
Artificial Intelligence

컴퓨터개론

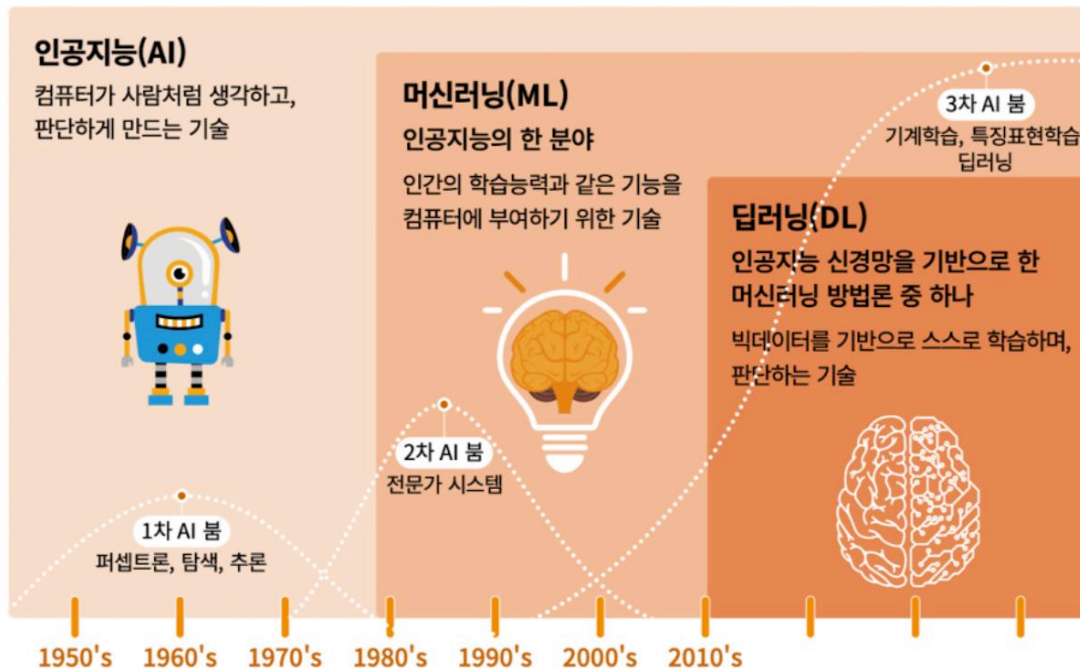
(Introduction to Computer Systems)

GEN1030

Machine Learning

인공지능, 머신러닝, 딥러닝

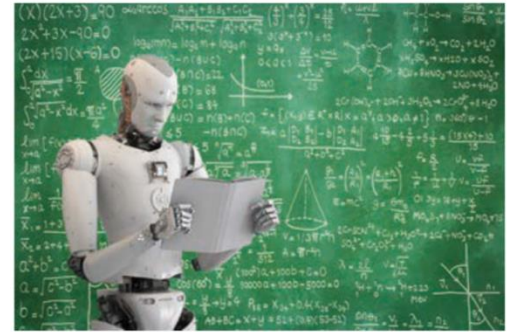
- 인공지능이 가장 큰 범주
- 머신러닝(Machine learning)
 - 데이터를 기반으로 기계 스스로 학습하는 인공지능의 한 분야
- 딥러닝(Deep learning)
 - 머신러닝의 분야 중 하나
 - 퍼셉트론으로 구성된 인공신경망으로 여러 단계의 심층 학습



머신러닝 정의

- 일반적 머신러닝 정의

- 주어진 데이터를 기반으로 기계가 스스로 학습, 성능을 향상시키거나 최적의 해답을 찾기 위한 지능적 학습 방법
- 명시적으로 프로그래밍을 하지 않아도 컴퓨터가 학습을 할 수 있도록 해주는 인공지능의 한 형태
- cf) 과거: 외부에서 학습 방법을 알려줌



- 데이터 중요

- 많은 데이터가 유입될 수록 더 많이 학습

머신러닝 vs. 전통적 프로그래밍

- Example) 강아지와 고양이를 분류하는 일



머신러닝 vs. 전통적 프로그래밍

- Example) 강아지와 고양이를 분류하는 일
- 전통적 프로그래밍
 - 명시적으로 프로그래밍(Explicit programming)
 - 생김새의 특징을 세부적으로 코딩해서 분류 (특징을 찾는 것 어려움)
 - 입력 데이터 사용, 프로그램을 실행하여 출력 생성
- 머신러닝
 - 스스로 데이터를 반복적으로 학습하여 기술을 터득하는 방식
 - 입력 데이터와 정답이 있는 출력이 알고리즘에 공급 → 프로그램(모델) 생성
 - 알고리즘이 데이터에서 학습을 해 자동으로 규칙 공식화



머신러닝 수행 과정

- 1) 데이터 수집
- 2) 데이터 전처리
- 3) 모델 학습
- 4) 성능 개선
- 5) 시각화

머신러닝 수행 과정

1) 데이터 수집

- 수집된 데이터의 품질과 양이 예측 모델 성능에 큰 영향

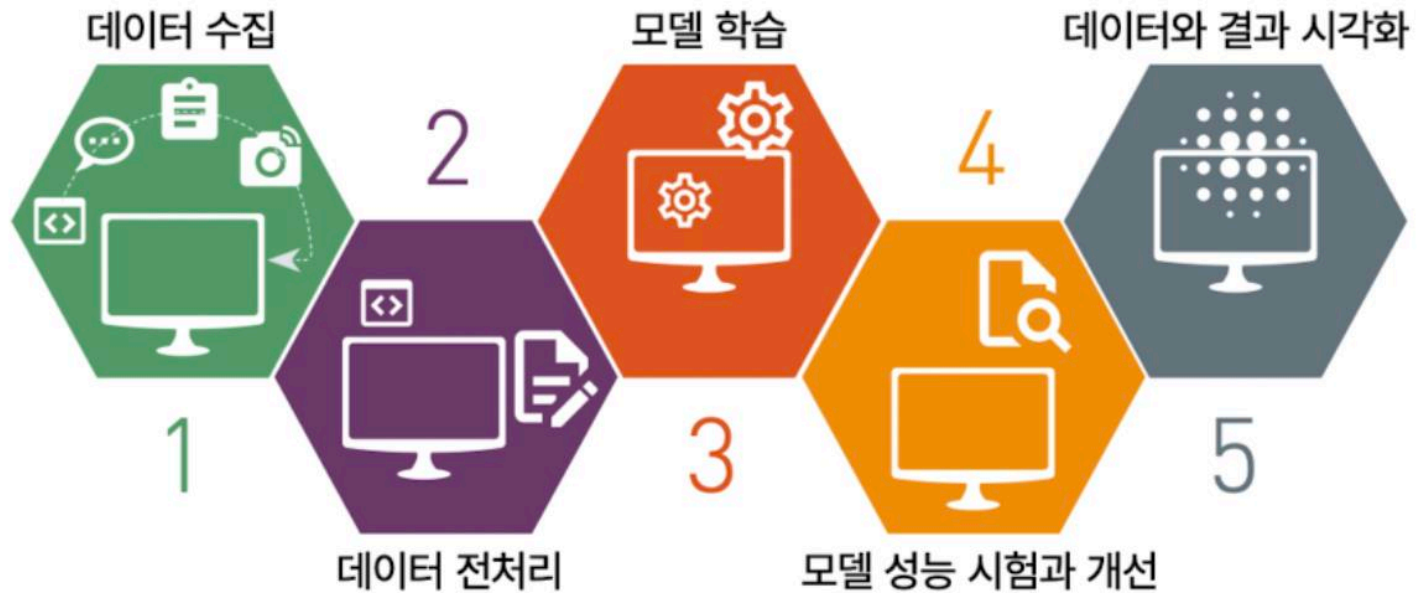
2) 데이터 전처리

- 잘못된 값 수정
 - 누락 값 채우거나 이상 값 수정/제거
- 정규화(Normalization)
 - 데이터 특성마다 차이가 큰 경우, 모델에 따라 값의 크기를 정규화
- 데이터 변환 작업
 - 모델 계산에 적합한 자료 구조로 변환

