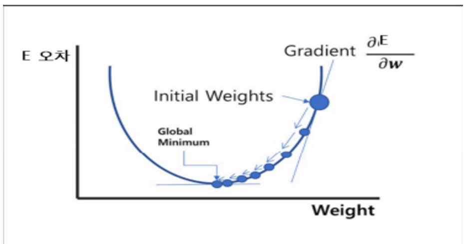
1교시

2022년 10월 26일 수요일   오전 8:56

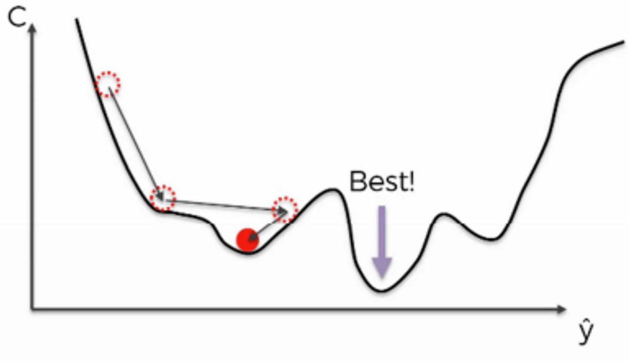
원노트 좋은데?
단점 - 타자가 빠르면 꼬인다. 변방의 작은나라의 요구따위는 듣지 않는다.

어제 - 뉴런세포를 흉내낸 퍼셉트론이 나오고, 퍼셉트론의 과정을 수식으로도 보고, 구현도 해봄.

경사하강법 (Grandient Descent)
- 개념

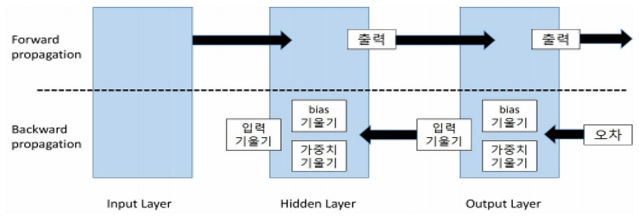


- W로편미분한다, 편미분한 값이 양수면 왼쪽으로이동(- 감소)
- 편미분값이 음수면 오른쪽으로이동 (+ 증가)
- 로컬미니멈에 빠져서 글로벌미니멈에 도달하지 못하는 경우 발생



- w = w - η(보폭) * ∂E/∂w // η = 에타 = 학습률(Learning Rate)
- b = b - η(보폭) * ∂E/∂b

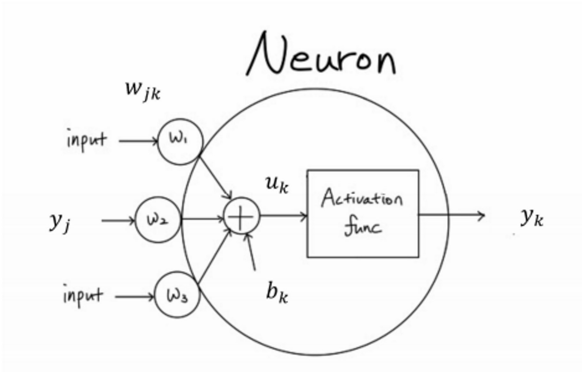
2. 기울기를 구하는 개념

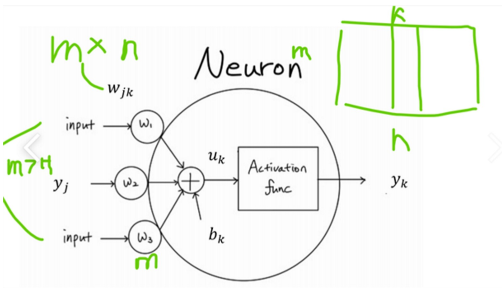


- Forward Propagation 순전파
- Backword propagation 역전파
- 오차에 오차를 편미분 하는 것을 반복함

3. 출력층에서 가중치와 bias 기울기

층	첨자	뉴런 수
입력	i	l
은닉	j	m
출력	k	n





- 출력층에 해당하는 입력은 m개, 따라서 하나는 Yj,
- K번째 출력을 만드는 뉴런 -> 출력값이 yk이므로,

Bias도 bk이고, 가중치는 (m,n)매트릭스여야 하므로, Wjk가 된다.

