

4장 딥러닝 시작

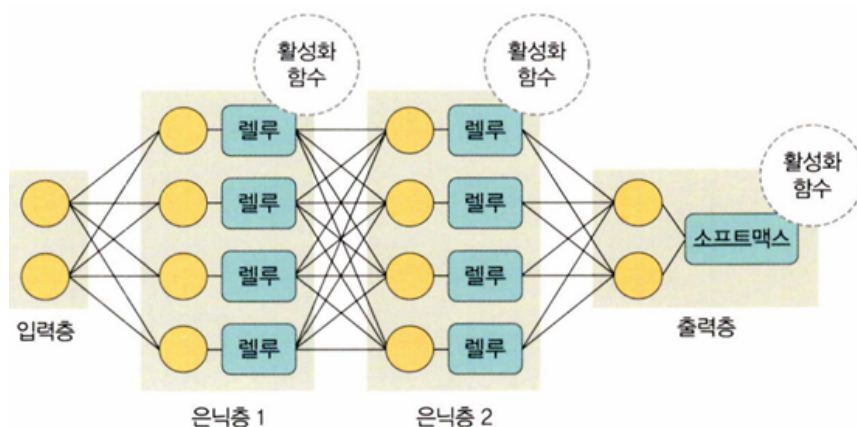
🕒 생성일	@2024년 9월 22일 오후 4:05
📅 주차	Week 2
☑ 완료여부	<input type="checkbox"/>

4.1 인공신경망의 한계와 딥러닝 출현

- 인공신경망의 기본 구조 → 퍼셉트론: 다수의 신호를 받아 하나의 신호를 출력 (1,0)
- AND연산 (모든 입력이 1일 때 작동). OR연산(입력에서 둘중 하나라도 1일 때 작동). XOR연산(두개 중 한개만 1일 때 작동)
- 이때 AND, OR 연산은 선형으로 분리되기 때문에 학습이 가능하지만 XOR의 경우 학습이 불가능함
- 따라서 입력, 출력 사이에 은닉층을 하나 이상 두어 비선형으로 분리되는 데이터에 대해서도 학습이 가능하도록 함 → DNN

4.2 딥러닝 구조

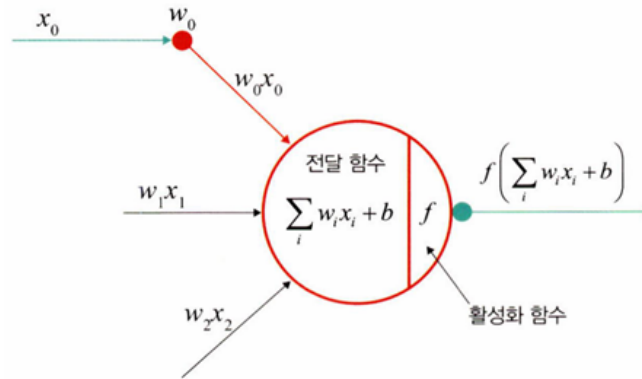
여러 층을 가진 인공신경망을 사용하여 학습을 수행하는 것



은닉층- 모든 입력 노드로부터 입력값을 받아 가중합을 계산하고 이를 활성화 함수를 통해 출력층으로

- 가중치: 입력값이 연산 결과에 미치는 영향력을 조절하는 요소로 w 가 0이거나 0과 가까운 값이라면 영향이 거의 없게 됨

- 가중합&전달함수: 각 노드에서 들어오는 신호에 가중치를 곱해서 다음 노드로 전달할 때 이 값을 모두 더한 합계
- 이를 활성화함수로 보내기 때문에 transfer function이라고도 함



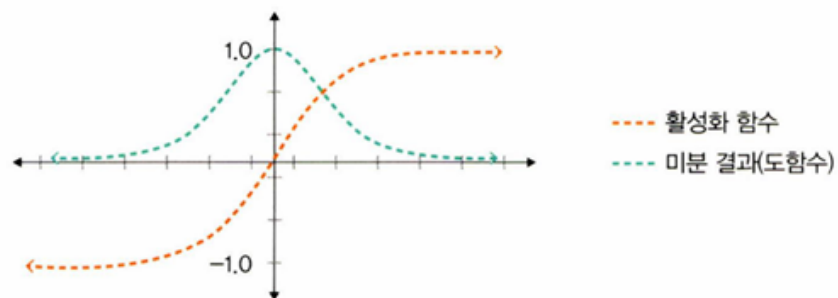
- 활성화 함수: 전달받은 값을 일정기준에 따라 출력시키는 비선형 함수
- 시그모이드 함수

