딥러닝의 통계적이해

4강. 딥러닝 모형의 구조와 학습(2)

1. 오차역전파법
2. 경사소실
3. 과대적합

한국방송통신대 이긍희 교수

오늘의 학습목표

1. 오차역전파법을 이해한다.
2. 경사소실을 이해한다.
3. 과대적합을 이해한다.

1. 오차역전파법

오차역전파법

◆ 경사하강법에서 연쇄 미분을 통해 손실함수의 경사를 효율적으로 구하여 신경망의 가중치를 갱신하는 것

- Rumelhart, Hinton, and Williams (1986)

→ 연산은 순방향과 역방향으로 구분

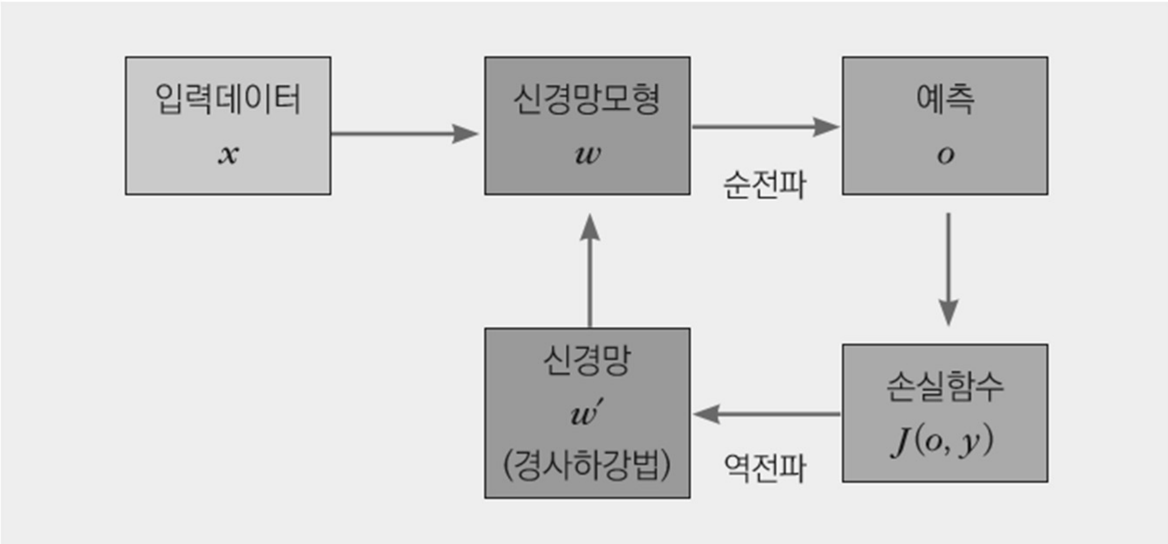
오차역전파법

- 순방향 연산 : 가중치의 초깃값을 정한 후 [입력데이터 → 신경망모형 → 예측] 순, 누적적·순차적으로 가중합과 비선형 연산 → 예측 오차로 손실함수 값을 구함

- 역방향 연산 : 손실함수 값 줄이도록 가중치 갱신
  - 순방향 연산의 역방향으로 경사(기울기, 미분값)을 구함

오차역전파법의 연산

- 순방향 연산시 미분을 저장해 두었다가 나중에 역전파 연산 시 이를 이용 → 합성함수 미분 이용, 계산 그래프 이용



합성함수의 미분과 계산그래프

- $z = f(t)$, $t = g(x, y)$, z의 x에 대한 편미분 → 연쇄 미분

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial t}\frac{\partial t}{\partial x}$$

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial \cancel{t}}\frac{\partial \cancel{t}}{\partial x}$$

합성함수의 미분과 계산그래프

- (예 3.1) $z = (x+y)^2$를 x에 대해 편미분하라.

