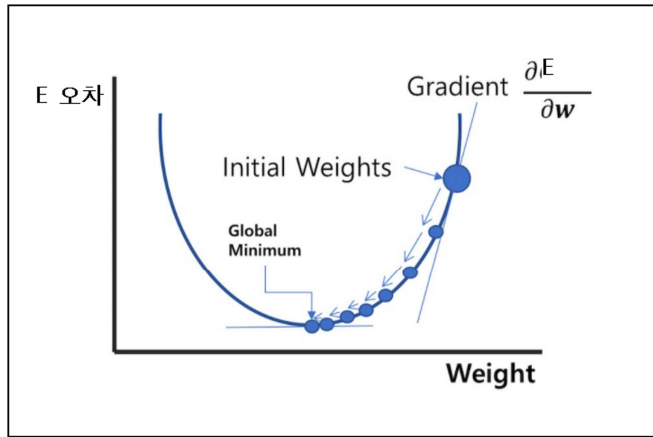


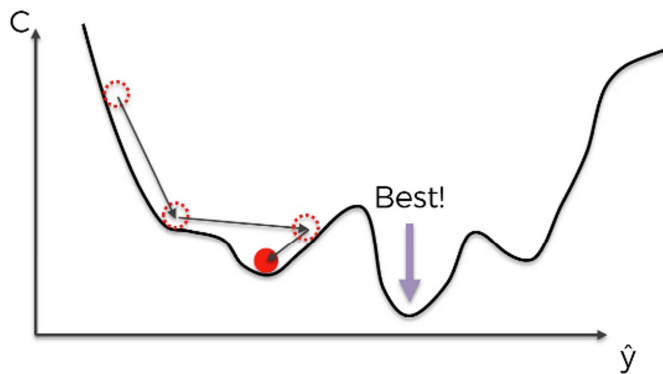
# 04.1. 경사하강법

2020년 12월 28일 월요일 오전 10:01

## 1. 개념



오차를 편미분 해서 양수면 - 음수면 +로 조절해야



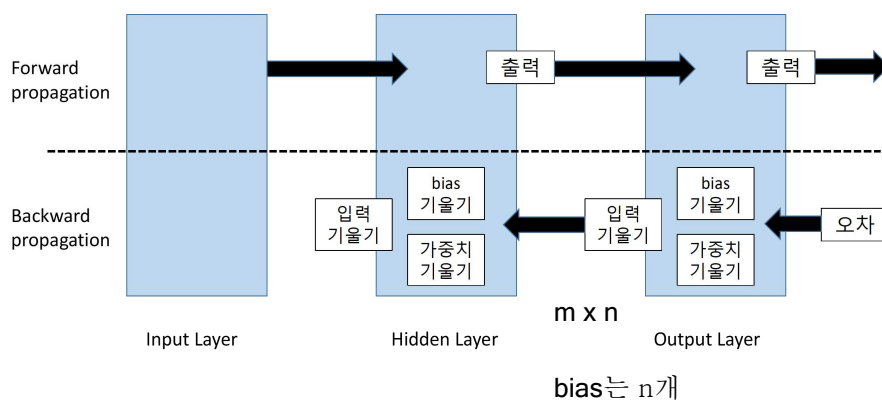
적절한 보폭을 설정해야 함

### ● 가중치와 bias를 수정하는 방법

$$w = w - \eta \frac{\partial E}{\partial w} \rightarrow \text{학습률}$$

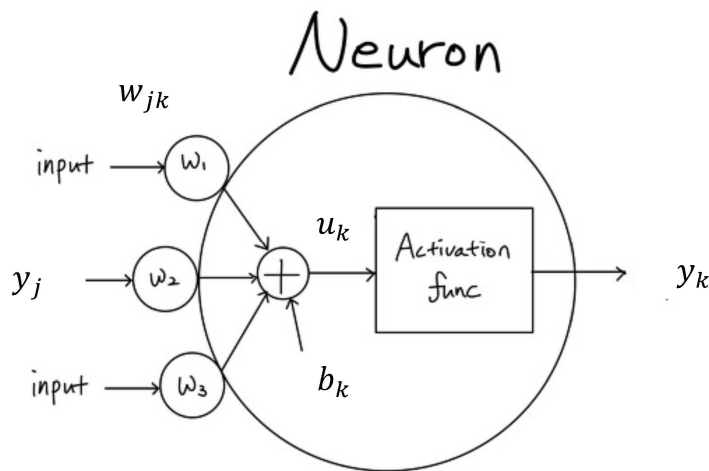
$$b = b - \eta \frac{\partial E}{\partial b}$$

## 2. 기울기를 구하는 개념



### 3. 출력층에서 가중치와 bias 기울기

층	첨자	뉴런 수
입력	i	l
은닉	j	m
출력	k	n



1) 오차를 가중치  $w_{jk}$  로 편미분한  $\frac{\partial E}{\partial w_{jk}}$  를 구함

가중치 기울기를  $\partial w_{jk}$  라고 하면 다음과 같음

$$\partial w_{jk} = \frac{\partial E}{\partial w_{jk}}$$

합성함수 미분 법칙을 이용하면 다음과 같음

$$\partial w_{jk} = \frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial u_k} \frac{\partial u_k}{\partial w_{jk}} \quad (1)$$

우변  $\frac{\partial u_k}{\partial w_{jk}}$  부분의 분자는 은닉층에 있는 여러 뉴런의 출력값과 가중치 곱의 합에 bias를 더한 것이므로 다음과 같음

$$\frac{\partial u_k}{\partial w_{jk}} = \frac{\partial (\sum_{q=1}^m y_q w_{qk} + b_k)}{\partial w_{jk}}$$

$$= \frac{\partial}{\partial w_{jk}} (y_1 w_{1k} + y_2 w_{2k} + \dots + y_j w_{jk} + y_m w_{mk} + b_k)$$

