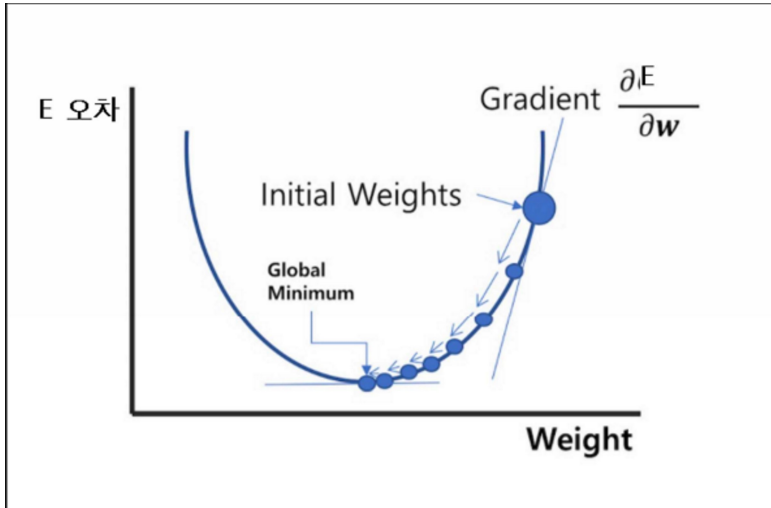


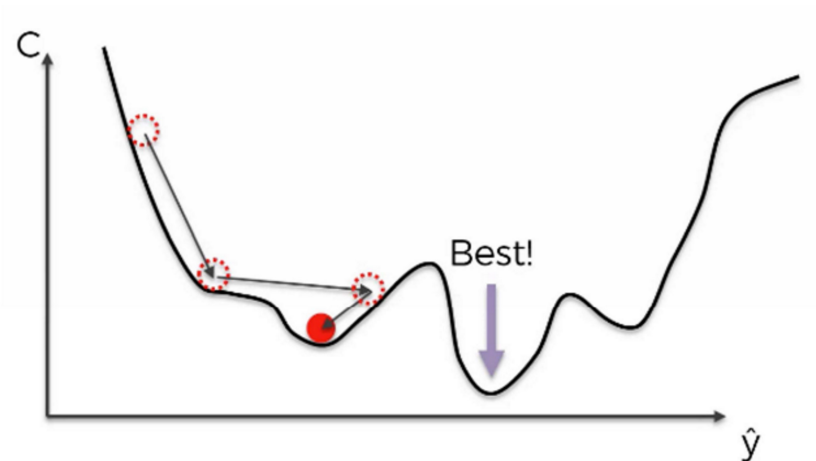
1번째

2022년 10월 25일 화요일 오후 2:03

- 어제는 인간의 뉴런 세포를 카피한 '퍼셉트론'에 대해 알아보았음
 - Pyplot으로 시각화
 - 활성화 함수 사용
- 오늘은 경사하강법에 대해서 알아볼 것 (딥러닝의 핵심)
 - Gradient Descent (통칭 GD)



- 가로는 수많은 파라미터중에 한 놈이고, 그 한놈의 값을 변경시켰을때 변하는 오차값을 파악
- 그러면 어떻게 파라미터 중 한 놈을 이동시키냐?
- 오차를 해당 파라미터로 편미분하여 접선의 기울기가 양수일경우는 감소, 음수일때는 증가시켜서 0일때로 보냄(Global minimum)
- 그런데 local minimum에 빠져서 global minimum(Best)으로 가기 전에 함수가 멈출 수 있다.



- 그래서 local minimum에 빠지지 않도록 보폭을 수정해줌

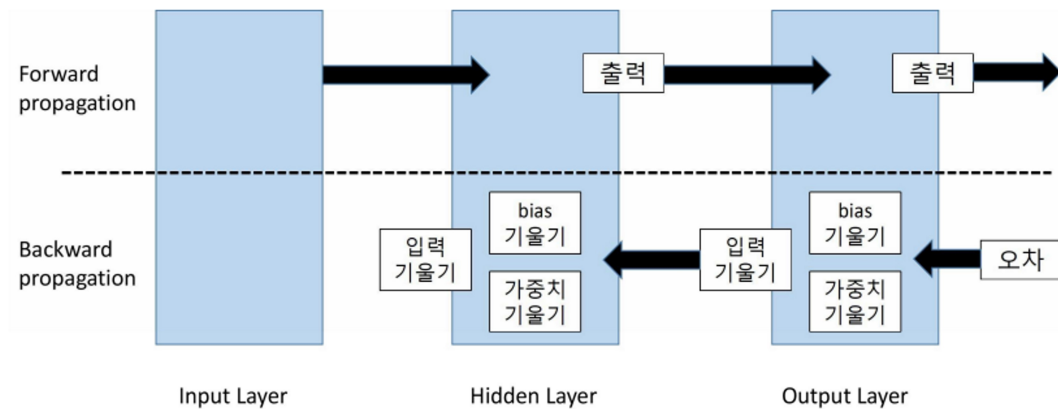
● 가중치와 bias를 수정하는 방법

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

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

- 보폭에 해당하는 에타(η) = 학습률(Learning Rate)이 커지면 global minimum까지 도달하는 시간이 감소하지만, 발산할 확률이 올라간다

