

인공 신경망의 한계와 딥러닝 출현

퍼셉트론

- 프랭크 로젠블라트가 1957년 고안
- 신호를 입력으로 받아 정보를 앞으로 전달하는 원리
- 단층 퍼셉트론 : **XOR 게이트**에 대해 제대로 된 분류가 어려움 (사유: 비선형적으로 분리 되는 데이터)
 - -> **다층 퍼셉트론**을 고안 : 입력층과 출력층 사이에 하나 이상의 중간층(은닉층)을 둬.
 - -> 이와 같이 은닉층이 여러 개 있는 신경망 : **심층 신경망, 딥러닝**

딥러닝 구조

딥러닝 용어

층

- **입력층** : 데이터를 받아들이는 층
- **은닉층** : 모든 입력 노드부터 입력 값을 받아 가중합을 계산하고, 이 값을 활성화 함수에 적용하여 출력층에 전달하는 층
- **출력층** : 신경망의 최종 결과값이 포함된 층

가중치

- 입력 값이 연산 결과에 미치는 영향력을 조절하는 요소

가중합, 전달 함수

- 각 노드에서 들어오는 신호에 가중치를 곱해서 다음 노드로 전달될 때, 이 값을 모두 더한 합계를 지칭.
- 가중합을 활성화 함수로 보내기 때문에 전달 함수라고도 함.

활성화 함수

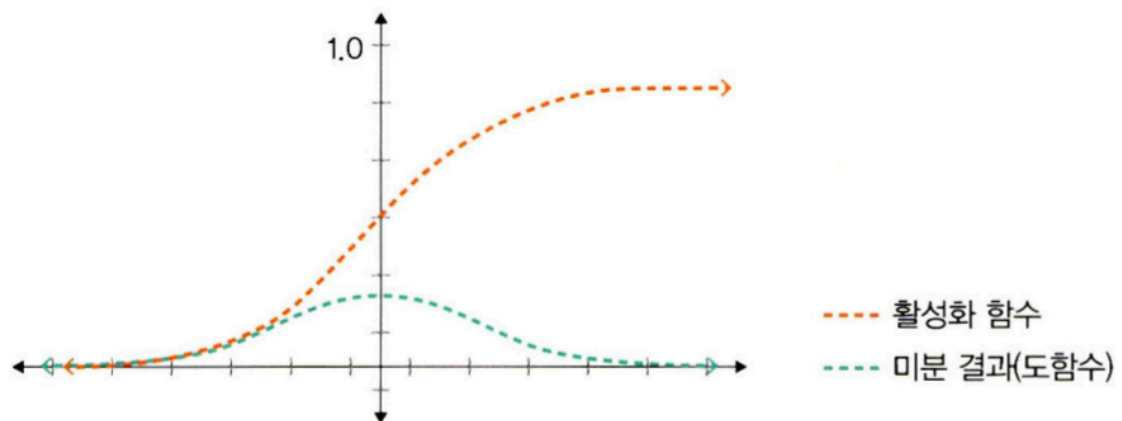
- 전달 함수에서 전달받은 값을 출력할 때 일정 기준에 따라 출력 값을 변화시키는 비선형 함수

시그모이드 함수

- 선형 함수의 결과를 0~1 사이에서 비선형 형태로 변형해 줌.
- 기울기 소멸 문제 (딥러닝 모델의 깊이가 깊어지면 기울기가 사라지는 문제)로 인해 딥러닝 모델에서는 잘 사용하지 않음.

