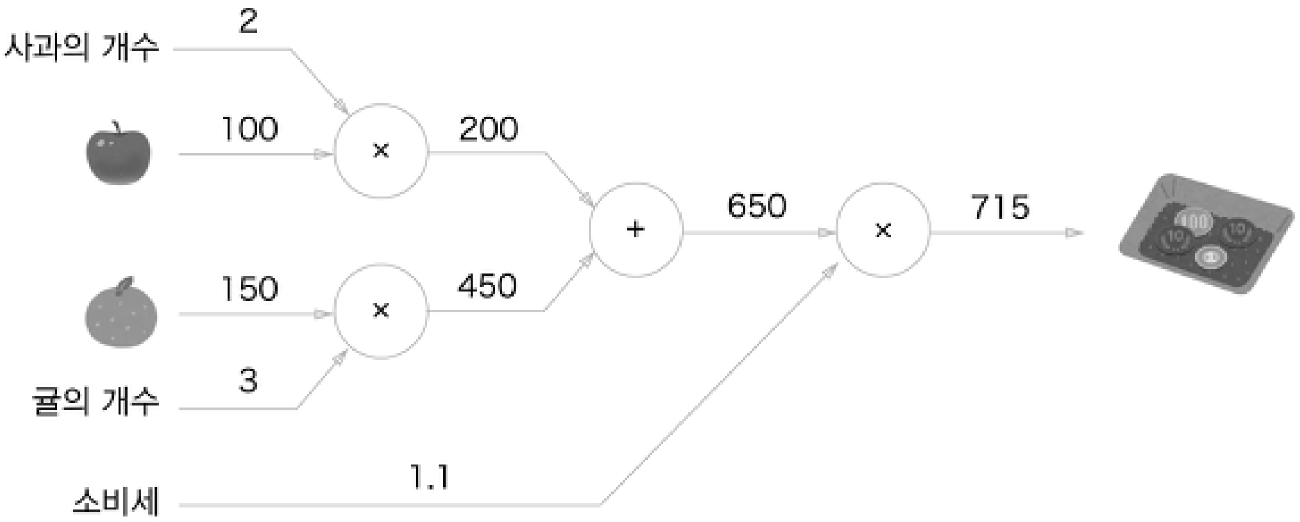
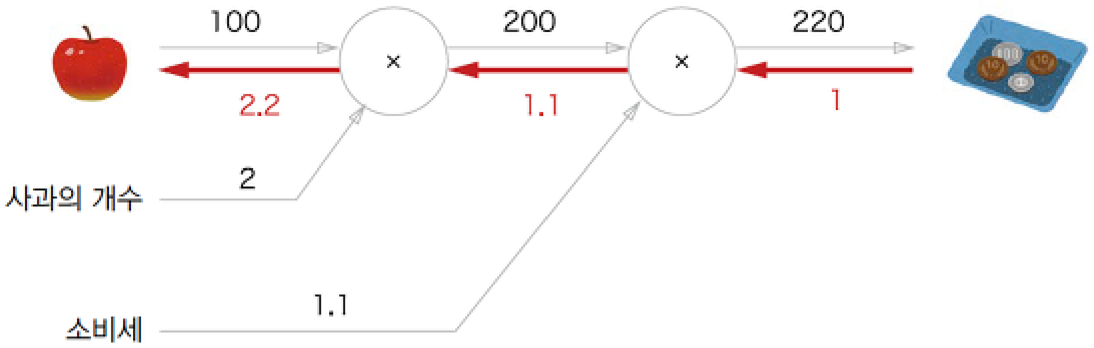
Computational Graph

- '계산을 왼쪽에서 오른쪽으로 진행'하는 단계를 순전파(forward propagation), 그 반대를 역전파(backward propagation)'
- 계산 그래프를 사용하면 국소적 계산'에 집중하여 문제를 단순화할 수 있습니다.
- 중간 계산 결과를 보관할 수 있고, 역전파를 통해 '미분'을 효율적으로 계산할 수 있다.





