1 신경망의 개요

1 신경망이란?

- 신경망은 인간 두뇌의 생물학적 뉴런의 작용을 모방한 모델
- 신경망은 인간 두뇌의 병렬성(parallelism)을 모델링
- 뉴런들로부터의 입력을 일정한 함수를 거쳐 출력
- '인공신경망'(Artificial Neural Networks)으로 부르기도 함
- 문자인식, 음성인식, 영상인식, 자연어 처리 등에 이용

2 신경망의 발상

- 인공지능을 기호주의와 연결주의로 나누었을 때 신경망은 연결주의의 대표적인 모델
- 인간 두뇌에 있는 뉴런의 연결을 모방
- 인간의 지능이 뉴런들 사이의 연결로부터 시작된다는 발상
- 두뇌가 어떤 원리에 따라 작동하는지가 주된 관심
- 병렬처리 구현에 중점을 둠
- 학습과 관련된 지능적인 역할을 훌륭하게 수행

3 신경망의 학습 기능

- 모든 신경망의 공통적인 주요 역할은 '학습(learning)' 기능
- 문자, 숫자, 음성, 영상, 동영상 등의 학습과 인식 능력이 필요한 인공지능
- 음성이나 영상 정보 등은 대규모 멀티미디어 정보
- 데이터가 크고 다루기가 매우 어려운 것을 학습하고 인식
- 선형 모델보다 해석이 어려운 비선형 모델을 구성

◁ 유튜브에서 추출한 고양이 사진

4 신경망의 시작과 발전 과정

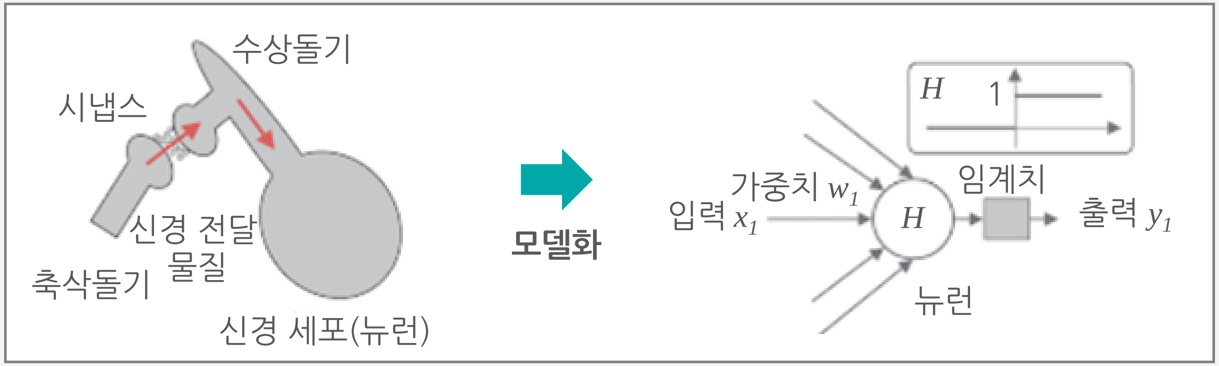
- 1943년 맥클럭-피츠 모델에서의 논리로 출발
- ‘헵’의 연결강도(weight) 조정을 위한 학습 규칙 나옴

헵의 법칙(Hebb’s Rule)과 형식 뉴런

- 유전 알고리즘과 함께 생명 현상에서 영감을 얻은 기법
- 신경 세포는 다른 신경 세포에게 전기 신호를 받았을 때 전기 신호가 일정 기준을 넘으면 다음 신경 세포로 신호를 전달
  - 이 과정을 수학적 모델(맥컬록-피츠 모델)로 고안한 것

4 신경망의 시작과 발전 과정

맥컬록-피츠 모델



△ 신경 세포는 축삭돌기 1개와 여러 개의 수상돌기로 구성
수상 돌기의 여러 개의 시냅스를 통해 축삭돌기에서 신경 전달 물질을 받음
자극은 화살표 방향으로 이동

