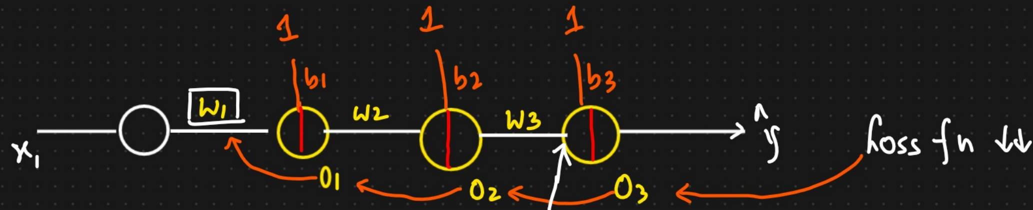
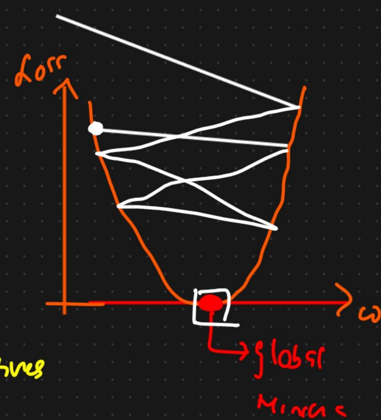
Exploding Gradient Problem $\Rightarrow$ Weight Initialization Technique



$\rightarrow z = (O_2 * w_3 + b_3) \rightarrow \sigma(z)$

Optimizers



$$W_{1new} = W_{1old} - \boxed{\eta \boxed{\frac{\partial L}{\partial W_{1old}}}}$$

$\left\{ \begin{array}{l} W_{1new} >>>> W_{1old} \\ W_{1new} <<<< W_{1old} \end{array} \right\}$

Chain Rule of Derivatives

$$\frac{\partial L}{\partial W_{1old}} = \frac{\partial L}{\partial O_3} * \boxed{\frac{\partial O_3}{\partial O_2}} * \frac{\partial O_2}{\partial O_1} * \frac{\partial O_1}{\partial W_{1old}}$$

big * big * big * big ⇒ big Value

Weight initialization
⇓
Very high value

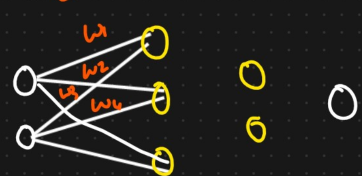
$$\frac{\partial O_3}{\partial O_2} = \boxed{\frac{\partial \sigma(z)}{\partial z}} * \frac{\partial z}{\partial O_2}$$

$$= [0 - 0.25] * \frac{\partial(O_2 * w_3 + b_3)}{\partial O_2}$$

$$= [0 - 0.25] * w_3 \Rightarrow 500 - 1000$$

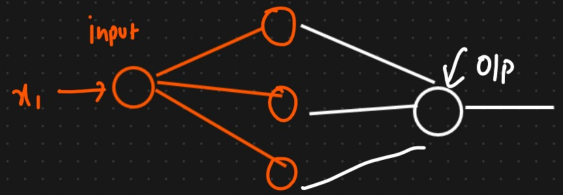
Weight Initializing Techniques

1. Uniform Distribution ✓
2. Xavier / Glorot Initialization ✓
3. Kaiming He Initialization ✓



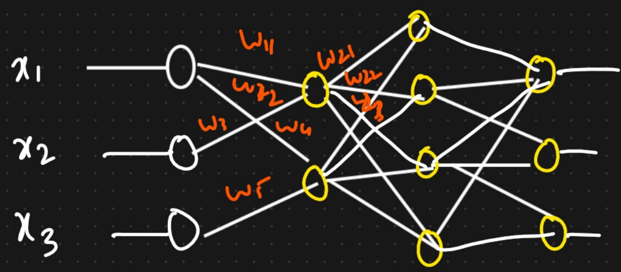
Key Points

1. Weights should be small ✓
2. Weights should not be same ✓
3. Weights should have good variance



input = 1

Output = 1



input = 3

O/P = 3

① Uniform Distribution

$$W_{ij} \approx \text{Uniform Distribution} \left[\frac{-1}{\sqrt{\text{input}}}, \frac{1}{\sqrt{\text{input}}} \right]$$

$$\left[\frac{-1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right]$$

② Xavier/Glorot Initialization

Researcher → Xavier Glorot

① Xavier Normal Init

$$W_{ij} \approx N(0, \sigma)$$

$$\sigma = \sqrt{\frac{2}{(\text{input} + \text{output})}}$$

② Xavier Uniform

$$W_{ij} \approx \text{Uniform Distribution} \left[\frac{-\sqrt{6}}{\sqrt{\text{input} + \text{output}}}, \frac{\sqrt{6}}{\sqrt{\text{i/p} + \text{o/p}}} \right]$$

③ Kaiming He Initialization

① He Normal

$$W_{ij} \approx N(0, \sigma)$$

$$\sigma = \sqrt{\frac{2}{\text{input}}}$$

② He uniform

$$W_{ij} \approx \underset{\text{Distribution}}{\text{Uniform}} \left[-\sqrt{\frac{6}{\text{input}}}, \sqrt{\frac{6}{\text{input}}} \right]$$