$$= y_j \quad (2)$$

② = j번째 들어오는 입력값

식 ①의 왼쪽부분  $\frac{\partial E}{\partial u_k}$  은 다음과 같이 전개 가능

$$\frac{\partial E}{\partial u_k} = \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial u_k} \rightarrow \text{활성함수 오차를 k번째 } y \text{로 편미분 한것과 그 출력값을 k번째 } u \text{로 편미분한 값의 곱}$$

k임 j아님

즉, 오차를 출력층 뉴런의 출력으로 편미분한 것과 그 출력값을  $u_k$  로 편미분한 것의 곱이 됨

이 때  $\frac{\partial E}{\partial y_k}$  은 손실 함수를 편미분해서 구할 수 있고

$\frac{\partial y_k}{\partial u_k}$  은 활성화 함수를 편미분해서 구할 수 있음

그 결과를  $\delta_k$  라고 하면 다음과 같다.

$$\delta_k = \frac{\partial E}{\partial u_k} = \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial u_k} \quad (3)$$

※  $\delta_k$  를 지정하는 이유는, 현재 상태에서 손실함수와 활성화 함수가 정해지지 않았기 때문

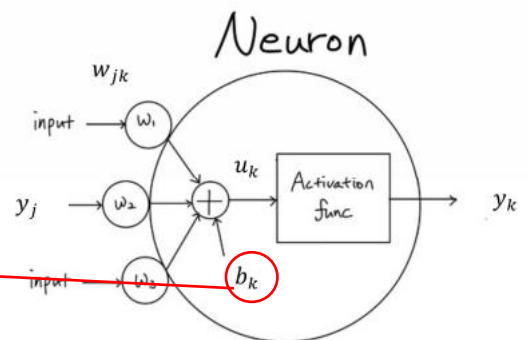
식 ②와 ③을 이용하여 식 ①을 다음과 정의

$$\partial w_{jk} = y_j \delta_k$$

가중치 기울기  $\partial w_{jk}$  를  $y_j$  와  $\delta_k$  의 곱으로 표시

2) bias의 기울기도 동일한 방법으로 구함

$$\partial b_k = \frac{\partial E}{\partial b_k}$$



활성함수 미분법칙에 의해 다음과 같이 변형 ①

$$\partial b_k = \frac{\partial E}{\partial b_k} = \frac{\partial E}{\partial u_k} \frac{\partial u_k}{\partial b_k}$$