△입력 x에 가중치 w를 붙여준 값의 총합이 임계치를 넘는다면, 출력 y를 1로 설정해 다음 뉴런에 값을 전달
H는 단위 계단 함수 사용

4 신경망의 시작과 발전 과정

- 1차 붐
  - 1950년대 퍼셉트론 개발로 신경망 붐을 주도
  - 단순한 XOR 문제의 한계에 도달
- 2차 붐
  - 1980년대 다층 퍼셉트론 개발과 오차 역전파법으로 붐을 주도
  - 다층 퍼셉트론 학습 시 기울기 소실 문제로 한계에 도달
- 3차 붐
  - 2000년대 활성 함수의 대체로 기울기 소실 문제를 해결
  - 이미지 인식 경연대회(ILSVRC)에서 알렉스넷으로 신경망; 딥러닝의 붐을 주도

5 기본 개념

- 실제 신경 세포 간의 전달 물질 이동은 시냅스를 통해 이뤄짐
- 받는 쪽 신경 세포의 세포막 안팎에 미세한 전위차(막 전위)를 발생시켜 전기 신호로 시각화
- 형식 뉴런(formal neuron; 소자): 신경 세포의 연결 모델
- 입력을 통해 얻은 값을 합산한 후 필터를 거쳐 값을 출력

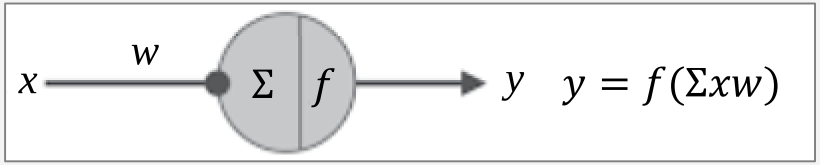
5 기본 개념

- 뉴런에서 값을 출력할 때 일정 기준은 단위 계산 함수 H로 결정
  - Unit Step Function; H

    x < 0, y = 0

    x > 0, y = 1

    x = 0, 0 <= y <= 1



[형식 뉴런]

5 기본 개념

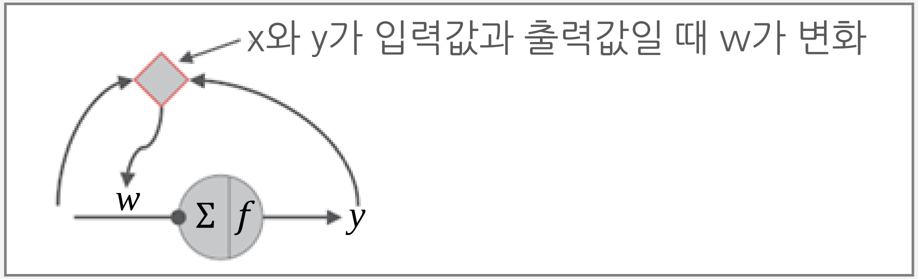
헵의 법칙

- 시냅스를 통한 신경 세포의 상호 작용이 증가하면 시냅스가 더 견고하게 강화
- 상호 작용이 감소하면 시냅스는 약해져 신경 회로를 사용하지 않음
- 시냅스의 유연한 연결, 즉 **가소성**있는 변화를 의미

