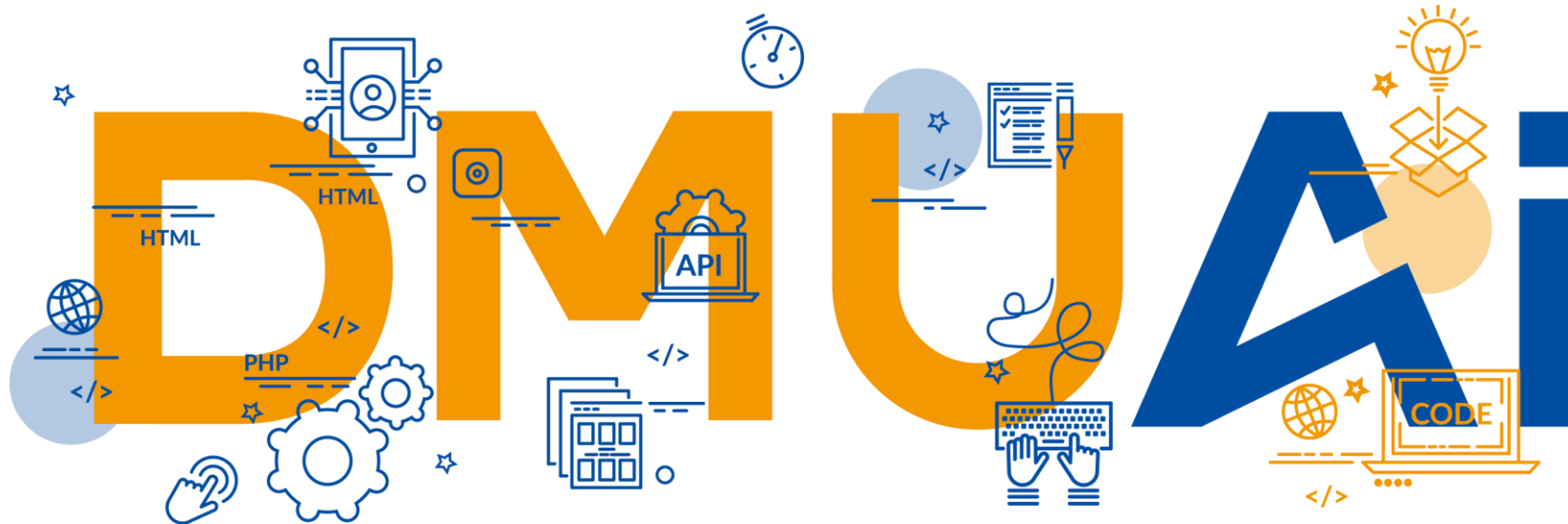


동양미래대학교 인공지능소프트웨어학과

Dongyang Mirae University
Dept. Of Artificial Intelligence

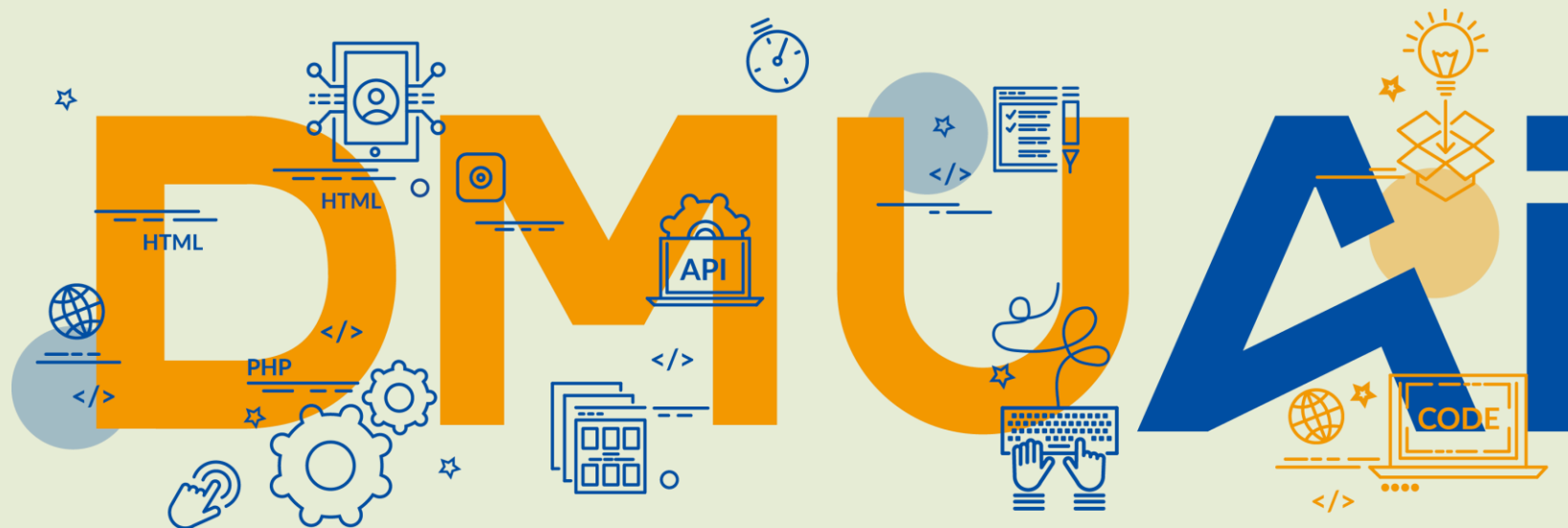
4차산업혁명 시대의 인공지능
“누구나 이해할 수 있는 인공지능”



동양미래대학교 인공지능소프트웨어학과

인공지능과 머신러닝

Dongyang Mirae University
Dept. Of Artificial Intelligence



인공지능과 머신러닝, 딥러닝

인공지능이라는 용어가 나올 때마다 항상 함께 나오는 머신러닝과 딥러닝은 무엇이고 어떻게 다를까?

- 인공지능이 가장 큰 범주

- 인간의 지능을 구현

- 머신러닝

- 데이터를 기반으로 기계 스스로 학습하는 인공지능의 한 분야

- 심층신경망인 딥러닝(deep learning)

- 머신러닝의 여러 분야 중에서
 - 2010년 이후 현재의 인공지능 붐을 주도하고 있는 기술
 - 퍼셉트론으로 구성된 인공신경망
 - 여러 단계의 심층 학습을 통하여 스스로 학습하는 기술

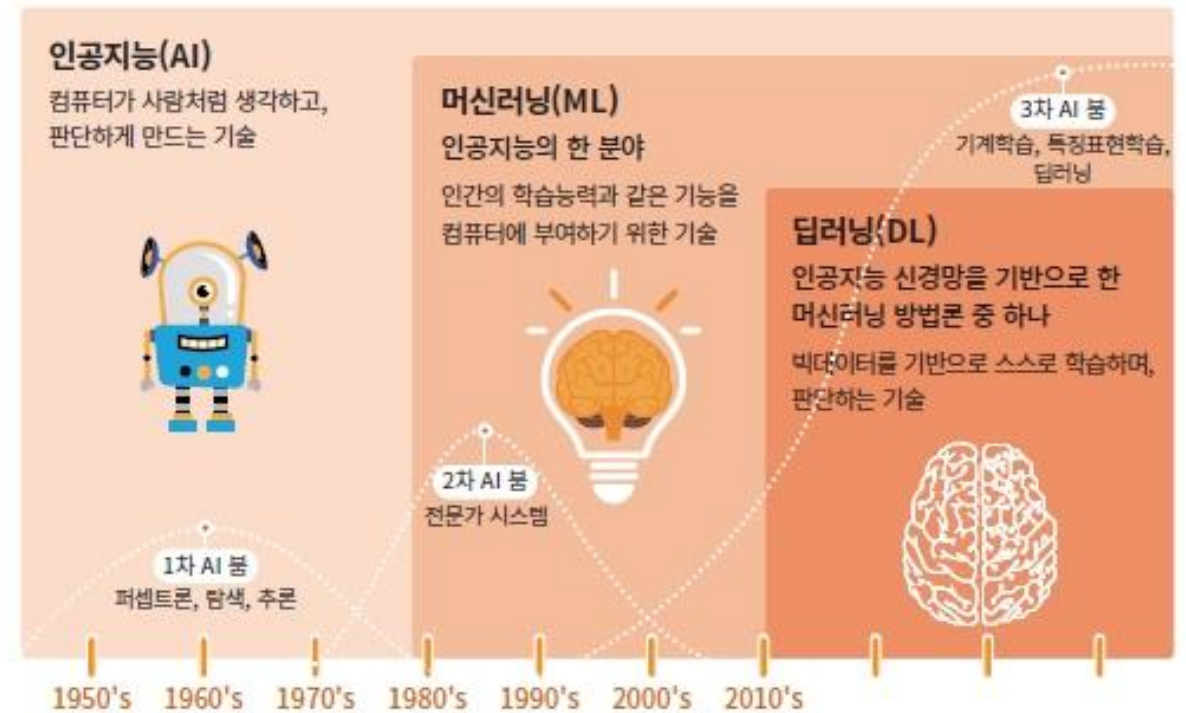


그림 6.23 ▶ 인공지능과 머신러닝, 딥러닝의 이해

머신러닝의 정의와 이해

1959년 아서 사무엘(Arthur Samuel)이 머신러닝(machine learning)이라는 용어를 처음 사용해 대중화시킴

• 사무엘의 머신러닝 정의

- "머신러닝은 컴퓨터가 인간처럼 학습하고 행동하도록 하는 과학이며,
- 관찰 및 실제 단어의 상호작용 형태로 데이터와 정보를 제공함으로써
- 시간이 지남에 따라 자율적으로 학습을 향상시키는 과학이다."

• 일반적 기계학습 정의

- 주어진 데이터를 기반으로 기계가 스스로 학습
 - 성능을 향상시키거나 최적의 해답을 찾기 위한 지능적 학습 방법
- 컴퓨터가 스스로 학습을 할 수 있도록 해주는 인공지능의 한 형태
 - 명시적(explicit)으로 프로그래밍을 하지 않음
- 데이터는 매우 중요
 - 더 많은 데이터가 유입되면 컴퓨터는 학습을 더 많이 수행
 - 시간이 흐르면서 스마트해져서 작업을 수행하는 능력과 정확도가 향상

개와 고양이 분류

사람에겐 너무 쉬운 일이나, 컴퓨터에게 개와 고양이 이의 분류를 맡긴다면 어떨까?

