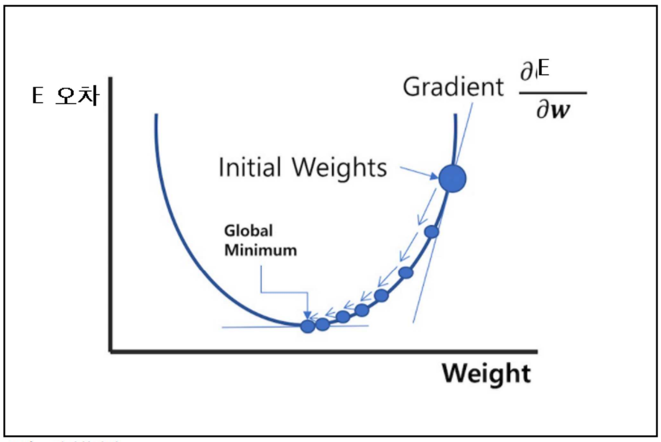
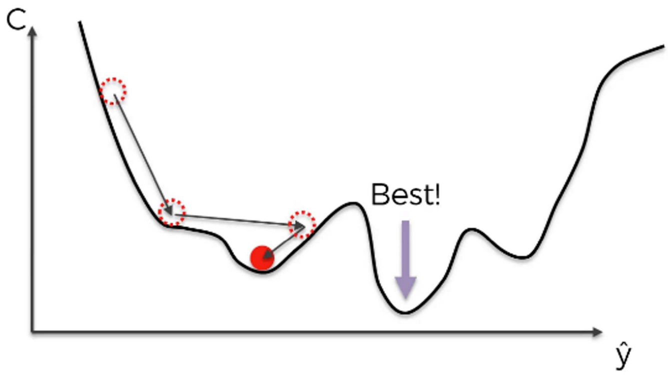
04.1. 경사하강법

2020년 12월 28일 월요일　　오전 10:01

1. 개념



오차 E와 가중치 간의 함수를 알고 있다고 가정하자
가중치를 조절해서 최소 오차값을 도출하자
가중치를 어떻게 조절할까?
편미분을 통해 특정 점에서 기울기를 보며 가중치를 조절하자



● 가중치와 bias를 수정하는 방법

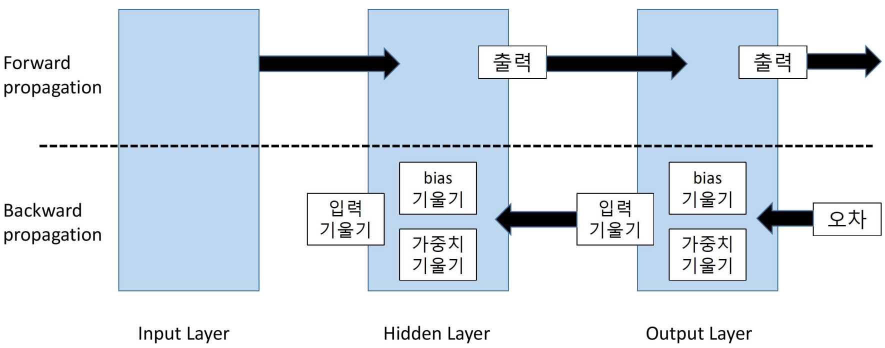
$$w = w - \eta \frac{\partial E}{\partial w}$$

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

에타 : 학습률

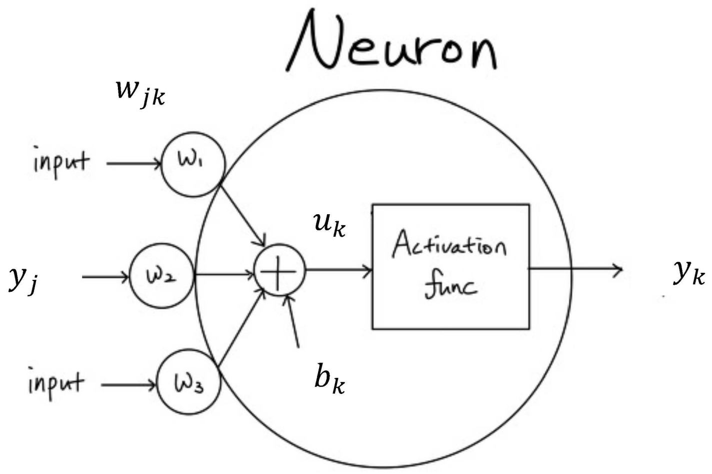
cf) w,b: 파라미터
에타: 하이퍼 파라미터
사용자 정의 파라미터

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 $\frac{\partial \mathrm{E}}{\partial w_{jk}}$ 를 구함