5 기본 개념

학습과의 연관성 증명

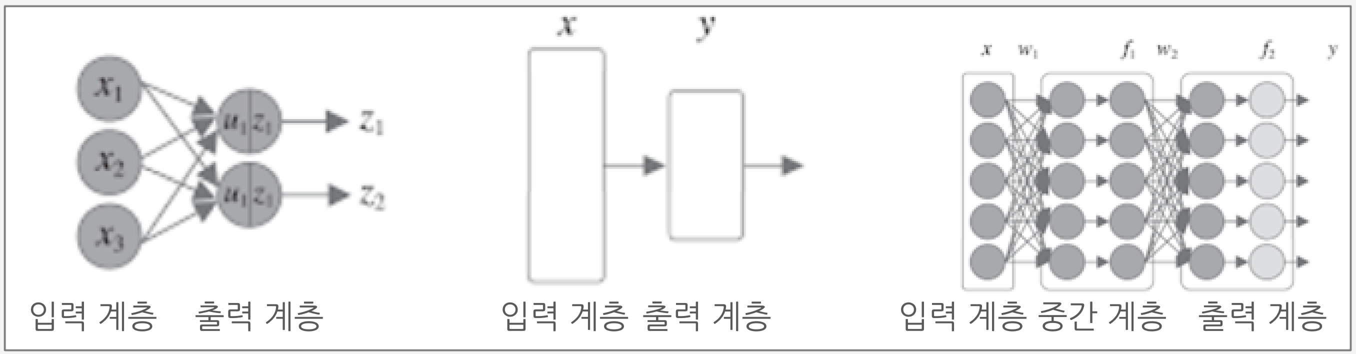
- 신경 세포의 연결이 더 견고하면 기억력 향상과 운동 능력 습득 같은 학습 작용에 도움을 줌
- 수학적 개념 도입 시 형식 뉴런에 헵의 법칙 적용에 따르면 입력값과 출력값에 의해 가중치 변화



[헵의 법칙을 적용한 형식 뉴런의 피드백]

5 기본 개념

- 형식 뉴런을 여러 개 이어서 ‘수학적인 신경회로’를 구성한 것
- 계층: 같은 종류의 형식 뉴런을 여러 개 병렬로 배열해 유닛을 형성하는 것



[다양한 신경망 예]

6 특징

- 분산성/병렬성(Dispersion/Parallelism)
  - 같거나 비슷한 뉴런끼리 구성된 다수 개의 노드 존재
  - 서로 연결되어 정보 교환
- 국소성(Locality)
  - 개별 뉴런이 받는 정보는 결합해 있는 다른 뉴런의 입력 신호 상태, 자신의 내부 상태, 출력 신호의 상태, 다음 연결부의 뉴런 상태가 됨
- 가중합(Weighted Sum)
  - 입력 받을 때 결합 상황에 따른 가중치(결합 하중) 연산 실행
  - 가중치가 적용된 입력의 합산 또는 합산한 값을 비선형 함수로 변환한 값을 내부 상태로 설정

6 특징

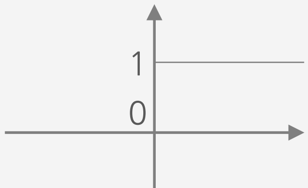
- 가소성(Plasticity)
  - 결합 하중은 뉴런이 얻는 정보로 변화
  - 학습과 자기 조직화에 활용
- 일반성(Generality)
  - 학습한 특정 상황에서 바람직한 행동을 하는 것은 물론이고 학습하지 않은 상황도 보간법과 외삽법 등으로 대응

7 활성화 함수

- 형식 뉴런에서 받은 입력 값을 출력할 때 일정 기준에 따라 출력 값을 변화시키는 비선형 함수

계단 함수

- 단위 계단 함수와 비슷하며 둘 다 디랙 델타 함수(Dirac delta function)와 관련된 (−∞ ~+∞) 적분을 계산한 결과
- 디랙 델타 함수(Dirac delta function)



[단위 계단 함수(계단 함수) 예]

$$y = f(x) = \begin{cases} 1, x > 0 \\ c, x = 0, 0 \leq c \leq 1 \\ 0, x < 0 \end{cases}$$

[단위 계단 함수 식]

$$y = f(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

[계단 함수 식]

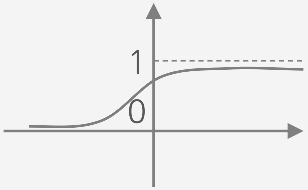
7 활성화 함수

시그모이드 함수

- $x = -\infty$ 일 때 0, $x = +\infty$ 일 때 1에 무한히 접근
- $x = 0$ 일 때 0.5가 되는 연속 함수
- 로지스틱 회귀에서 이용했던 로지스틱 함수의 역함수와 같은 방식으로 계산
- Sigmoid(x) 또는 $\sigma(x)$와 같은 식으로 표기
- 유사한 함수로 $x = -\infty$ 일 때 0, $x = +\infty$ 일 때 1에 무한히 접근하며 $x = 0$ 일 때 0이 되는 tanh(x)(쌍곡선 탄젠트 함수)도 사용