- 전통적 프로그래밍 방식으로 해결

- 개와 고양이 생김새의 특징을 세부적 코딩
- 외모 특징 자체를 찾는 것도 어려운 문제
 - 이러한 전통적 방식의 코딩 결과는 좋지 않음
 - 명시적으로 프로그래밍(explicit programming)
 - 입력 데이터를 사용하고 프로그램을 실행하여 출력을 생성



그림 6.26 ▶ 전통적 프로그래밍과 머신러닝의 차이

머신러닝 방법

사람에겐 너무 쉬운 일이나, 컴퓨터에게 개와 고양이 이의 분류를 맡긴다면 어떨까?

- **인간이 어린 시절 부모 도움의 학습으로 개와 고양이를 자연스럽게 분류하는 방법**
 - 개와 고양이 사진에서 반복되는 패턴을 인지해 스스로 학습하는 방식
- **스스로 데이터를 반복적으로 학습하여 기술을 터득하는 방식**
 - 입력 데이터와 정답이 있는 출력이 알고리즘에 공급되어 프로그램(모델)을 생성
 - 알고리즘이 데이터에서 학습을 해 자동으로 규칙을 공식화
 - 즉 머신러닝 모델을 통해 미래의 결과를 예측



그림 6.26 ▶ 전통적 프로그래밍과 머신러닝의 차이

머신러닝의 수행 과정 1/2

• 데이터 수집

- 수집된 데이터의 품질과 양에 따라
 - 예측 모델의 성능이 결정
- 머신러닝 과정에서 가장 중요한 단계

• 데이터 전처리

- 잘못된 값은 수정하는 데이터 정리 필요
 - 누락 값을 채우고
 - 이상 값을 수정하거나 제거
- 정규화(normalization) 과정도 필요
 - 데이터 특성마다 차이가 너무 크면
 - 머신러닝 모델에 따라 값을 비슷한 크기로 조정
- 데이터 변환 작업이 필요
 - 모델의 계산에 적합한 자료구조로
 - 기초 자료로 새로운 자료를 생성

• 모델 학습

- 적절한 머신러닝 모델을 생성해 데이터로 학습

• 성능 개선

- 학습된 모델의 성능을 시험하고 보다 좋은 결과를 도출

• 시각화

- 데이터와 머신러닝 과정 그리고 예측 결과를 보기 좋게 시각화하는 과정

머신러닝의 수행 과정 2/2



그림 6.27 ▶ 머신러닝 수행 과정

데이터와 특징

머신러닝은 데이터에 숨겨진 정보를 찾는 분야로 데이터가 무엇보다 중요

- 데이터 집합(data set)

- 머신러닝의 데이터

- 중고 자동차 데이터 예

- 엑셀의 테이블 형태 자료

- 9개의 열

- 제조사와 모델, 색상, 사용
기간, 배기량, 주행거리,
연료, 신차가격, 중고가격

- 표본(sample)

- 인스턴스(instance)

- 데이터 포인터(data
pointer)

- 실제 중고 자동차 개
개의 자료인 행

- 표본 수(# of samples)

- 데이터 수인 행 수

특징 X_1, X_2, X_3, \dots									정답(label) Y
	제조사	모델	색상	사용기간	배기량	주행거리	연료	신차가격	중고가격
1	현대	제네시스	흰색	5.6	3,200	78,000	가솔린	50,000,000	12,000,000
2	기아	카니발	검정색	3.0	2,100	30,000	디젤	45,000,000	1,300,000
3	르노삼성	SM6	회색	2.2	1,300	2,000	가솔린	37,000,000	1,000,000
4	쌍용	티볼리	흰색	7.3	1,450	63,000	가솔린	41,000,000	5,300,000
...									
100	쉐보레	말리부	빨간색	4.5	1,320	44,000	가솔린	39,000,000	2,700,000

표본 수 100

그림 6.28 ▶ 머신러닝에서 사용되는 전형적인 테이블 형태의 데이터

특징과 레이블

특이값이나 결측값은 머신러닝 성능에 영향을 미치므로 사전 처리가 필요

- **특이값(outlier) 또는 이상치**

- 정상적인 범주를 벗어난 값

- **결측값(missing value) 또는 결측치**

- 빠진 데이터 값

- **중고가격 예측 예**

- 특성(features) 또는 특성 벡터(features vector)

- 예측값을 위한 속성인 제조사와 색상 등 8개

- 예측값 또는 목표(target) 값

- 마지막 열(정답)인 중고가격

- 레이블(label)

- 이미 아는 정답

- 정답이 확보된 데이터를 레이블 데이터(labeled data)

- **특징과 목표값, 수학 집합으로 표시**

- 중고 자동차 데이터에서는 목표 값이 중고가격인 y_1 하나인 경우

특징 표현이 차수가 높으면
여러 개로 표현

$$X = \{x_1, x_2, x_3, \dots, x_n\}, Y = \{y_1, y_2, y_3\}$$

그림 6.29 ▶ 데이터의 특성과 정답(레이블)

머신러닝 모델

주어진 문제를 해결하기 위해 데이터로 알고리즘을 학습시키면 바로 머신러닝 모델이 생성

머신러닝 모델

- 내부의 알고리즘으로 특정 유형의 패턴을 인식
 - 문제를 해결
- 훈련(학습) 과정을 통해
 - 데이터에 가장 적합한 매개변수로 설정된 알고리즘 준비
- 처음엔 몰랐던 매개 변수 값을
 - 스스로 학습을 통해 값을 결정
 - 정답이 붙은 학습 데이터로 모델을 학습
- 머신러닝 모델의 예측
 - 새로운 테스트 데이터에 대한 결과를 예측

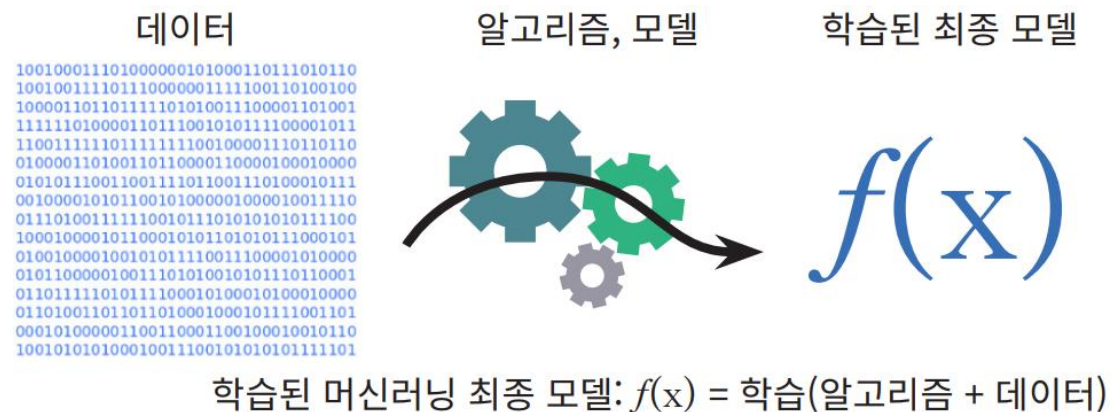


그림 6.30 ▶ 머신러닝 모델

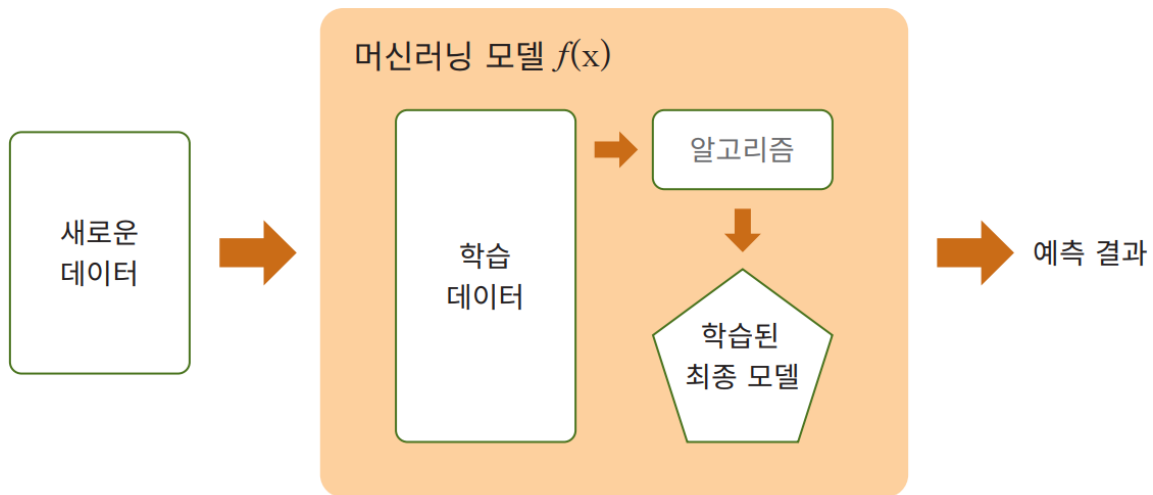


그림 6.31 ▶ 머신러닝 모델과 데이터

학습 데이터와 테스트 데이터 1/2

현재 머신러닝 모델이 제대로 학습되고 있는지를 판단하려면 훈련에 사용되지 않은 검증 데이터로 모델의 정확도(accuracy)나 손실함수(loss function)를 평가해 지속적으로 모델의 성능을 개선

• 데이터 분리

- 학습(훈련) 데이터(training data)와 테스트(검사 또는 시험) 데이터(test data)로 나눔

• 학습 데이터

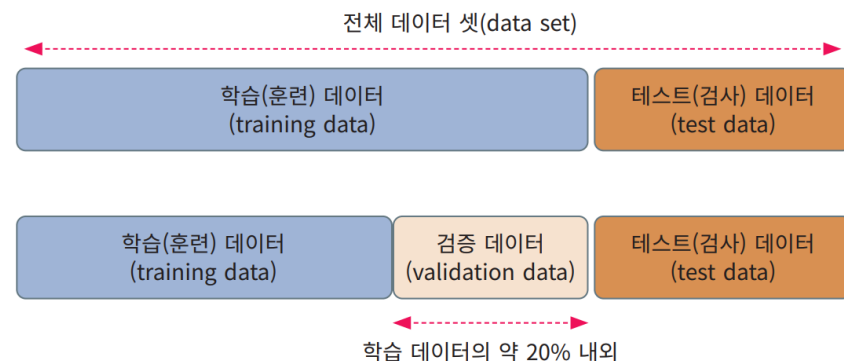
- 다시 학습 데이터(training data)와 검증 데이터(validation data)로 나눌 수 있음
- 학습에 사용되는 데이터
 - 참고서 연습문제 시험에 비유
- 검증 데이터
 - 학습 향상을 점검하기 위한 데이터, 모의고사에 비유