이 때 우변의  $\frac{\partial u_k}{\partial b_k}$  부분은 아래와 같음

$$\frac{\partial u_k}{\partial b_k} = \frac{\partial (\sum_{q=1}^m y_q w_{qk}) + b_k}{\partial b_k}$$

$$= \frac{\partial}{\partial b_k} (y_1 w_{1k} + y_2 w_{2k} + \dots + y_j w_{jk} + y_m w_{mk} + b_k)$$

$$= 1$$

식 ①의  $\frac{\partial E}{\partial u_k}$  부분은 가중치의 기울기 경우와 같기 때문에  $\delta_k$  라고 한다면 다음과 같이 정리

$$\partial b_k = \delta_k$$

즉, bias 의 기울기는  $\delta_k$  와 동일한 값

#### 4. 출력층에서 입력값 기울기

1) 출력층에서의 입력값 기울기는  $\frac{\partial E}{\partial y_j}$ , 즉 은닉층의 출력기울기를 말하며 다음과 같이 줄여서 표현

$$\partial y_j = \frac{\partial E}{\partial y_j} \quad j\text{번째 } y\text{의 영향력을 계산하기 위해서}$$

합성함수 미분 공식에 의해 다음과 같이 변형 가능

①

$$\partial y_j = \frac{\partial E}{\partial y_j} = \sum_{r=1}^n \frac{\partial E}{\partial u_r} \frac{\partial u_r}{\partial y_j}$$

출력 뉴런 수? 모든 u의 계산에 영향력이 있기때문에

은닉층의 뉴런 1개는 출력층의 n개의 모든 뉴런에 영향을 주므로 위와 같이 다 더해야 함

식 ①의 우변 중  $\frac{\partial u_r}{\partial y_j}$  는 다음과 같이 구할 수 있음

$$\begin{aligned} \frac{\partial u_r}{\partial y_j} &= \frac{\partial (\sum_{q=1}^m y_q w_{qr} + b_r)}{\partial y_j} \\ &= \frac{\partial}{\partial y_j} (y_1 w_{1r} + y_2 w_{2r} + \dots + y_j w_{jr} + y_m w_{mr} + b_r) \\ &= w_{jr} \end{aligned}$$