7 활성화 함수

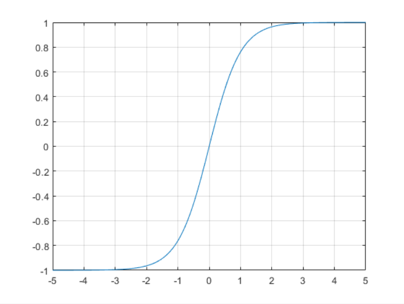
시그모이드 함수



[시그모이드 함수]

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

[시그모이드 함수 식]

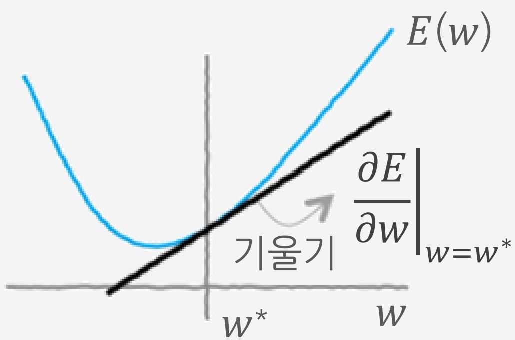


[Tanh(x) 함수]

7 활성화 함수

오차 함수(목적 함수)

- gradient descent 방식
  - $E(w)$의 편미분 $\left.\frac{\partial E}{\partial w}\right|_{w=w^*}$를 계산하여 마이너스 방향으로 갱신



[Gradient Descent 방식]

7 활성화 함수

오차 함수(목적 함수)

▶ 근사화

- $w = w^*$의 위치에서 아주 작은 값(ε) 의 ± 한 위치의 근사적 기울기 산출
- ε이 0에 가까울 수록 미분과 동일한 값을 산출
- $\left.\frac{\partial E}{\partial w}\right|_{w=w^*} \cong \frac{E(w^*+\varepsilon)-E(w^*-\varepsilon)}{2\varepsilon}$

7 활성화 함수

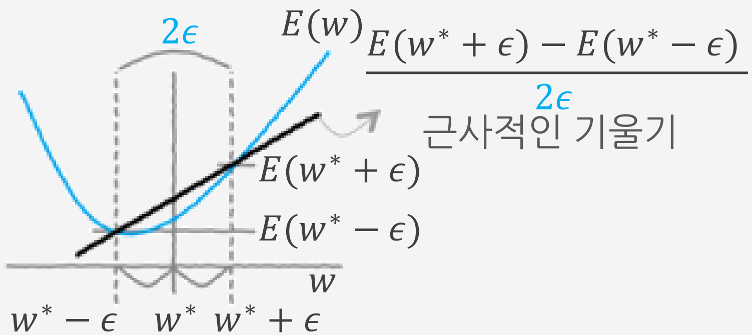
오차 함수(목적 함수)

▶ 근사화

- 매개 변수가 증가하면 하나의 매개 변수를 제외한 매개 변수들을 고정하여 편미분 근사화

$$\left.\frac{\partial E}{\partial w}\right|_{w_0^*, w_1^*, w_2^*} \cong \frac{E(w^* + \varepsilon, w_1^*, w_2^*) - E(w^* - \varepsilon, w_1^*, w_2^*)}{2\varepsilon}$$

- 근사화 방법은 정밀도 보다 계산 비용이 크다는 단점



[근사화 방식]

2 초기의 신경망

1 퍼셉트론

- 맥컬록-피츠 모델을 기반으로 두는 학습 기계
- 헵의 법칙을 적용하여 출력값에 따라 플러스와 마이너스로 진동하도록 가중치 업데이트
- 가중치 업데이트
  - 출력값 $w\varphi(x)$에 긍정적인 예라면 +
  - 부정적인 예라면 -
  - 학습률(learning rate) η로 하여 가중치를 업데이트