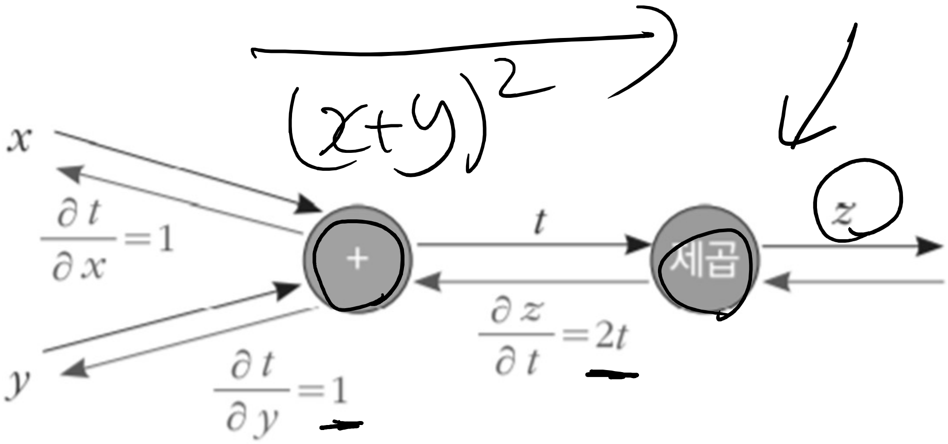
$$z = t^2, \quad t = (x+y)$$

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial t} \cdot \frac{\partial t}{\partial x}$$

$$= 2t \cdot 1$$

$$= 2(x+y) \checkmark$$

합성함수의 미분과 계산그래프



$$\frac{\partial z}{\partial x}=\frac{\partial t}{\partial x}\frac{\partial z}{\partial t}=2(x+y)$$

1. 오차역전파법

합성함수의 미분과 계산그래프

- 합성함수 : $y = a_3\left(a_2\left(a_1(x)\right)\right)$
- 계산그래프와 합성함수의 미분

$\cdot \dfrac{\partial y}{\partial h_2} = a_3{}'$ $\cdot \dfrac{\partial y}{\partial h_1} = a_3{}' a_2{}'$ $\cdot \dfrac{\partial y}{\partial x} = a_3' a_2' a_1'$

$\dfrac{\partial y}{\partial x} = a_3{}' \cdot a_2{}' \cdot a_1{}'$

$= \dfrac{\partial y}{\partial h_2} \dfrac{\partial h_2}{\partial h_1} \dfrac{\partial h_1}{\partial x}$

