

머신러닝(Machine Learning)

목 차

1-1 machine learning(기계학습)이란?

1-2 머신러닝으로 무엇이 가능한가?

1-3 용어 이해하기 - 모델, 모델링, 샘플, 데이터 포인터

1-4 머신러닝의 구분

1-5 머신러닝과 딥러닝

01 machine learning(머신러닝)이란?



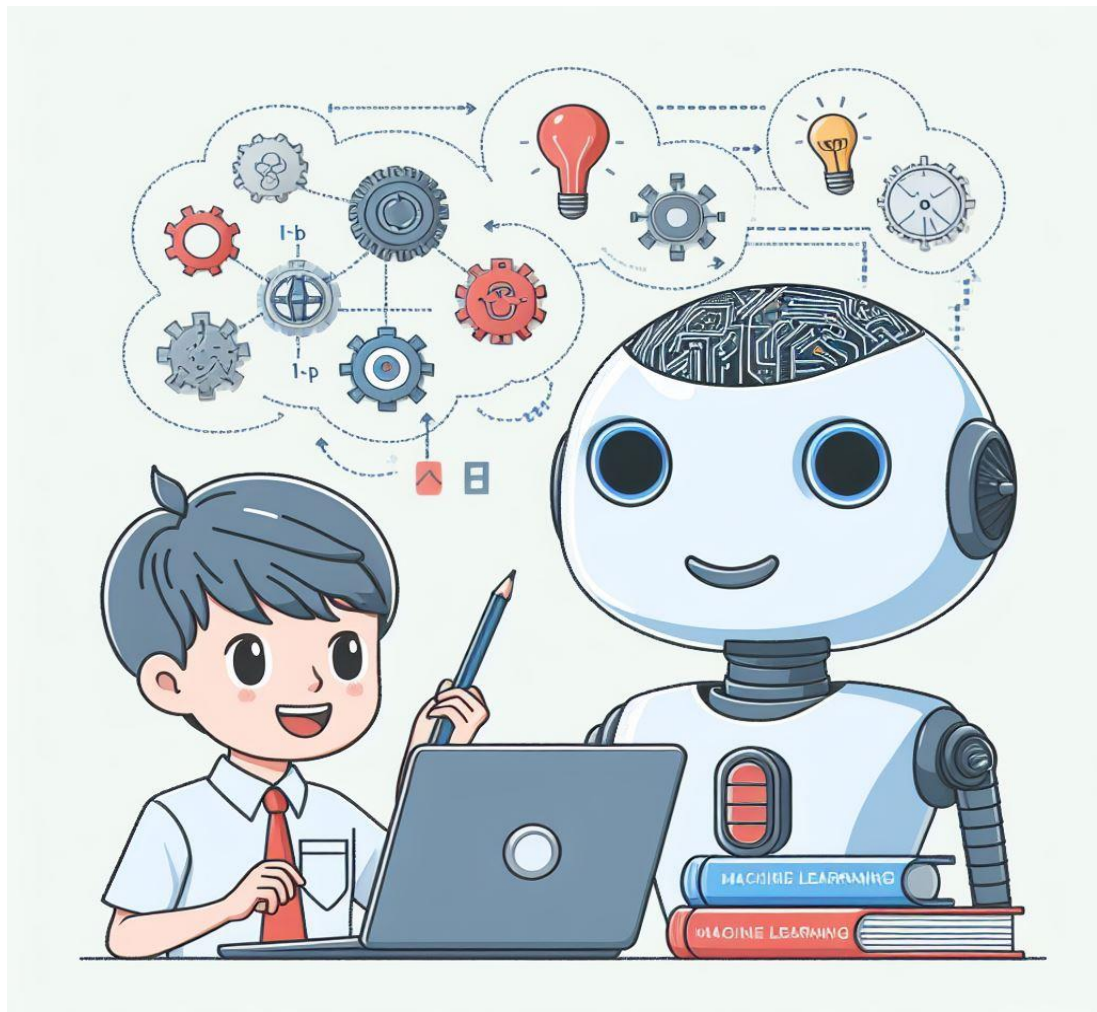
▶ 머신러닝은 실험적인 데이터를 기반으로 프로그램의 실행 동작을 개선시키는 알고리즘의 설계와 개발을 고려하는 분야이다.

(위키 피디아)

▶ 머신러닝은 컴퓨터가 데이터를 이용해 향후 적용 가능한 규칙을 생성해 낸다. 인공지능의 한 계열로 볼 수 있다.