$$f(x) = \frac{1}{1 + e^{-x}}$$

함수의 결과를 0-1사이 비선형 형태로 변형해줌

→ 하이퍼볼릭 탄젠트 함수

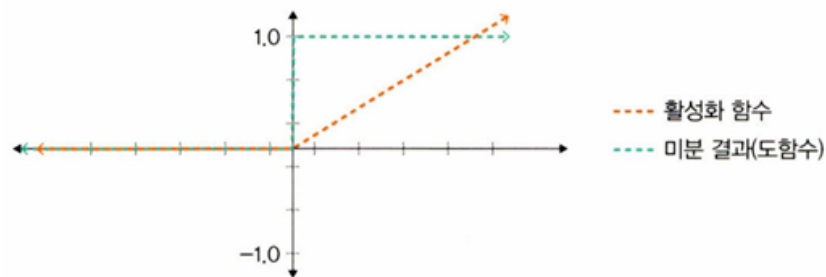
♥ 그림 4-9 하이퍼볼릭 탄젠트 활성화 함수와 미분 결과



-1~1 사이의 비선형 형태로 변형해줌. 하지만 시그모이드와 마찬가지로 기울기 소실 문제 발생함.

→ 렐루 함수

▼ 그림 4-10 렐루 활성화 함수와 미분 결과

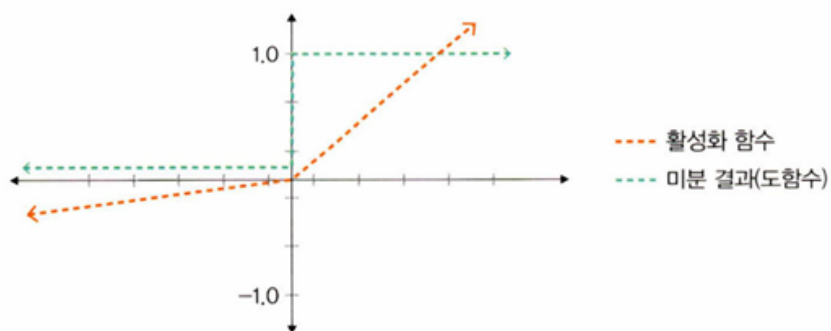


경사하강법에 영향을 주지 않아 학습속도가 빠르고 기울기 소멸 문제가 발생하지 않음

- hidden layer에서 주로 사용되며 위 함수 대비 학습속도가 6배 빠름
- 이때 음수값을 입력받으면 항상 0을 출력하기 때문에 학습 능력이 감소한다는 단점

→ Leaky ReLU 함수

▼ 그림 4-11 리키 렐루 활성화 함수와 미분 결과



입력값이 음수면 0이 아닌 0과 가까운 매우 작은 수를 반환함

→ Softmax 함수

- 입력값을 0-1사이에 출력하도록 정규화하여 출력값들의 합이 총 1이 되게 함
- 출력노드의 활성화함수로 주요 사용됨

$$y_k = \frac{\exp(a_k)}{\sum_{i=1}^n \exp(a_i)}$$

- 경사하강법은 학습률과 손실함수간 순간 기울기를 이용하여 가중치를 업데이트하는 방법 → 미분의 기울기를 이용하여 오차를 비교하고 최소화하는 방법으로 이동시키는 방법 → 이때 오차를 구하는 방법이 손실함수