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

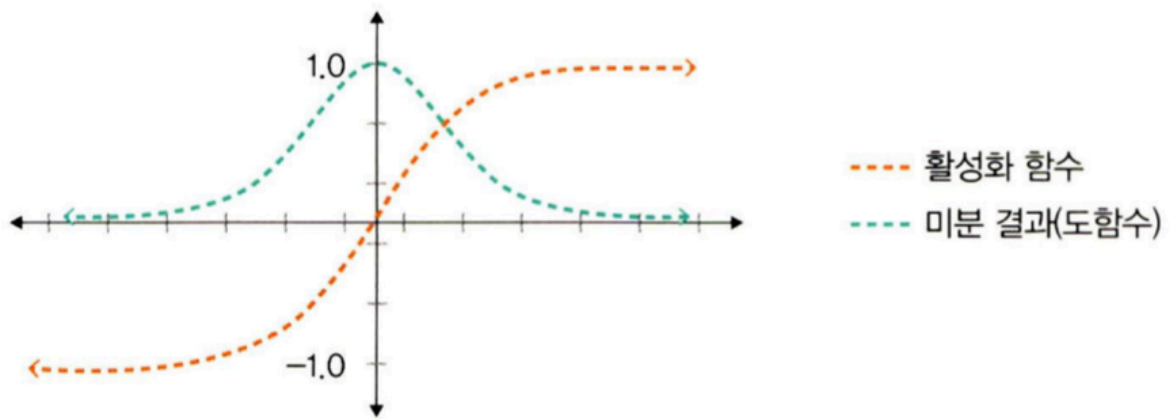
▼ 그림 4-8 시그모이드 활성화 함수와 미분 결과



하이퍼볼릭 탄젠트 함수

- 선형 함수의 결과를 -1~1 사이에서 비선형 형태로 변형해 줌.
- 시그모이드의 결과값의 평균이 0이 아닌 양수로 편향된 문제를 해결
- 기울기 소멸 문제는 해결 X

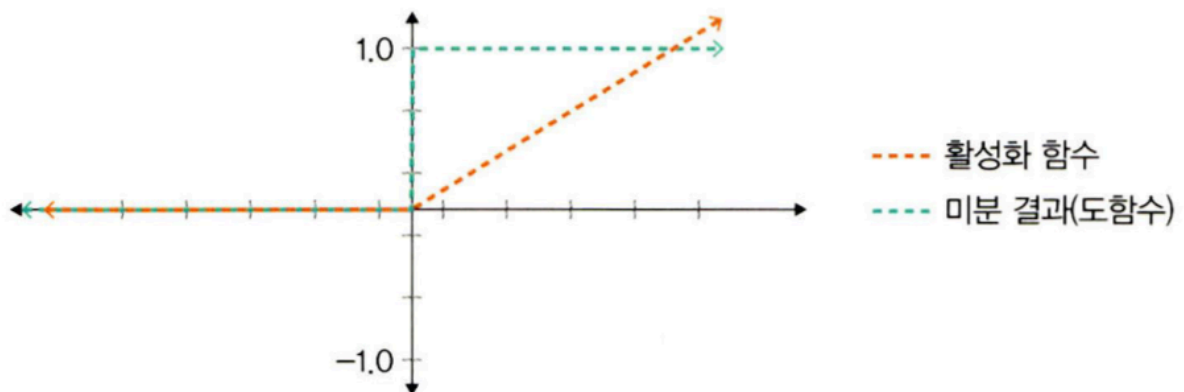
▼ 그림 4-9 하이퍼볼릭 탄젠트 활성화 함수와 미분 결과



렐루 함수

- 입력이 음수일 때는 0, 양수일 때는 입력값을 출력
- 학습 속도가 빠르고, 기울기 소멸 문제가 발생하지 않음.
- 음수값을 입력받을 시 항상 0을 출력하는 문제점 존재

