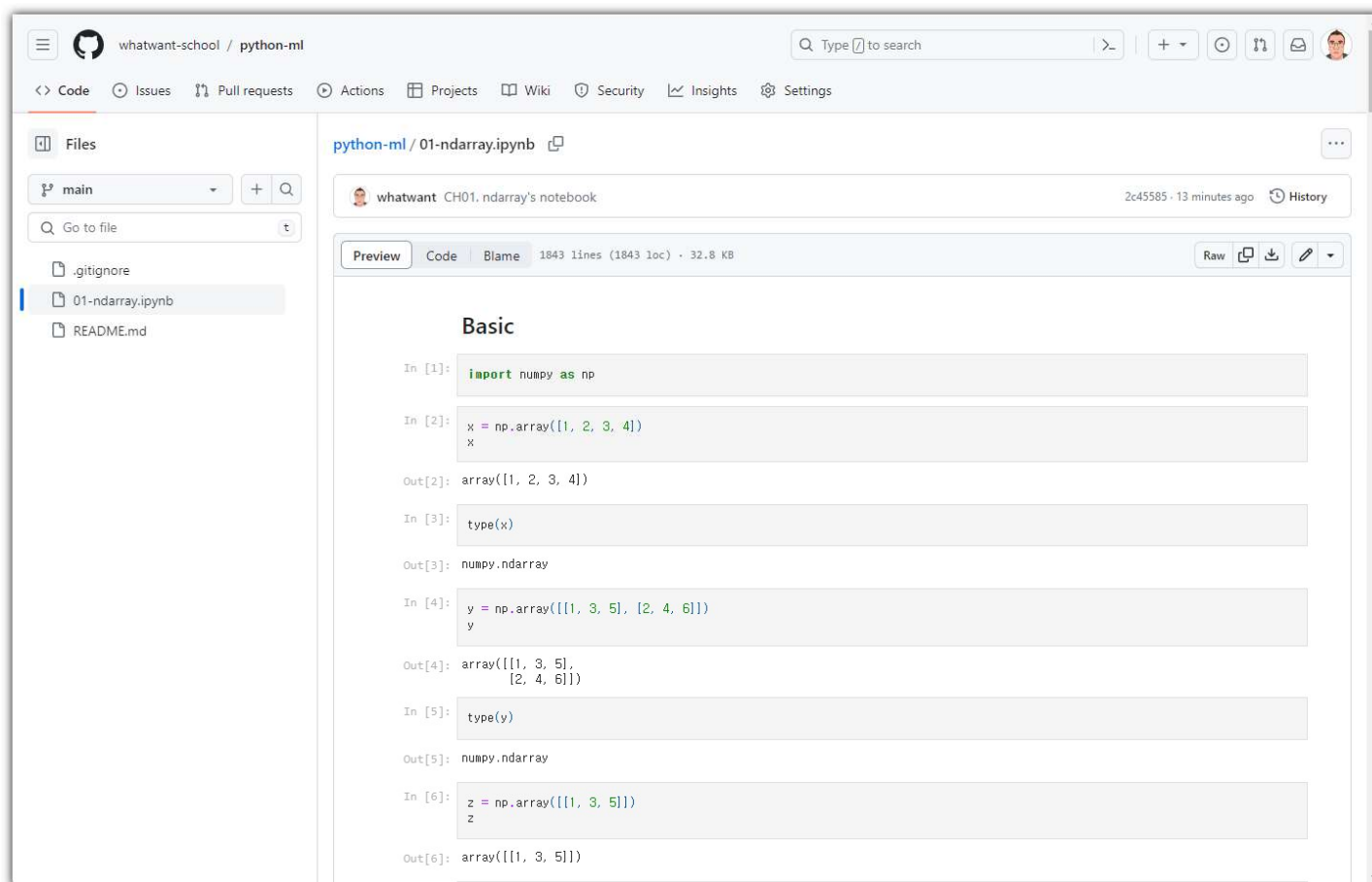
- . Python의 list보다 빠름 ← 데이터의 보관이 아닌 처리 목적
- . Python의 list보다 적은 메모리 사용
- . 선형대수 및 통계 관련 여러 함수를 내장(빌트인 함수)