머신러닝 수행 과정

- 1) 데이터 수집
- 2) 데이터 전처리
- 3) 모델 학습
 - 적합한 머신러닝 모델 생성하고 데이터 학습
- 4) 성능 개선
 - 학습된 모델의 성능 시험 및 개선
- 5) 시각화
 - 데이터, 머신러닝 과정, 결과를 시각화

머신러닝 수행 과정



Data, Feature

- 데이터 집합(dataset)

- 행

- 표본(Sample)
- 인스턴스(Instance)
- 데이터 포인터(Data pointer)

sample 수: 100

- 열

- 특징/특성(feature)

feature 수: 8

특징 X_1, X_2, X_3, \dots

	제조사	모델	색상	사용기간	배기량	주행거리	연료	신차가격
1	현대	제네시스	흰색	5.6	3,200	78,000	가솔린	50,000,000
2	기아	카니발	검정색	3.0	2,100	30,000	디젤	45,000,000
3	르노삼성	SM6	회색	2.2	1,300	2,000	가솔린	37,000,000
4	쌍용	티볼리	흰색	7.3	1,450	63,000	가솔린	41,000,000
...								
100	쉐보레	말리부	빨간색	4.5	1,320	44,000	가솔린	39,000,000

Data, Feature

- 특이값(Outlier)
 - 정상적인 범주를 벗어난 값
- 결측값(Missing value)
 - 빠진 데이터 값
- 특이값이나 결측값은 머신러닝 성능에 영향을 줌
→ 전처리 필요

특징 X_1, X_2, X_3, \dots

	제조사	모델	색상	사용기간	배기량	주행거리	연료	신차가격
1	현대	제네시스	흰색	5.6	3,200	78,000	가솔린	50,000,000
2	기아	카니발	검정색	3.0	2,100	30,000	디젤	45,000,000
3	르노삼성	SM6	회색	2.2	1,300	2,000	가솔린	37,000,000
4	쌍용	티볼리	흰색	7.3	1,450	63,000	가솔린	41,000,000
...								
100	쉐보레	말리부	빨간색	4.5	1,320	44,000	가솔린	39,000,000

Feature, Label

- Example) 중고가격 예측
 - 8개의 특징(feature) 이용하여 중고가격 예측
 - Feature vector: 8개 특징 묶은 벡터
 - Label (정답): 중고가격
 - Labeled data: 정답(label) 포함된 데이터
 - 예측값(Prediction): 머신러닝 모델이 예측한 중고 가격

특징 X_1, X_2, X_3, \dots

	제조사	모델	색상	사용기간	배기량	주행거리	연료	신차가격
1	현대	제네시스	흰색	5.6	3,200	78,000	가솔린	50,000,000
2	기아	카니발	검정색	3.0	2,100	30,000	디젤	45,000,000
3	르노삼성	SM6	회색	2.2	1,300	2,000	가솔린	37,000,000
4	쌍용	티볼리	흰색	7.3	1,450	63,000	가솔린	41,000,000
...								
100	쉐보레	말리부	빨간색	4.5	1,320	44,000	가솔린	39,000,000

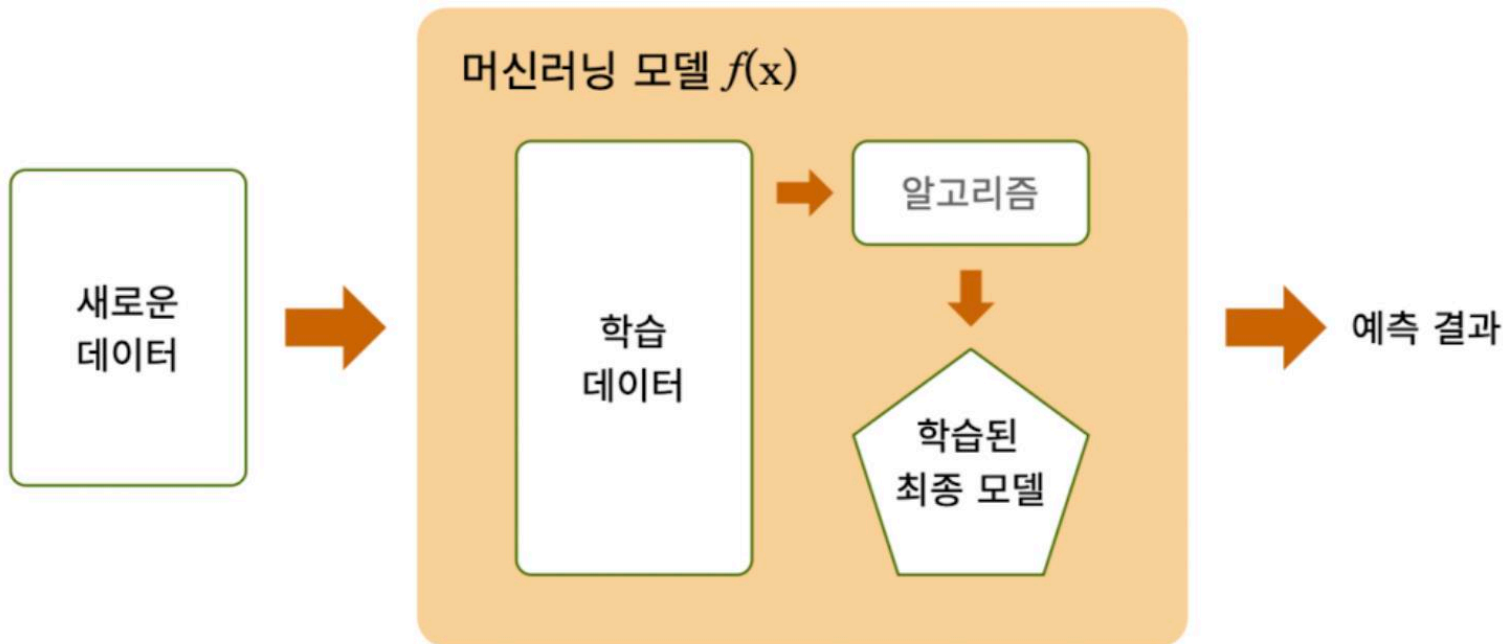
머신러닝 모델

- 내부의 알고리즘으로 특정 유형의 패턴을 인식해 문제 해결
 - 주어진 문제 해결을 위해 데이터로 알고리즘을 학습 시킨 결과
- 훈련 과정을 통해 데이터에 가장 적합한 매개변수로 설정된 알고리즘 준비
 - 처음에는 몰랐던 매개변수 값을 스스로 학습을 통해 결정



