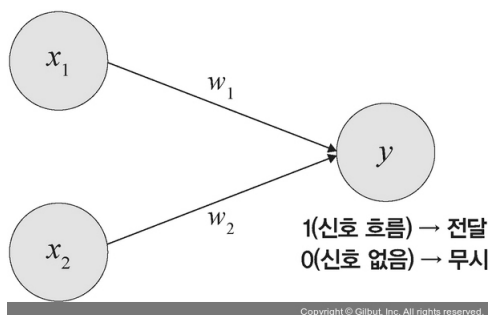


4. 딥러닝 시작

4.1 인공 신경망의 한계와 딥러닝 출현

** 퍼셉트론

- 오늘날 신경망(딥러닝)의 기원이 되는 알고리즘
- 다수의 신호(흐름이 있는)를 입력으로 받아 하나의 신호를 출력
- 이 신호를 입력으로 받아 '흐른다/안흐른다'는 정보를 앞으로 전달하는 원리



1. AND 게이트

- 모든 입력이 1일 때 작동
- 입력 중 어떤 하나라도 0을 갖는다면 작동을 멈춤

x1	x2	y
0	0	0
1	0	0
0	1	0
1	1	1

2. OR 게이트

- 입력에서 둘 중 하나만 1이거나 둘 다 1일 때 작동
- 입력 모두가 0을 갖는 경우를 제외한 나머지가 모두 1

x1	x2	y
0	0	0
1	0	1
0	1	1
1	1	1

3. XOR 게이트

- 입력 두 개 중 한개만 1일 때 작동하는 논리 연산

x1	x2	y
0	0	0
1	0	1
0	1	1
1	1	0



데이터가 비선형적으로 분리되기 때문에 제대로 된 분류가 어려움

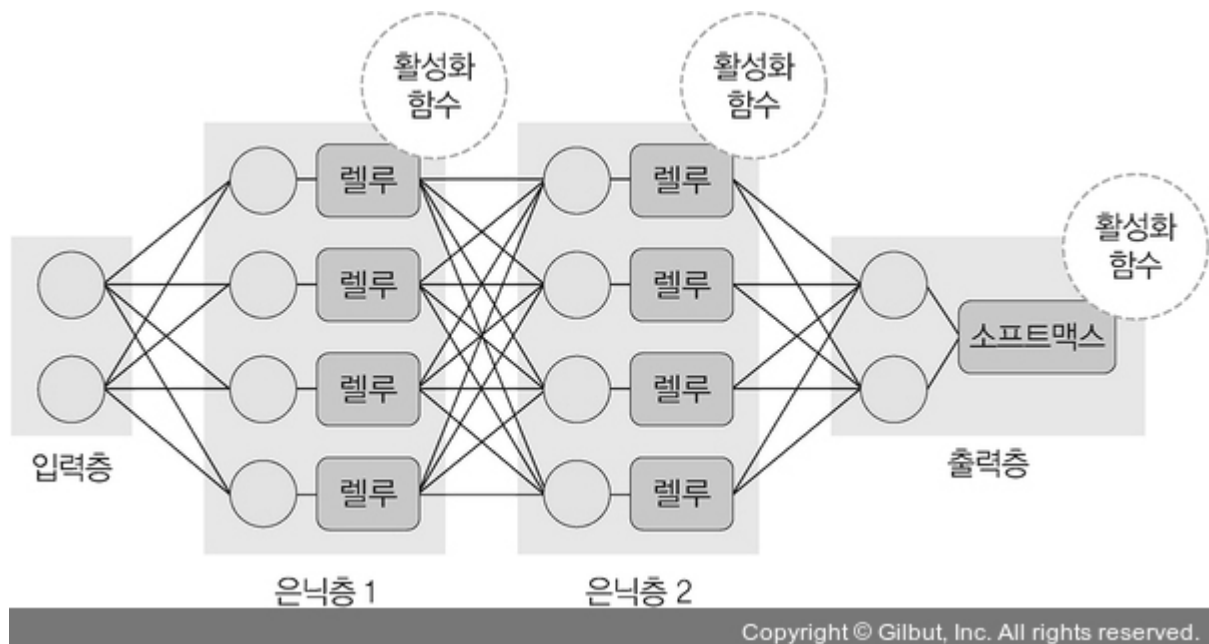
→ 이를 극복하는 방안으로 입력층과 출력층 사이에 하나 이상의 중간층(은닉층)을 두어 비선형적으로 분리되는 데이터에 대해서도 학습이 가능하도록 **다층 퍼셉트론(multi-layer perceptron)**을 고안

