

#1

군집화

: 비지도학습

: 데이터 집단 대표할 수 있는 **중심점**을 찾는 것으로 데이터 마이닝을 하는 방법

K-means

: K는 군집의 수

: 분류할 클러스터의 수를 미리 측정함

: 각 군집 중심의 초기값을 랜덤하게 정하기 때문에, 초기값 위치에 따라 원하는 결과가 나오지 않을 수도 있음

EM(Expectation-Maximization)

: 관측되지 않는 잠재변수에 의존하는 확률 모델

: 최대우도값을 갖는 매개변수를 찾는 반복적인 알고리즘

: 해가 수렴하거나 사용자가 정한 반복 수를 채우게 되면 학습이 완성됨

- **1단계: Initialization Step**
 - 매개변수 θ 를 임의의 값으로 초기화
- **2단계: Expectation Step (E-Step)**
 - 주어진 매개변수 값에 관한 잠재변수 z 값을 추정
- **3단계: Maximization Step (M-Step)**
 - 2단계에서 얻은 잠재변수 값을 이용하여 매개변수 θ 값을 다시 추정
- **4단계: Convergence Step**
 - 매개변수 θ 값과 잠재변수 z 값이 수렴할 때까지 2,3단계를 반복

DBSCAN

: 밀도 기반 데이터 클러스터링

: 밀집되어 있는 지역을 하나의 군집으로 정의

: 밀도가 높은 지역에 대한 정의

: 지정거리

: 데이터 개수

- ① 거리 척도, e (지정 거리), n (필요 최소 샘플 수) 설정을 통해 밀도 높은 지역 정의
- ② 밀도 높은 지역을 만족하는 core point를 찾고 그 지역을 군집으로 할당
- ③ 해당 밀도 높은 지역 안에 core point를 만족하는 데이터가 있다면 그 지역을 포함하여 군집 확장
- ④ 해당 밀도 높은 지역 안에 더 이상 core point를 정의할 수 없을 때까지 2~3 단계 반복
- ⑤ 어떤 군집에도 해당되지 않은 데이터 noise point로 정의

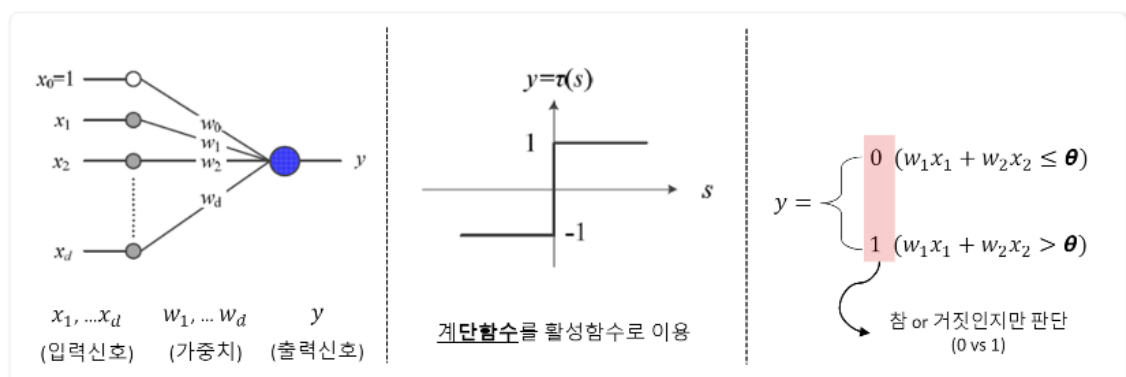
Neural Network

Neural Network

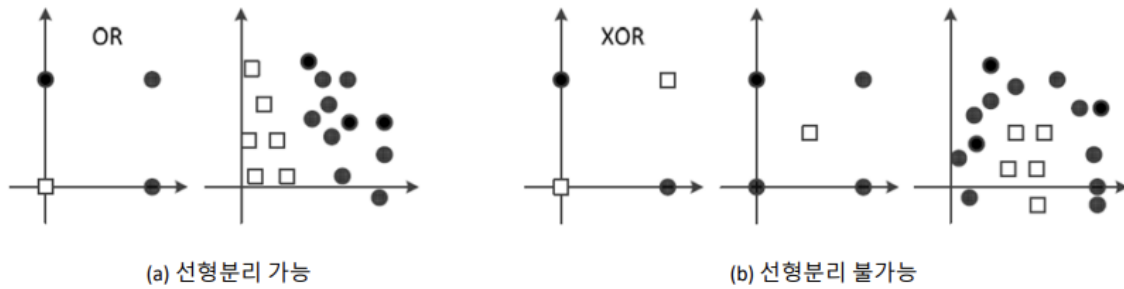
- 가장 오래됨
 - 현재 가장 다양한 형태를 가짐
- 퍼셉트론 → 다층 퍼셉트론 → 딥러닝

Perceptron

- 1세대 딥러닝
- 다수의 신호(n 개)를 입력으로 받아 특정한 연산을 거쳐 하나의 값을 출력하는 방식
- 입력층은 연산 $x \rightarrow$ 단일 층 구조로 간주
- 출력층은 한개의 노드
- 단점 : 간단한 xor 연산도 학습 못함



- 동작은 ppt 보기
- 한계 : 선형분리가 가능한 상황과 불가능한 상황



다중 Perceptron

- 2세대 딥러닝
- 입력과 출력 사이에 하나 이상의 은닉층을 추가해 학습
- Key Point
 - 은닉층의 존재
 - 시그모이드 활성화함수
 - 오류 역전파 알고리즘 사용
- 특징 공간 변환

Activation function

- 신경망이 사용하는 다양한 활성화함수

Name	Function	1th a derived function	Range
<i>Step</i>	$r(s) = \begin{cases} 1 & s \geq 0 \\ -1 & s < 0 \end{cases}$	$r'(s) = \begin{cases} 0 & s \neq 0 \\ \text{불가} & s = 0 \end{cases}$	-1과 1
<i>Sigmoid</i>	$r(s) = \frac{1}{1 + e^{-as}}$	$r'(s) = ar(s)(1 - r(s))$	(0~1)
<i>Tanh</i>	$r(s) = \frac{2}{1 + e^{-as}} - 1$	$r'(s) = \frac{a}{2}(1 - r(s)^2)$	(-1~1)
<i>ReLU</i>	$r(s) = \max(0, s)$	$r'(s) = \begin{cases} 0 & s < 0 \\ 1 & s > 0 \\ \text{불가} & s = 0 \end{cases}$	(0~ ∞)