(dictionary.com)

01 machine learning(머신러닝)이란?



- ▶ 머신러닝은 데이터에서 지식을 추출하는 작업
- ▶ 머신러닝은 통계학, 인공지능, 컴퓨터 과학이 얹혀 있는 분야이다.
- ▶ 머신러닝은 인공지능의 한 분야로, 컴퓨터가 명시적인 프로그래밍 없이 데이터를 통해 스스로 학습하고 개선하는 기술.

02 머신러닝으로 무엇이 가능한가?

(1) 편지 봉투에 손으로 쓴 우편번호 숫자 판별

(2) 의료 영상 이미지에 기반한 종양 판단

(3) 의심되는 신용카드 거래 감지

=> 신용카드 거래 내역이 입력이 되고 부정 거래인지가 출력이 된다.

(4) 블로그 글의 주제 구분

=> 많은 양의 텍스트 데이터를 요약하고 그 안에 담긴 핵심 주제를 찾기.

02 머신러닝으로 무엇이 가능한가?

(5) 고객들을 취향이 비슷한 그룹으로 묶기

=> 어떤 고객들의 취향이 비슷한지 비슷한 취향의 고객을 그룹으로 묶고 싶을 때,

(6) 비정상적인 웹 사이트 접근 탐지

=> 정상 패턴과 비정상 패턴을 찾아본다.

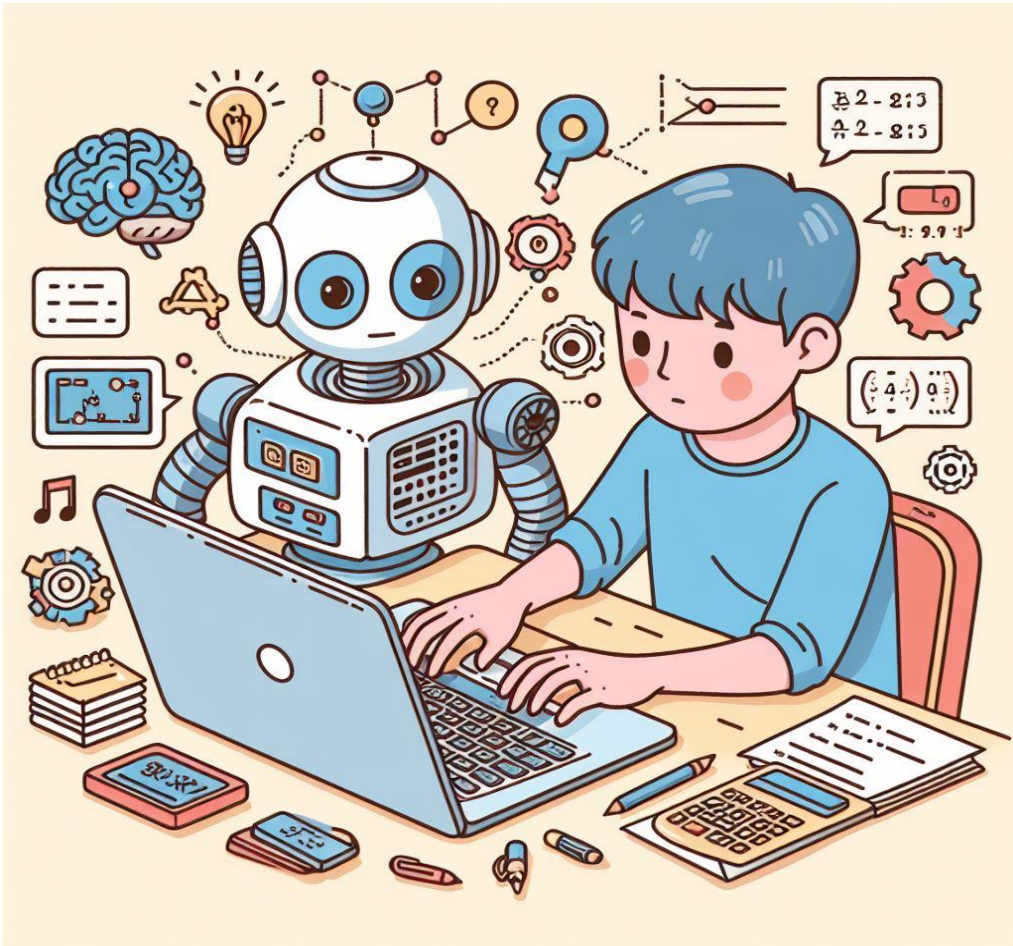
(7) 영화 추천에서 음식 주문, 쇼핑, 맞춤형 온라인 라디오 방송 등

(8) 객체 인식 - 얼굴, 눈 등 인식

=> 스마트폰 얼굴 인식. 픽셀 데이터를 이용한 학습을 통해 사람의 얼굴 확인 가능.