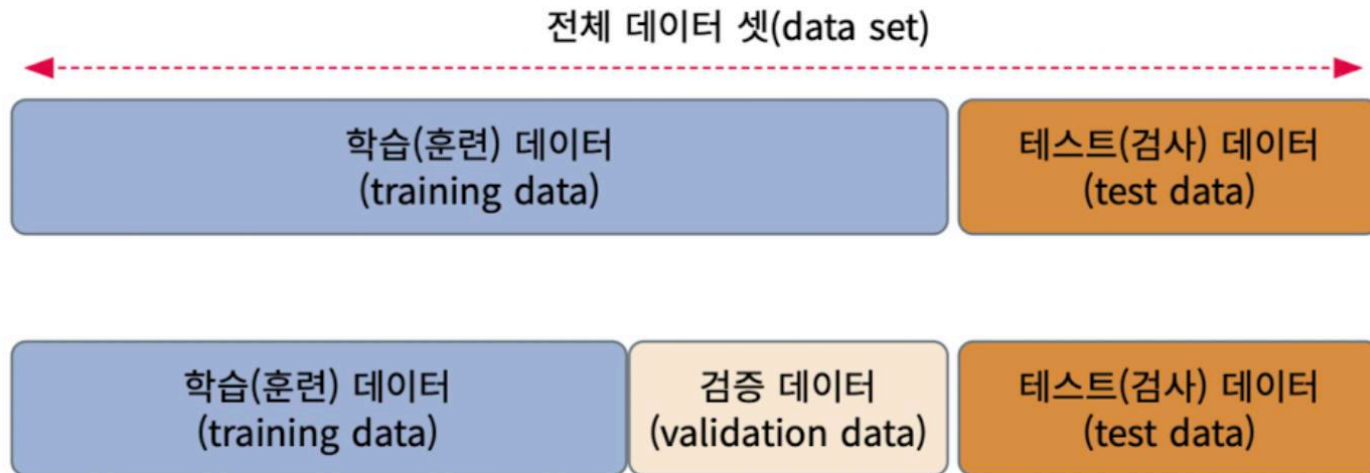
머신러닝 모델

- 정답이 있는 학습 데이터(labeled data)로 모델 학습 → 새로운 데이터에 대해 결과 예측



학습 데이터, 테스트 데이터

- 데이터 분리
 - 학습(훈련) 데이터(training data)와 테스트(시험) 데이터(test data)
 - 학습 데이터
 - 다시 학습 데이터와 검증 데이터(validation data)로 나눌 수 있음
 - 테스트 데이터
 - 최종 모델 성능 평가를 위해 활용
 - 학습이 잘 되었는지 최종 검사



머신러닝 학습 방식에 따른 분류

지도 학습(Supervised Learning)

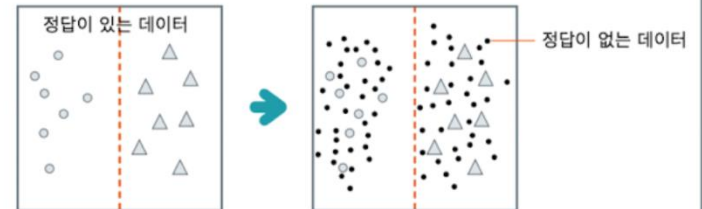
문제와 정답을 모두 알려주고 공부시키는 방법



▶ 예측, 분류

반지도 학습(Semi supervised Learning)

지도 학습과 비지도 학습을 섞어서 사용하는 학습 방식



비지도 학습(Unsupervised Learning)

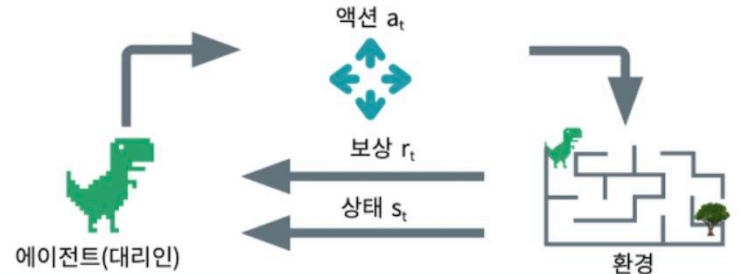
답을 가르쳐주지 않고 공부시키는 방법



▶ 연관규칙, 군집

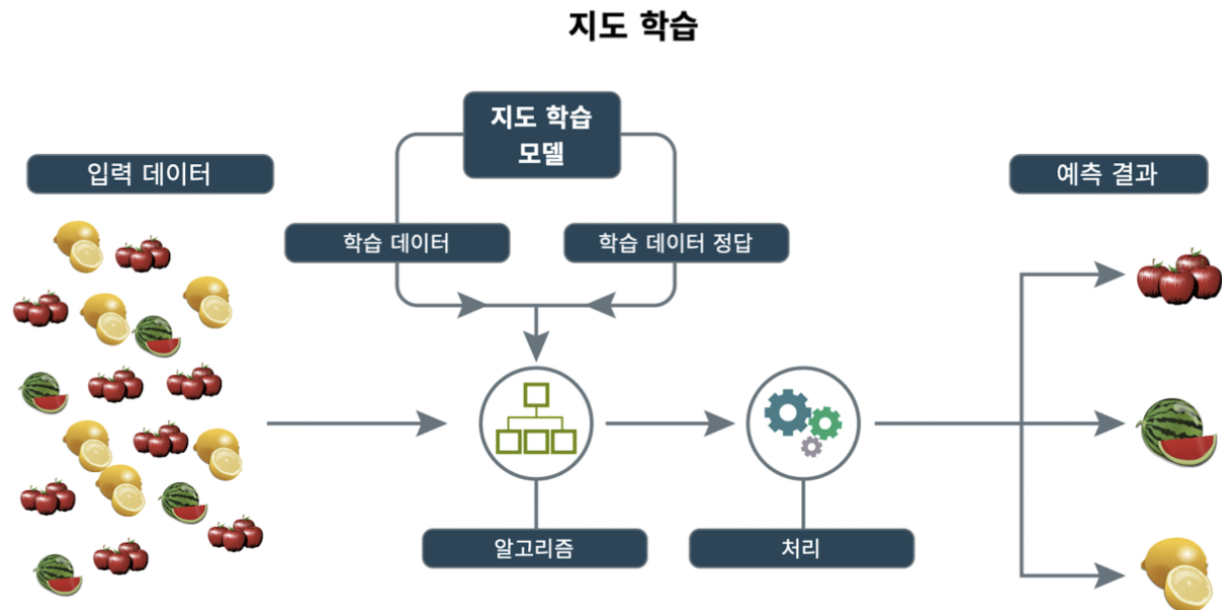
강화 학습(Reinforcement Learning)