https://codingneedsinsanity.tistory.com/24

미분 공식

$y=c$ (c는 상수) $\Rightarrow$ $y'=0$

$y=x^n$ (n는 상수) $\Rightarrow$ $y'=nx^{n-1}$

$y=cf(x)$ (c는 상수) $\Rightarrow$ $y'=cf'(x)$

$y=f(x)\pm g(x)$ $\Rightarrow$ $y'=f'(x)\pm g'(x)$ (복부호 동순)

$y=f(x)g(x)$ $\Rightarrow$ $y'=f'(x)g(x)+f(x)g'(x)$

$y=f(x)g(x)h(x)$ $\Rightarrow$

$$y'=f'(x)g(x)h(x)+f(x)g'(x)h(x)+f(x)g(x)h'(x)$$

미분 공식

$$y = \frac{f(x)}{g(x)} \Rightarrow y' = \frac{f'(x)g(x) - f(x)g'(x)}{\{g(x)\}^2}$$

($g(x) \neq 0$)에서 두 함수 $f(x)$, $g(x)$가 미분가능임)

$$y = f(g(x)) \Rightarrow y' = f'(g(x))g'(x)$$

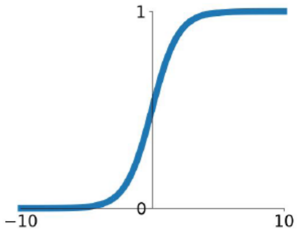
(단, $y = f(u)$, $u = g(x)$가 미분가능)

$y = \{f(x)\}^n$(n은 정수) $\Rightarrow$ $y' = n\{f(x)\}^{n-1}f'(x)$

Sigmoid 함수 미분

Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



$$Sigmoid \Rightarrow \quad F(x) = \frac{1}{(1+e^{-x})}$$

$$quotient \ \ rule \Rightarrow$$

$$F(x) = \frac{f(x)}{g(x)} \quad \Rightarrow \ Sigmoid : f(x) = 1 \ , \ g(x) = (1+e^{-x})$$

$$F'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$$

$$f'(x) = 0$$
$$g'(x) = -e^{-x}$$

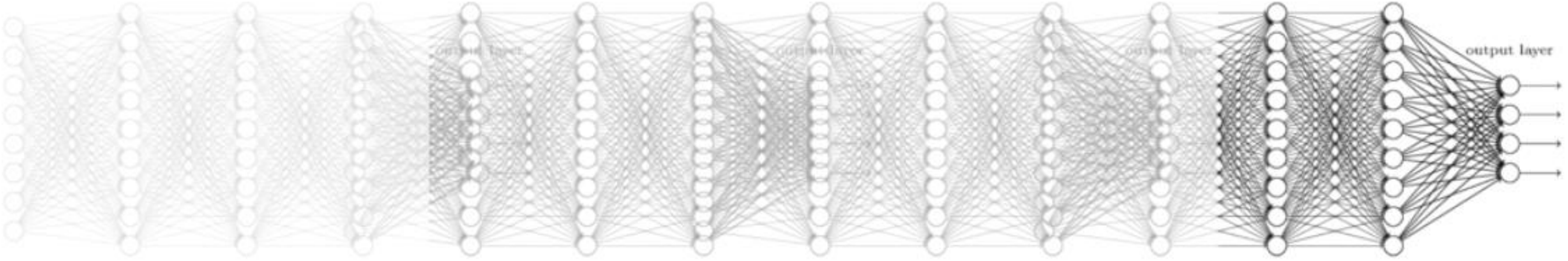
$$F'(x) = \frac{0 \cdot (1+e^{-x}) \ - \ 1 \cdot -e^{-x}}{(1+e^{-x})^2} = \frac{0 \cdot (1+e^{-x}) \ + \ e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{e^{-x} \ +1 \ -1}{(1+e^{-x})^2} = \frac{+1 \ + \ e^{-x}}{(1+e^{-x})^2} + \frac{-1}{(1+e^{-x})^2} = \frac{1}{(1+e^{-x})} - \frac{1}{(1+e^{-x})^2}$$