• 테스트 데이터

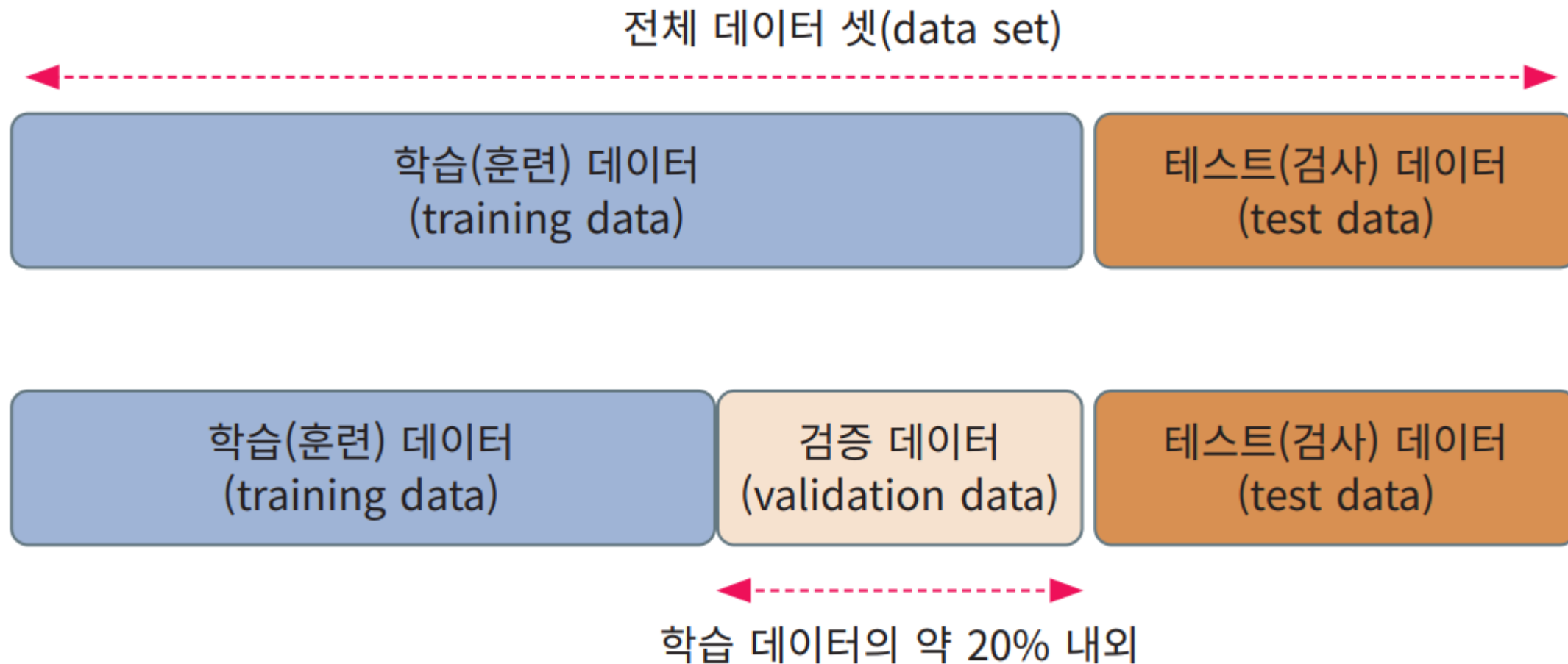
- 최종 모델 성능에 활용되는 데이터
 - 학습이 잘 되었는지 최종 검사에 사용, 수능 시험에 비유

• 지속적으로 모델의 성능을 개선에 필요한 기준 값(뒤에서 좀 더 자세히 설명됨)

- 손실함수(loss function)
 - 현재 모델의 예측의 오류 정도를 나타내는 값으로 작을수록 좋은 함수 값
- 정확도(accuracy)
 - 예측값과 정답의 일치하는 수준이라고 이해



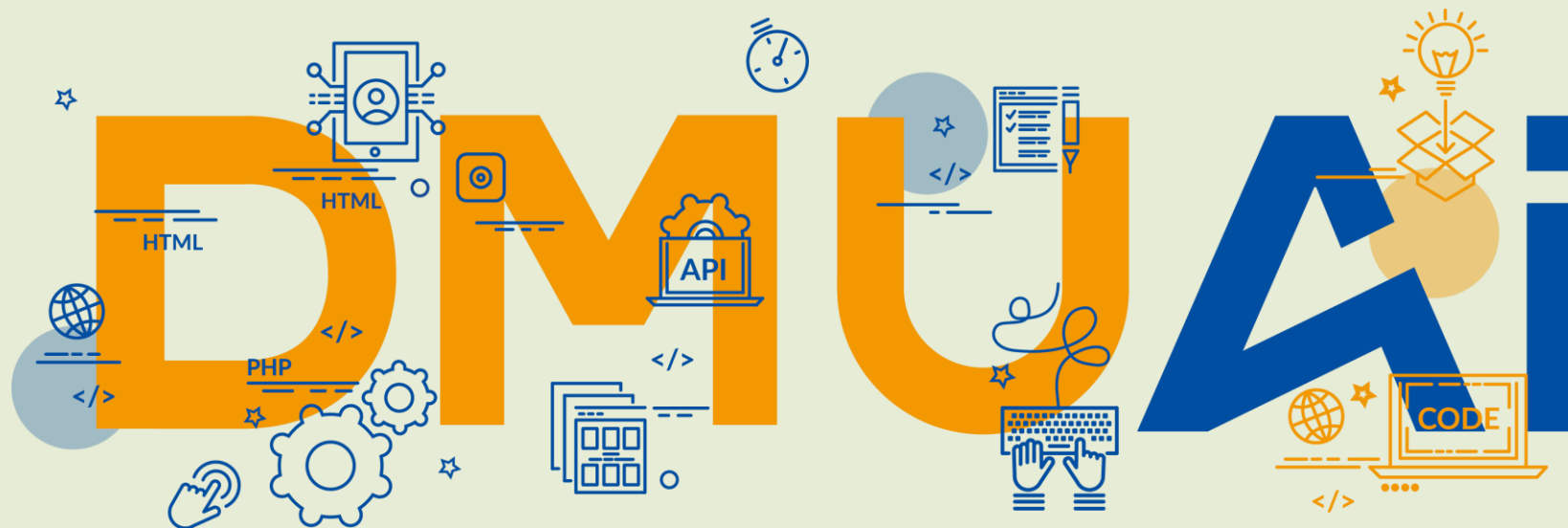
학습 데이터와 테스트 데이터 2/2



동양미래대학교 인공지능소프트웨어학과

머신러닝 종류(구분)

Dongyang Mirae University
Dept. Of Artificial Intelligence



지도 학습과 비지도 학습, 강화 학습

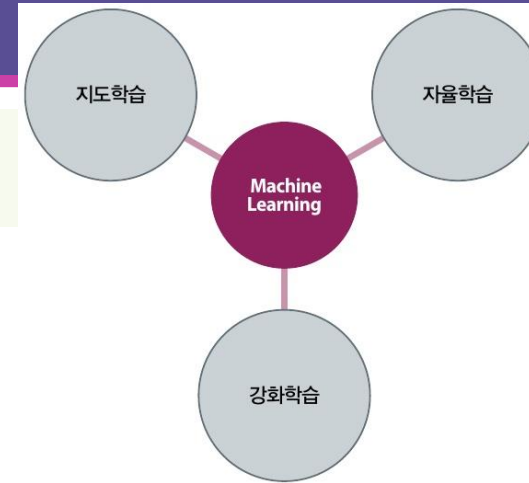
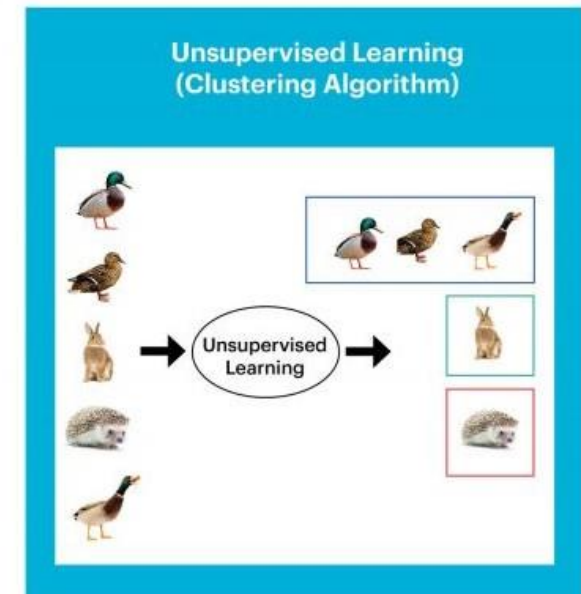
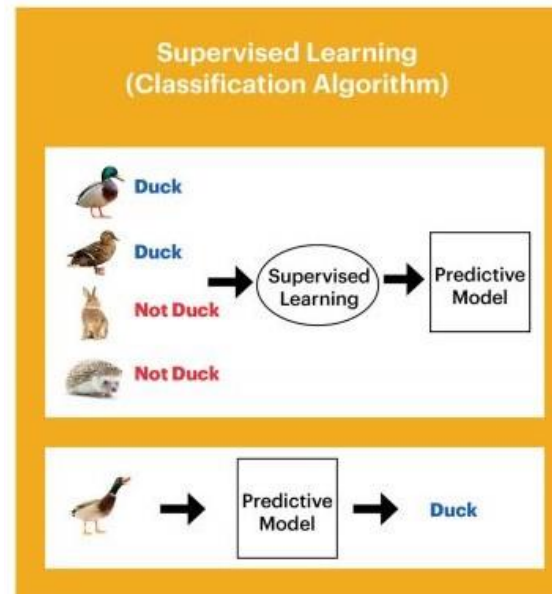