보상을 얻기 위한 정책을 다음 행동에 반영시키는 방법



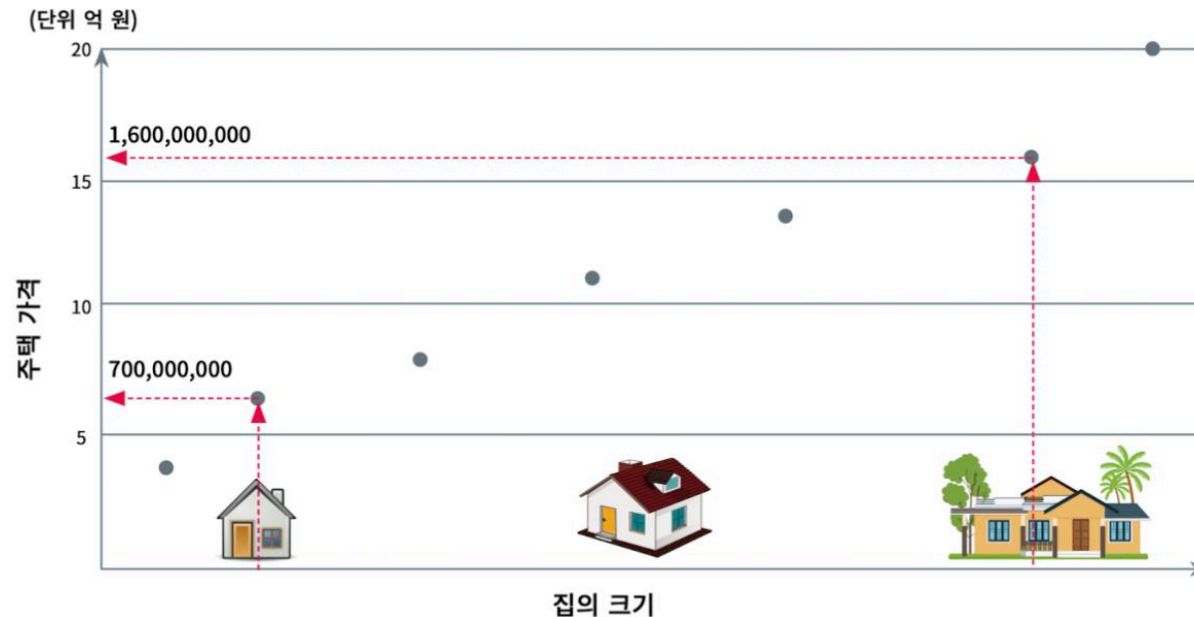
지도 학습(Supervised Learning)

- 정답이 있는 훈련 데이터(labeled data)로부터 입출력 간의 함수인 모델을 학습
 - 분류와 회귀
- 분류(Classification)
 - 주어진 데이터를 정해진 정답 클래스에 따라 분류
 - 클래스(class): 구분해야 할 유형 (ex. 사과와 배)



지도 학습(Supervised Learning)

- 정답이 있는 훈련 데이터(labeled data)로부터 입출력 간의 함수인 모델을 학습
 - 분류와 회귀
- 회귀(Regression)
 - 입력 데이터로부터 연속된 출력값(연속값, continuous value) 예측
 - Ex) 가격, 급여, 온도

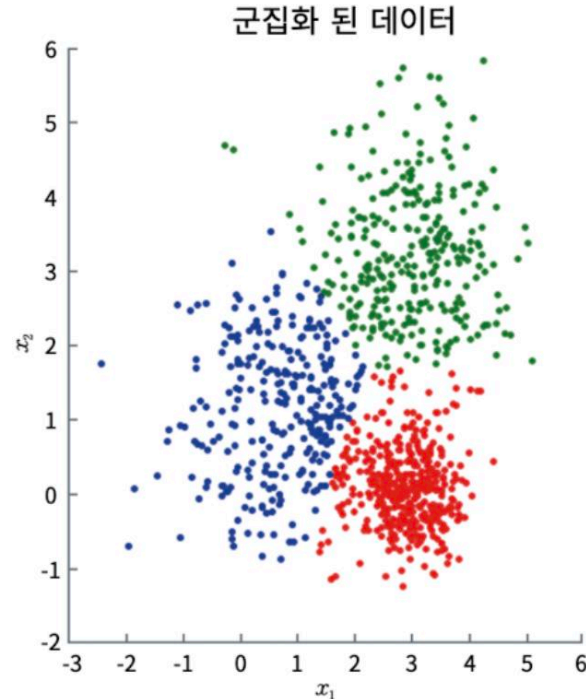
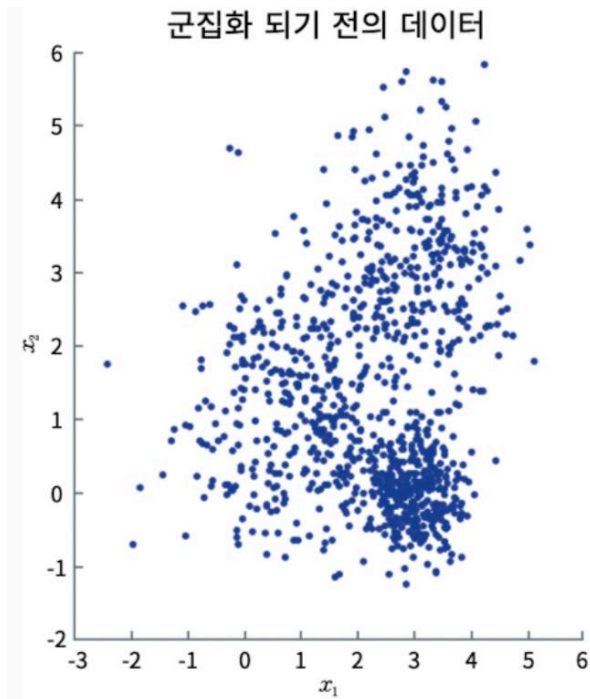


지도 학습: 회귀와 분류

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 경계를 찾는 과정
종류	<ul style="list-style-type: none"> • 선형 회귀 • 비선형 회귀 	<ul style="list-style-type: none"> • 이중 분류(Binary Classifier) • 다중 분류(Multi-class Classifier)
활용	<ul style="list-style-type: none"> • 날씨 예측 • 주가 예측 • 주택 가격 예측 • 중고 자동차 가격 예측 	<ul style="list-style-type: none"> • 스팸메일 식별 • 음성 인식 • 암세포 식별 • 손글씨 숫자 인식

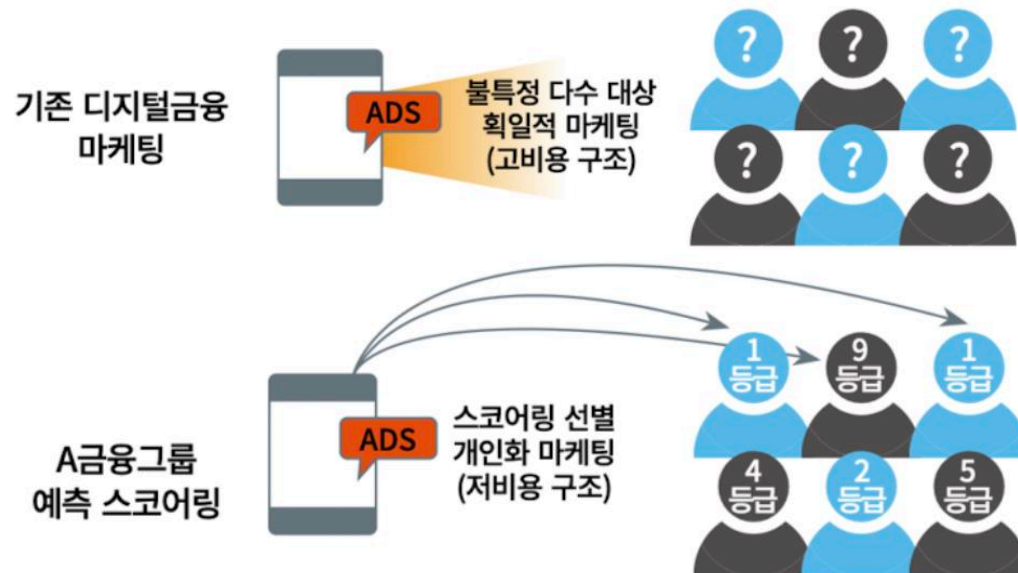
비지도 학습(Unsupervised Learning)