플러스일 때	마이너스일 때
$w \Leftarrow w + \eta \cdot \varphi(x)$	$w \Leftarrow w - \eta \cdot \varphi(x)$

[가중치의 업데이트 식]

1 퍼셉트론

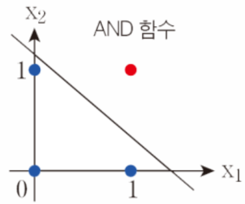
- 단순 퍼셉트론
  - 퍼셉트론 수렴 정리(perceptron convergence theorem): 긍정적인 예와 부정적인 예에서 학습 데이터를 선형 분리 가능하다면 반드시 유한 횟수를 반복 실행해 평면 분리 상태를 찾을 수 있다는 개념
- 데이터 선형 분리가 불가능하면 퍼셉트론 학습 적용 불가
- 선형 분리할 수 있더라도 평면 분리 상태로 만드는데 시간이 오래 소요될 한계 포함

2 초기의 신경망

1 퍼셉트론

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

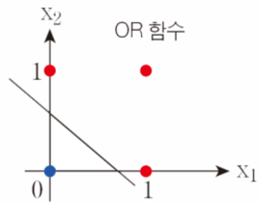
(AND 함수)



(선형 분리 가능)

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

(OR 함수)



(선형 분리 가능)

[선형 분리 가능한 예]

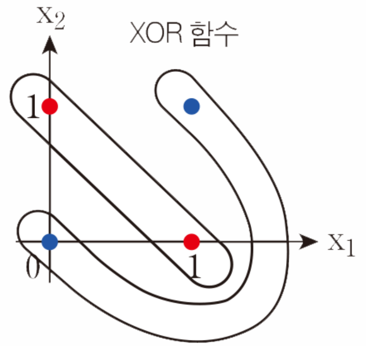
2 초기의 신경망

1 퍼셉트론

- 선형 함수 불가능 예
  - Exclusive-Or(XOR) 함수는 선형 분리가 불가능
  - 하나의 직선이 아닌 곡선에 의해서만 분리가 가능

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

(XOR 함수)



(선형 분리 불가능)

[선형 분리 불가능 예]

1 퍼셉트론

- 로젠블럿의 단층 퍼셉트론
  - 최초의 신경망 모델인 ‘퍼셉트론’은 단층으로 구성
  - 신경망 하드웨어 장치인 ‘마크 I 퍼셉트론’ 1957년에 제작
  - 마크 I(one) 퍼셉트론은 A, B, C 등의 문자를 인식
  - 20 × 20개의 화소(pixel)를 가진 마크 I 퍼셉트론 화면
  - 마크 I 퍼셉트론에서 연결선으로 연결강도 조정
  - 학습을 위해 몇 km나 되는 연결선을 사용

2 초기의 신경망

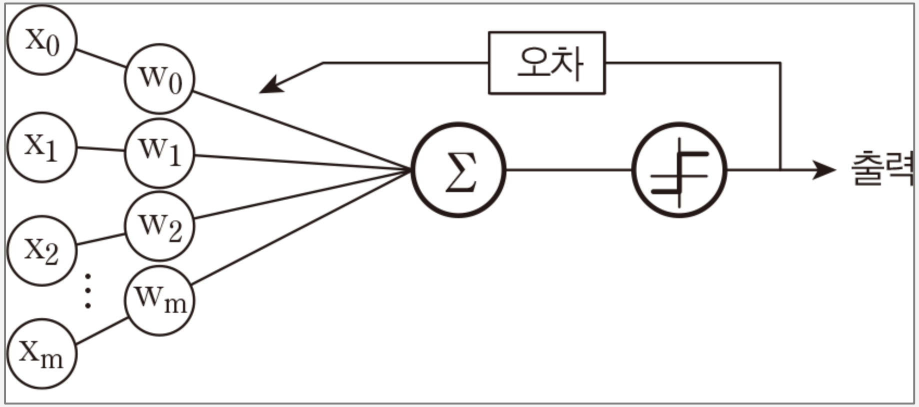
1 퍼셉트론

- 로젠블럿의 단층 퍼셉트론

[Mark I Perceptron]