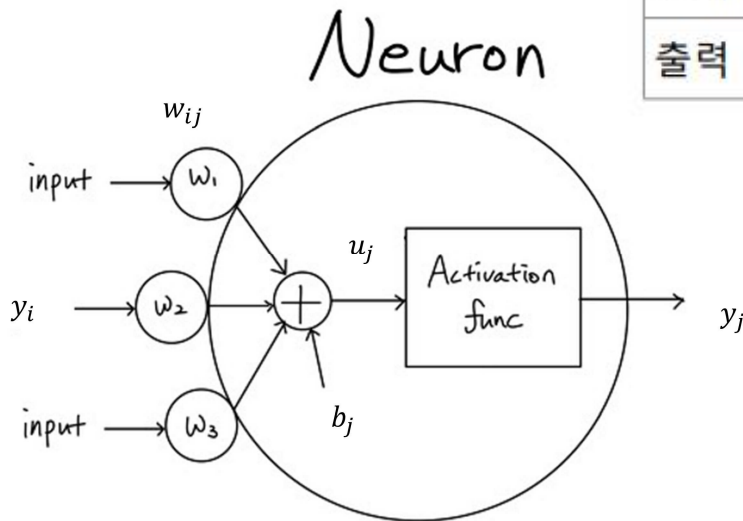
q번째 행의 r번째 열

여기서  $\delta_r = \frac{\partial E}{\partial u_r}$  로 설정하면 식 ①은 다음과 같이 정리됨

$$\partial y_j = \sum_{r=1}^n \delta_r w_{jr}$$

## 5. 은닉층 기울기

층	첨자	뉴런 수
입력	i	l
은닉	j	m
출력	k	n



### 1) 가중치 기울기

출력층의 경우와 마찬가지로 다음의 관계가 성립

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial u_j} \frac{\partial u_j}{\partial w_{ij}} \quad (1)$$

위 식에서 우변의  $\frac{\partial u_j}{\partial w_{ij}}$  부분은 다음과 같이 구할 수 있음

$$\begin{aligned} \frac{\partial u_j}{\partial w_{ij}} &= \frac{\partial (\sum_{p=1}^l y_p w_{pj} + b_j)}{\partial w_{ij}} \\ &= \frac{\partial}{\partial w_{ij}} (y_1 w_{1j} + y_2 w_{2j} + \dots + y_i w_{ij} + y_l w_{lj} + b_j) \\ &= y_i \quad (2) \end{aligned}$$

위 결과는 출력층의 경우와 동일함

식 (1)의 왼쪽부분  $\frac{\partial E}{\partial u_j}$  은 다음과 같이 전개 가능

$$\frac{\partial E}{\partial u_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial u_j}$$

이 식의 우변  $\frac{\partial y_j}{\partial u_j}$  은 활성화 함수 미분으로 구할 수 있고,  $\frac{\partial E}{\partial y_j}$  부분은 은닉층의 출력 기울기이고 이것은 이전에 출력층에서 구한  $\partial y_j$  이다. 이  $\partial y_j$  를

이용해 위 식을 다음과 같이  $\delta$  로 나타냄

$$\delta_j = \frac{\partial E}{\partial u_j} = \partial y_j \frac{\partial y_j}{\partial u_j} \quad (3)$$

이처럼  $\delta_j$  를 구하기 위해서는 출력층에서 구한  $\partial y_j$  를 사용함, 즉 신경망을 거슬러 올라가는 것임

식 ①과 ③을 대입하면 다음과 같은 식이 완성됨

$$\partial w_{ij} = y_i \delta_j$$

## 2) bias 기울기

bias의 기울기  $\partial b_j$  는 다음과 같이 나타낼 수 있음 ①

$$\partial b_j = \frac{\partial E}{\partial b_j} = \frac{\partial E}{\partial u_j} \frac{\partial u_j}{\partial b_j}$$

위 식에서 우변의  $\frac{\partial u_j}{\partial b_j}$  부분은 다음과 같이 정리됨

$$\begin{aligned} \frac{\partial u_j}{\partial b_j} &= \frac{\partial (\sum_{p=1}^l y_p w_{pj} + b_j)}{\partial b_j} \\ &= \frac{\partial}{\partial b_j} (y_1 w_{1j} + y_2 w_{2j} + \cdots + y_i w_{ij} + y_l w_{lj} + b_j) \end{aligned}$$

$$= 1$$

위 값을 식 ①에 적용 하면 다음과 같이 나타낼 수 있음

$$\partial b_j = \delta_j$$

이 층의 앞에 은닉층이 더 있을 경우 다음과 같이  $\partial y_i$  를 구해서 전파시킴

$$\partial y_i = \sum_{q=1}^m \delta_q w_{iq}$$

## 6. 기울기 최종 정리

### 1) 출력층

$$\delta_k = \frac{\partial E}{\partial u_k} = \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial u_k}$$