2. 기울기를 구하는 개념

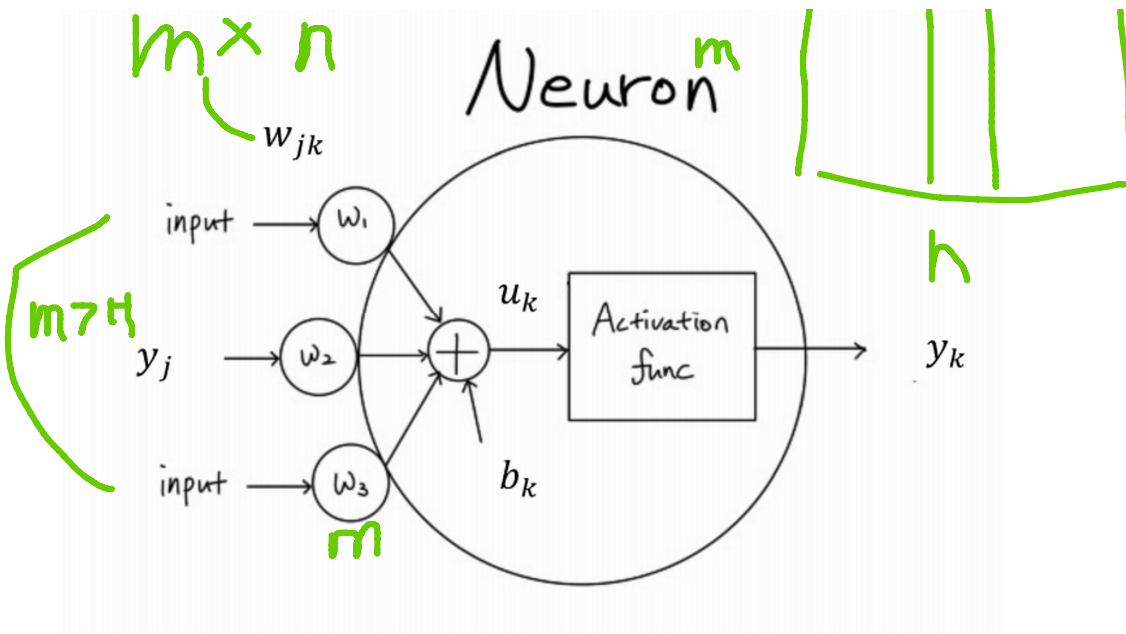


3. 출력층에서 가중치와 bias 기울기

층	첨자	뉴런 수
입력	i	l
은닉	j	m
출력	k	n

- l개중에 하나를 i, m개중에 하나를 j, n개중에 하나를 k라고 지칭
- 출력층을 먼저 구현 (k번째 일때)





- 출력층은 은닉층(hidden layer)에서 입력이 오기때문에 m개 m개가 들어와서 n개가 나가야하니까 $m \times n$ 이고 그 중에 k번째 열 출력층에 해당하는 입력은 m개, 따라서 하나는 y_j , k번째 출력을 만드는 뉴런 \rightarrow 출력값이 y_k 이므로, Bias도 b_k 이고, 가중치는 (m,n) 매트릭스여야 하므로, w_{jk} 가 된다.

- 실제로 편미분을 할 것은 아니고, 수식을 간소화 시켜서 파이썬 코드로 구현할 것임

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

가중치 기울기를 ∂w_{jk} 라고 하면 다음과 같음

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

합성함수 미분 법칙을 이용하면 다음과 같음

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

- 1번식의 오른쪽
- 합성함수 미분법칙을 이용한 ∂u_k = 행렬 곱의 합 + bias

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

- 풀어서 보면

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

$$= y_j \quad (2)$$

- 1번식의 왼쪽

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

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

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 δ_k 라고 하면 다음과 같다.

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

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

• 식 2번, 3번을 사용하여 1번을 재정의 할 수 있음

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

가중치 기울기 ∂w_{jk} 를 y_j 와 δ_k 의 곱으로 표시

- Bias 도 동일한 방법으로 구함 (다음 타임)