$$= \frac{1}{(1+e^{-x})}\left(1 \ - \ \frac{1}{(1+e^{-x})}\right)$$

$$= F(x)(1 \ - \ F(x))$$

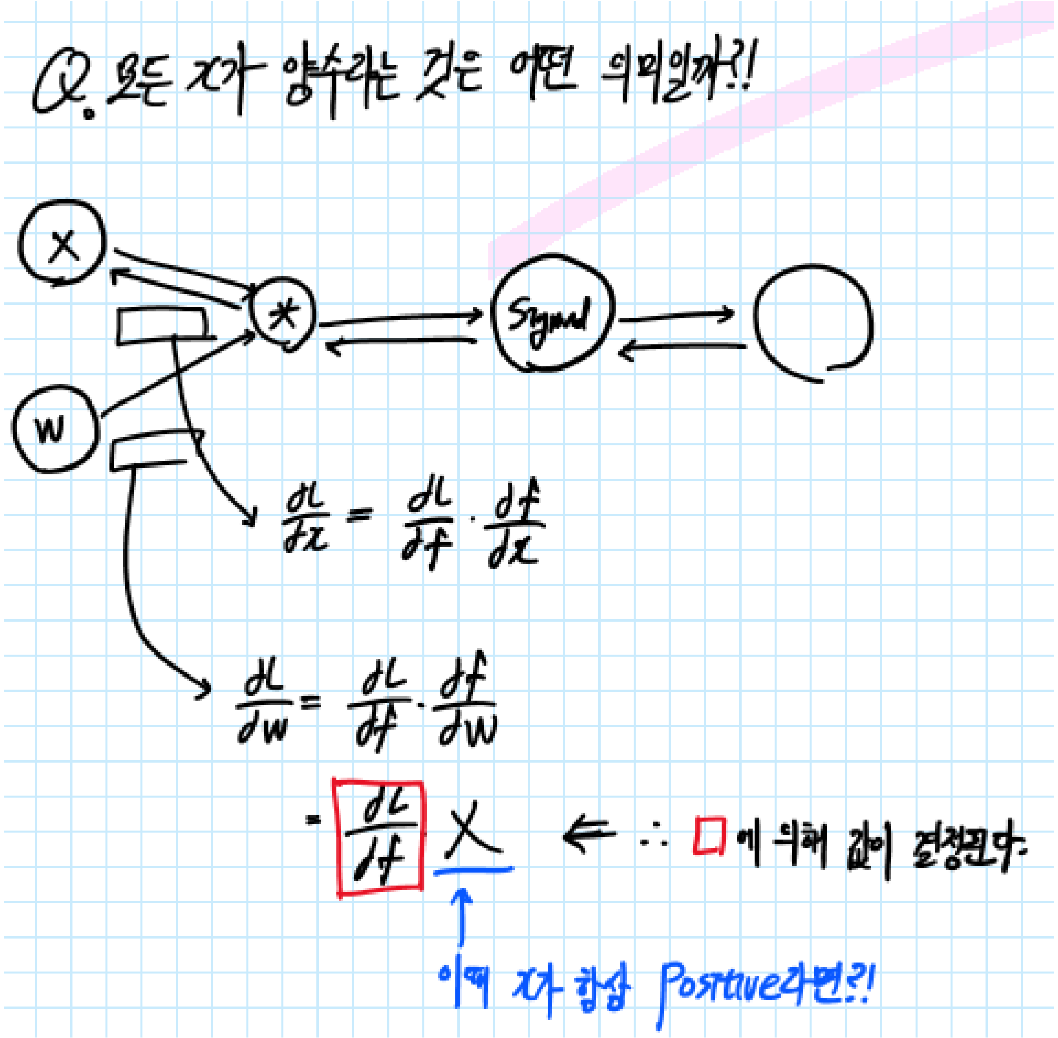
Vanishing gradient (NN winter2: 1986-2006)