가중치 기울기를 $\partial \mathrm{w}_{jk}$ 라고 하면 다음과 같음

$\partial \mathrm{w}_{jk} = \frac{\partial \mathrm{E}}{\partial w_{jk}}$

합성함수 미분 법칙을 이용하면 다음과 같음

$\partial \mathrm{w}_{jk} = \frac{\partial \mathrm{E}}{\partial w_{jk}} = \frac{\partial \mathrm{E}}{\partial \mathrm{u}_k}\frac{\partial u_k}{\partial \mathrm{w}_{jk}}$ ①

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

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

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

$= y_j$ ②

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

$\frac{\partial \mathrm{E}}{\partial \mathrm{u}_k} = \frac{\partial \mathrm{E}}{\partial y_k}\frac{\partial y_k}{\partial u_k}$

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

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

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

그 결과를 δ_k 라고 하면 다음과 같다.

$\delta_k = \frac{\partial \mathrm{E}}{\partial \mathrm{u}_k} = \frac{\partial \mathrm{E}}{\partial y_k}\frac{\partial y_k}{\partial u_k}$ ③

※ δ_k 를 지정하는 이유는, 현재 상태에서 손실함수와 활성화 함수가 정해지지 않았기 때문

식 ②와 ③을 이용하여 식 ①을 다음과 정의

$\partial \mathrm{w}_{jk} = y_j \delta_k$

가중치 기울기 $\partial \mathrm{w}_{jk}$ 를 y_j 와 δ_k 의 곱으로 표시

2) bias의 기울기도 동일한 방법으로 구함

$$\partial \mathrm{b}_k = \frac{\partial \mathrm{E}}{\partial \mathrm{b}_k}$$

합성함수 미분법칙에 의해 다음과 같이 변형 ①

$$\partial \mathrm{b}_k = \frac{\partial \mathrm{E}}{\partial \mathrm{b}_k} = \frac{\partial \mathrm{E}}{\partial \mathrm{u}_k}\frac{\partial u_k}{\partial \mathrm{b}_k}$$

이 때 우변의 $\frac{\partial u_k}{\partial \mathrm{b}_k}$ 부분은 아래와 같음

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

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

$= 1$

식 ① 의 $\frac{\partial \mathrm{E}}{\partial \mathrm{u}_k}$ 부분은 가중치의 기울기 경우와 같기 때문에 δ_k 라고 한다면 다음과 같이 정리

$\partial \mathrm{b}_k = \delta_k$

즉, bias 의 기울기는 δ_k 와 동일한 값

4. 출력층에서 입력값 기울기

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

$$\partial \mathrm{y}_j = \frac{\partial \mathrm{E}}{\partial \mathrm{y}_j}$$

합성함수 미분 공식에 의해 다음과 같이 변형 가능

①

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

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

$$\frac{\partial u_r}{\partial y_j} = \frac{\partial (\sum_{q=1}^{\mathrm{m}} y_q\, w_{qr} + b_r)}{\partial y_j}$$

