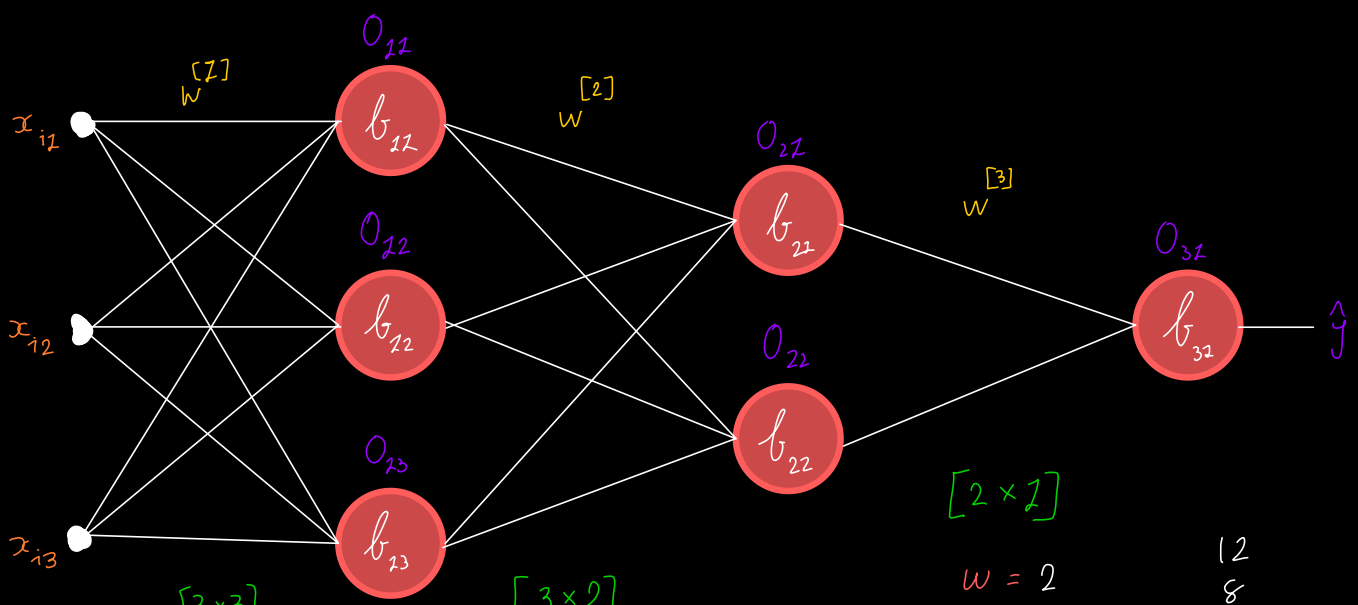
# Memonization in Neural Network

## Memonization :

in computing, memonization is optimization technique used primarily to speed-up computer programs by speed-up storing the results of expensive function calls & returning the cached results when the same input occurs again.

## MLP - Memonization




$[3 \times 3]$

$w = 9$

$b = 3$

$12$

$[3 \times 2]$

$w = 6$

$b = 2$

$8$

$[2 \times 1]$

$w = 2$

$b = 1$

$3$

$23$

trainable params

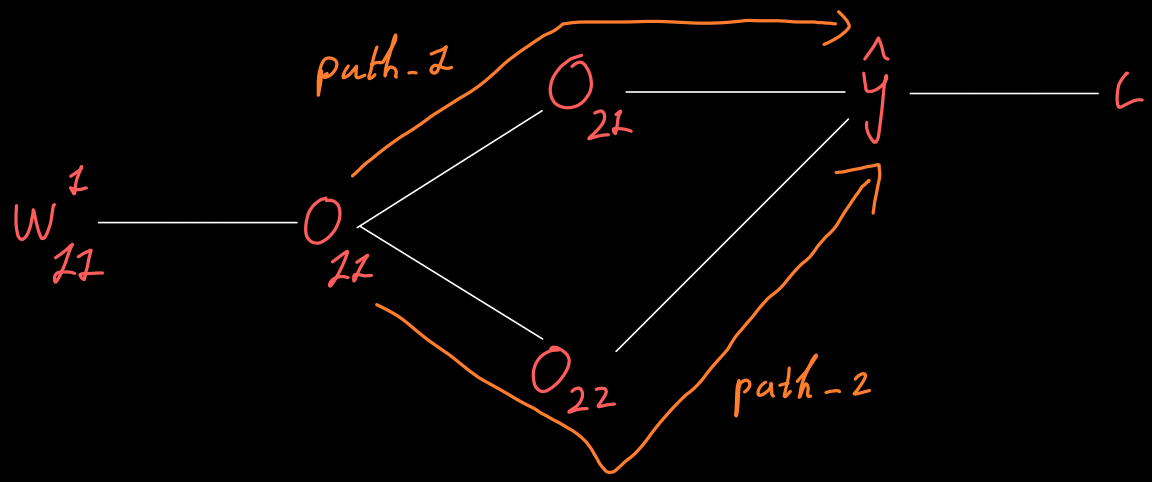
→ Based on above network structure, we can see that $W^{[1]}$ derivative calculation process will be much-more complex than $W^{[2]}$ and $W^{[3]}$.

$$\frac{\partial L}{\partial w^3_{11}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w^3_{11}}$$

$$L \to \hat{y} \to O_{21} \to O_{11} \to w^2_{11}$$

$$\frac{\partial L}{\partial w^2_{11}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_{21}} \frac{\partial O_{21}}{\partial w^2_{11}}$$

Created with Notewise




$$\frac{\partial L}{\partial w^1_{11}} = \frac{\partial L}{\partial \hat{y}} \times \left[ \underbrace{\frac{\partial \hat{y}}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial w^1_{11}}}_{\text{path-1}} + \underbrace{\frac{\partial \hat{y}}{\partial O_{22}} \times \frac{\partial O_{22}}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial w^1_{11}}}_{\text{path-2}} \right]$$

→ We can imagine that how things will be scary in big neural networks & how it will be difficult to calculate these values, hence we need to use memorization technique.

→ As we will go backward in neural-networ. & we need those derivatives which are already calculated. that's why we will store those values & reuse it.

Back propagation = Chain Rule + Memorization

Created with Notewise

$$\frac{\partial L}{\partial w_{11}^{1}} = \frac{\partial L}{\partial \hat{y}} \times \left[ \frac{\partial \hat{y}}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial w_{11}^{1}} + \frac{\partial \hat{y}}{\partial O_{22}} \times \frac{\partial O_{22}}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial w_{11}^{1}} \right]$$

→ As we can see in above formula that in both paths, $\frac{\partial O_{11}}{\partial w_{11}^{1}}$ is common hence we won't re-calculate it again & again.

→ We will store the o/p of $\frac{\partial O_{11}}{\partial w_{11}^{1}}$ in a memory and just use it.

Created with Notewise