1 퍼셉트론

- 단층 퍼셉트론의 구조
  - 단 1개 층의 연결 강도 조정
  - 오차에 대한 피드백 학습



[단층 퍼셉트론의 구조]

1 퍼셉트론

- 퍼셉트론의 학습 과정
  - 연결 강도를 조정하며 학습

[1]	연결강도들과 임계값을 초기화
[2]	새로운 입력과 기대되는 출력을 제시
[3]	실제 출력값을 계산
[4]	연결강도를 재조정
[5]	더 이상 조정이 없을 때까지 [2] 단계로 가서 반복 수행

1 퍼셉트론

- 단층 퍼셉트론의 2가지 제한점
  - 단층 퍼셉트론의 출력은 0 또는 1(1 또는 -1)만 가짐
  - 선형 분리가 가능한 집합만을 분리 가능
- XOR 함수의 문제점
  - 선형 분리가 불가능
  - 한 직선으로 두 집합을 교차하지 않고 나눌 수 없음
  - 이 점은 단층 퍼셉트론 학습에서 매우 심각한 문제점
  - 1980년 중반, 다층 퍼셉트론은 XOR 문제부터 해결

1 퍼셉트론

- 단층 퍼셉트론의 한계점과 기여
  - XOR 문제 해결 불가, 10여 년 동안 관심이 멀어짐
  - 단층 퍼셉트론은 학습 모델로서는 적절하지 않음
  - 1980년대 중반 다층 퍼셉트론 모델의 기반이 됨
    - 문자 인식을 비롯한 여러 분야에 폭넓게 응용되었음
    - 신경망 연구의 새로운 장을 열게 된 결정적인 계기

3 다층 퍼셉트론

1 새로운 신경망 시대의 도래

- 1969년 이후 신경망 연구가 10여 년간 침체
- 1980년대 중반에 다층 퍼셉트론 모델이 제안됨
- 단층 퍼셉트론 모델에 하나 이상의 은닉층을 추가로 사용
- 단층 퍼셉트론의 XOR 문제를 해결하여 비선형 문제에 도입

2 다층 퍼셉트론

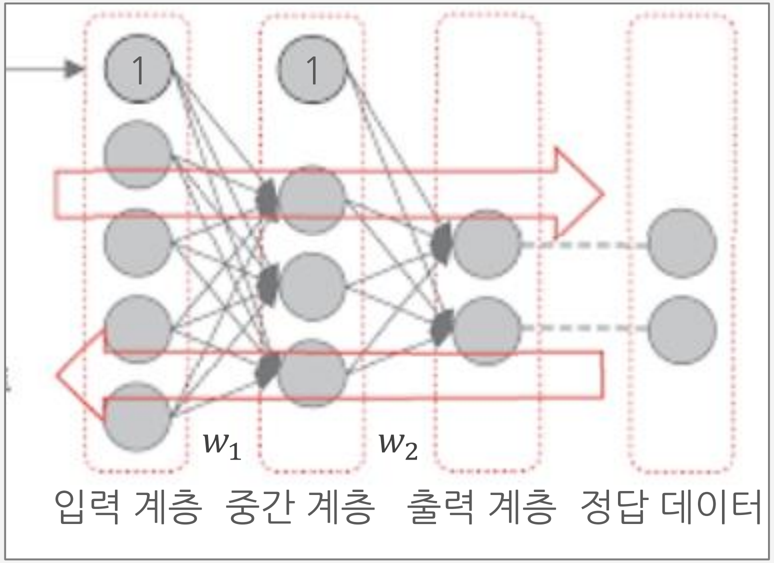
- 입력 계층과 출력 계층만 있는 단순 퍼셉트론의 단점은 선형 분리 데이터에만 적용 가능하며 계산 시간이 오래 걸림
- 단순 퍼셉트론을 여러 번 서로 연결하여 비선형으로 보이는 분포를 선형 분리할 수 있는 분포로 변환
- 순방향으로 계층마다 출력을 계산한 후 오차 역전파법을 이용해 출력 계층에서 역방향으로 가중치 업데이트 실행
- 네트워크 학습 전제와 정답 데이터 제공 시 동작
- 정답 데이터와의 조율은 최소 제곱 오차 등을 반영한 오차 함수를 만든 후 경사 하강법 이용