02 머신러닝으로 무엇이 가능한가?

▶ 군사 훈련 및 교육



(1) 개인 맞춤형 교육

개인의 학습 능력과 특성에 맞춘 교육 프로그램을 제공

(2) 훈련 시뮬레이션

실제 전투 환경과 유사한 훈련 시뮬레이션 개발

(3) 전문가 시스템

전문가의 지식을 기반으로 의사 결정하는 시스템 개발

(4) 군사 교육 자료 개발

머신러닝 기술을 이용하여 군사 교육 자료 개발

03 용어 이해하기

▶ 모델, 모형, 모델링(Model, Modeling)

(가) 정보 시스템 모델링

가. 데이터 모델링 : 현실세계의 복잡한 데이터들을 컴퓨터 정보 구조로 변환시키는 과정

(나) 수학적 모델링

가. 시스템의 변화를 나타내는(예측하는) 수학적 모델이 방정식으로 표현된다.

(정보통신기술용어해설 참조)

03 용어 이해하기

▶ 샘플(sample), 데이터 포인트(data point), 특성(feature)

(가) 샘플(sample) 또는 데이터 포인트

하나의 개체 또는 행을 샘플이라고 말한다.

(나) 특징(feature or variable)

샘플의 속성, 즉 열을 말한다.

(다) 특성 추출(feature extraction) or feature engineering

좋은 입력 데이터를 만들어 내는 것.

03 용어 이해하기

▶ 특징, 피처(features)

(가) 다음과 같이 다른 용어로 사용되기도 한다.

예측 인자(Predictors), 독립변수(independent variables), 인풋(Input), 특징 및 특성

03 용어 이해하기

▶ 클래스(class)와 레이블

(가) 레이블(label)

데이터 포인트가 속하는 클래스를 나타내는 값.

(예) 고양이 사진의 레이블은 고양이, 강아지 사진의 레이블은 강아지

(나) 클래스(class)

- 머신러닝 모델이 학습하고자 하는 대상의 범주. 레이블의 범주를 클래스라 한다.
- 분류 문제에서 가능한 카테고리. 고양이 강아지 사진 분류 문제에서는 클래스는 고양이와 강아지

03 용어 이해하기

▶ 클래스(class)와 레이블의 예시

(가) 문제

이메일이 스팸인지 아닌지를 분류하는 문제

(나) 클래스(class)

"스팸", "정상 "

(다) 레이블(Label)

스팸 이메일 : "스팸"

정상 이메일 : "정상 "

04 머신러닝의 구분

▶ 지도학습(supervised learning)

(가) 예측하고자 하는 **목표(Target)가 존재**한다.

학습하고자 하는 데이터의 정답이 있다.

우리는 이 정답을 레이블(label)이라 한다. 교사의 역할이 존재.

▶ 비지도학습(unsupervised learning) or 자율학습

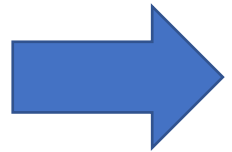
(나) **목표(Target)가 존재**하지 않는다. 교사 역할이 없음.

04 머신러닝의 구분

▶ 지도학습(supervised learning) 구분

(가) Regression – (회귀) – 레이블이 수치형 변수

(나) Classification – (분류) – 레이블이 범주형 변수



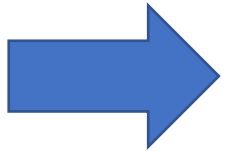
공통점 : 입력 및 특성(feature) 값을 이용하여
주어진 입력변수에 대한 **타겟(target, 목표변수)**의 값을 예측하는
모델을 구축한다.