그림 10.26 ▶ 머신러닝 분야

• 머신러닝의 구분

- 지도 학습
 - 정답이 있는 예측
- 비지도(자율) 학습
 - 군집화(클러스터링) 알고리즘
- 강화 학습
 - 보상과 벌을 활용한 학습



머신러닝(지도학습) 대상 문제: 회귀와 분류

머신러닝을 통해 해결하고자 하는 문제는 크게 회귀(regression)와 분류(classification)로 나뉨

• 회귀

- 연속값(continuous value)을 예측하는 문제
 - 가격, 급여, 연령, 온도 등

• 분류

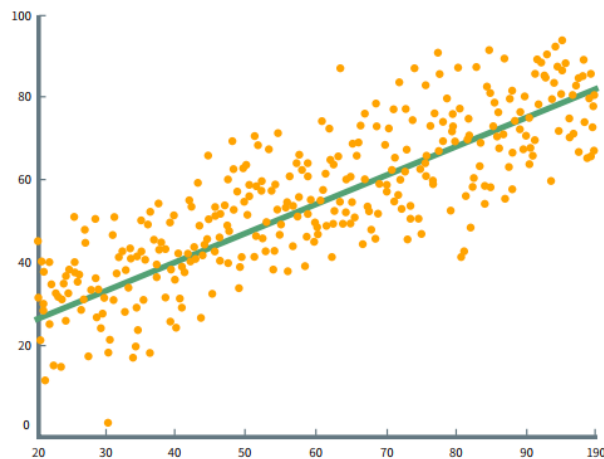
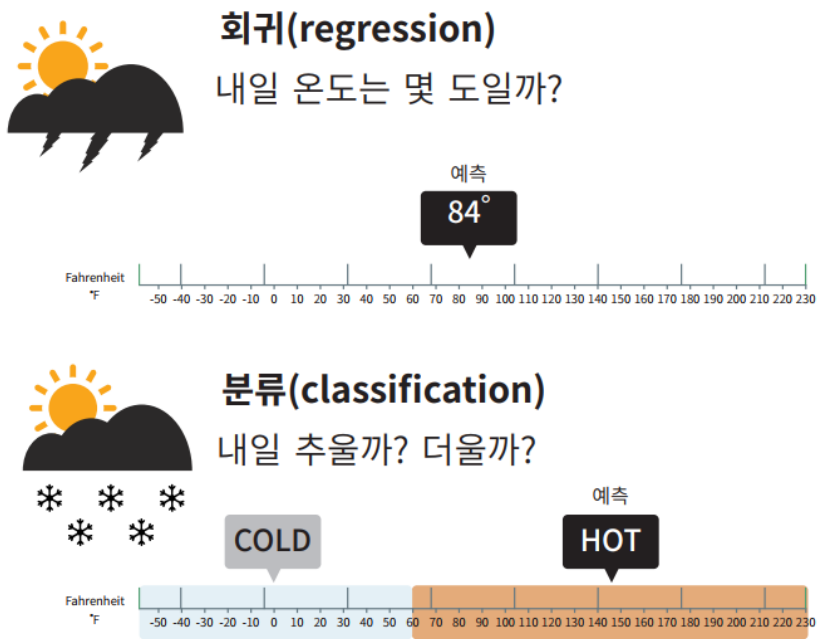
- 불연속적인 이산값(discrete value)을 예측

- 사과와 배, 남성과 여성,
 - 춥거나 더움, 참 또는 거짓
 - 클래스(class)

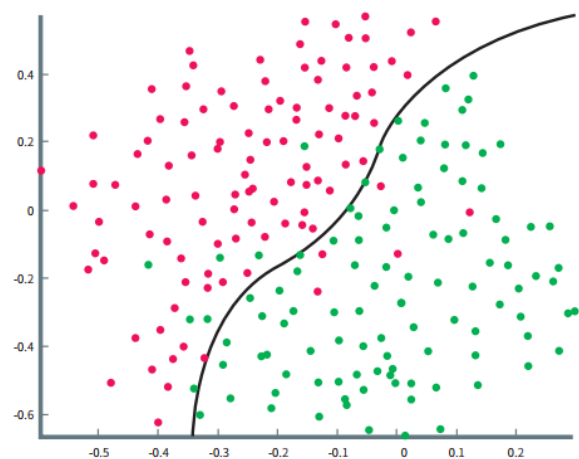
- 구분해야 할 유형

- 유형 수(# of classes)

- 남성과 여성을 분류해야 한다면 2



회귀(regression)



분류(classification)

머신러닝(지도학습) 대상 문제: 회귀와 분류

정답이 있는 학습 데이터

- 회귀: 일기예보

- 데이터를 잘 표현할 수 있는 최적의 선(line)을 찾는 과정

- 분류: 스팸(spam) 메일 필터링

- 데이터를 분류할 수 있는 결정 경계(decision boundary)를 찾는 과정

구분	회귀	분류
예측 값	연속 값	불연속 값
의미	목표 값을 더 정확하게 예측할 수 있는 최적의 선을 찾는 과정	데이터 세트를 여러 클래스로 나눌 수 있는 결정 경계 를 찾는 과정
종류	• 선형 회귀 • 비선형 회귀	• 이중 분류(Binary Classifier) • 다중 분류(Multi-class Classifier)
활용	• 날씨 예측 • 주가 예측 • 주택 가격 예측 • 중고 자동차 가격 예측	• 스팸메일 식별 • 음성 인식 • 암세포 식별 • 손글씨 숫자 인식

머신러닝 비지도 학습 문제: 군집화

처음 주어진 정답이 없는 왼쪽의 파란 데이터에서 오른쪽과 같이 주어진 데이터들을 가장 잘 설명하도록 그룹을 찾아내는 것이 군집화의 목적

• 군집화(clustering)

- 정답이 없는 학습인 비지도 학습의 일종
 - 군집(cluster)으로 나누는 작업
 - 주어진 데이터들을 유사한 특성을 가진 그룹으로 분리
- 그룹인 클러스터의 수
 - 알고리즘에 따라 미리 결정된 수

• 군집화 활용

- 고객의 성향을 분석해
 - 비슷한 성향의 고객끼리 군집으로 나누는 고객 세분화

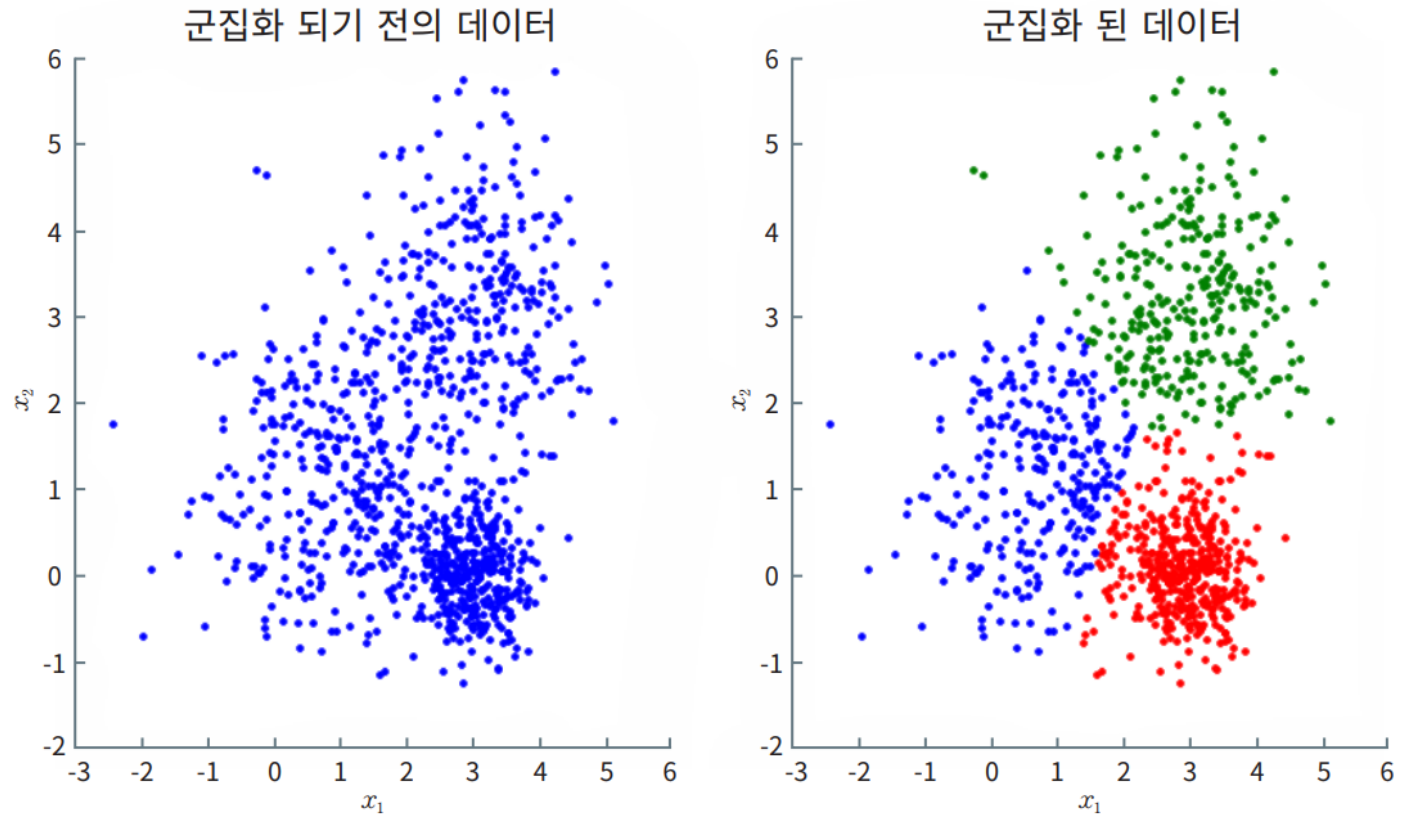


그림 6.34 ▶ 군집화 과정(i.stack.imgur.com/cIDB3.png)

머신러닝 종류

머신러닝은 용어 의미대로 '컴퓨터가 스스로 학습하여 문제를 해결하는 방법'

• 지도 학습과 비지도 학습, 반지도 학습 그리고 강화 학습으로 나뉨

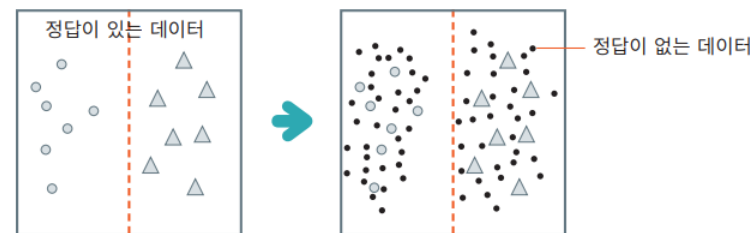
지도 학습(Supervised Learning)

문제와 정답을 모두 알려주고 공부시키는 방법



반지도 학습(Semi supervised Learning)

지도 학습과 비지도 학습을 섞어서 사용하는 학습 방식



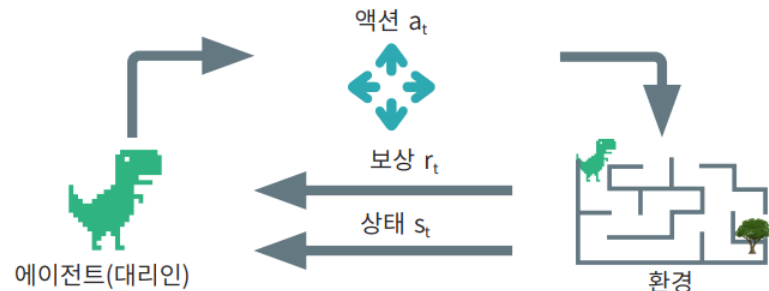
비지도 학습(Unsupervised Learning)