나온 출력값 Yk의 오차를 다시 가중치 W로 편미분함 -> ∂E / ∂Wjk = 기울기

-> 기울기 : ∂Wjk = (∂E / ∂uk) * (∂uk / ∂Wjk) -> 1번

∂uk = (Y1W1k + bk) + (Y2W2k + bk) + (Y3W3k + bk) + ... + (YmWmk + bk) = Yj -> 2번

1번의 왼쪽부분 ∂E / ∂uk는 다음과 같이 전개가능∂E / ∂uk = (∂E / ∂Yk) * (∂Yk / ∂uk)

여기서 현재상태에서 손실함수와 활성화함수가 정의되지 않았기 때문에 더 뽀갤수 없고, 이 때 값을 델타k라고 정의한다.

δk = ∂E / ∂uk -> 3번

2번과 3번식을 이용해서 1번식을 다음과같이 정의함

∂Wjk = yj * δk

가중치 기울기 ∂Wjk 를 Yj 와 델타k의 곱으로 표시

Bias의 기울기도 동일한 방법으로 구함

∂bk = ∂E / ∂bk

합성함수 미분법칙에 의해 다음과 같이 변형

∂bk = ∂E / ∂bk = (∂E / ∂uk) * (∂uk / ∂bk)

이 때 우변의 ∂uk / ∂bk 부분은 아래와 같음

∂uk / ∂bk = ∂ / ∂bk * (y1W1k + y2W2k + ... + yjWjk + YmWmk + bk) = 1

-> bk로 편미분하기때문에 다른 상수들은 다 사라지고 bk의앞에 값, 따라서 1
여기에 수식을 입력하십시오.

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

출력층에서 입력값 기울기

4. 출력층에서 입력값 기울기

1) 출력층에서의 입력값 기울기는 $\frac{\partial \mathrm{E}}{\partial \mathrm{y}_j}$, 즉 은닉층의 출력기울기을 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

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

= 최종오차에 영향을 준 양 (은닉층의 뉴런1개가 출력층에 모든값에 영향을 주었기 때문에)

=1 결과 (bk 편미분)

식 ① 의 $\frac{\partial \mathrm{E}}{\partial \mathrm{u}_k}$ 부분은 가중치의 기울기 경우와 같기 때문에 δ_k 라고 한다
면 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 출력기울기을 말하며 다
음과 같이 줄여서 표현

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

합성함수 미분 공식에 의해 다음과 같이 변형 가능
①

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

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

오른쪽 식 풀어헤치기

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

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

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

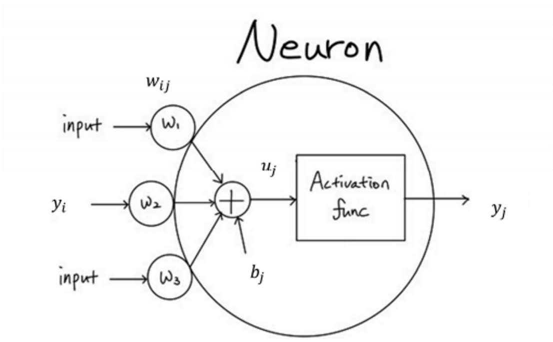
$$= w_{jr}$$

왼쪽꺼

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

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

은닉층 기울기



1) 가중치 기울기

출력증의 경우와 마찬가지로 다음의 관계가 성립

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

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

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

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

$$= y_i \quad ②$$

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

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

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

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

$$= 1$$

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

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