이때 입력층과 출력층 사이에 은닉층이 여러 개 있는 신경망을 **심층 신경망(Deep Neural Network, DNN)**이라고 하며, 심층 신경망을 다른 이름으로 **딥러닝**이라고 함

4.2 딥러닝 구조

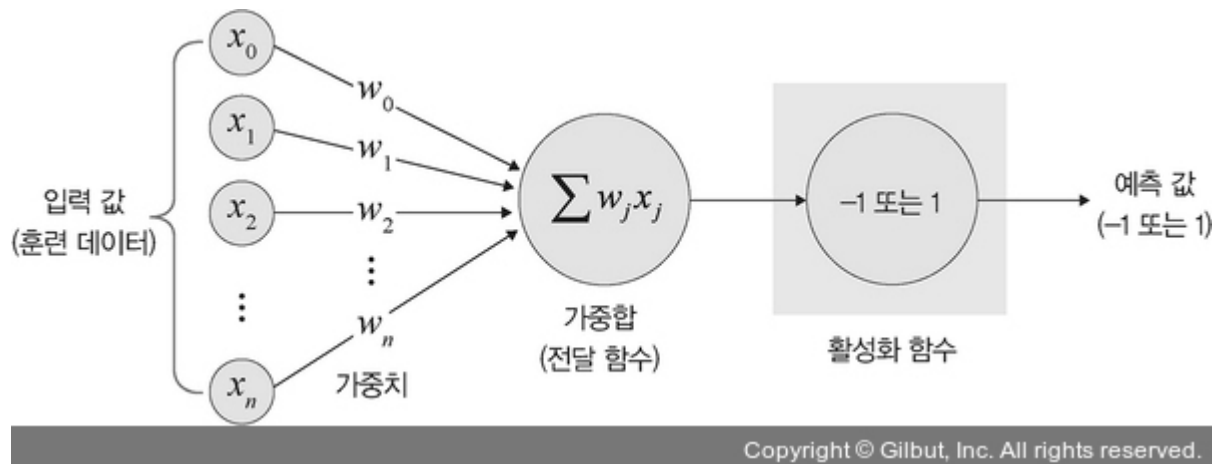
딥러닝 : 여러 층을 가진 인공 신경망을 사용하여 학습을 수행하는 것

4.2.1 딥러닝 용어

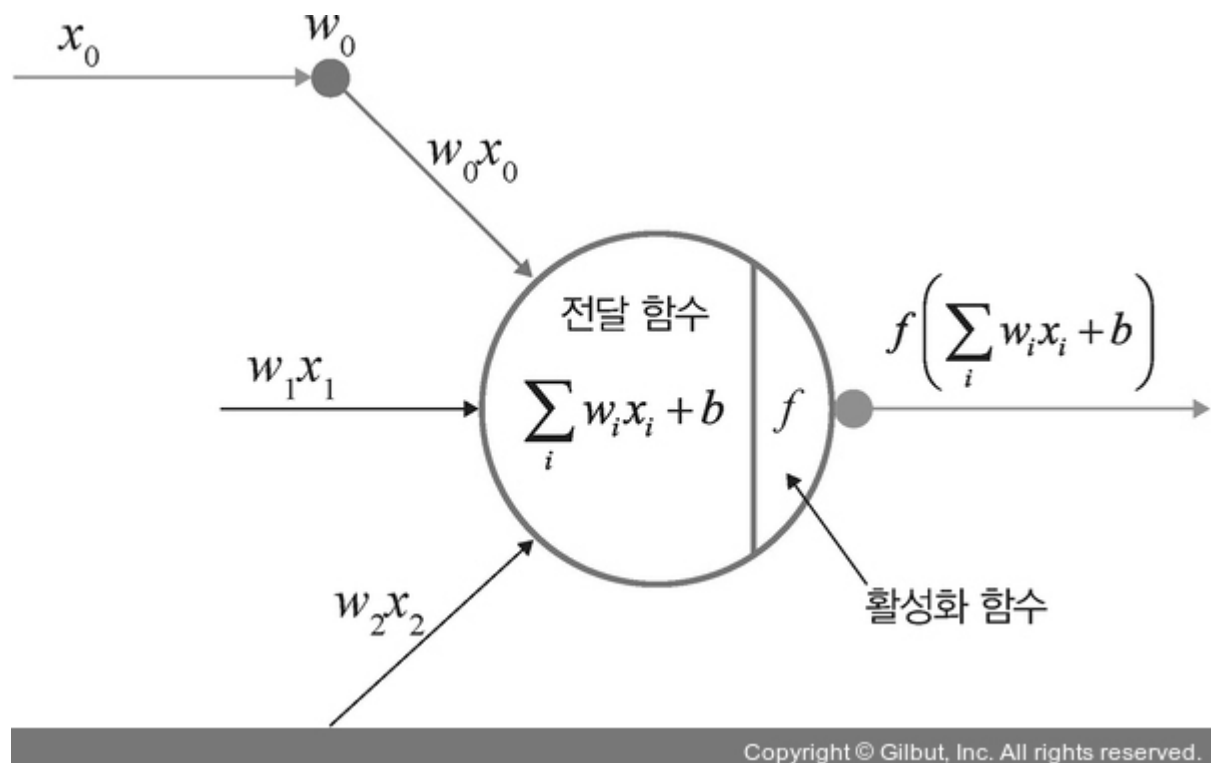


구분	구성 요소	설명
층	입력층(input layer)	데이터를 받아들이는 층
	은닉층(hidden layer)	모든 입력 노드부터 입력 값을 받아 가중합을 계산하고, 이 값을 활성화 함수에 적용하여 출력층에 전달하는 층
	출력층(output layer)	신경망의 최종 결괏값이 포함된 층
가중치(weight)		노드와 노드 간 연결 강도
바이어스(bias)		가중합에 더해 주는 상수로, 하나의 뉴런에서 활성화 함수를 거쳐 최종적으로 출력되는 값을 조절하는 역할을 함
가중합(weighted sum), 전달 함수		가중치와 신호의 곱을 합한 것
함수	활성화 함수(activation function)	신호를 입력받아 이를 적절히 처리하여 출력해 주는 함수
	손실 함수(loss function)	가중치 학습을 위해 출력 함수의 결과와 실제 값 간의 오차를 측정하는 함수

가중치 : 입력 값이 연산 결과에 미치는 영향력을 조절하는 요소



가중합 또는 전달 함수: 각 노드에서 들어오는 신호에 가중치를 곱해 다음 노드로 전달, 이 값을 모두 더한 합계



가중합 공식 :

$$\sum_i w_i x_i + b$$

(w : 가중치, b : 바이어스)

Copyright © Gilbut, Inc. All rights reserved.

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

- ex) 시그모이드, 하이퍼볼릭 탄젠트, 렐루

손실 함수: 오차를 구하는 방법, 학습을 통해 얻은 데이터의 추정치가 실제 데이터와 얼마나 차이가 나는지 평가하는 지표이며 0에 가까울수록 완벽하게 추정할 수 있음.

- ex) 평균 제곱 오차(MSE), 크로스 엔트로피 오차(CEE)