$\dfrac{\partial y}{\partial h_1}$

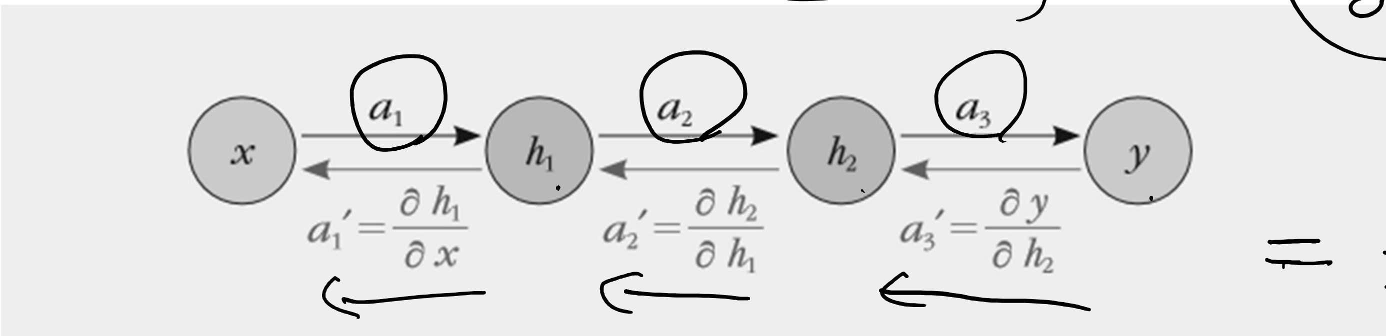
$= a_3{}' \, a_2{}' \cdot a_1{}'$

1. 오차역전파법

합성함수의 미분과 계산그래프

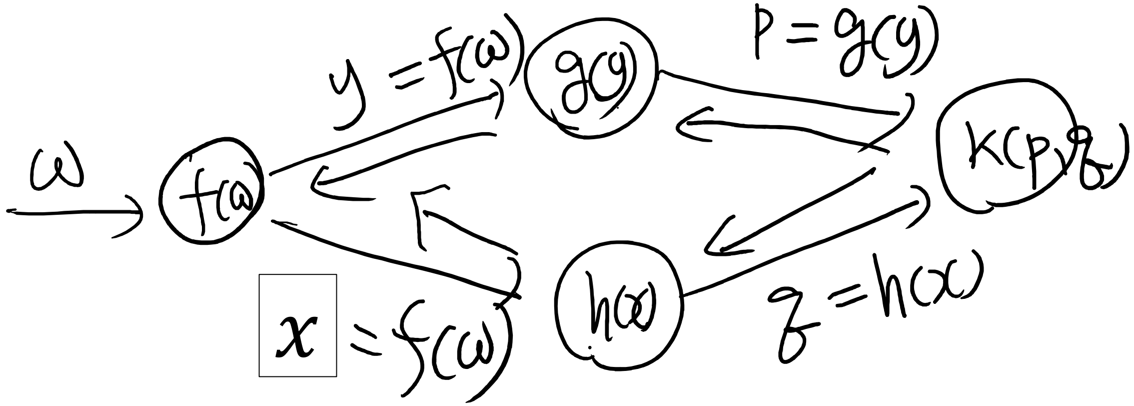
$$o = k(p, q)$$

순방향



$$\frac{\partial o}{\partial w} = \frac{\partial o}{\partial p} \cdot \frac{\partial p}{\partial w} + \frac{\partial o}{\partial q} \cdot \frac{\partial q}{\partial w}$$

$$= \frac{\partial o}{\partial p} \cdot \frac{\partial p}{\partial y} \cdot \frac{\partial y}{\partial w} + \frac{\partial o}{\partial q} \cdot \frac{\partial q}{\partial x} \cdot \frac{\partial x}{\partial w}$$



1. 오차역전파법

로지스틱 회귀모형

$$y = \frac{1}{1+e^{-(w_1x_1 + w_2x_2)}}$$

◆ 입력변수 2개, 로지스틱 회귀모형

- $y = \frac{1}{1+e^{-z}}$, $z = w_1x_1 + w_2x_2$, J : 손실함수

- $\frac{\partial J}{\partial z}$, $\frac{\partial J}{\partial w_1}$, $\frac{\partial J}{\partial w_2}$

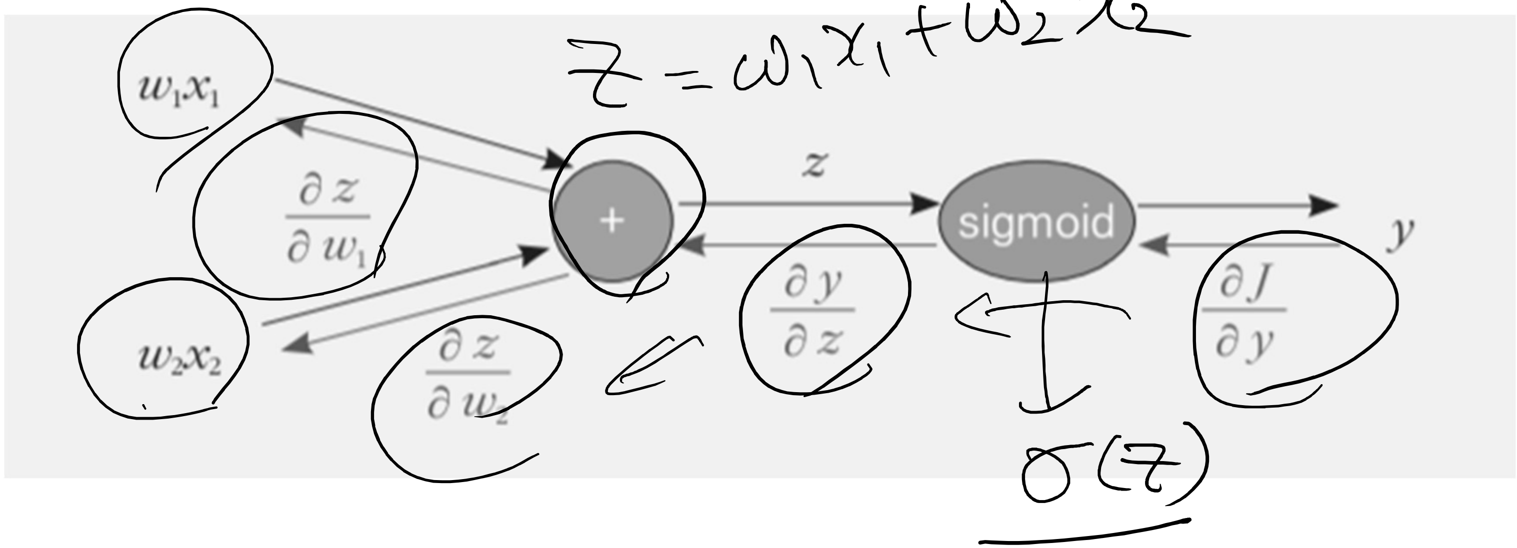
① $\frac{\partial J}{\partial z} = \frac{\partial J}{\partial y} \cdot \frac{\partial y}{\partial z} = \frac{\partial J}{\partial y} y(1-y)$