- 정답 레이블이 없는 데이터에서 비슷한 특징을 가진 데이터들을 그룹화하거나 숨겨진 패턴을 찾는 방법
 - 데이터 내에 숨어있는 관계(ex. 패턴)를 찾아 냄
- 군집화(Clustering)



비지도 학습(Unsupervised Learning)

- 정답 레이블이 없는 데이터에서 비슷한 특징을 가진 데이터들을 그룹화하거나 숨겨진 패턴을 찾는 방법
 - 데이터 내에 숨어있는 관계(ex. 패턴)를 찾아 냄
- 군집화(Clustering)
 - Example) 고객의 투자 성향이 비슷한 고객들을 같은 군집으로 묶고, 군집별 맞춤 광고를 제공

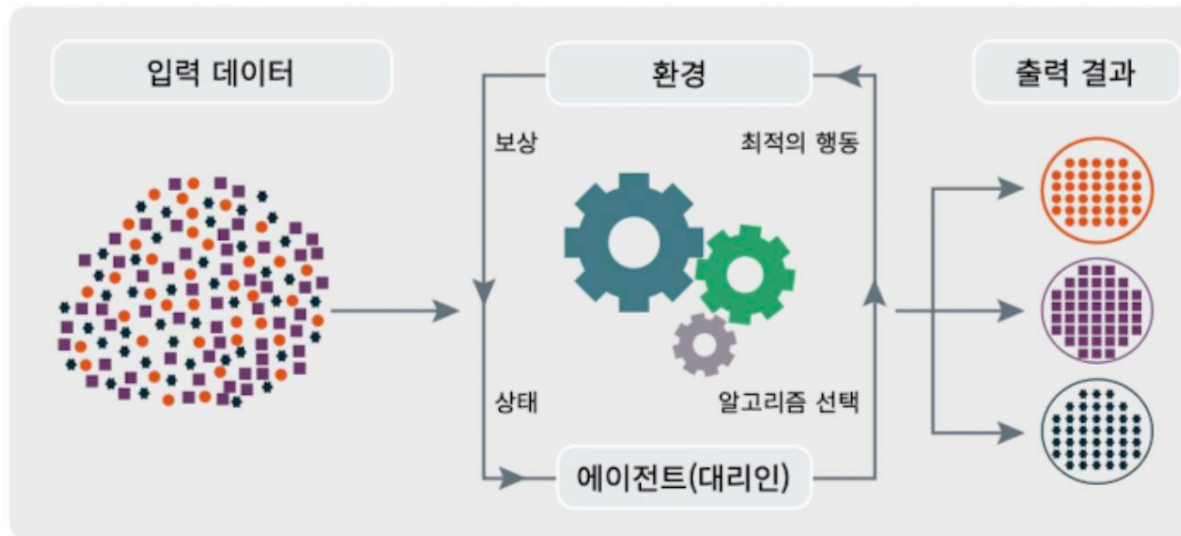


강화 학습(Reinforcement Learning)

- 주어진 현재 상태(state)에서 보상(reward)을 최대화 하도록 다음 행동(action)을 학습시키는 방법
- 구성
 - Agent: 학습하는 주체(ex. 로봇)
 - Environment: Agent가 행동하는 환경(ex. 로봇이 움직이는 공간)
 - Observation: Agent가 현재 환경에서 관찰하는 정보(ex. 카메라 영상)
 - Action: Agent가 선택하는 행동(ex. 물체 집기, 이동)
 - Reward: 행동 결과에 대한 점수(ex. 목적지 도착 +10, 충돌 -100)
- 비유
 - 아기가 걸음을 걷는 과정, 자전거 배우는 과정
- 특징
 - 정답이 주어지지 않고 시행착오(trial and error)를 통해 학습

강화 학습(Reinforcement Learning)

- Agent가 현재 환경(Environment) 관찰(observation)
 - Agent가 행동(action) 선택
 - 환경은 행동 결과 반영
 - 행동에 따라 보상(reward)
 - Agent는 정책(Policy) 업데이트하여 더 큰 보상을 얻도록 행동

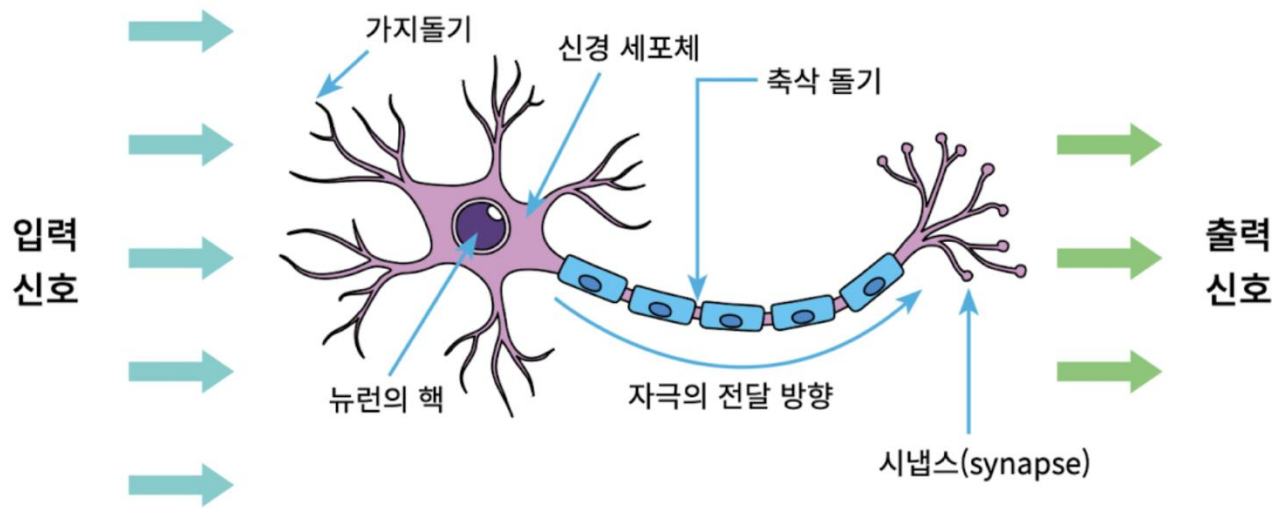


Artificial Neural Network and Deep Learning

뉴런과 퍼셉트론

- 뉴런(Neuron)