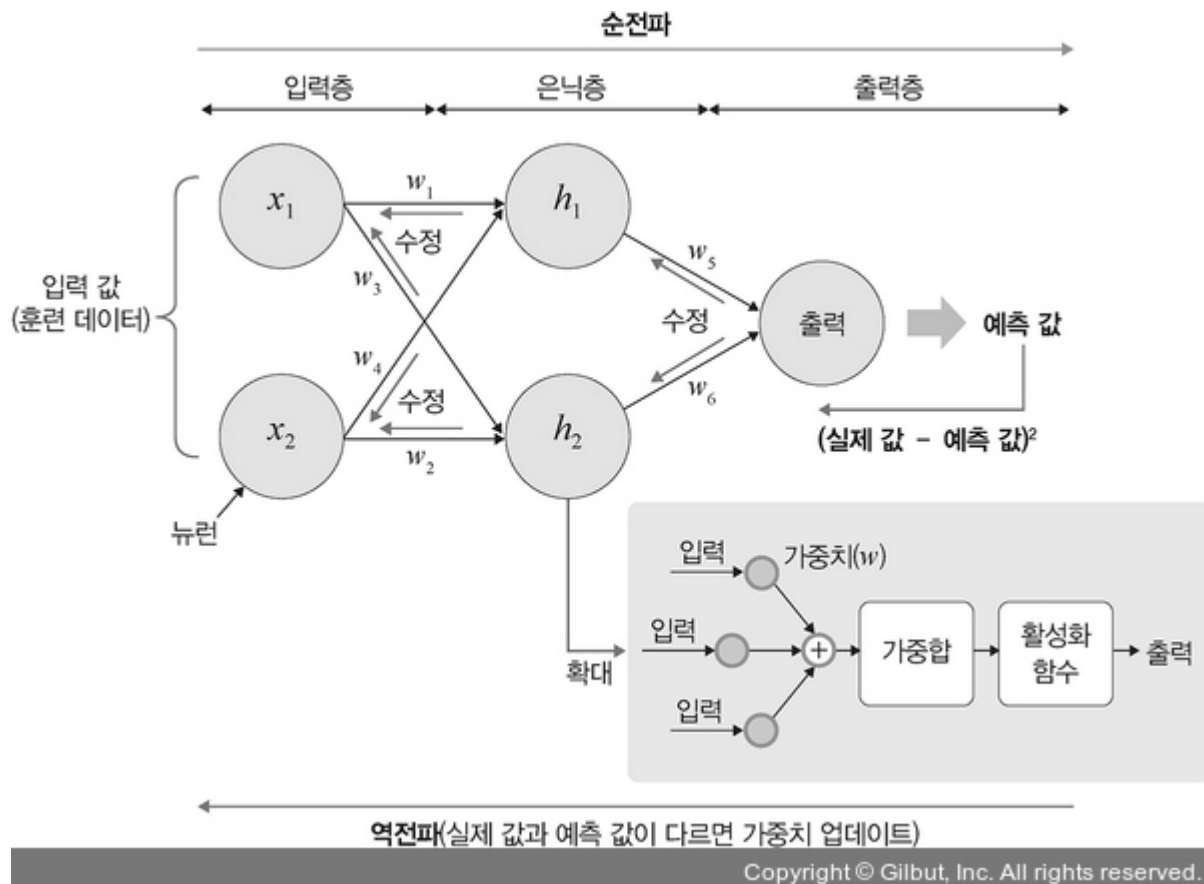
1. 평균 제곱 오차

- 실제 값과 예측 값의 차이를 제곱하여 평균을 낸 것
- 실제, 예측 값의 차이가 클수록 평균 제곱 오차의 값도 커짐
- 값이 작을수록 예측력이 좋다는 것
- 회귀의 손실 함수로 주로 사용

2. 크로스 엔트로피 오차

- 분류에서 원-핫 인코딩했을 때에만 사용할 수 있는 오차 계산법
- 일반적으로 분류 문제에서 데이터의 출력을 0,1로 구분하기 위해 시그모이드 함수를 사용, 시그모이드 함수에 포함된 자연 상수 e 때문에 평균 제곱 오차를 적용하면 매끄럽지 못한 그래프(울퉁불퉁한 그래프)가 출력
- 이 함수를 사용하면 손실 함수를 적용할 경우 경사 하강법 과정에서 학습이 지역 최소점에서 멈출 수 있음. 이를 방지하고자 자연 상수 e 에 반대되는 자연 로그를 모델의 출력 값에 취함.

4.2.2 딥러닝 학습



1. 순전파

- 훈련데이터가 네트워크에 입력 되었을 때, 전체 신경망을 교차해 나가는 과정
- 모든 뉴런이 이전 층의 뉴런에서 수신한 정보에, 가중합과 활성화 함수를 적용하여 다음층의 뉴런으로 전송