② $\frac{\partial J}{\partial w_1} = \frac{\partial J}{\partial y} \cdot \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial w_1} = \frac{\partial J}{\partial y} y(1-y)\, x_1$

③ $\frac{\partial J}{\partial w_2} = \frac{\partial J}{\partial y} \cdot \frac{\partial J}{\partial z} \cdot \frac{\partial z}{\partial w_2} = \frac{\partial J}{\partial y} y(1-y)\, x_2$

1. 오차역전파법

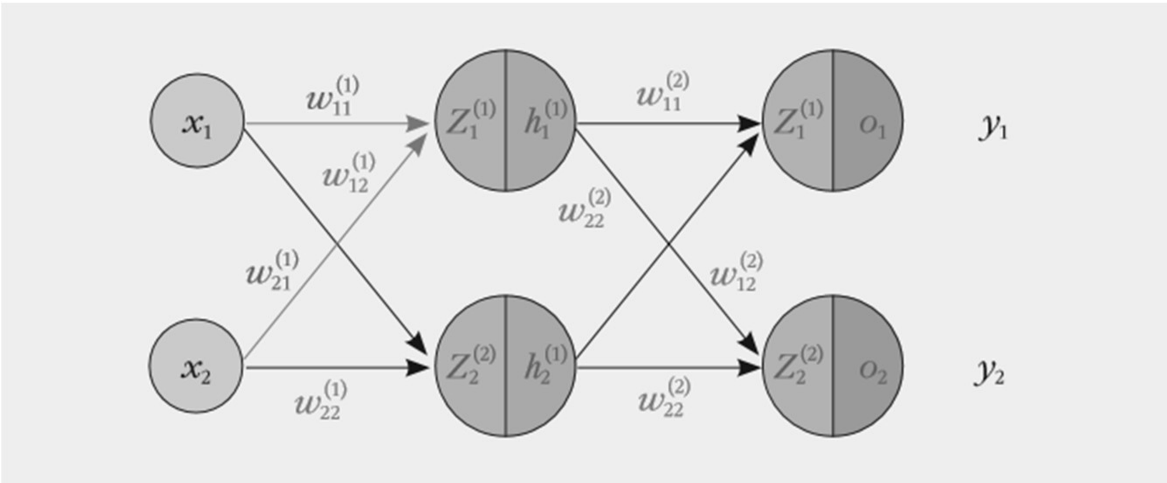
로지스틱 회귀모형의 계산그래프



오차역전파법의 예(1)

◆ 입력층, 은닉층과 출력층 : 2개의 뉴런, 은닉층 1개

- 은닉층의 활성화 함수 : 시그모이드 함수
- 출력층의 활성화 함수 : 항등함수



오차역전파법의 예(1)

- 손실함수 : $J(w) = \frac{1}{2}[(y_1 - o_1)^2 + (y_2 - o_2)^2] = L_{1w} + L_{2w}$

- 층별 가중치의 갱신 : $w_{ij}^{(l)} := w_{ij}^{(l)} - \eta \frac{\partial J(w)}{\partial w_{ij}^{(l)}}$

층

학습률

신경망 초기값 → O_i

$(y_i - O_i)$

1. 오차역전파법

신경망의 순방향 연산

$$\checkmark\ z_1^{(1)} = w_{11}^{(1)} x_1 + w_{12}^{(1)} x_2$$

$$\checkmark\ z_2^{(1)} = w_{12}^{(1)} x_1 + w_{22}^{(1)} x_2$$

$$h_j^{(1)} = a(z_j^{(1)}) \qquad j = 1, 2$$

$$z_1^{(2)} = w_{11}^{(2)} h_1^{(1)} + w_{12}^{(2)} h_2^{(1)}$$

$$z_2^{(2)} = w_{12}^{(2)} h_1^{(1)} + w_{22}^{(2)} h_2^{(1)}$$

$$o_j = (z_j^{(\boxed{2})}) \qquad j = 1, 2$$

신경망의 역방향 연산

- $w_{11}^{(2)}$에 대한 손실함수의 미분

$$\frac{\partial J}{\partial w_{11}^{(2)}} = \frac{\partial J}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}}$$

