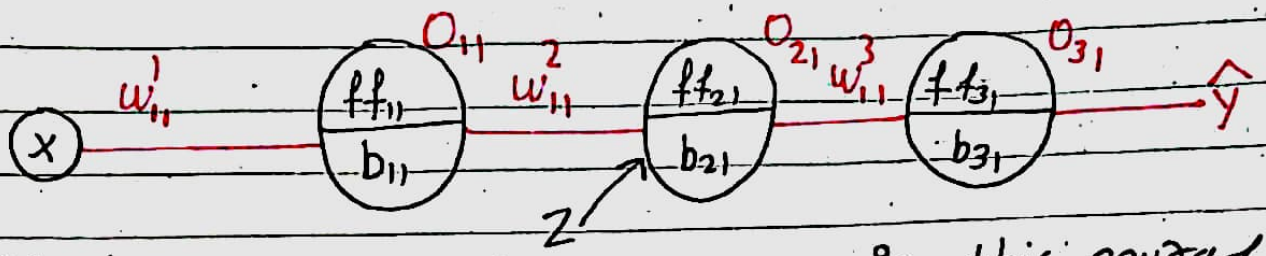
Exploding Gradient Problem

$ff = \text{Sigmoid}(z)$
$z = wx + b$



At first we feed our row in this neural network. It calculates $\hat{y}$, then loss. After calculating loss it performs back propagation to update parameters.

let's do for w'_{11} update

$$w'_{11_{new}} = w'_{11_{old}} - \alpha \underbrace{\frac{d\,loss}{dw'_{11_{old}}}}_{A}$$

$$\therefore A = \frac{d\,loss}{dw'_{11\,old}}$$

$[w'_{11} \to O_{11} \to O_{21} \to O_{31}(\hat{y}) \to loss]$

$$\Rightarrow \frac{d\,loss}{dw'_{11\,old}} = \frac{d\,loss}{dO_{31}} \times \frac{dO_{31}}{dO_{21}} \times \underbrace{\frac{dO_{21}}{dO_{11}}}_{B} \times \frac{dO_{11}}{dw'_{11}}$$

let's see for B for now,

$$\therefore B = \frac{d\,O_{21}}{d\,O_{11}}$$

$$O_{21} = \phi_{21}(z)$$

$$\phi_{21} = \frac{1}{1+e^{-z}} \quad [z = w_{11}^2 \cdot O_{11} + b_{21}]$$

$$\phi_{21} = \frac{1}{1+e^{-(w_{11}^2 \cdot O_{11} + b_{21})}}$$

$$\Rightarrow \frac{dO_{21}}{dO_{11}}$$

$$[O_{11} \rightarrow z \rightarrow O_{21}] \sim [\phi_{11} \rightarrow z \rightarrow \phi_{21}]$$

$$\Rightarrow \frac{dO_{21}}{dO_{11}} = \frac{d\phi_{21}}{dz} \times \frac{dz}{d\phi_{11}}$$

$$= [0 - 0.25] \times \frac{d(w_{11}^2 \cdot O_{11} + b_2)}{dO_{11}}$$

$$= 0.25 \times (w_{11}^2 + 0)$$

$$= 0.25 \times w_{11}^2$$

$$\therefore [0.25 \times w_{11}^2]$$

Suppose I initilize w_{11}^2 valve as 500 the valve belomes

$$= 0.25 \times 500$$

$$= 125$$

$$\therefore \frac{dO_{21}}{dO_{11}} = 125$$

now, Similaly if we have $\frac{dO_{31}}{dO_{21}} = 100$, $\frac{dO_{11}}{dw_{11}^1} = 50$, $\frac{dLOSS}{dO_{31}} = 150$

$$\Rightarrow \frac{dLOSS}{dw_{11}^1 old} = \frac{dLOSS}{dO_{31}} \times \frac{dO_{31}}{dO_{21}} \times \frac{dO_{21}}{dO_{11}} \times \frac{dO_{11}}{dw_{11}^1}$$

$$= 150 \times 100 \times 125 \times 50$$

$$= 93750000$$

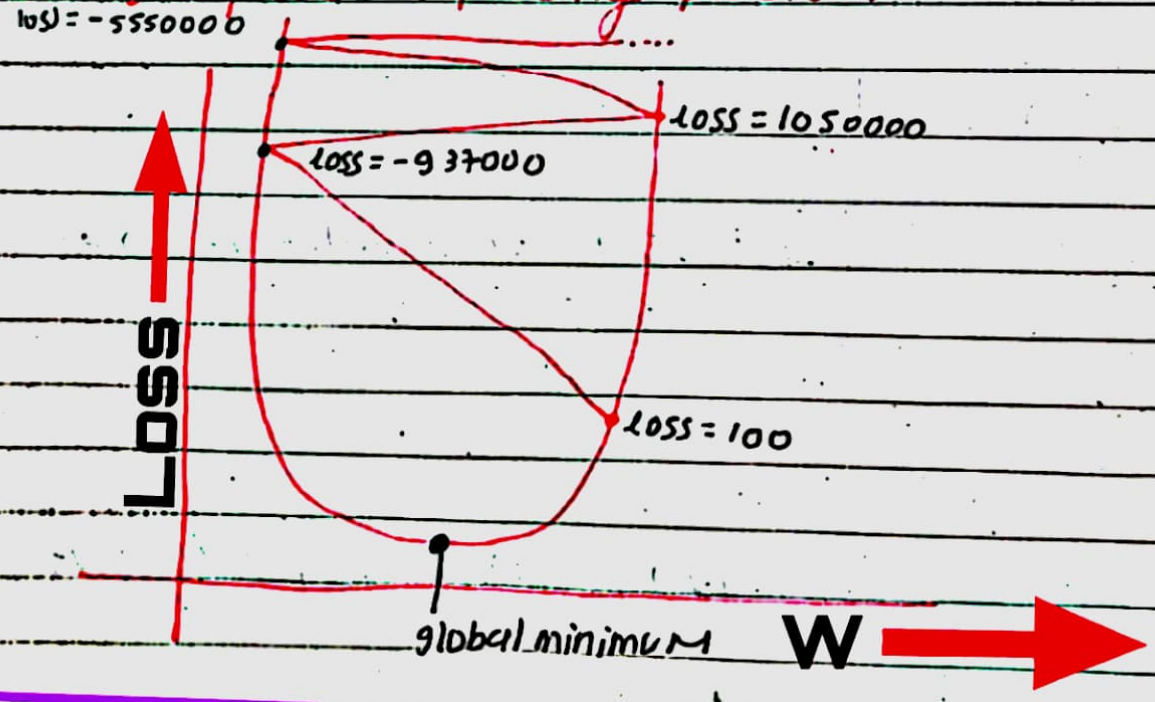
$$w'_{11\,new} = w'_{11\,old} - \alpha \frac{d\,loss}{d\,w'_{11\,old}}$$

$w'_{11\,new} = 500 - (0.01 \times 93750000)$

$w'_{11\,new} = 500 - 937500$

$w'_{11\,new} = -937000$

As, we can see our weight value is very high so, we will also get high loss value. due to high weight we will never reach to the global minimum we will face Exploding problem.



Conclusion: If we initilize the weights (also for bias) values very high we face exploding gradient problem. due to this reason, we have to be very careful during weight initilize and we various weight initilization techniques.