- 손실함수: 학습을 통해 얻은 데이터의 추정치가 실제 데이터와 얼마나 차이가 나는지 평가하는 지표

→ Mean Squared Error: 실제-예측값을 제곱하여 평균낸것으로 주로 회귀에서 사용됨

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

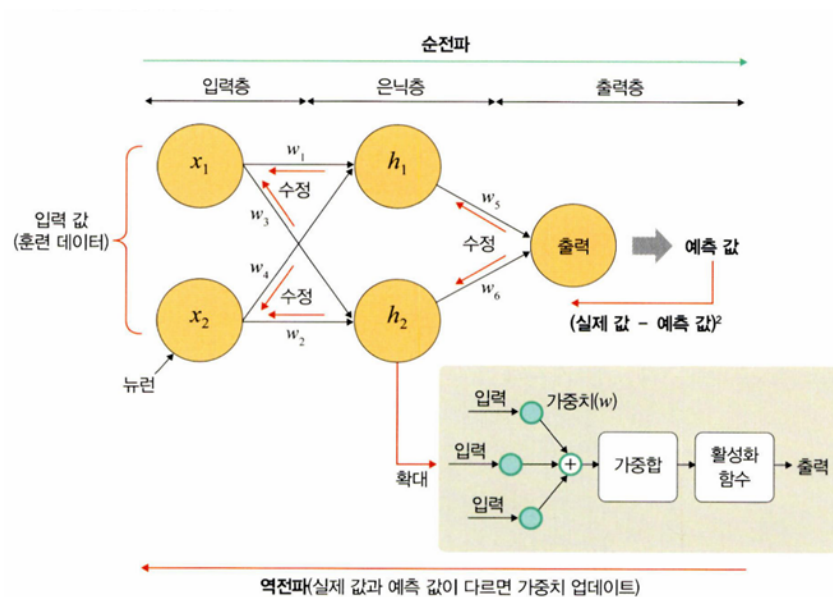
\hat{y}_i : 신경망의 출력(신경망이 추정한 값)
 y_i : 정답 레이블
 i : 데이터의 차원 개수

→ Cross Entropy Error: 분류에서 원핫인코딩 했을 때 사용할 수 있는 오차계산법으로 분류문제의 출력 함수는 보통 시그모이드 함수로 여기 포함된 e로 지역 최소점에 멈추는 걸 방지하고자 자연 로그를 취함

$$CrossEntropy = - \sum_{i=1}^n y_i \log \hat{y}_i$$

\hat{y}_i : 신경망의 출력(신경망이 추정한 값)
 y_i : 정답 레이블
 i : 데이터의 차원 개수

4.2.2 딥러닝 학습



순전파→

모든 뉴런이 이전 층의 뉴런에서 수신한 정보에 변환(가중합/활성화 함수)을 적용하여 다음 층의 뉴런으로 전송하는 방식 → 손실함수로 오차 추정 (이때 이상비용은 0),

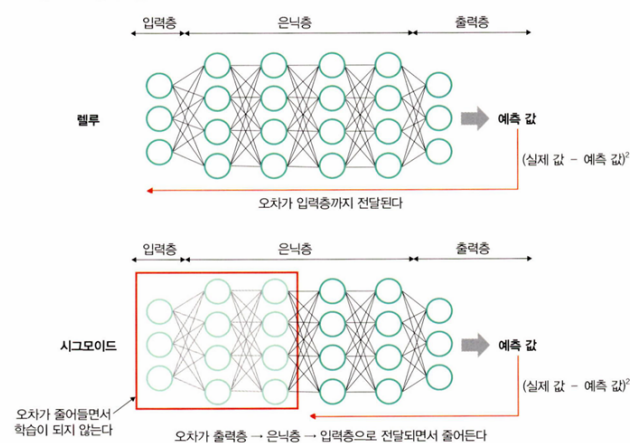
→ 오차가 계산되면 그 정보는 역으로 전파되어 가중치를 조정 → 역전파 → (예측-실제값)을 뉴런의 가중치로 미분한 후 기초 가중치에서 뺌

4.2.3 딥러닝의 문제점과 해결방안

딥러닝의 핵심은 활성화 함수가 적용된 여러개의 은닉층을 겹바하여 비선형 영역을 표시하는 것인데, 은닉층 개수가 많을 수록 데이터 분류가 잘 됨 → 하지만 이때 문제점

1. 과적합 문제 → dropout을 통해 랜덤하게 일부 노드를 학습에서 제외시킴
2. 기울기 소멸 문제: 출력 은닉층으로 전달되는 오차가 크게 줄어들어 학습이 되지 않는 현상 → ReLU 활성화 함수를 사용하여 해결