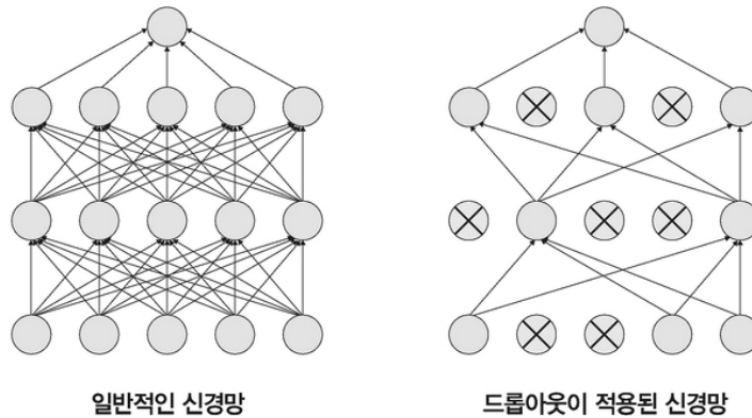
2. 역전파

- 순전파가 완료되면, 최종 출력층에서 도출된 예측값과 실제값의 차이를 손실함수로 추정하고, 손실함수가 결과적으로 0에 가깝도록 하기 위해 모델이 훈련을 반복하며 가중치를 조정한다.
- 오차가 계산되면 출력층 → 은닉층 → 입력층 순으로 정보가 역으로 전파 된다.

4.2.3 딥러닝의 문제점과 해결 방안

1. 과적합

- 훈련 데이터를 과하게 학습해서 새로운 데이터에서는 학습이 잘 안되는 현상 (검증 데이터에는 오차가 증가하게 됨)
- Solution : **Dropout**



2. 기울기 소멸

- 출력층에서 은닉층으로 전달되는 오차가 크게 줄어들어 학습이 되지 않는 현상
- 기울기가 소멸되어 학습이 더디게 진행된다 오차를 더 줄이지 못하고 그 상태로 수렴하는 현상
- Solution : 렐루 활성화 함수를 사용하기

3. 성능이 나빠지는 문제 발생

- 경사 하강법은 손실 함수의 비용이 최소가 되는 지점을 찾을 때까지 기울기가 낮은 쪽으로 계속 이동시키는 과정 반복, 이 과정에서 성능이 나빠지는 문제 발생
- 이를 해결하기 위해 확률적 경사 하강법, 미니 배치 경사 하강법 사용



1. 배치 경사 하강법

- 전체 데이터셋에 대한 오류를 구한 뒤, 기울기를 한 번만 계산하여 모델의 파라미터를 업데이트하는 방법
- 전체 훈련 데이터셋에 대해 가중치를 편미분

2. 확률적 경사 하강법

- 임의로 선택한 데이터에 대해 기울기를 계산하는 방법으로 적은 데이터를 사용하므로 빠른 계산이 가능
- 파라미터 변경 폭이 불안정하고, 때로는 배치 경사 하강법보다 정확도가 낮을 수 있지만 속도가 빠르다는 장점

3. 미니 배치 경사 하강법

- 전체 데이터셋을 미니 배치 여러 개로 나누고, 미니 배치 한 개마다 기울기를 구한 후 그것의 평균 기울기를 이용해 모델을 업데이트해서 학습하는 방식
- 전체 데이터를 계산하는 것보다 빠르며, 확률적 경사 하강법보다 안정적이라는 장점이 있기 때문에 실제로 가장 많이 사용