2 다층 퍼셉트론

회귀분석 시와 같이
편의상 1 대입

입력 계층에서
순방향으로 계산

출력 계층에서 역방향으로
계산을 실행해 w를 업데이트



[다층 퍼셉트론]

학습 데이터 존재 시
출력 결과를 정답 데이터와 비교해
오차가 적은 방향으로 조작

3 오차역전파 법

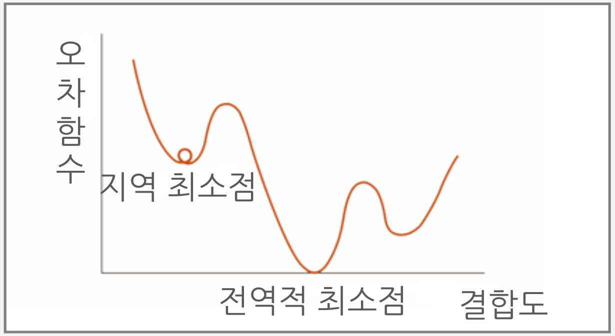
- 중간 계층을 갖는 신경망은 출력 계층에서 학습 데이터와 값의 오차를 이용해 중간 계층 뉴런의 특성을 변화시킬 수 있음 → 오차 역전파 법(back propagation)
- 다층 퍼셉트론의 학습 규칙으로 적용
- 진행 흐름이 한 방향으로 되어 있는 신경망에서도 피드백 메커니즘 생성 가능
- 과거 단층 퍼셉트론의 제한점 극복
- 특히 XOR 함수의 선형 분리 문제 등 해결
- 침체에 빠졌던 신경망 연구가 새롭게 활기를 띠게 됨
- 역전파 알고리즘은 입력층에서 은닉층을 거쳐 출력층, 다시 반대 방향으로 되돌아오면서 학습

4 다층 퍼셉트론의 학습 과정과 규칙들

- 델타규칙, 최급 하강법, 일반화 델타규칙 등이 사용됨
- 델타규칙(delta rule)은 출력과 목표 출력값과의 오차 제곱의 총합을 최소로 하도록 연결 강도를 조정하는 규칙
- 최급 하강법(gradient descent method)은 곡면에서 오차의 제곱이 가장 많이 감소하는 방향으로 기울기를 따라가며 변화하는 방법

5 역전파 학습 알고리즘과 지역최소점(local minima) 문제

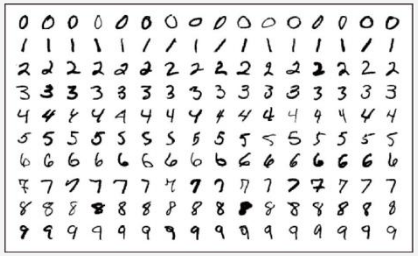
- 역전파 학습 알고리즘의 단점
- 오랜 학습 시간, 낮은 확률이나 지역 최소점 봉착 가능
- 특히 최급 하강법은 지역 최소점에 머물 가능성이 큼
- 우리는 지역 최소점(Local Minima)이 아닌 전역적 최소점(Global Minima)을 추구
- 그러나 신경망이나 인공지능에서는 다소 불가피함



[지역 최소점과 전역 최소점]

6 다층 퍼셉트론 모델의 기타 응용

- 패리티 문제, 부호화 문제, 대칭성 문제 등
- 텍스트를 음성으로 변환하는 네트워크 시스템 개발
- 주식시장의 예측, 다른 언어들 간의 번역
- 공장자동화, 실시간 음성인식, 로봇 등



[숫자 인식을 위한 학습 데이터]

[영문자의 인식을 위한 학습 데이터]