- 수상돌기(dendrite)에 들어온 여러 신호는 하나로 통합되어, 특정 임계 값을 초과하면 축삭돌기를 통해 다음 신경세포로 전달

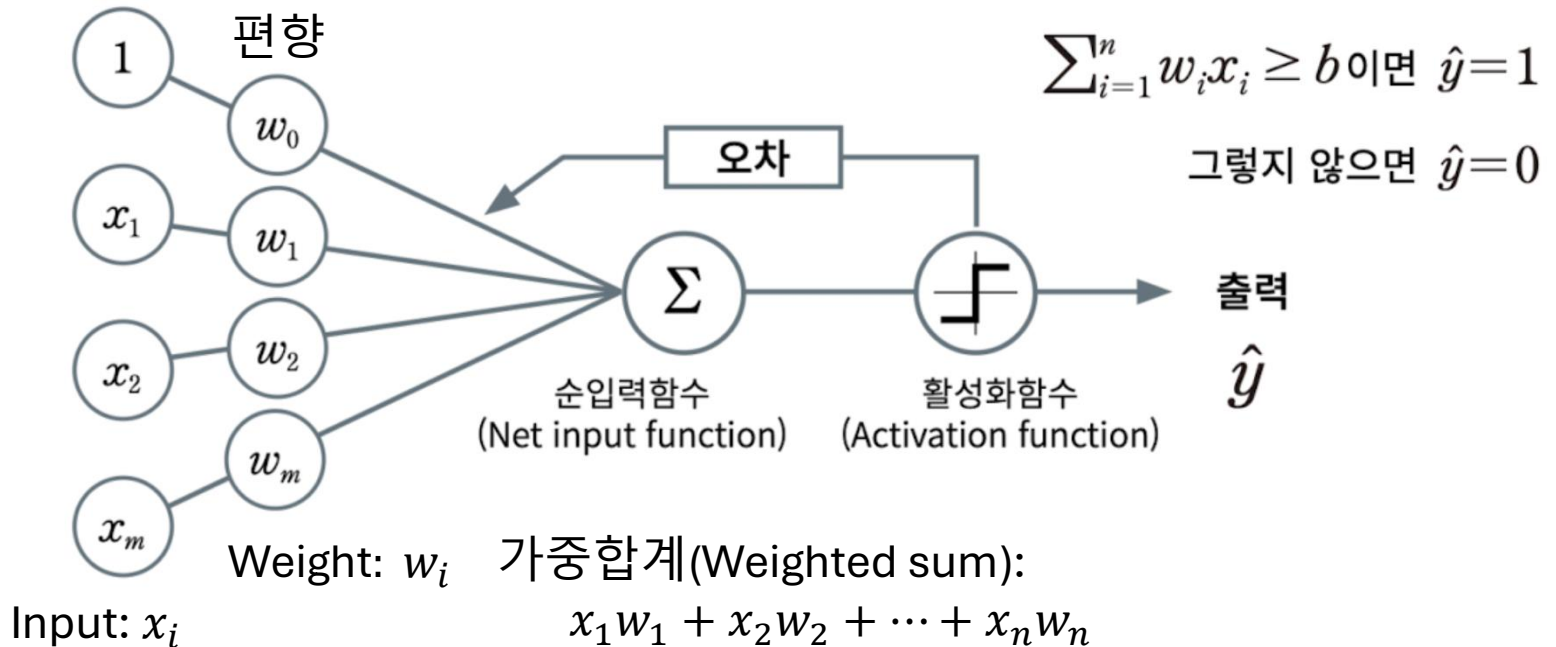


뉴런과 퍼셉트론

- 퍼셉트론(Perceptron)
 - 기본적인 인공 뉴런 모델
 - 다수 입력(input)
 - 하나의 출력(output)
 - 가중치(weights)
 - 입력의 강도 표현
 - 입력이 결과에 얼마나 중요한지 표현
 - 초기 가중치는 임의의 값으로 지정
 - 결과를 목표값과 비교, 오차(error)를 기반으로 다음 단계 가중치 조정

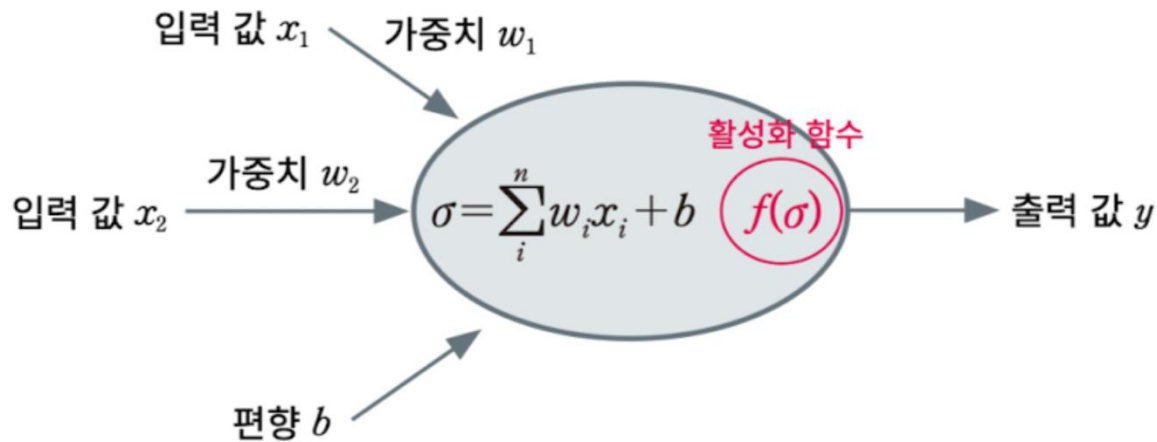
뉴런과 퍼셉트론

- 퍼셉트론(Perceptron)