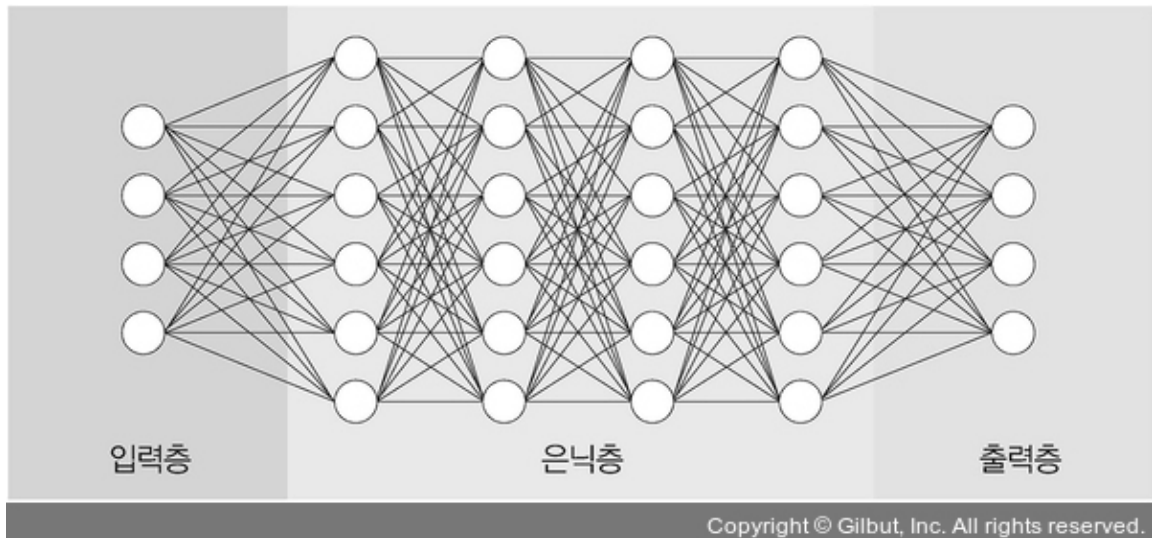
4. 옵티마이저

- 확률적 경사 하강법의 파라미터 변경 폭이 불안정한 문제를 해결하기 위해 학습 속도와 운동량을 조정하는 옵티마이저 적용 가능

4.3 딥러닝 알고리즘

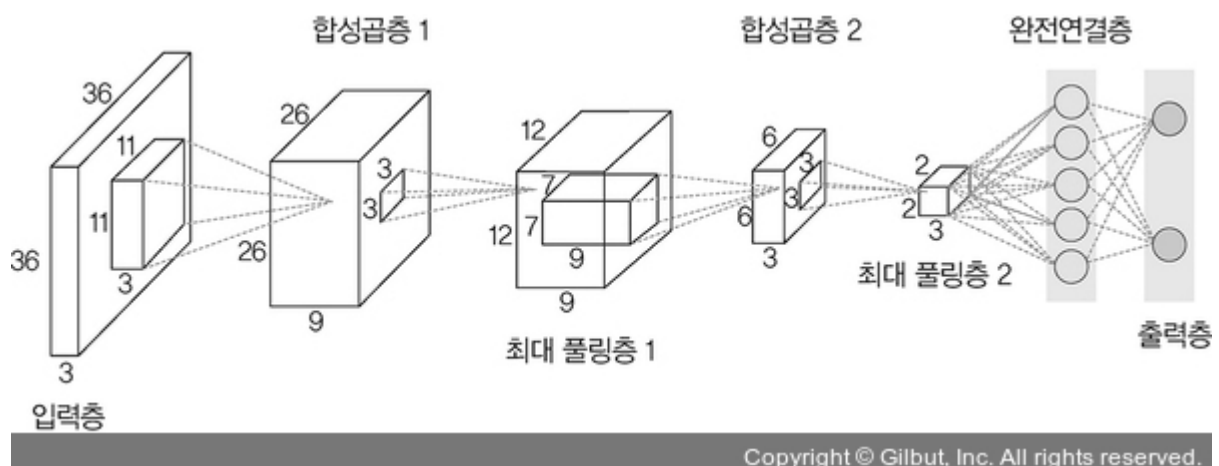
4.3.1 심층 신경망

- 입력층과 출력층 사이 다수의 은닉층을 포함하는 인공 신경망
 - 다수의 은닉층을 추가했기 때문에 별도의 트릭 없이 비선형 분류가 가능
 - 다양한 비선형적 관계를 학습할 수 있는 장점이 있지만, 학습을 위한 연산량이 많고, 기울기 소멸 문제가 발생할 수 있음
- 이를 해결하기 위해 드롭아웃, 렐루 함수, 배치 정규화 등을 적용해야 함.