## □ Practice

- <https://github.com/whatwant-school/python-ml/blob/main/01-ndarray.ipynb>



The screenshot displays the GitHub interface for the repository 'whatwant-school / python-ml'. The left sidebar shows the file structure with '01-ndarray.ipynb' selected. The main content area shows the notebook's preview, titled 'python-ml / 01-ndarray.ipynb'. The notebook content includes the following code cells:

```
In [1]: import numpy as np

In [2]: x = np.array([1, 2, 3, 4])
x
Out[2]: array([1, 2, 3, 4])

In [3]: type(x)
Out[3]: numpy.ndarray

In [4]: y = np.array([[1, 3, 5], [2, 4, 6]])
y
Out[4]: array([[1, 3, 5],
               [2, 4, 6]])

In [5]: type(y)
Out[5]: numpy.ndarray

In [6]: z = np.array([[1, 3, 5]])
z
Out[6]: array([[1, 3, 5]])
```

Next Week ~ ?

## 01

파이썬 기반의 머신러닝과  
생태계 이해

01. 머신러닝의 개념	1
머신러닝의 분류	2
데이터 전쟁	3
파이썬과 R 기반의 머신러닝 비교	4
02. 파이썬 머신러닝 생태계를 구성하는 주요 패키지	6
파이썬 머신러닝을 위한 S/W 설치	7
03. 넘파이	13
넘파이 ndarray 개요	14
ndarray의 데이터 타입	16
ndarray를 편리하게 생성하기 - <code>arange</code> , <code>zeros</code> , <code>ones</code>	18
ndarray의 차원과 크기를 변경하는 <code>reshape()</code>	20
넘파이의 ndarray의 데이터 세트 선택하기 - 인덱싱(indexing)	23
행렬의 정렬 - <code>sort()</code> 와 <code>argsort()</code>	33
선형대수 연산 - 행렬 내적과 전치 행렬 구하기	37
04. 데이터 핸들링 - 판다스	39
판다스 시작 - 파일을 DataFrame으로 로딩, 기본 API	40
DataFrame과 리스트, 딕셔너리, 넘파이 ndarray 상호 변환	49
DataFrame의 칼럼 데이터 세트 생성과 수정	53
DataFrame 데이터 삭제	55
Index 객체	58
데이터 셀렉션 및 필터링	62
정렬, Aggregation 함수, GroupBy 적용	73
결손 데이터 처리하기	77
apply lambda 식으로 데이터 가공	80
05. 정리	84

## 02

사이킷런으로 시작하는  
머신러닝

01. 사이킷런 소개와 특징	85
02. 첫 번째 머신러닝 만들어 보기 - 붓꽃 품종 예측하기	86
03. 사이킷런의 기본 프레임워크 익히기	91
Estimator 이해 및 <code>fit()</code> , <code>predict()</code> 메서드	91
사이킷런의 주요 모듈	92
내장된 예제 데이터 세트	94
04. Model Selection 모듈 소개	98
학습/테스트 데이터 세트 분리 - <code>train_test_split()</code>	98
교차 검증	100
GridSearchCV - 교차 검증과 최적 하이퍼 파라미터 튜닝을 한 번에	111
05. 데이터 전처리	116
데이터 인코딩	116
피쳐 스케일링과 정규화	122
StandardScaler	123
MinMaxScaler	125
학습 데이터와 테스트 데이터의 스케일링 변환 시 유의점	126
06. 사이킷런으로 수행하는 타이타닉 생존자 예측	129
07. 정리	143
01. 정확도(Accuracy)	146
02. 오차 행렬	150
03. 정밀도와 재현율	154
정밀도/재현율 트레이드오프	157
정밀도와 재현율의 명점	165

## 03

## 평가

Who ~ ?

See you Next Weekend ~ ?