답을 가르쳐주지 않고 공부시키는 방법



강화 학습(Reinforcement Learning)

보상을 얻기 위한 정책을 다음 행동에 반영시키는 방법



머신러닝 종류 개요

- 머신러닝은 지도학습과 자율학습, 그리고 강화학습으로 분류

- 지도학습(supervised learning)

- 올바른 입력과 출력의 쌍으로 구성된 정답의 훈련 데이터(labeled data)로부터
입출력 간의 함수를 학습시키는 방법

- k-최근접 이웃 (k-Nearest Neighbors)
 - 선형 회귀 (Linear Regression)
 - 로지스틱 회귀 (Logistic Regression)
 - 서포트 벡터 머신 (Support Vector Machines (SVM))
 - 결정 트리 (Decision Tree)와 랜덤 포레스트 (Random Forests)

- 비지도(자율)학습(unsupervised learning)

- 정답이 없는 훈련 데이터(unlabeled data)를 사용하여 데이터 내에 숨어있는
어떤 관계를 찾아내는 방법

- clustering

- 강화학습(reinforcement learning)

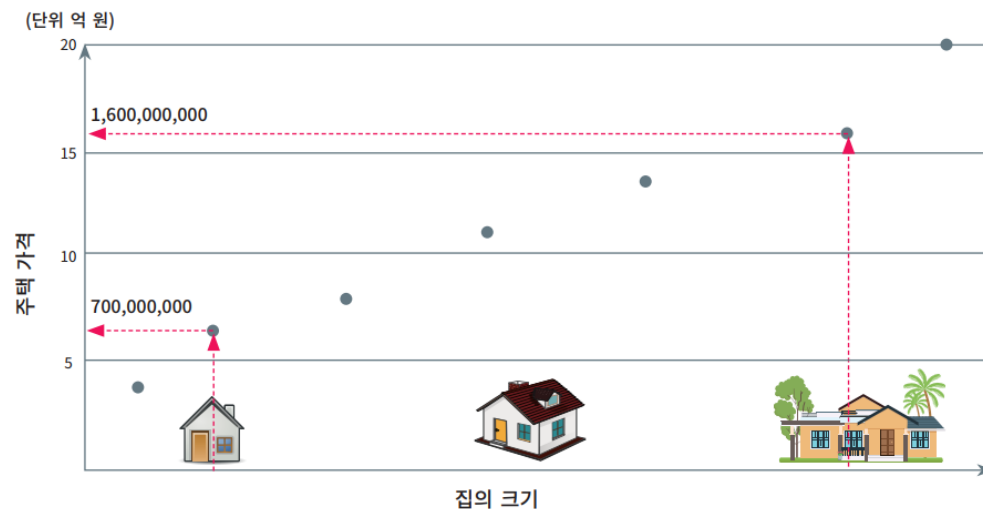
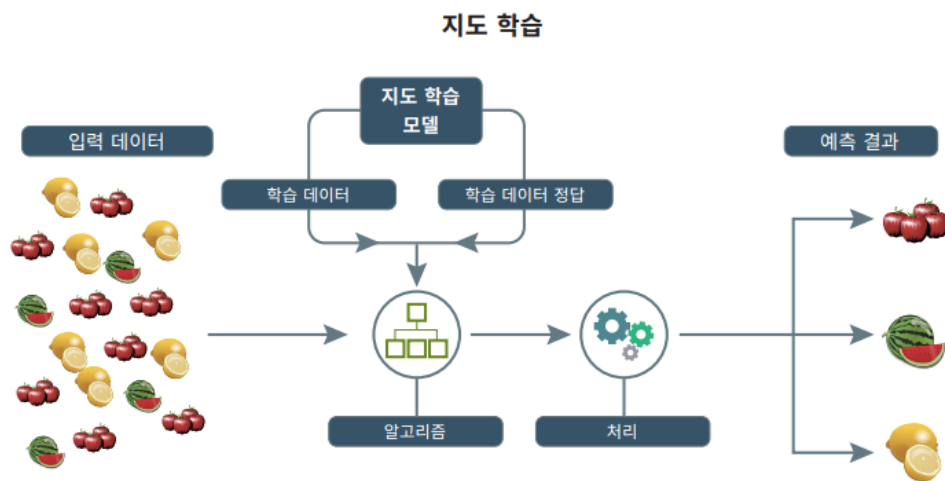
- 잘한 행동에 대해 보상을 주고 잘못된 행동에 대해 벌을 주는 경험을 통해 지식을 학습하는 방법

- 딥마닝의 알파고
 - 자동 게임분야

지도 학습(supervised learning)과 활용

훈련 과정에서 입력값 X_{data} 가 주어지면 입력값에 대한 정답(label)인 Y_{data} 로 학습

- 정답이 있는 훈련 데이터(labeled data)로부터 입출력 간의 함수인 모델을 학습
- 분류와 회귀 문제
 - 분류(classification): 주어진 데이터를 정해진 정답 유형에 따라 분류하는 문제
 - 사과와 배, 오렌지를 분류하는 지도 학습
 - 회귀(regression): 입력 데이터로부터 연속된 출력값을 예측하는 문제
 - 키로부터 몸무게를 예측하거나 지역과 평형으로 아파트 값을 예측
- 신용카드 부정사용 감지: 회귀 or 분류 어느 것일까요?



비지도 학습과 활용

정답 레이블이 없는 데이터를 비슷한 특징끼리 군집화해 새로운 데이터에 대한 결과를 예측하는 방법

- **데이터 내에 숨어있는 어떤 관계를 찾아 내는 방법**

- 정답이 없는 데이터로부터 패턴이나 형태를 찾아야 하기 때문에 지도 학습보다 더 어려움

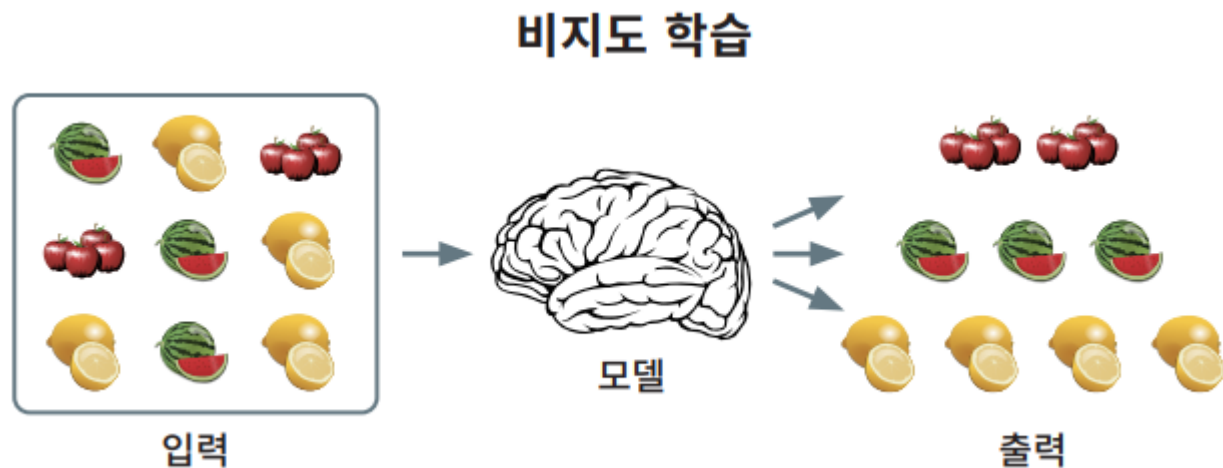
- **준지도 학습: 반지도 학습(semi-supervised learning)**

- 지도 학습과 비지도 학습을 섞어서 사용하는 학습 방식

- **비지도 학습의 대표적인 종류는 군집화(clustering)**

- 사과와 배, 오렌지 등으로 군집화

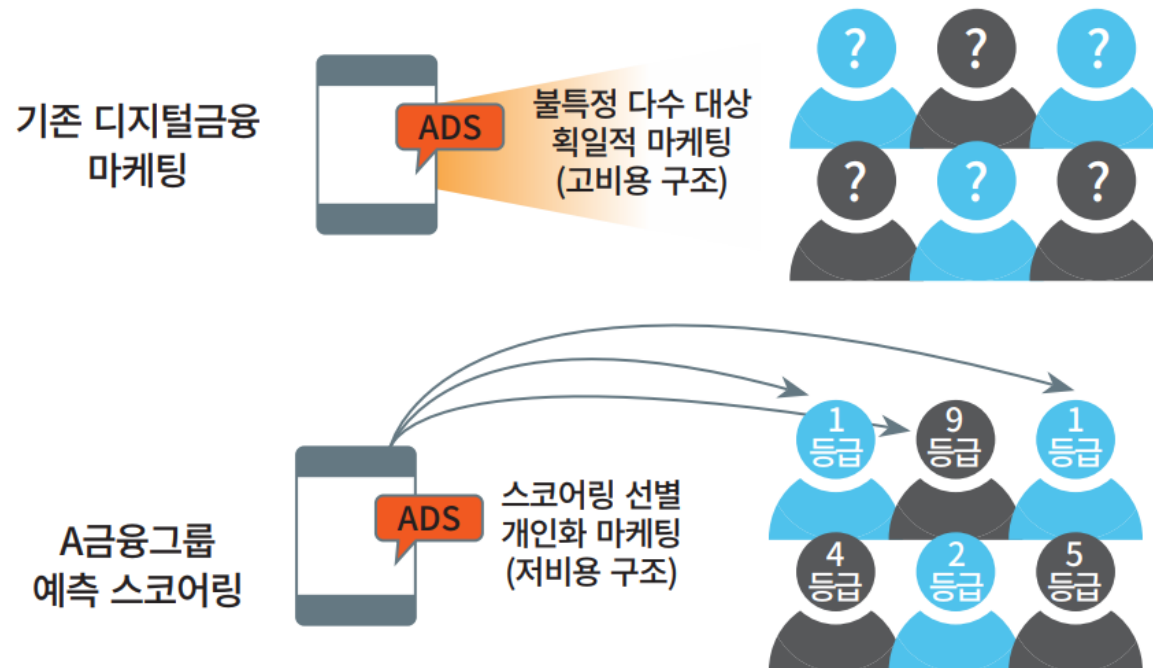
- **과일의 사진에서 각각의 사진 데이터에 대해 색깔이 무엇인지, 모양이 어떠한 지 등에 대한 특징**



비지도 학습의 군집화 활용 사례

• 고객의 성향 군집화

- 금융 회사는 고객의 투자 성향을 몇 개로 군집화
- 다른 광고 메시지를 보낸다면 마케팅 효과를 증진
 - **상품 가입 확률이 높은 잠재 고객에게만 목적성 광고 메시지를 보냄**
 - 광고비를 줄이고 수익성을 높일 수 있음