$$= -(y_1 - o_1)\, 1 \cdot h_1^{(1)}$$

$$J = \frac{1}{2}\left[(y_1 - o_1)^2 + \cancel{(y_2 - o_2)^2} \right]$$

$$\frac{\partial J}{\partial o_1} = -(y_1 - o_1)$$

$$z_1^{(2)} = o_1$$

$$o_1 = z_1^{(2)}$$

$$z_1^{(2)} = w_{11}^{(2)} h_1^{(1)} + w_{12}^{(2)} h_2^{(1)}$$

신경망의 역방향 연산

◆ $w_{11}^{(1)}$에 대한 손실함수의 미분

$$J = L_1 w + L_2 w = \frac{1}{2}(y_1 - o_1)^2 + \frac{1}{2}(y_2 - o_2)^2$$

시그모이드 함수

$$z_1^{(1)} = w_{11}^{(1)} x_1 + w_{12}^{(1)} x_2$$

$$\frac{\partial J(w)}{\partial w_{11}^{(1)}} = \left[\frac{\partial L_1 w}{\partial h_1^{(1)}} + \frac{\partial L_2 w}{\partial h_1^{(1)}} \right] \frac{\partial h_1^{(1)}}{\partial z_1^{(1)}} \frac{\partial z_1^{(1)}}{\partial w_{11}^{(1)}}$$

$$= [\ ① + ②\ ] \; a(z_1^{(1)})(1 - a(z_1^{(1)}))\; x_1$$

$$\frac{\partial L_1 w}{\partial h_1^{(1)}} = \frac{\partial L_1 w}{\partial o_1} \frac{\partial o_1}{\partial z_1^{(2)}} \frac{\partial z_1^{(2)}}{\partial h_1^{(1)}}$$

$$= -(y_1 - o_1) \cdot 1 \cdot w_{11}^{(2)}$$

$$\frac{\partial L_2 w}{\partial h_1^{(1)}} = -(y_2 - o_2) \cdot 1\ w_{12}^{(2)}$$

신경망의 역방향 연산

◆ $w_{11}^{(1)}$에 대한 손실함수의 미분

$$\frac{\partial J(w)}{\partial w_{11}^{(1)}} = -\sum_{i=1}^{2} (y_i - o_i)\, w_{1i}^{(2)} \cdot a(z_1^{(1)})(1 - a(z_1^{(1)}))\, x_1$$

신경망의 역방향 연산

- $w_{11}^{(2)}$와 $w_{11}^{(1)}$: 경사하강법을 통한 갱신

  - $w_{11}^{(2)} := w_{11}^{(2)} - \eta \frac{\partial J(w)}{\partial w_{11}^{(2)}}$

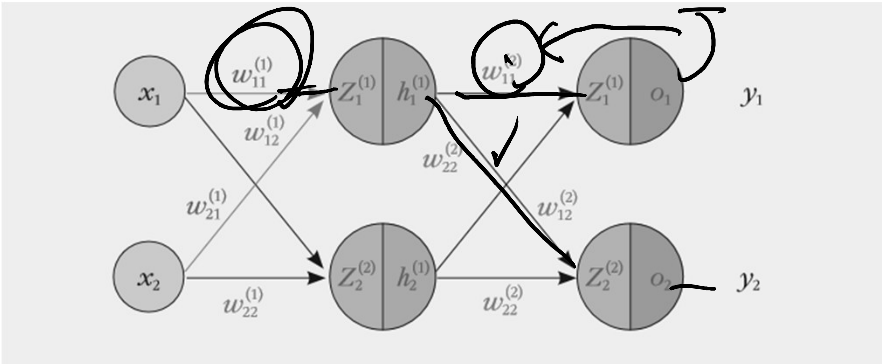
  - $w_{11}^{(1)} := w_{11}^{(1)} - \eta \frac{\partial J(w)}{\partial w_{11}^{(1)}}$

