

[01.머신러닝의 개념]

-데이터를 기반으로 패턴을 학습하고 결과를 예측하는 알고리즘 기법을 통칭

- 머신러닝 알고리즘은 데이터를 기반으로 통계적인 신뢰도를 강화하고 예측 오류를 최소화하기 위한 다양한 수학적 기법을 적용해 데이터 내의 패턴을 스스로 인지하고 신뢰도있는 예측 결과를 도출
- 머신러닝은 데이터를 관통하는 패턴을 학습하고 이에 기반한 예측을 수행하면서 데이터 분석 영역에 새로운 혁신을 가져옴

-머신러닝의 분류

- 지도학습
 - 분류
 - 회귀
 - 추천 시스템
 - 시각, 음성 감지, 인지
 - 텍스트 분석, nlp
- 비지도 학습
 - 클러스터링
 - 차원 축소
 - 강화학습
- 강화학습

-데이터 전쟁

- 데이터&머신러닝 알고리즘
- 머신러닝의 단점
 - 데이터에 매우 의존적
 - 가비지 인, 가비지 아웃

-파이썬과 r기반의 머신러닝 비교

- r은 통계 전용 프로그램임
- 파이썬이 r에 비해 뛰어난 점
 - 쉽고 뛰어난 개발 생산성
 - 오픈 소스 계열의 전폭적인 지원-> 많은 라이브러리
 - 등등

[02.파이썬 머신러닝 생태계를 구성하는 주요 패키지]

-파이썬 기반의 머신러닝을 익히기 위해 필요한 패키지

- 머신러닝 패키지
 - Scikit-Learn
- 행렬,선형대수, 통계 패키지
 - 대표적인 행렬과 선형대수 다루는 패키지: NumPy, SciPy
- 데이터 핸들링
 - 판다스: 2차원 데이터 처리에 특화되어있으며
- 시각화
 - 맷플롯립
 - 시본

[03.넘파이]

-넘파이의 의미: **Numerical Python**을 의미 / 파이썬에서 선형대수 기반의 프로그램을 쉽게 만들 수 있도록 지원하는 대표적인 패키지

- 루프를 사용하지 않고 대량 데이터의 배열 연산을 가능하게 함
- 넘파이는 매우 빠른 배열 연산을 보장해주지만 파이썬 언어 자체가 가지는 수행 성능의 제약이 있으므로 수행 성능이 매우 중요한 부분은 **c/c++** 기반의 코드로 작성하고 이를 넘파이에서 호출하는 방식으로 쉽게 통합 가능
- 배열 기반이니 연산은 물론이고 다양한 데이터 핸들링 기능을 제공
 - 넘파이는 파이썬의 대표적인 데이터 처리 패키지인 판다스의 편리성에는 미치지 못한다

-넘파이 ndarray 개요

- **import numpy as np**
 - 넘파이의 기반 데이터 타입은 **ndarray**
 - **ndarray**를 이용해서 넘파이에서 다차원 배열을 쉽게 생성하고 다양한 연산을 수행
- **array()** 함수
 - 파이썬의 리스트와 같은 다양한 인자를 입력 받아서 **ndarray**로 변환하는 기능 수행
 - 생성된 **ndarray** 배열의 **shape** 변수는 **ndarray**의 크기, 즉 행과 열의 수를 튜플 형태로 가지고 있으며 이를 통해서 배열의 차원까지 알 수 있음
 - **np.array**(여기에 변환을 원하는 객체)
 - 배열 이름.**shape**
 - 배열 이름.**ndim**

-ndarray의 데이터 타입

- **ndarray**내의 데이터값은 숫자, 문자열, 불 값등이 모두 가능
- 숫자형은
 - **int**
 - **unsigned int**
 - **float**
 - **complex**
- **ndarray** 내의 데이터 타입은 같은 데이터 타입만 가능
 - 한개의 **ndarray** 객체에 **int** 와 **float**가 함께 있을 수 없다
- **ndarray**내의 데이터 타입은 **dtype** 속성으로 확인 가능
- 만약 다른 데이터 유형이 섞여 있는 리스트를 **ndarray**로 변경하면 데이터 크기가 더 큰 데이터 타입으로 형 변환을 일괄 적용
 - **int**와 문자열 섞임-> 문자열로
 - **int**와 **float** 섞임-> **float**로
- 데이터 타입 변경 **astype()**
 - **dlswkfh dnjsgksms xkdlqdmf answkduffh wlwjdgkaus ehla**

-ndarray를 편리하게 생성하기 - **arange, zeros, ones**

- 특정 크기와 차원을 가진 **ndarray**를 연속값이나 **0**또는 **1**로 초기화해 쉽게 생성해야할 필요가 있는 경우에 사용
- 테스트용으로 데이터를 만들거나 대규모의 데이터를 일괄적으로 초기화해야할 경우에 사용됨
 - **arange()**
 - **array**를 **range()**로 표현하는 것
 - **0**부터 함수 인자값 **-1**까지의 값을 순차적으로 **ndarray**의 데이터값으로 변환해줌

■