강화 학습과 활용

주어진 현재 상황(situation)에서 보상(reward)을 최대화 하도록 다음 행동(action)을 학습시키는 방법으로 개에게 좋아하는 먹이를 주며 개를 훈련시키는 방법에 비유

• 강화 학습(reinforcement learning)

- 어느 주어진 환경(environments)에서 학습 주체인 대리인(agent)에게 지시한 명령에 대해
 - 대리인을 관찰(observation)
- 다음 행동(action)을 하면
 - 잘못된 행동에 대해서는 벌
 - 잘한 행동에 대해 보상
 - 대리인은 가장 큰 보상을 얻기 위한 정책(policy)을 통해 스스로 학습하는 방법

• 강화 학습 비유

- 아기가 걸음을 걷는 학습 과정
- 어릴 적 자전거를 배우는 과정

• 강화 학습 사례

- 게임이나 자율주행 등에 주로 활용되는 기법
 - '궁극적인 목표는 있지만, 그에 이르는 과정은 정답이 없는 상황'
- 알파고(AlphaGo)

강화 학습(deep Q-learning) 적용 게임

- 구글 딥마인드의 인공지능 벽돌 깨기와 팩맨 게임 모습
 - <https://www.youtube.com/watch?v=V1eYniJ0Rnk>
 - <https://www.youtube.com/watch?v=QilHGSYbjDQ>
- 자율주행 강화학습 aws
 - <https://www.youtube.com/watch?v=OBSIOIZ1yM8>

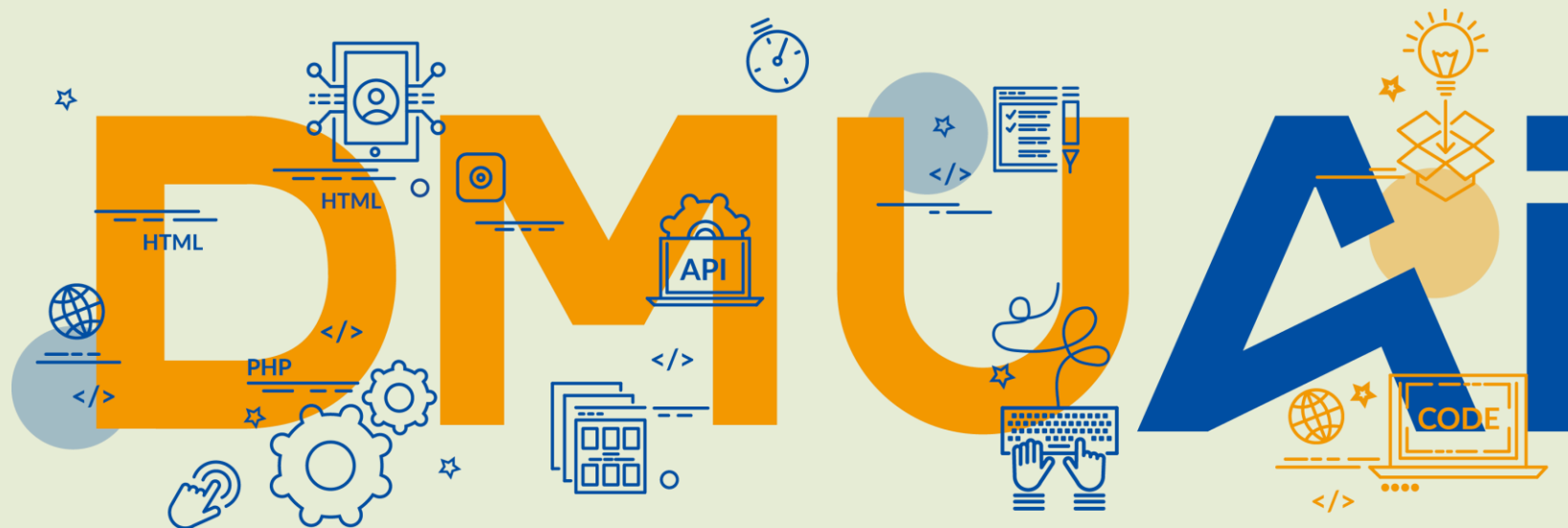


그림 6.42 ▶ 구글 딥마인드 Atari 벽돌 깨기 게임과 팩맨 게임

동양미래대학교 인공지능소프트웨어학과

머신러닝과 딥러닝 비교

Dongyang Mirae University
Dept. Of Artificial Intelligence



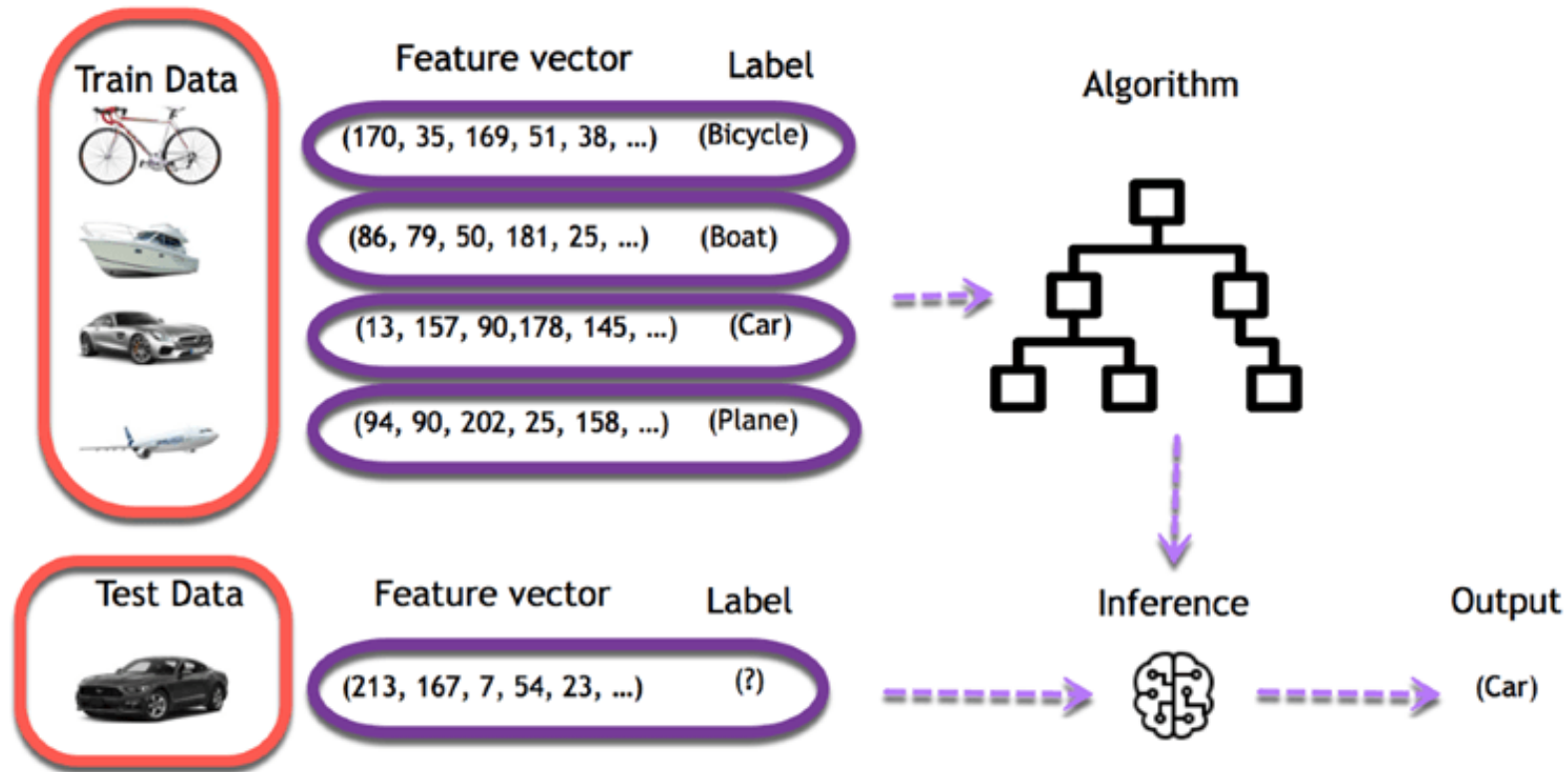
머신러닝과 딥러닝

- 기계학습이라고도 부르는 머신러닝(machine learning)
 - 주어진 데이터를 기반으로
 - 기계가 스스로 학습하여
 - 성능을 향상시키거나 최적의 해답을 찾기 위한 학습 지능 방법
- 딥러닝
 - 인공신경망 기반의 머신러닝