▼ 그림 4-16 기울기 소멸 문제



3. 성능이 나빠지는 문제 → 확률적 경사하강법과 미니배치 경사 하강법을 사용함

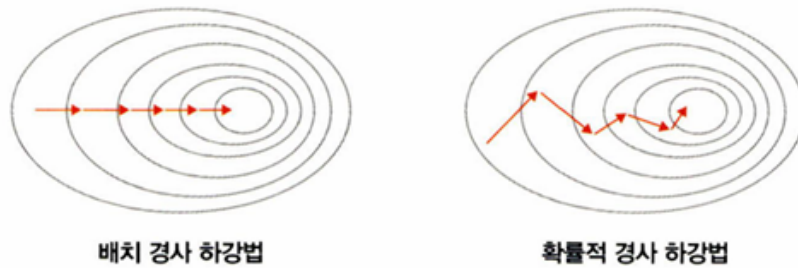
배치경사하강법 Batch Gradient Descent: 전체 데이터셋에 대한 오류를 구한 후 기울기를 한번만 계산하여 모델의 파라미터를 업데이트 하는 방법 (가중치 편미분)

$$W = W - a \nabla J(W, b)$$

(a : 학습률, J : 손실 함수)

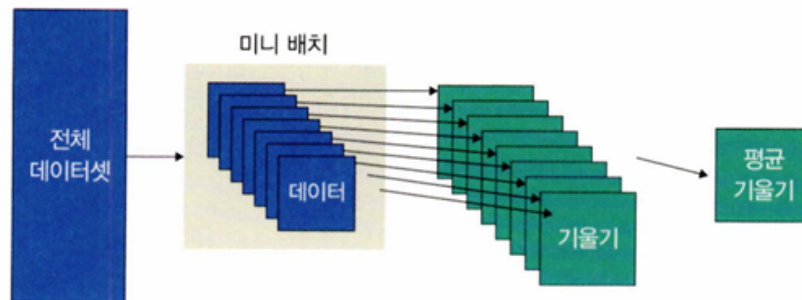
확률적경사하강법 Stochastic Gradient Descent: 임의로 선택한 데이터에 대해 기울기를 계산하는 방식으로 적은 데이터를 사용해 빠른 연산이 가능

▼ 그림 4-19 배치 경사 하강법과 확률적 경사 하강법



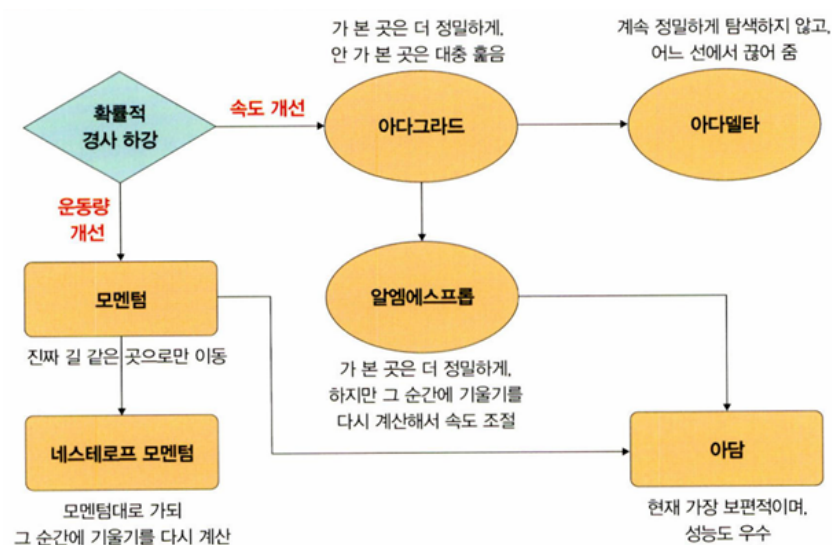
미니배치경사하강법 Mini-Batch Gradient Descent: 전체 데이터를 미니배치로 나누고 미니배치 한개마다 기울기를 구한 후 그것의 평균 기울기를 이용해 모델을 업데이트

▼ 그림 4-20 미니 배치 경사 하강법



: 확률적경사하강법에 비해 변경폭이 안정적이고 속도가 빠르기 때문에 주로 사용됨

옵티마이저: 확률적경사하강법의 파라미터 변경 폭이 불안정한 문제를 해결하기 위해 학습 속도와 운동량을 조정하는 옵티마이저



4.2.4 딥러닝을 사용할 때 이점

특징추출로 컴퓨터가 데이터에서 패턴이나 규칙을 찾아내려면 컴퓨터가 인지할 수 있는 데이터로 변환해주어야함, 도메인에 대한 지식이 필요했던 이전 머신러닝과 달리 은닉층을 깊게 쌓는 방식으로 특성추출알고리즘을 통합시킴. 따라서 빅데이터를 더욱 효율적으로 활용할 수 있게 됨

