

파머완 1장

■ 상태	완료
■ 담당자	 시현 이
■ 마감일	@09/07/2025
▣ 작업 유형	개념정리 및 필사
■ 설명	파이썬 머신러닝 완벽 가이드_개정2판_제1장 개념 정리+ 필사 링크
■ 업데이트 시간	@September 7, 2025 7:18 PM

01. 머신러닝의 개념

머신러닝(ML)이란?

:애플리케이션 수정 없이 데이터를 기반으로 패턴을 학습하고 결과를 예측하는 알고리즘 기법
데이터를 기반으로 숨겨진 패턴을 인지해 해결. 예측 및 의사 결정 도출

머신러닝의 분류

:지도학습(분류, 회귀, 추천 시스템, 시각/음성 감지/인지, 텍스트 분석, NLP)

:비지도학습(clustering, 차원 축소)

:강화학습

데이터 VS ML 알고리즘

데이터 전쟁-머신러닝은 데이터에 의존적인 성향을 지님(Garbage In → Garbage Out)
양질의 데이터가 필수적인 시대.

python VS R

대표적인 머신러닝 오픈소스 프로그램 언어: python, R

R: 통계 전용 프로그램(SPSS, SAS, MATLAB)

python: 객체지향, 함수형 프로그래밍 모두 포괄, 다양한 라이브러리

-딥러닝 프레임워크(TensorFlow, Keras, PyTorch 등)

02. Python ML 주요 패키지

- ML 패키지: Scikit-Learn(사이킷런)-데이터 마이닝 기반의 머신러닝
- 행렬/선대/통계 패키지: Numpy(넘파이)-행렬기반 데이터처리에 특화/SciPy(사이파이)
- 데이터 핸들링: Pandas(판다스)-2차원 데이터 처리에 특화
- 시각화: Matplotlib(matplotlib)/Seaborn(시본)
- IPython(아이파이썬)-대화형 파이썬 터

03. Numpy

선형대수 기반의 프로그램을 만드는 패키지

루프 사용x → 대량 데이터의 배열 연산 가능케함

C/C++과 같은 저수준 언어 기반의 호환 API 제공

[단점]

일반적으로 데이터는 2차원 형태 → numpy는 이러한 데이터 처리에 특화X

ndarray의 데이터 타입

: int, unsigned int, float, complex 타입 제공

: ndarray 내의 데이터 타입은 그 연산의 같은 데이터 타입만 가능→한개의 ndarray 객체에 int와 float가 함께 있을 수 X

: dtype 속성으로 ndarray 내의 데이터 타입 확인 할 수 있음.

04. Data handling - Pandas

: 2차원 데이터를 고수준으로 가공/처리할 수 있는 API를 제공하는 라이브러리

python의 리스트, 컬렉션, numpy 등의 내부데이터 + CSV 등의 파일을 쉽게 DataFrame으로 변경

DataFrame:

2차원 데이터를 담는 구조체(칼럼이 여러개)

→ index: 주식별자

→ Series: 칼럼이 하나뿐인 데이터 구조체

Pandas의 API

1. `read_csv()` : CSV(칼럼을 ','로 구분한 파일 포맷) 파일 포맷을 위한 API
2. `read_table()`: '\t'으로 필드를 구분
3. `read_fwf()`

Google Colab 필사

☞ [파머완_1장_코드필사.ipynb](https://colab.research.google.com/drive/1KZMdUtUODY8pZ6dF-cPE8x4asY62f9nm)