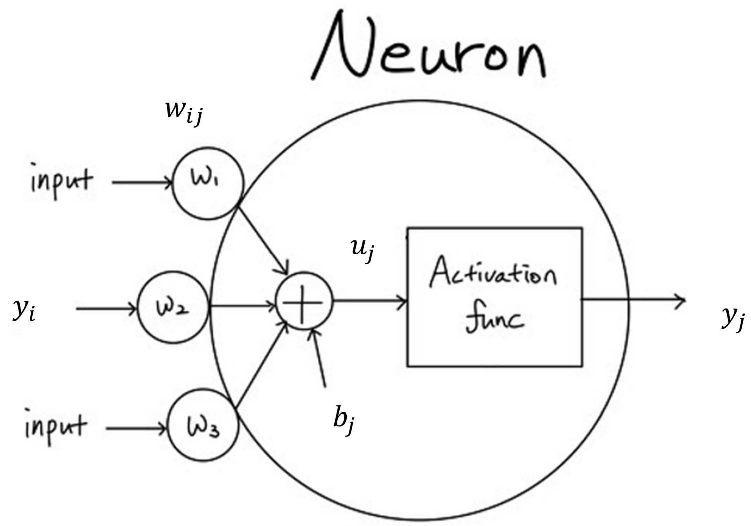
$$= \frac{\partial}{\partial y_j}(y_1 w_{1r} + y_2 w_{2r} + \cdots + y_j w_{jr} + y_m w_{mr} + b_r)$$

$$= w_{jr}$$

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

$$\partial \mathrm{y}_j = \sum_{r=1}^{n} \delta_r w_{jr}$$

5. 은닉층 기울기



1) 가중치 기울기

출력층의 경우와 마찬가지로 다음의 관계가 성립

$\partial \mathrm{w}_{ij} = \frac{\partial \mathrm{E}}{\partial w_{ij}} = \frac{\partial \mathrm{E}}{\partial \mathrm{u}_j}\frac{\partial u_j}{\partial \mathrm{w}_{ij}}$ ①

위 식에서 우변의 $\frac{\partial u_i}{\partial \mathrm{w}_{ij}}$ 부분은 다음과 같이 구할 수 있음

$$\frac{\partial u_j}{\partial \mathrm{w}_{ij}} = \frac{\partial(\sum_{p=1}^{l} y_p\, w_{pj} + b_j)}{\partial w_{ij}}$$

$$= \frac{\partial}{\partial w_{ij}}(y_1 w_{1j} + y_2 w_{2j} + \cdots + y_i w_{ij} + y_l w_{lj} + b_j)$$

$$= y_i \quad ②$$

위 결과는 출력층의 경우와 동일함

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

$\frac{\partial \mathrm{E}}{\partial \mathrm{u}_j} = \frac{\partial \mathrm{E}}{\partial y_j}\frac{\partial yj}{\partial u_j}$

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

이용해 위 식을 다음과 같이 δ 로 나타냄

$$\delta_j = \frac{\partial \mathrm{E}}{\partial \mathrm{u}_j} = \partial \mathrm{y}_j \frac{\partial yj}{\partial u_j} \quad ③$$

이처럼 δ_j 를 구하기 위해서는 출력층에서 구한 $\partial \mathrm{y}_j$ 를 사용함, 즉 신경망을 거슬러 올라가는 것임

식 ①과 ③을 대입하면 다음과 같은 식이 완성됨

$$\partial \mathrm{w}_{ij} = y_i \delta_j$$

2) bias 기울기

bias의 기울기 $\partial \mathrm{b}_j$ 는 다음과 같이 나타낼 수 있음 ①

$$\partial \mathrm{b}_j = \frac{\partial \mathrm{E}}{\partial \mathrm{b}_j} = \frac{\partial \mathrm{E}}{\partial \mathrm{u}_j} \frac{\partial u_j}{\partial \mathrm{b}_j}$$

위 식에서 우변의 $\frac{\partial u_j}{\partial \mathrm{b}_j}$ 부분은 다음과 같이 정리됨

$$\frac{\partial u_{\mathrm{j}}}{\partial \mathrm{b}_j} = \frac{\partial \left(\sum_{p=1}^{\mathrm{l}} y_p w_{pj} + b_j\right)}{\partial b_j}$$

$$= \frac{\partial}{\partial b_j}(y_1 w_{1j} + y_2 w_{2j} + \cdots + y_i w_{ij} + y_l w_{lj} + b_j)$$

$$= 1$$

위 값을 식 ①에 적용 하면 다음과 같이 나타낼 수 있음

$$\partial \mathrm{b}_j = \delta_j$$

이 층의 앞에 은닉층이 더 있을 경우 다음과 같이 $\partial \mathrm{y}_i$ 를 구해서 전파시킴

$$\partial \mathrm{y}_i = \sum_{q=1}^{m} \delta_q w_{iq}$$

6. 기울기 최종 정리

1) 출력층

$$\delta_k = \frac{\partial \mathrm{E}}{\partial \mathrm{u}_k} = \frac{\partial \mathrm{E}}{\partial y_k}\frac{\partial y_k}{\partial u_k}$$