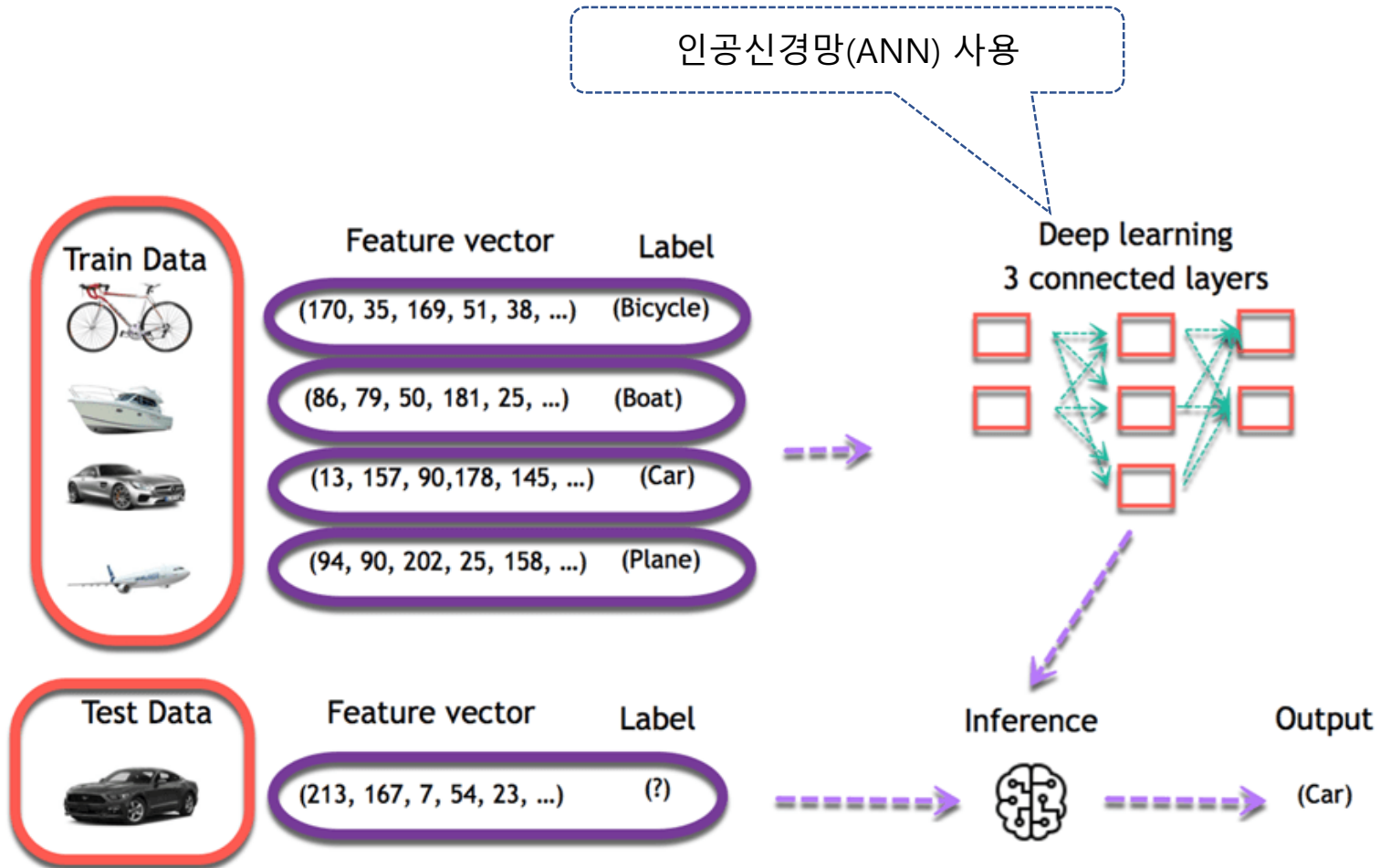
머신러닝과 딥러닝 비교 1/7

• 머신러닝



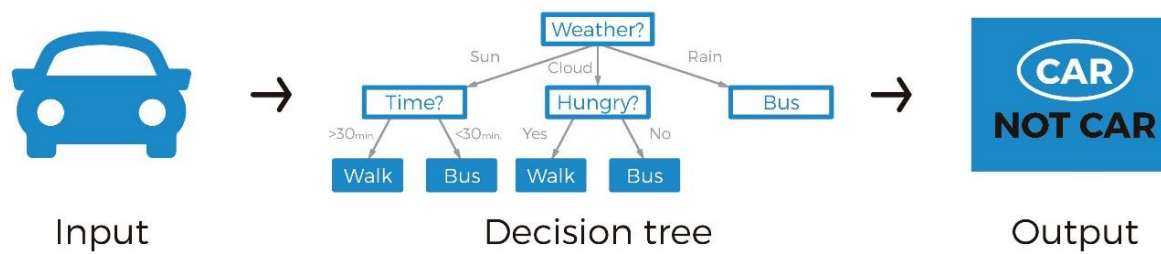
머신러닝과 딥러닝 비교 2/7

• 딥러닝

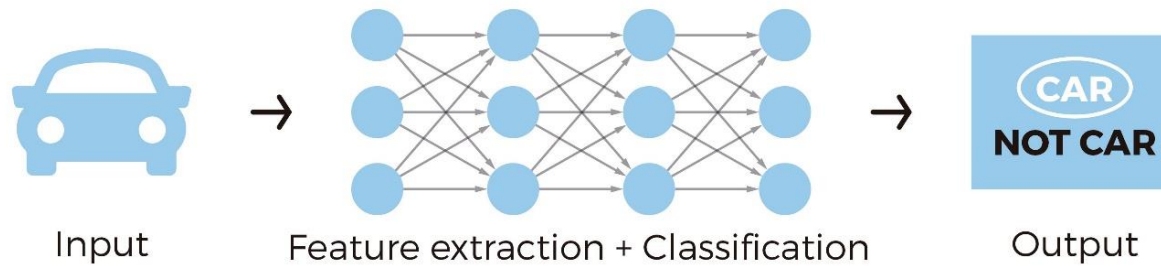


머신러닝과 딥러닝 비교 3/7

Machine Learning



Deep Learning



머신러닝과 딥러닝 비교 4/7

- 머신러닝 (Machine Learning)

- 알고리즘을 이용해 데이터를 분석하고, 분석을 통해 학습하며, 학습한 내용을 기반으로 판단이나 예측

- 딥러닝 (Deep Learning)

- 인공신경망에서 발전한 형태의 인공지능으로, 인간 뇌의 뉴런과 유사한 입력 계층, 은닉 계층, 출력 계층을 활용해 데이터를 학습

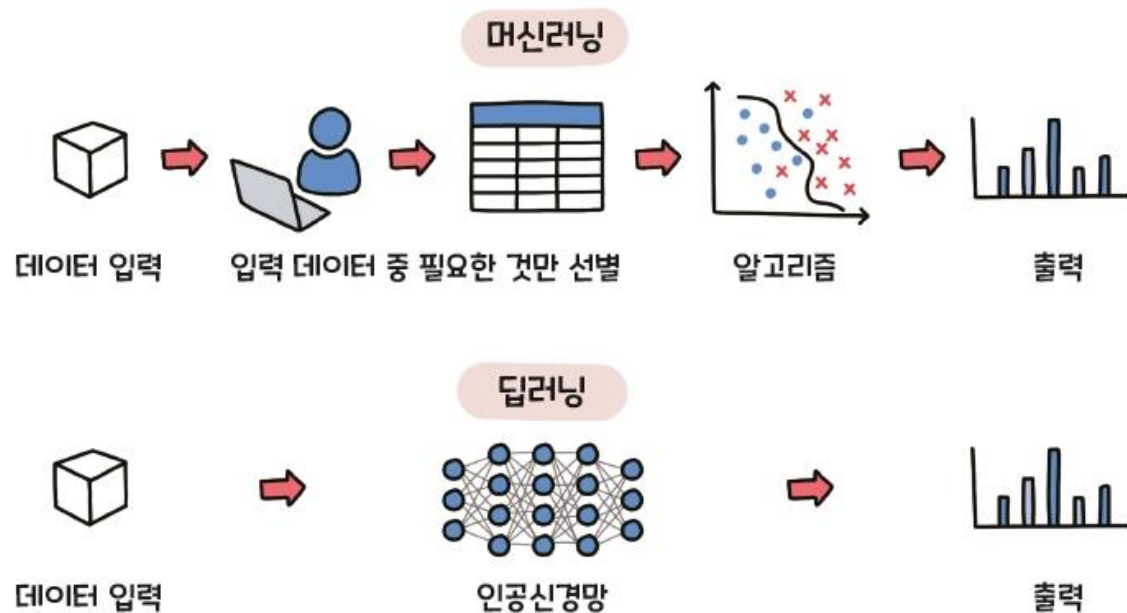


그림 1-28 머신러닝과 딥러닝의 학습 차이

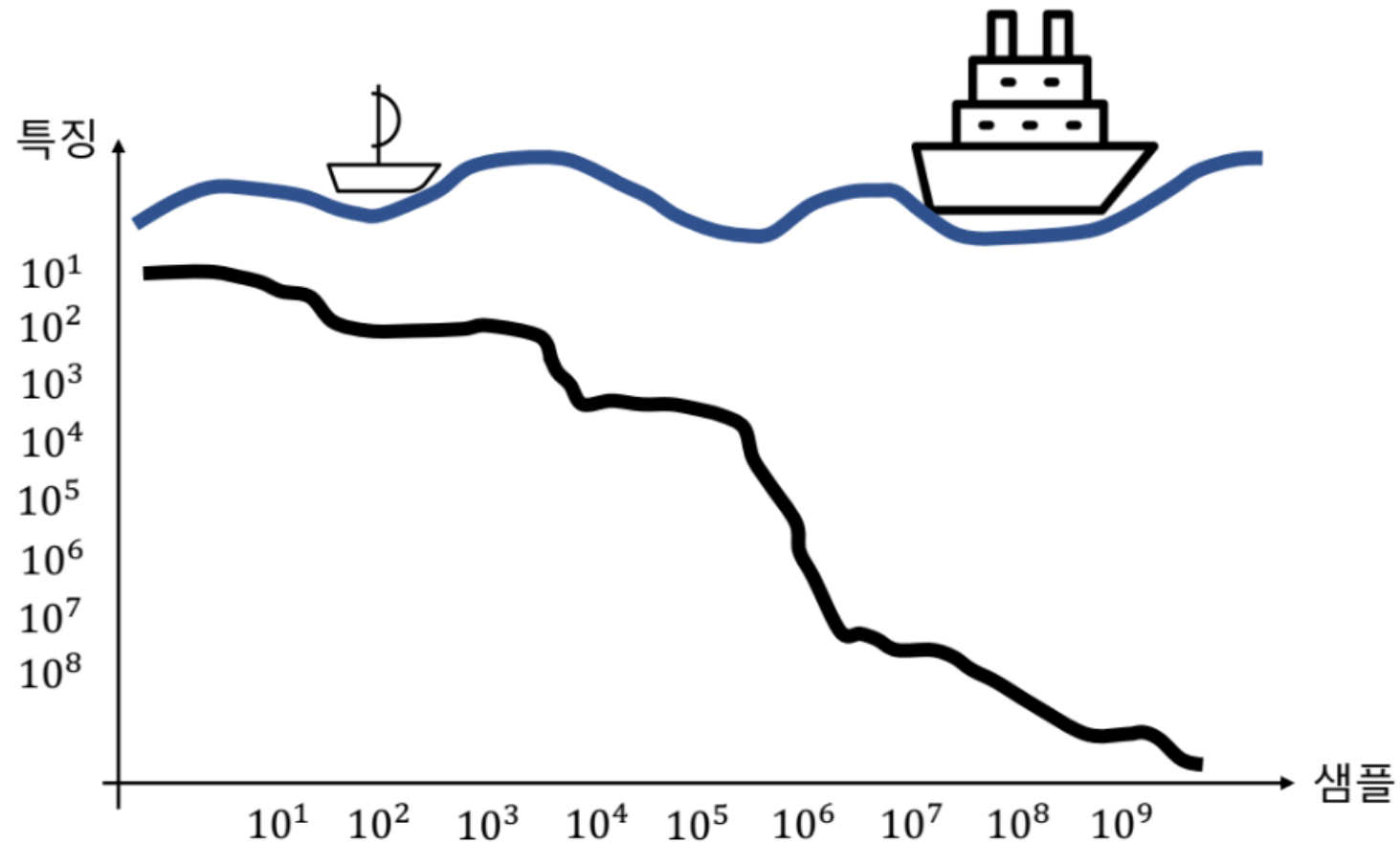
머신러닝과 딥러닝 비교 5/7

머신러닝과 딥러닝의 차이점

	기계 학습	딥러닝
데이터의 존성	중소형 데이터 세트에서 탁월한 성능	큰 데이터 세트에서 뛰어난 성능
하드웨어 의존성	저가형 머신에서 작업하십시오.	GPU가있는 강력한 기계가 필요합니다. DL은 상당한 양의 행렬 곱셈을 수행합니다.
기능 공학	데이터를 나타내는 기능을 이해해야 함	데이터를 나타내는 최고의 기능을 이해할 필요가 없 습니다
실행 시간	몇 분에서 몇 시간	최대 몇 주. 신경망은 상당한 수의 가중치를 계산해 야합니다.