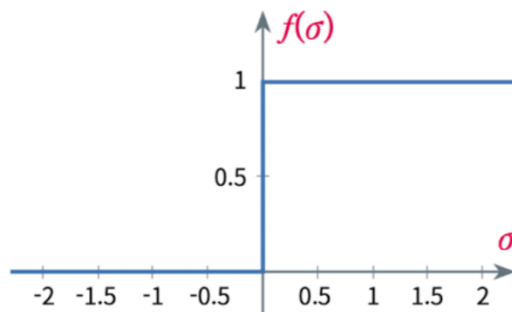


활성화 함수

- Activation function
 - 퍼셉트론에서 최종 값을 결정하는 함수



- Ex) Unit step function

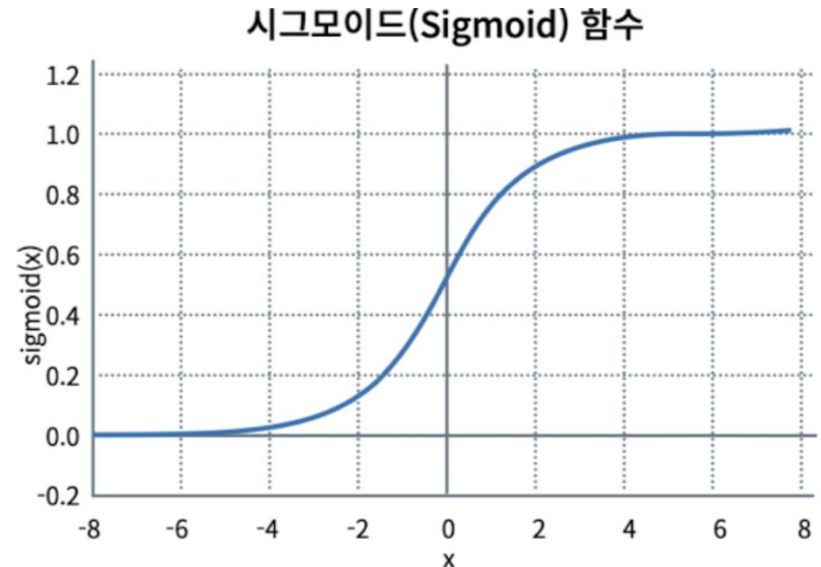
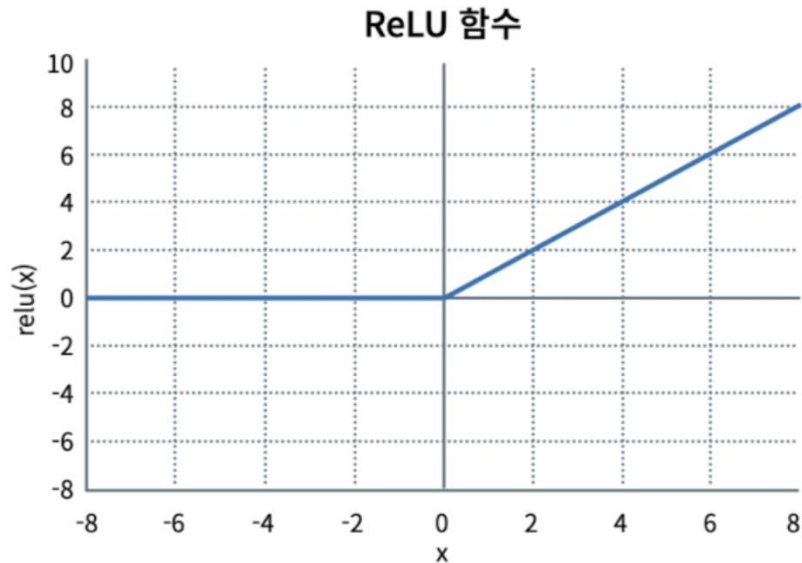


$$f(\sigma) = \begin{cases} \sigma \geq 0 \text{이면 } 1 \\ \sigma < 0 \text{이면 } 0 \end{cases}$$

활성화 함수

- 다양한 활성화 함수 존재

ReLU (Rectified Linear Unit)



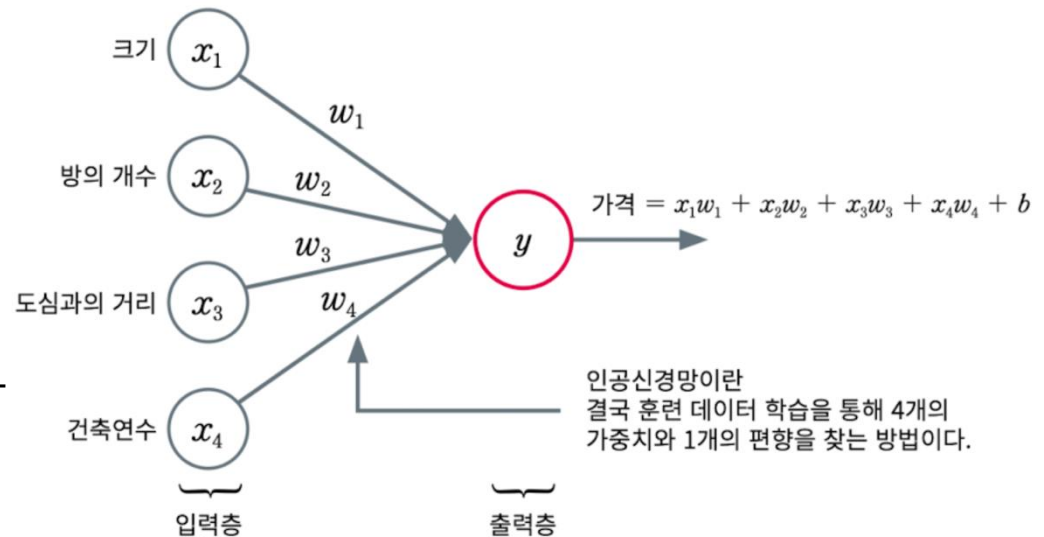
인공신경망

- Artificial Neural Network (ANN)

- 사람 뇌의 뉴런 구조를 참고해서 만든 머신러닝 모델
- 여러 인공 뉴런(퍼셉트론) 연결해서 데이터의 패턴 학습 및 예측

- Example) 아파트 가격 예측하는 문제

- 퍼셉트론 1개 사용
- 활성화 함수
 - 항등 함수
 - 입력 x 가중치 곱의 합 + 편향 = 아파트 가격



- 학습

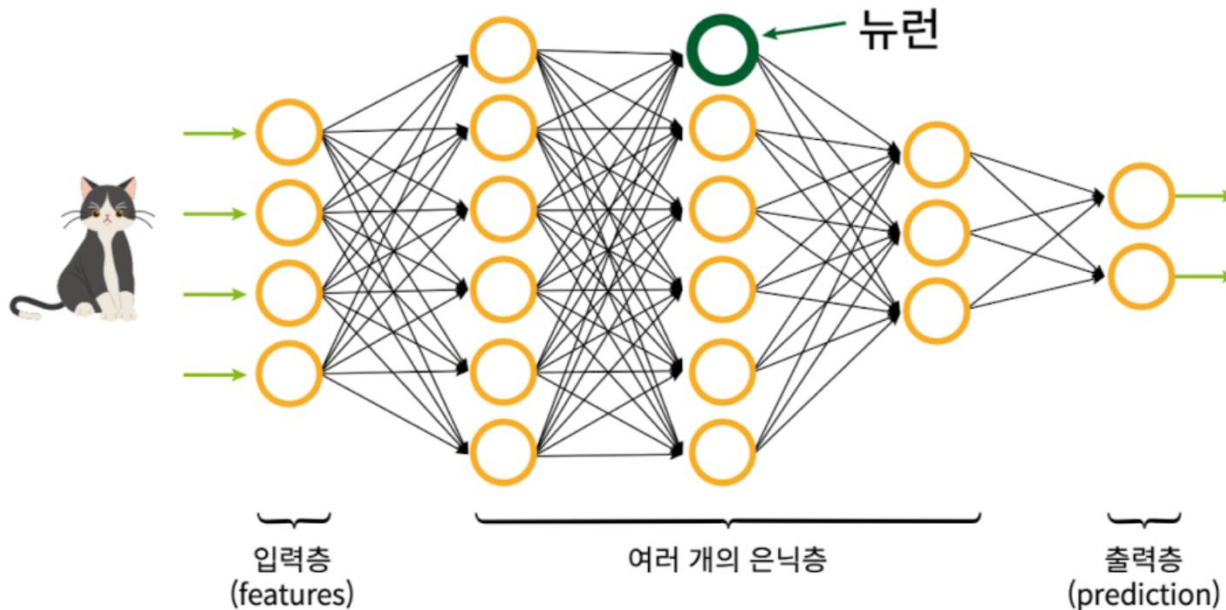
- 훈련 데이터로 가중치(w), 편향(b) 결정

- 예측

- 테스트 데이터로 아파트 가격 예측

심층신경망

- 심층신경망(Deep Neural Network, DNN)
 - 입력층과 출력층 사이에 여러 개의 은닉층(hidden layer)들로 이뤄진
인공신경망