$$\partial \mathrm{w}_{jk} = y_j \delta_k$$

$$\partial \mathrm{b}_k = \delta_k$$

$$\partial \mathrm{y}_j = \sum_{r=1}^{n} \delta_r w_{jr}$$

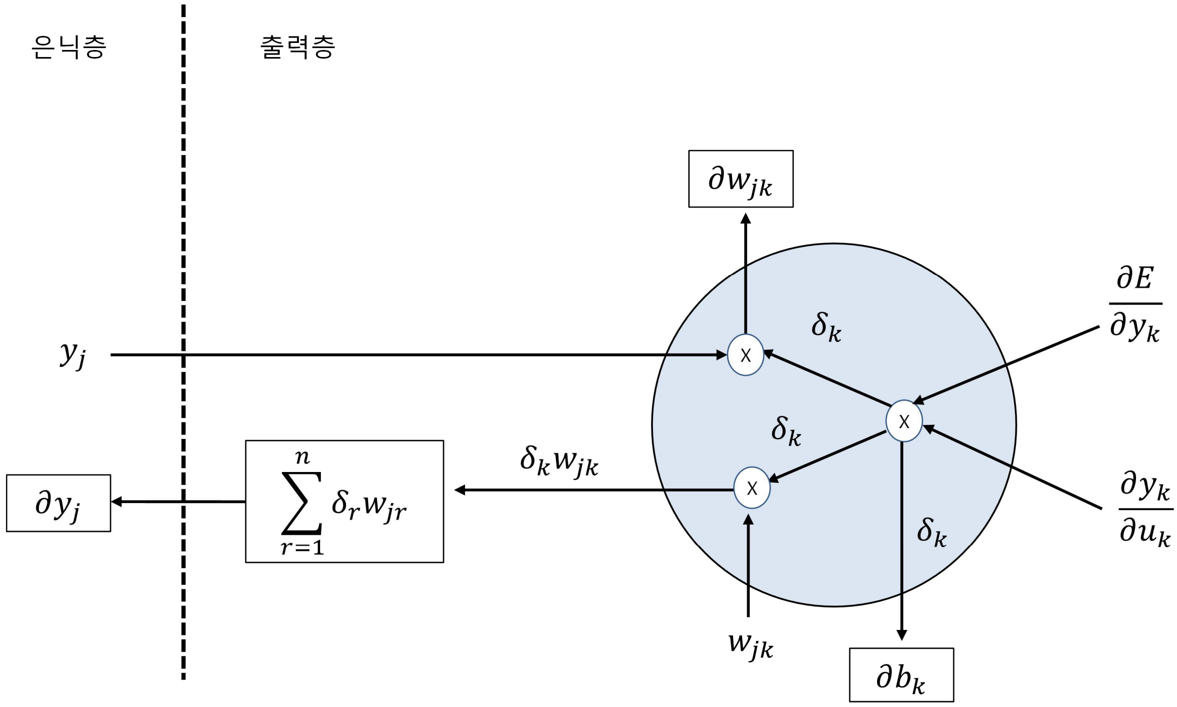
$$\partial \mathrm{w}_{jk} = \frac{\partial \mathrm{E}}{\partial w_{jk}}$$

$$\partial \mathrm{b}_k = \frac{\partial \mathrm{E}}{\partial \mathrm{b}_k}$$

$$\partial \mathrm{y}_j = \frac{\partial \mathrm{E}}{\partial \mathrm{y}_j}$$

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

출력층 뉴런에서의 역전파를 그림으로 나타내면 아래와 같음



2) 은닉층

$$\delta_j = \frac{\partial \mathrm{E}}{\partial \mathrm{u}_j} = \frac{\partial \mathrm{E}}{\partial y_j}\frac{\partial y_j}{\partial u_j} = \partial \mathrm{y}_j \frac{\partial y_j}{\partial u_j}$$