머신러닝과 딥러닝 비교 6/7

- 특징과 데이터가 많을수록 딥러닝에 적합



머신러닝과 딥러닝 비교 7/7

- 딥러닝 외 대부분의 머신러닝 알고리즘

- 데이터의 양이 일정 수준을 넘어가면 더 이상 성능이 향상되지 않는 한계를 보임

- 딥러닝

- 데이터가 많을수록 성능이 좋아지는 것이 특징

- “얼마나 더 많은 양의 데이터를 확보하느냐”에 따라 인공지능 (딥러닝) 역량에 차이가 커짐

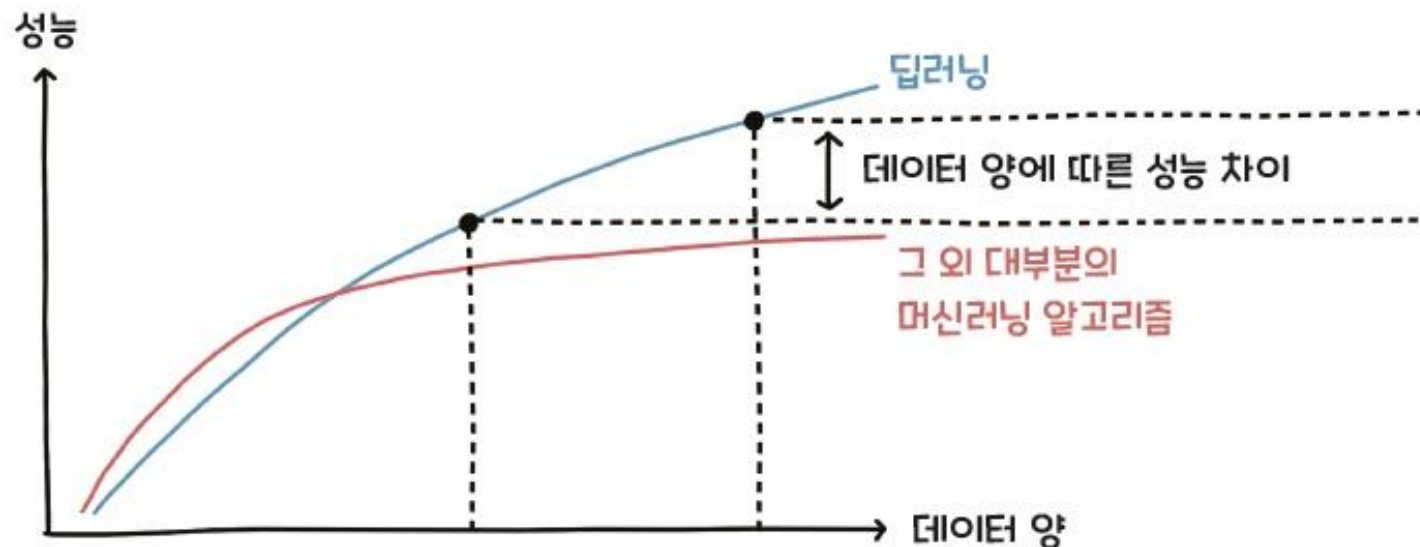
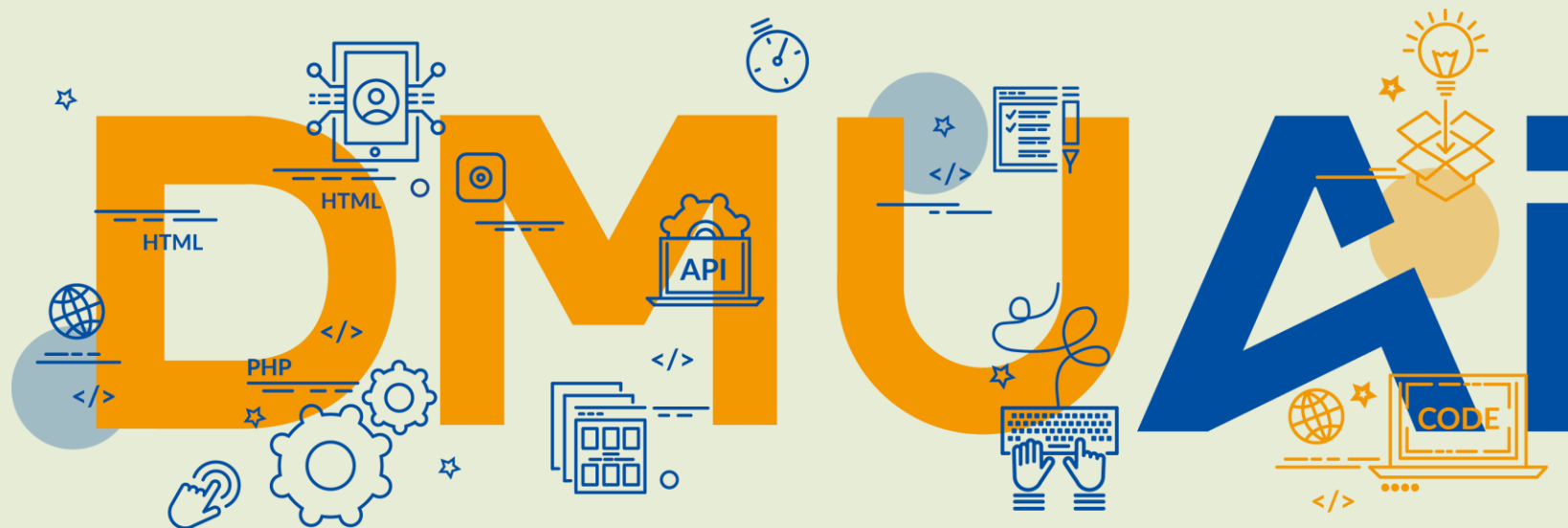


그림 3-12 딥러닝의 특징 2: 데이터의 양이 많아질수록 성능이 지속적으로 향상

동양미래대학교 인공지능소프트웨어학과

CPU GPU TPU

Dongyang Mirae University
Dept. Of Artificial Intelligence



그래픽처리 장치 GPU의 인기

GPU란 용어는 1999년 엔비디아(Nvidia)에서 처음 사용

• 그래픽 처리 장치 GPU

- Graphics Processing Unit
- 그래픽 연산 처리를 하는 전용 프로세서

• GPGPU

- General Purpose Graphic Processing Unit
 - 일반 CPU 프로세서를 돕는 보조프로세서 (coprocessor)로서의 GPU
- 중앙 처리 장치(CPU)가 맡았던 응용 프로그램들의 계산에 GPU를 사용하는 기술
 - GPU 컴퓨팅이란 GPGPU를 연산에 참여
 - 고속의 병렬처리
 - 대량의 행렬과 벡터를 다루는 데 뛰어난 성능을 발휘
- 딥러닝의 심층신경망에서 빅데이터를 처리
 - 대량의 행렬과 벡터를 사용

• GPU 사용이 매우 효과적

- 12개 GPU가 2,000개의 CPU와 비슷한 계산 능력

• GPU 병렬 처리 동영상

- <https://www.youtube.com/watch?v=P28LKWTzrl>



엔비디아 리드텍 쿼드로 그래픽카드 RTX8000
(리더스시스템즈)
★★★★★ 2개 상품평

2% 8,772,970원 ①
8,572,970 원 즉시할인가 로켓배송

🕒 최대 50,000원 적립

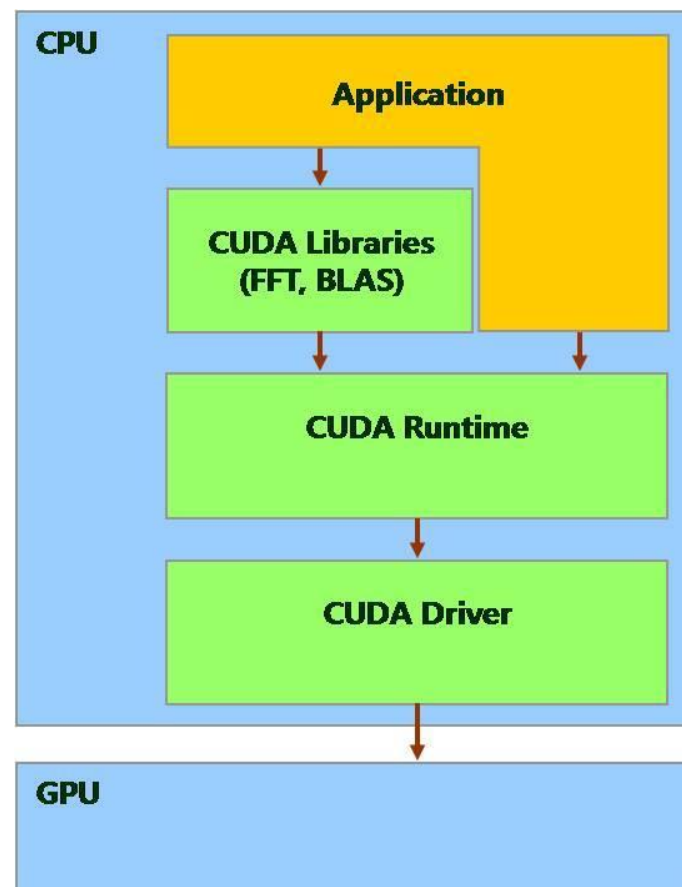
무료배송 [다른 판매자 보기\(2\)](#)

내일(월) 2/14 도착 보장 (11시간 28분 내 주문 시 / 서울·경기 기준)

모델명/품번: RTX8000



- GPU 업체인 NVIDIA의 GPU를 사용하기 위한 라이브러리 소프트웨어
 - Compute Unified Device Architecture의 약자



구글의 TPU

- **구글은 2016년**

- 텐서 처리 장치(Tensor Processing Unit)를 발표
- 텐서란 벡터·행렬을 의미
- TPU는 데이터 분석 및 딥러닝용 칩으로서 벡터·행렬연산의 병렬처리에 특화
- 텐서플로(TensorFlow)
 - **TPU를 위한 소프트웨어**