04 머신러닝의 구분

▶ 지도학습(supervised learning)

(가) Regression – (회귀) – 레이블이 수치형 변수

(나) Classification – (분류) – 레이블이 범주형 변수



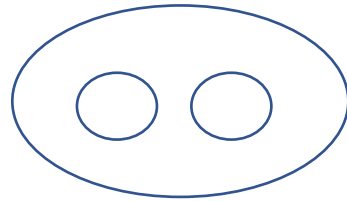
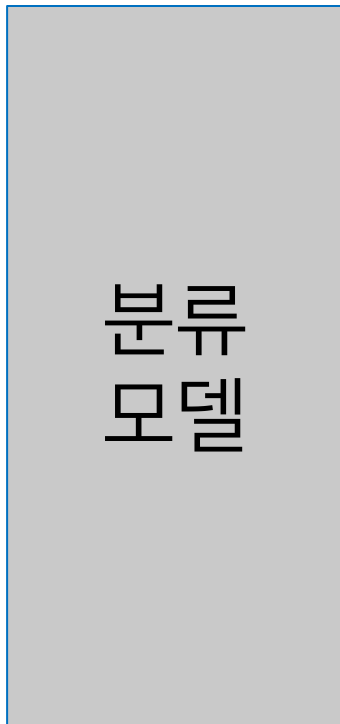
차이점 :

A. 목표 변수의 형태가 회귀의 경우 **연속형**이다.

B. 분류의 경우는 **범주형**이다.(고정되어 있음)

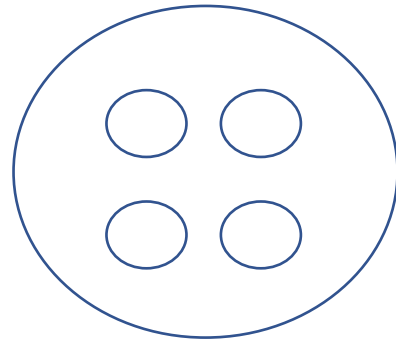
04 머신러닝의 구분

▶ Classification(분류)의 구분 - 이항분류와 다항분류



이항분류

목표값이 2개



다항(or 다중)분류

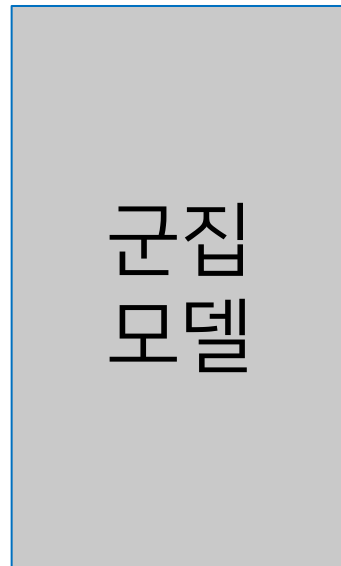
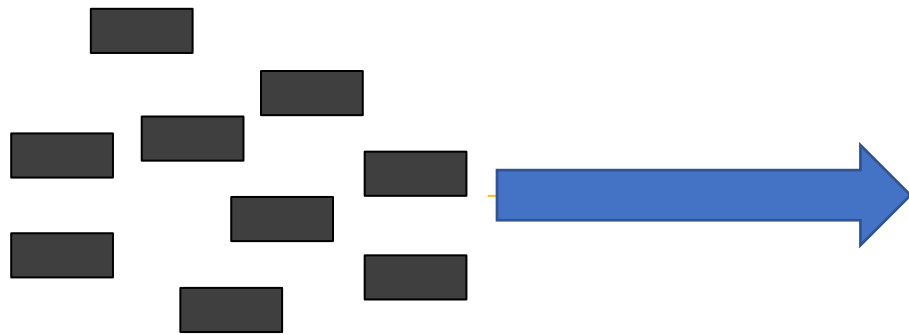
목표값이 3개 이상

04 머신러닝의 구분 - 비지도학습(unsupervised learning)

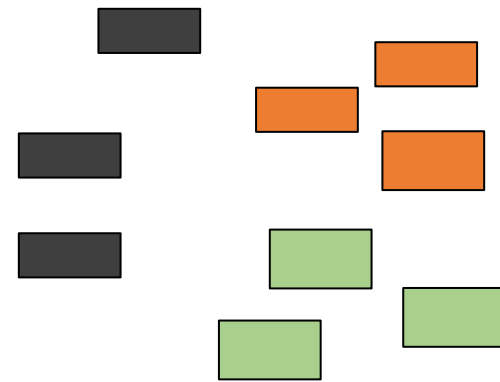
군집은 레이블이 없다.

레이블(목표 변수)가 없다.

군집은 레이블 없이 확보된
데이터의 특성을 분석



서로 유사한 특성을 가진
데이터끼리 그룹화



05 머신러닝(Machine Learning)과 딥러닝(Deep Learning)

Machine Learning



Deep Learning