4.3 딥러닝 알고리즘

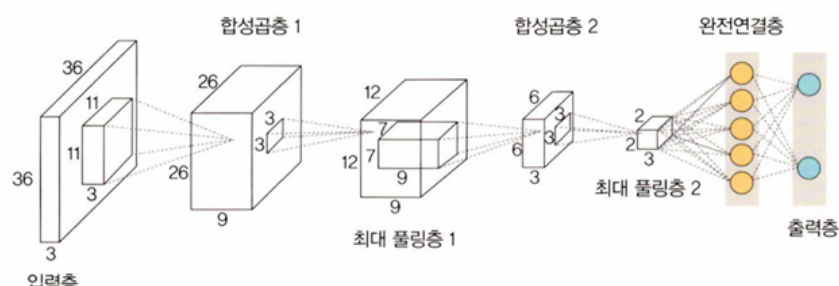
4.3.1 심층 신경망 DNN

입력층과 출력층 사이에 다수의 은닉층을 포함 → 연산량이 많고 기울기 소멸 문제 발생할 수 있음 → dropout, batchnorm, relu 등을 적용해야함

4.3.2 합성곱 신경망 CNN

합성곱층 Convolutional layer와 풀링층을 포함하는 이미지 처리 성능이 좋은 인공신경망 객체 탐지, 위치를 파악하는데 유용함

▼ 그림 4-25 합성곱 신경망



예시) LeNet, AlexNet, VGG, GoogleNet, ResNet

CNN만의 차별점

- 각 층의 입출력 형상 유지 → 이미지의 공간 정보 유지
- 복수의 필터를 사용해 이미지의 특징 추출
- 풀링층을 통해 추출한 이미지의 특징을 통합시킴
- 공유 파라미터 필터를 사용하기 때문에 일반적인 DNN에 비해 학습 파라미터가 매우 적음

4.3.3 순환 신경망 RNN

Recurrent Neural Network는 시계열 데이터와 같은 시간의 흐름에 따라 변화하는 데이터를 학습하기 위한 인공신경망

RNN만의 차별점

- 시간에 따라 내용이 변하므로 데이터는 동적이고 길이는 가변적임

기울기 소멸 문제가 제대로 학습되지 않는 문제가 있음 → 이를 해결하고자 LSTM이 주로 사용됨

4.3.4 제한된 볼츠만 머신

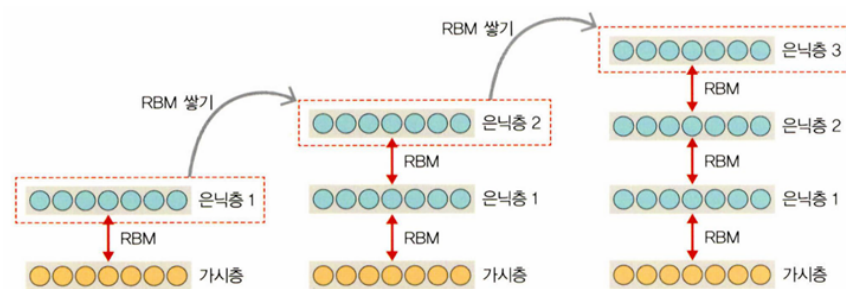
Boltzmann machine은 visible&hidden layer로 구성된 모델로 가시층끼리, 은닉층끼리는 연결이 없는게 Restricted Boltzmann Machine임

Restricted Boltzmann Machine의 차별점

- 기울기 소멸 문제를 해결하기 위해 pre-trained 용도로 활용이 가능함
- DBN 심층 신뢰 신경망의 요소로 활용됨

4.3.5 심층 신뢰 신경망

Deep Belief Network는 pre-trained된 제한된 볼츠만 머신을 층층이 쌓아올린 구조로 레이블이 없는 데이터에 대한 비지도 학습이 가능함.



DBN 특징 1) 순차적으로 DBN을 학습시켜 계층적 구조 생성함, 2) 위로 올라갈 수록 추상적인 특징을 추출함, 3) 이때 학습된 가중치를 다층퍼셉트론의 가중치 초기값으로 사용