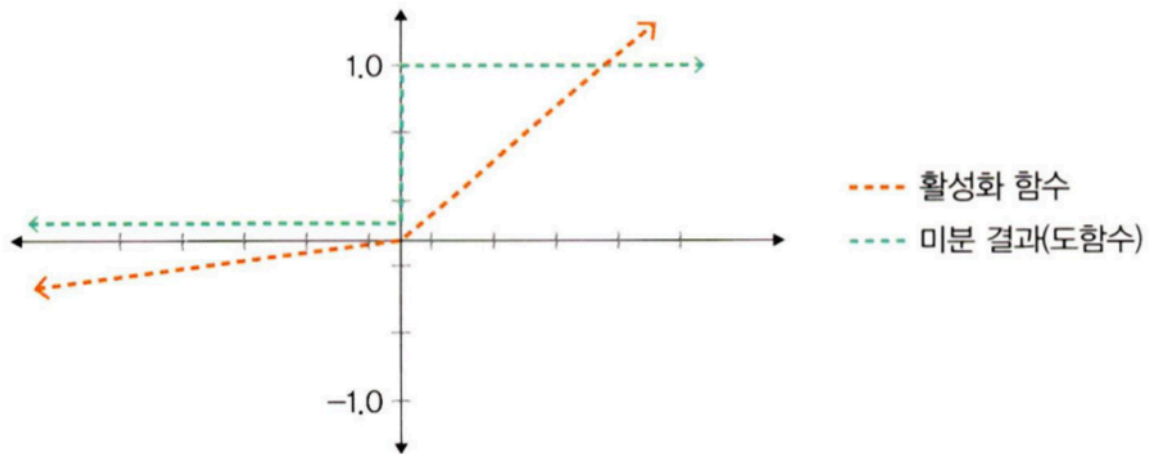
▼ 그림 4-10 렐루 활성화 함수와 미분 결과



리키 렐루 함수

- 입력값이 음수일 시 0.001처럼 매우 작은 수를 반환 -> 입력 값이 수렴되는 구간이 제거

▼ 그림 4-11 리키 렐루 활성화 함수와 미분 결과



소프트맥스 함수

- 입력 값을 0~1 사이에 출력되도록 정규화하여 출력 값들의 총합이 항상 1이 되도록 함.
- 출력 노드의 활성화 함수로 많이 사용

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

손실 함수

- **경사 하강법** : 학습률과 손실 함수의 순간 기울기를 이용하여 가중치를 업데이트 하는 방법
 - 미분의 기울기를 이용하여 오차를 비교하고 최소화하는 방향으로 이동시키는 방법
- 오차를 구하는 방법 : **손실 함수**

평균 제곱 오차

- 실제 값과 예측 값의 차이를 제곱하여 평균을 낸 것
- 값이 작을수록 예측력이 좋음
- 회귀에서 손실 함수로 주로 사용

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

$$\left(\begin{array}{l} \hat{y}_i: \text{신경망의 출력(신경망이 추정한 값)} \\ y_i: \text{정답 레이블} \\ i: \text{데이터의 차원 개수} \end{array} \right)$$

•

크로스 엔트로피 오차

- 분류 문제에서 원-핫 인코딩 했을 때만 사용할 수 있는 오차 계산법

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

$$\left(\begin{array}{l} \hat{y}_i: \text{신경망의 출력(신경망이 추정한 값)} \\ y_i: \text{정답 레이블} \\ i: \text{데이터의 차원 개수} \end{array} \right)$$

•

딥러닝 학습

- **순전파**
 - 네트워크에 훈련 데이터가 들어올 때 발생
 - 데이터를 기반으로 예측 값을 계산하기 위해 전체 신경망을 교차해 지나감.
 - 모든 뉴런이 이전 층의 뉴런에서 수신한 정보에 변환을 적용하여 다음 층의 뉴런으로 전송하는 방식
- **역전파**
 - 순전파 이후 손실 함수로 네트워크와 예측 값의 실제 값의 차이를 추정
 - 해당 정보는 역으로 전파 -> 역전파
 - 뉴런이 원래 출력에 기여한 상대적 기여도에 따라 뉴런의 값이 달라짐

딥러닝의 문제점과 해결 방안

- 은닉층이 많을 수록 다음과 같은 문제점이 생김.
- **과적합**
 - 훈련 데이터를 과하게 학습해서 발생
 - 실제 데이터에 대한 오차가 증가하는 현상
 - **드롭아웃**을 통해 해결 : 학습 과정 중 일부 노드들을 학습에서 제외.
- **기울기 소멸 문제 발생**
 - 출력층에서 은닉층으로 전달되는 오차가 크게 줄어들어 학습이 되지 않는 현상
 - 렐루 활성화 함수를 사용함으로써 해결
- **성능이 나빠지는 문제 발생**
 - 경사 하강법의 과정 중 성능이 나빠지는 문제가 발생.
 - **배치 경사 하강법**
 - 전체 데이터셋에 대한 오류를 구한 후 기울기를 한 번만 계산하여 모델의 파라미터를 업데이트 하는 방법
 - **확률적 경사 하강법**
 - 임의로 선택한 데이터에 대해 기울기를 계산하는 방법
 - 빠른 계산이 가능
 - **미니 배치 경사 하강법**
 - 전체 데이터셋을 미니 배치 여러 개로 나누고, 미니 배치 한 개마다 기울기를 구한 후 그것의 평균 기울기를 이용하여 모델을 업데이트해서 학습
 - 전체 데이터를 계산하는 것보다 빠르며, 확률적 경사 하강법보다 안정적