4.3.2 합성곱 신경망

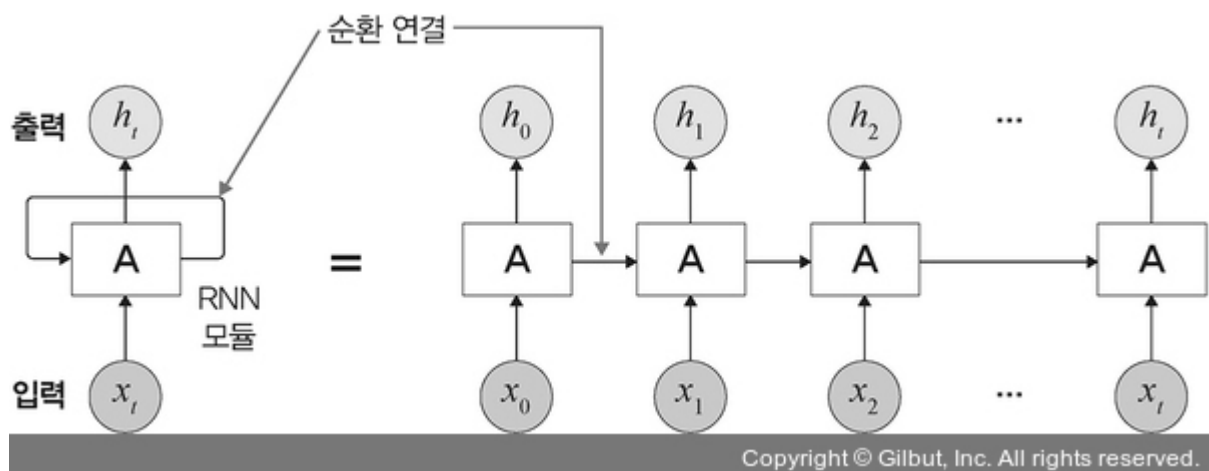
- 합성곱층, 풀링층을 포함하는 이미지 처리 성능이 좋은 인공 신경망 알고리즘
- 영상 및 사진이 포함된 이미지 데이터에서 객체를 탐색하거나 객체 위치를 찾아내는데 유용한 신경망
- 이미지에서 객체, 얼굴, 장면을 인식하기 위해 패턴을 찾는데 특히 유용 ex) LeNet-5, AlexNet, VGG, GoogLeNet, ResNet



4.3.3 순환 신경망

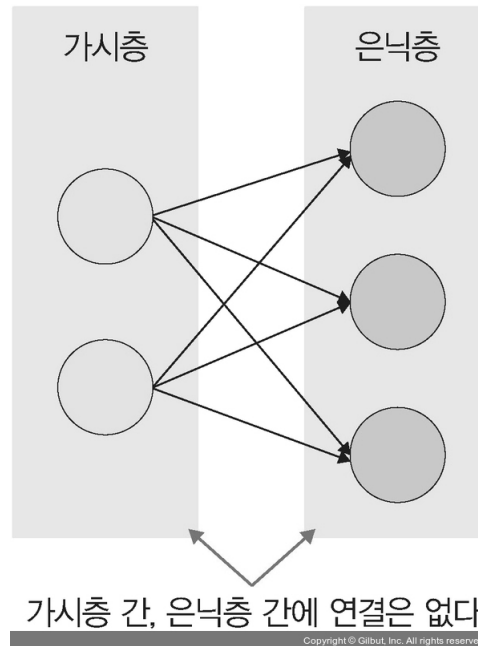
- 시계열 데이터 (음악, 영상 등) 같은 시간 흐름에 따라 변화하는 데이터를 학습하기 위한 인공 신경망
- 순환 : 자기 자신을 참조, 현재 결과가 이전 결과와 연관이 있다는 의미
- 순환 신경망의 특징

- 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 진행
- 기울기 소멸 문제로 학습이 제대로 되지 않는 문제가 있는데, 메모리 개념을 도입한 LSTM으로 보완
- 자연어 처리 분야와 궁합이 잘 맞음 ex) 언어 모델링, 텍스트 생성, 자동 번역, 음성 인식, 이미지 캡션



4.3.4 제한된 볼츠만 머신

- 가시층 (visible layer), 은닉층 (hidden layer)으로 구성된 모델
- 가시층은 은닉층과만 연결되는데(가시층과 가시층, 은닉층과 은닉층 사이 연결이 없는 층) 이것이 제한된 볼츠만 머신
- 제한된 볼츠만 머신의 특징
 - 차원 감소, 분류, 선형 회귀 분석, 협업 필터링, 특성 값 학습, 토픽 모델링에 사용
 - 기울기 소멸 문제를 해결하기 위해 사전 학습 용도로 활용 가능
 - 심층 신뢰 신경망 (DBN)의 요소로 활용



4.3.5 심층 신경망

- 입력층과 은닉층으로 구성된 제한된 볼츠만 머신을 블록처럼 여러 층으로 쌓은 형태로 연결된 신경망
- 사전 훈련된 제한된 볼츠만 머신을 층층이 쌓아 올린 구조로, 레이블이 없는 데이터에 대한 비지도 학습이 가능
- 부분적인 이미지에서 전체를 연상하는 일반화와 추상화 과정을 구현할 때 사용하면 유용