$$\partial w_{jk} = y_j \delta_k$$

$$\partial b_k = \delta_k$$

$$\partial y_j = \sum_{r=1}^n \delta_r w_{jr}$$

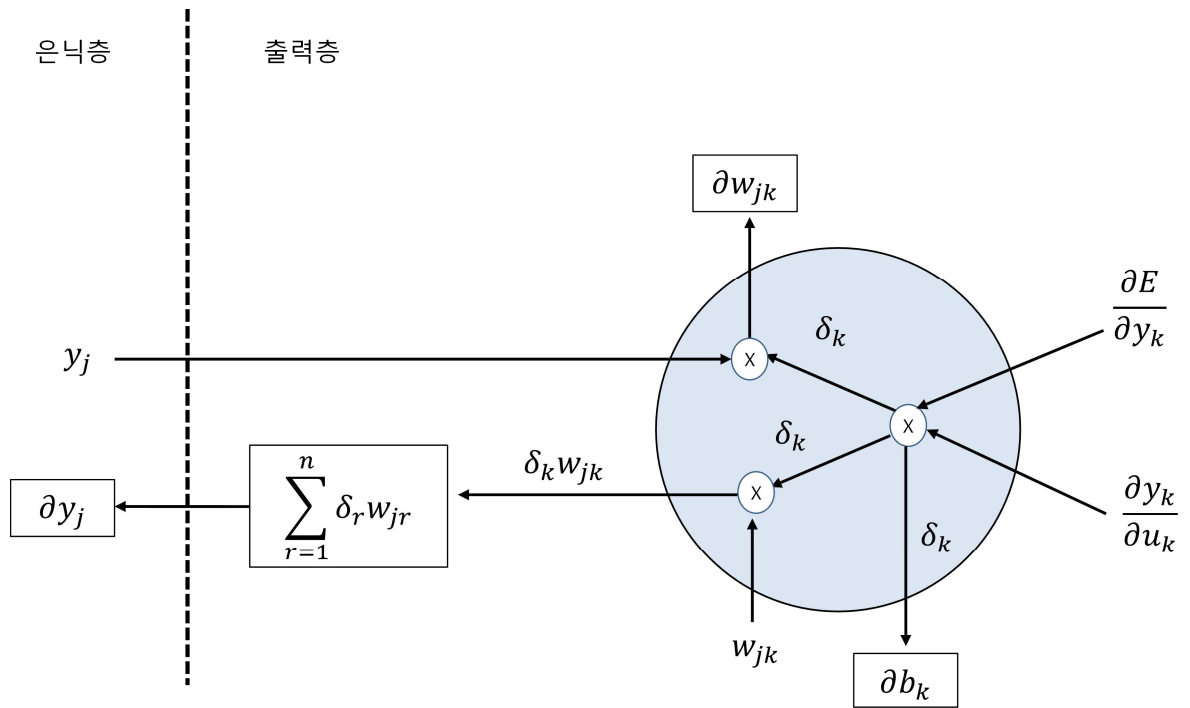
$$\partial w_{jk} = \frac{\partial E}{\partial w_{jk}}$$

$$\partial b_k = \frac{\partial E}{\partial b_k}$$

$$\partial y_j = \frac{\partial E}{\partial y_j}$$

$\delta_k$  를 구하는 방법은 손실 함수와 활성화 함수의 조합에 따라 다르며,  $\delta_k$  의 수는 출력층 뉴런의 수와 동일

출력층 뉴런에서의 역전파를 그림으로 나타내면 아래와 같음



## 2) 은닉층

$$\delta_j = \frac{\partial E}{\partial u_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial u_j} = \delta y_j \frac{\partial y_j}{\partial u_j}$$

$$\partial w_{ij} = y_i \delta_j$$

$$\partial b_j = \delta_j$$

$$\partial y_i = \sum_{q=1}^m \delta_q w_{iq}$$

$$\partial w_{ij} = \frac{\partial E}{\partial w_{ij}}$$

$$\partial b_j = \frac{\partial E}{\partial b_j}$$

$$\partial y_i = \frac{\partial E}{\partial y_i}$$