- 딥러닝(Deep Learning)
 - 심층신경망을 이용해 데이터에서 복잡한 패턴을 학습하는 방법

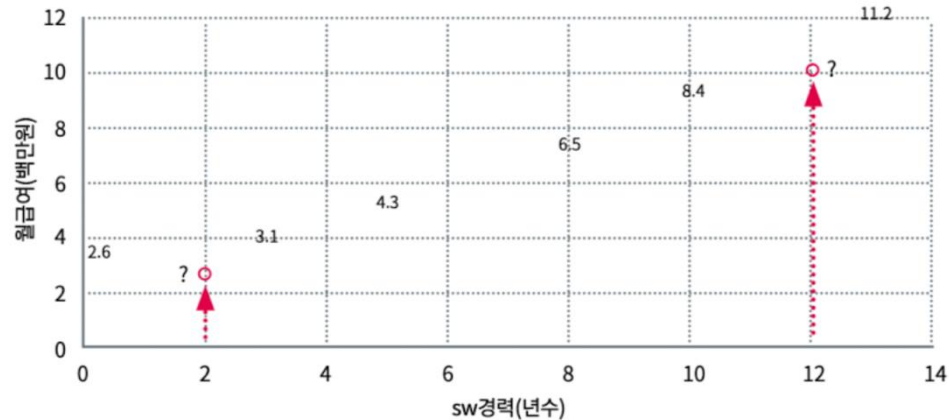
인공신경망의 학습

- Example) SW 경력에 따른 월 급여 예측
 - 6개의 데이터로 학습
 - 원하는 연수의 월 급여 예측

소프트웨어 개발 분야 급여 수준 자료

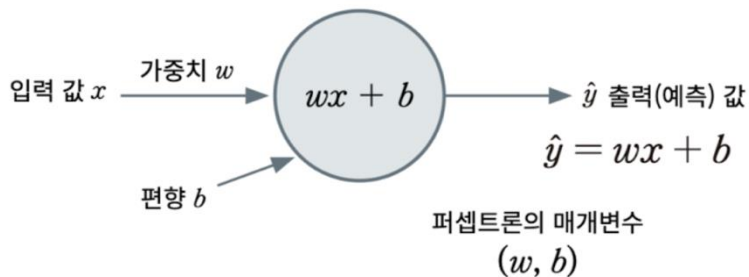
		신입사원	초급기능사	고급기능사	초급기술자	고급기술자	기술사
월급여(백만원)	y	2.6	3.1	4.3	6.5	8.4	11.2
sw경력(년수)	x	0	3	5	8	10	13

소프트웨어 개발 분야 급여 수준



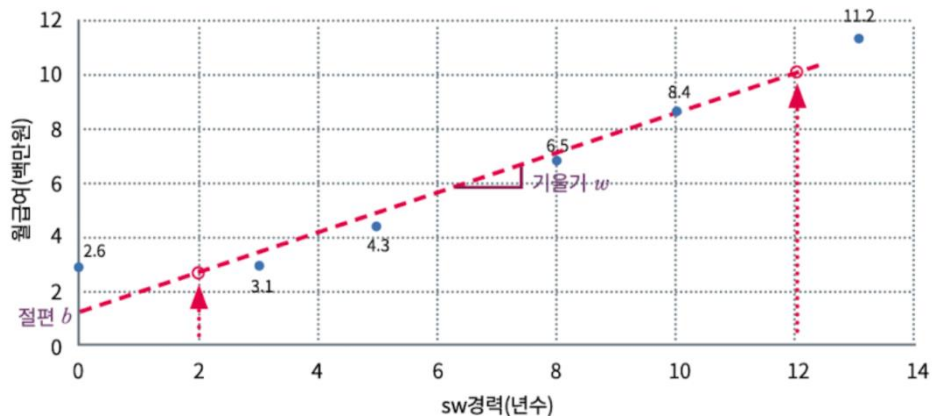
인공신경망의 학습

- Example) SW 경력에 따른 월 급여 예측
 - 6개의 데이터로 학습
 - 원하는 연수의 월 급여 예측
- 간단한 인공신경망 정의
 - 입력 x 가 퍼셉트론으로 들어가 출력값 \hat{y} 생성



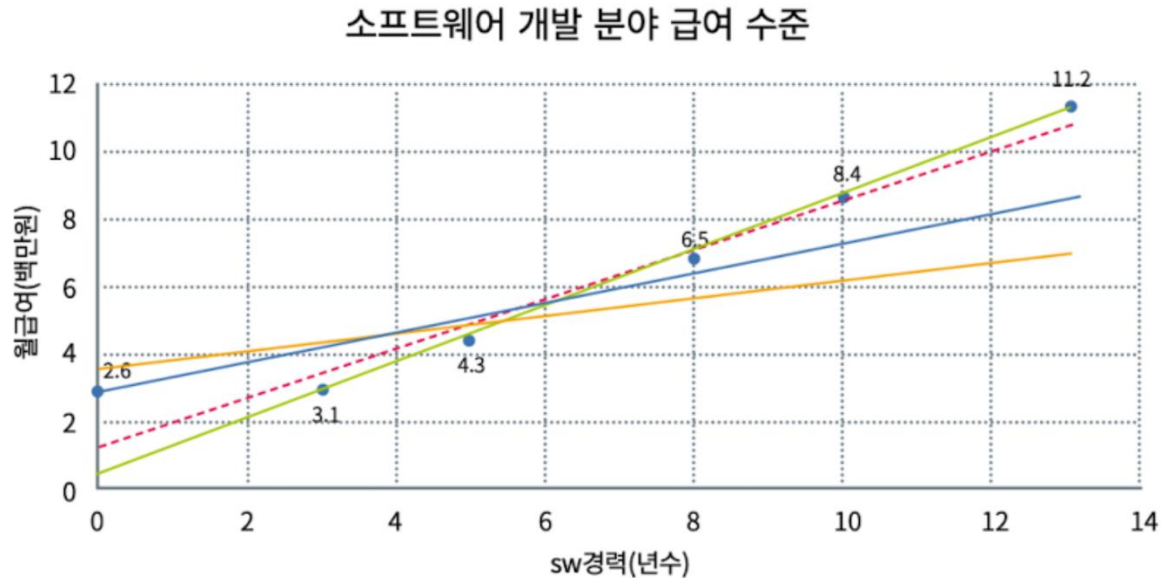
학습 통해 최적의 w 와 b 구함

소프트웨어 개발 분야 급여 수준



손실함수

- $\hat{y} = wx + b$
 - 어떻게 w 와 b 찾을 수 있을까?
 - 무엇이 가장 좋은 직선인지의 기준 필요



- 실제값 y 와 예측값 \hat{y} 차이(오류)를 최소화하는 직선

손실함수

- Loss Function

- 모델의 예측이 얼마나 틀렸는지를 나타내는 함수
- Ex) Mean Squared Error (MSE)
 - $Loss = (실제값 - 예측값)^2$ 의 평균
 - Loss가 작을수록 예측 정확

- 신경망 학습

- 가중치와 편향을 조정해서 loss를 최소화하는 과정

입력 데이터



Summary

- Artificial Intelligence
- Machine Learning
- Artificial Neural Network
- Deep Learning