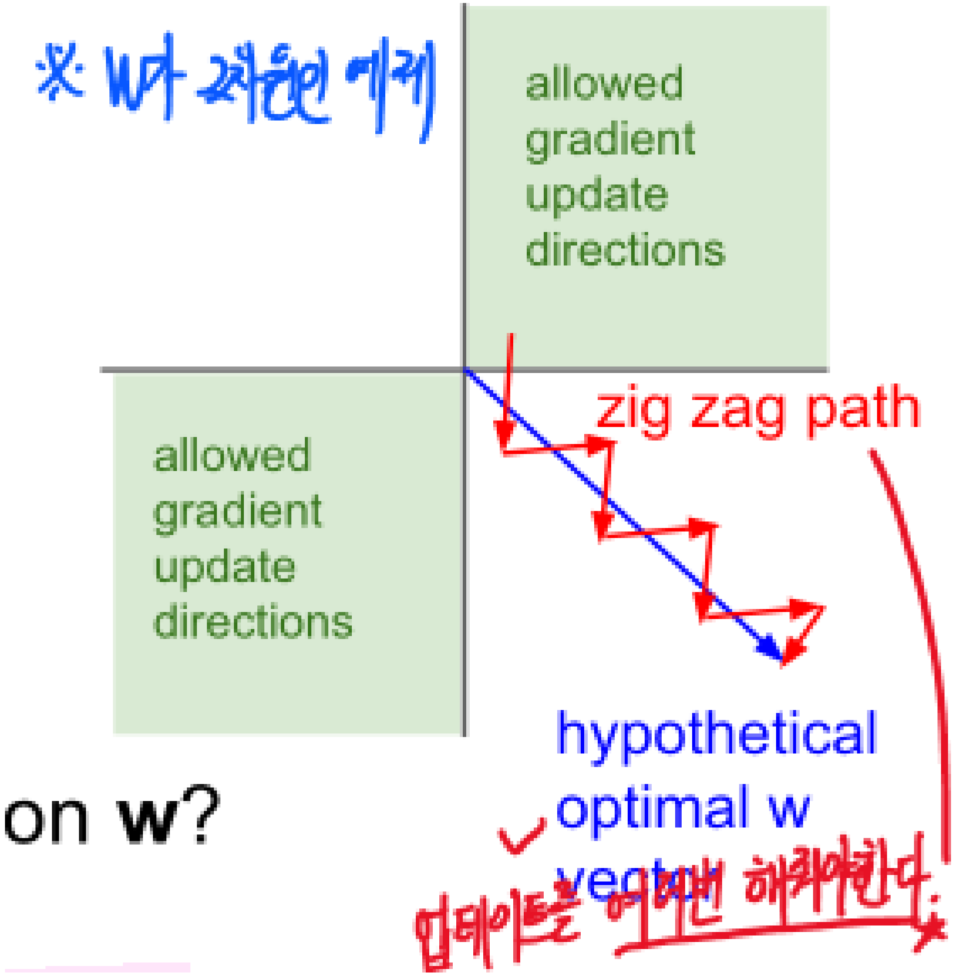


- mlp에서 역전파를 써도 10개 층처럼 깊은 중간층을 구성하면 gradient가 사라지는 끝까지 전달되지 못하고 점차 사라져 가는 문제가 발생했습니다. 이를 vanishing gradient문제라고 합니다.
- gradient값은 미분한 값이므로 특정 값이 어떠한 결과에 미치는 영향을 말합니다.
- 즉 gradient 값이 갈수록 원래 가진 값보다 작아지면서 전달되는 것은 원래의 영향력이 제대로 결과에 미치지 못하는 결과를 초래하는 것을 말합니다.

https://smartstuartkim.wordpress.com/2019/02/09/vanishing-gradient-problem/

not zero-centerd 의 문제점





https://deepinsight.tistory.com/113