$$\partial \mathrm{w}_{ij} = y_i \delta_j$$

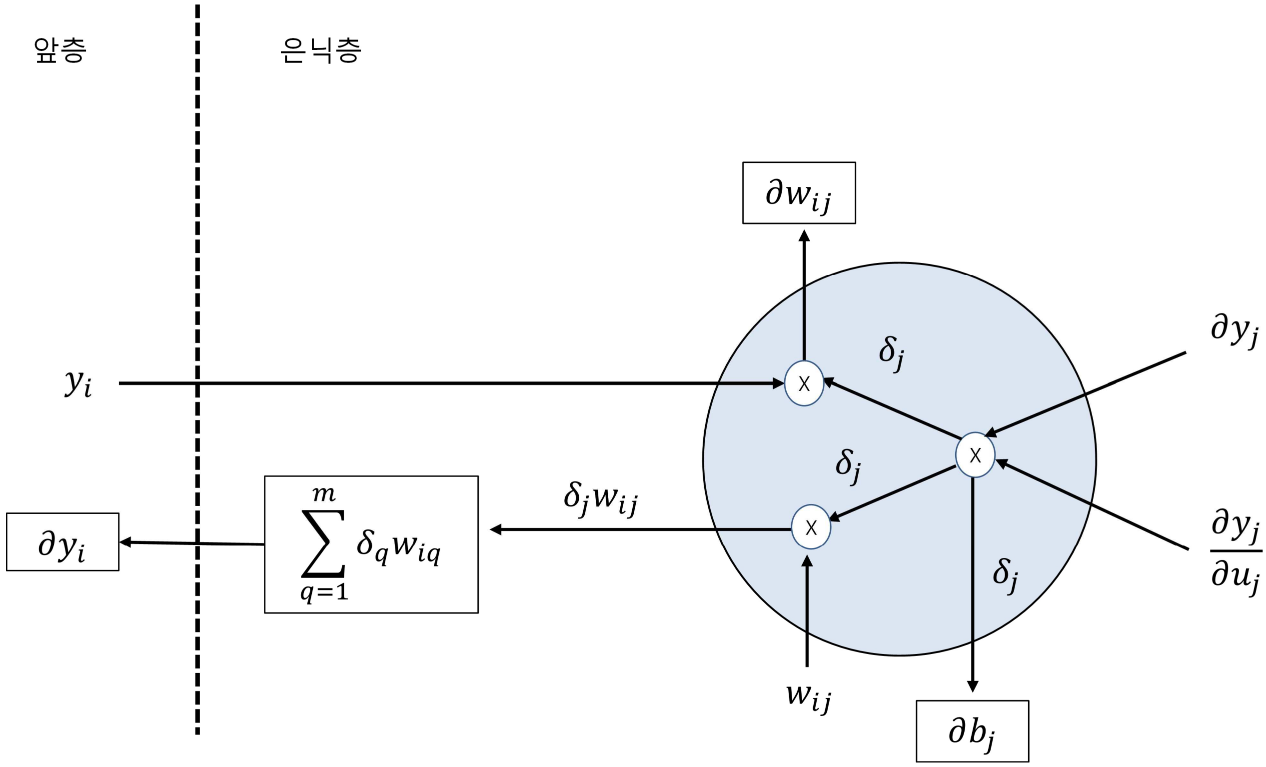
$$\partial \mathrm{b}_j = \delta_j$$

$$\partial \mathrm{y}_i = \sum_{q=1}^{m} \delta_q w_{iq}$$

$$\partial \mathrm{w}_{ij} = \frac{\partial \mathrm{E}}{\partial w_{ij}}$$

$$\partial \mathrm{b}_j = \frac{\partial \mathrm{E}}{\partial \mathrm{b}_j}$$

$$\partial \mathrm{y}_i = \frac{\partial \mathrm{E}}{\partial \mathrm{y}_i}$$


앞층
은닉층
∂w_{ij}
y_i
δ_j
∂y_j
δ_j
$\delta_j w_{ij}$
$\sum_{q=1}^{m} \delta_q w_{iq}$
∂y_i
$\frac{\partial y_j}{\partial u_j}$
δ_j
w_{ij}
∂b_j