딥러닝을 사용할 때 이점

특성 추출

- **특성 추출** : 데이터별로 어떤 특징을 가지고 있는지 찾아내고, 그것을 토대로 데이터를 벡터로 변환하는 작업
- 딥러닝 : 은닉층을 깊게 쌓는 방식으로 파라미터를 늘린 모델 구조 덕분에 특성 추출 과정을 알고리즘에 통합.

빅데이터의 효율적 활용

- 빅데이터를 통해 더 높은 성능으로 특성 추출이 가능.

딥러닝 알고리즘

심층 신경망(DNN)

- 입력층과 출력층 사이에 다수의 은닉층을 포함하는 인공 신경망
- **장점**
 - 다양한 비선형적 관계를 학습 가능
- **단점**
 - 학습을 위한 연산량이 많음
 - 기울기 소멸 문제 등이 발생할 수 있음

합성곱 신경망

- 합성곱층, 풀링층을 포함하는 이미지 처리 성능이 좋은 인공 신경망 알고리즘
- 이미지에서 객체, 얼굴, 장면을 인식하기 위해 패턴을 찾는 데 특히 유용
- **차별성**
 - 각 층의 입출력 형상을 유지
 - 이미지의 공간 정보를 유지하면서 인접 이미지와 차이가 있는 특징을 효과적으로 인식
 - 복수 필터로 이미지의 특징을 추출하고 학습
 - 추출한 이미지의 특징을 모으고 강화하는 풀링층의 존재
 - 필터를 공유 파라미터로 사용하기 때문에 일반 인공 신경망과 비교하여 학습 파라미터가 매우 적음

순환 신경망

- 시계열 데이터를 학습하기 위한 인공 신경망
- 자기 자신을 참조하면서 학습, 현재 결과가 이전 결과와 연관이 있음.
- **특징**
 - 시간성을 가진 데이터가 많음
 - 시간성 정보를 이용하여 데이터의 특징을 잘 다룸
 - 시간에 따라 내용이 변하므로 데이터는 동적이고, 길이가 가변적
 - 매우 긴 데이터를 처리하는 연구가 활발히 진행되고 있음
- 기울기 소멸 문제 존재 -> LSTM이 많이 사용됨.
- 자연어 처리 분야에서 주로 사용.

제한된 볼츠만 머신

- **볼츠만 머신** : 가시층과 은닉층으로 구성된 모델
- 가시층과 가시층, 은닉층과 은닉층 사이에 연결이 없는 모델
- **특징**
 - 차원 감소, 분류, 선형 회귀 분석, 협업 필터링, 특성 값 학습, 주제 모델링에 사용
 - 기울기 소멸 문제를 해결하기 위해 사전 학습 용도로 활용 가능
 - 심층 신뢰 신경망의 요소로 활용

심층 신뢰 신경망

- 제한된 볼츠만 머신을 여러 층으로 쌓은 형태로 연결된 신경망
- 레이블이 없는 데이터에 대한 비지도 학습이 가능
- **학습 과정**
 1. 가시층과 은닉층 1에 제한된 볼츠만 머신을 사전 훈련
 2. 첫 번째 층 입력 데이터와 파라미터를 고정하여 두 번째 층 제한된 볼츠만 머신을 사전 훈련
 3. 원하는 층 개수만큼 제한된 볼츠만 머신을 쌓아 올려 전체 DBN을 완성
- **특징**
 - 순차적으로 심층 신뢰 신경망을 학습시켜 가면서 계층적 구조를 생성
 - 비지도 학습으로 학습
 - 위로 올라갈수록 추상적 특성을 추출
 - 학습된 가중치를 다층 퍼셉트론의 가중치 초기값으로 사용