오차역전파법의 예(2)

$$J(\omega) = -\sum_{i=1}^{2} \left(y_i \ln o_i + (1-y_i) \ln (1-o_i) \right)$$

- 입력층, 은닉층과 출력층에 각각 2개의 뉴런, 은닉층이 1개인 신경망

 - 은닉층의 활성화 함수 : 시그모이드 함수, 출력층의 활성화 함수로는 시그모이드 함수



신경망의 순방향 연산

$$z_1^{(1)} = w_{11}^{(1)} x_1 + w_{21}^{(1)} x_2$$

$$z_2^{(1)} = w_{12}^{(1)} x_1 + w_{22}^{(1)} x_2$$

$$h_j^{(1)} = a(z_j^{(1)}) \quad j=1,2$$

$$z_1^{(2)} = w_{11}^{(2)} h_1^{(1)} + w_{21}^{(2)} h_2^{(1)}$$

$$z_2^{(2)} = w_{12}^{(2)} h_1^{(1)} + w_{22}^{(2)} h_2^{(1)}$$

$$o_j = \underline{a_o}(z_j^{(2)}) \quad j=1,2$$

신경망의 역방향 연산

$$J = -\sum_{i=1}^{2} \left(y_i \ln a_i + (1-y_i) \ln(1-a_i) \right)$$

$$o_1 = a_o(z_1^{(2)})$$

◆ $w_{11}^{(2)}$에 대한 손실함수의 미분

$$\frac{\partial J(w)}{\partial w_{11}^{(2)}} = \frac{\partial J(w)}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_1^{(1)}} \cdot \frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}}$$

$$= -\left(\frac{y_1}{o_1} - \frac{1-y_1}{1-o_1} \right) o_1(1-o_1)\, h_1^{(1)}$$

$$\frac{y_1 - \cancel{y_1 o_1} - o_1 + \cancel{y_1 o_1}}{\cancel{o_1(1-o_1)}} \cancel{o_1(1-o_1)}\, h_1^{(1)}$$

$$= -(y_1 - o_1)\, h_1^{(1)}$$

1. 오차역전파법

신경망의 역방향 연산

- $w_{11}^{(1)}$에 대한 손실함수의 미분

시그모이드

$$\frac{\partial J(\omega)}{\partial \omega_{11}^{(1)}} = \left[\frac{\partial L_{1\omega}}{\partial h_1^{(1)}} + \frac{\partial L_{2\omega}}{\partial h_1^{(1)}} \right] \frac{\partial h_1^{(1)}}{\partial z_1^{(1)}} \frac{\partial z_1^{(1)}}{\partial \omega_{11}^{(1)}}$$

$$x_1$$

$$a(z_1^{(1)}) \cdot (1 - a(z_1^{(1)}))$$

$$\frac{\partial L_{1\omega}}{\partial h_1^{(1)}} = \frac{\partial L_{1\omega}}{\partial o_1} \frac{\partial o_1}{\partial z^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial h_1^{(1)}}$$

$$= -(y_1 - o_1)\, \omega_{11}^{(2)}$$

신경망의 역방향 연산

◆ $w_{11}^{(1)}$에 대한 손실함수의 미분

$$\frac{\partial L_{2}(w)}{\partial h_2} = -(y_2 - o_2)\, w_{12}^{(2)}$$

$$\frac{\partial J(w)}{\partial w_{11}^{(1)}} = -\sum_{i=1}^{2} (y_i - o_i)\, w_{1i}^{(2)}\, a(z_1^{(1)}) \times (1 - a(z_1^{(1)}))\, x_1$$

2. 경사소실

2. 경사소실

다층신경망

◆ 입력층 $x_0, x_1, \dots, x_p$, 출력 o_j인 딥러닝 모형

- 제 l은닉층의 뉴런 : $h_j^{(l)}$, 활성화함수 : a, a_0

층	뉴런	뉴런의 가중합
입력층	x_j	
제1 은닉층	$h_{j_1}^{(1)}=a(z_{j_1}^{(1)})$	$z_{j_1}^{(1)}=\sum_i w_{ij_1}^{(2)} x_i$
제2 은닉층	$h_{j_2}^{(2)}=a(z_{j_2}^{(2)})$	$z_{j_2}^{(2)}=\sum_i w_{ij_2}^{(2)} h_i^{(1)}$
⋮	⋮	⋮
제 $L-1$ 은닉층	$h_{j_{L-1}}^{(L-1)}=a(z_{j_{L-1}}^{(L-1)})$	$z_{j_{L-1}}^{(L-1)}=\sum_i w_{ij_{L-1}}^{(L-1)} h_i^{(L-2)}$
출력층	$o_{j_o}=a_o(z_{j_o}^{(L)})$	$z_{j_o}^{(L)}=\sum_i w_{ij_o}^{(L)} h_i^{(L-1)}$

다층신경망과 활성화 함수

◆ 신경망 하나의 함수 표현 → 활성화 함수 a의 누적 적용

- $o_{j_0} = a_0\left(a\left(\sum w_{i_L j_0}^{(L)} a\left(\sum w_{i_{L-1} j_0}^{(L-1)} a \ldots\right)\right)\right)$

2. 경사소실

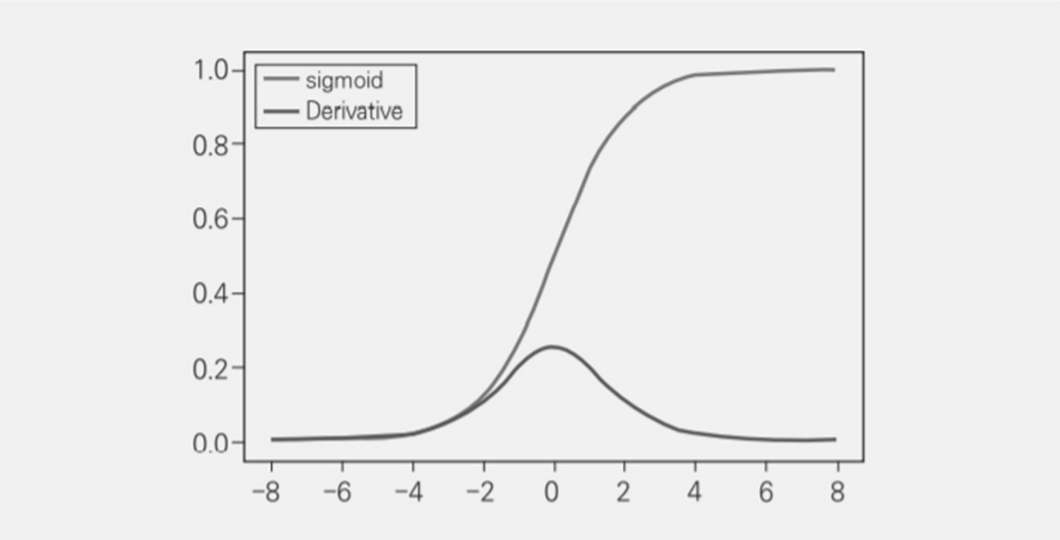
다층신경망

- 딥러닝 모형에서 역전파로 가중치가 갱신될 때 편미분 값이 전달

  - 시그모이드 함수와 이의 미분

시그모이드 함수의 미분

- $a(1-a)$의 최댓값은 0.25



경사소실 현상

- 다층 신경망에서 역전파로 경사가 전달되어 가는 과정 중
  → 연쇄법칙으로 인해 활성화 함수 미분값이 누적·계산되어
  아래층 뉴런으로 전파

- 은닉층의 수 k, 활성화 함수를 중첩 이용, 그 최댓값은 0.25^k,
  k가 커진다면 그 값은 0에 가까워짐
  → 손실함수 값이 전파되지 않아 가중치가 수정되지 않는
  현상 발생

3. 과대적합

과대적합의 문제

- 신경망에서 층을 깊이 쌓으면 가중치의 수가 많아져서 과대적합(overfitting) 문제 발생
  - 과대적합의 결과 : 신경망이 훈련 데이터에는 잘 적합, 새로운 데이터에 대한 예측력은 낮음

과대적합의 문제

- ◆ 1990년대말 서포트벡터머신(SVM)이나 부스팅과 같은 앙상블 방법 등이 신경망에 비해 계산이 간단, 좋은 성과를 보이면서 신경망 연구가 지연됨

  - 2010년대 빅데이터, 컴퓨팅 능력, 알고리즘 혁신으로 오차역전파의 경사소실, 과대적합 문제 해결

학습정리

- ✓ 오차역전파법은 신경망의 가중치를 연쇄미분을 이용하여 반복적으로 구하는 방법이다.
- ✓ 경사소실은 다층 신경망에서 역전파로 경사의 전달 과정에서 활성화 함수의 미분값이 누적, 계산됨으로써 손실함수 값의 변화가 아래층 뉴런으로 전파되지 않는 현상이다.

학습정리

✓ 신경망에서 층을 깊이 쌓으면 가중치의 수가 많아져서 과대적합(overfitting) 문제가 발생한다.

다음시간안내

5강.
딥러닝의 제 문제와 발전