# Optimisation

Step-3 (optimise weight for single Neural N/w) :-

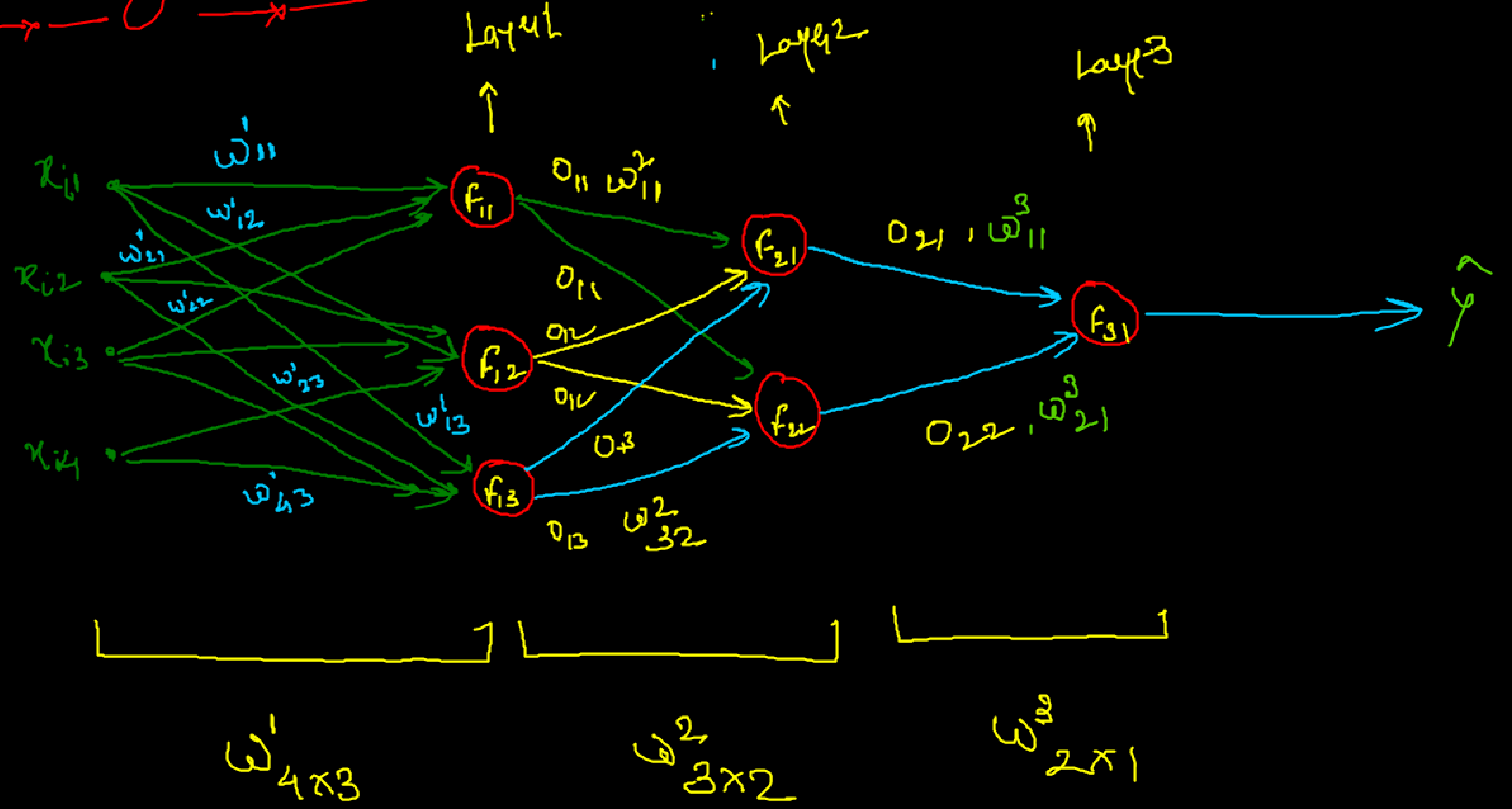
(a) initialize the weight $w_i$'s $\longrightarrow$ Randomaly.

(b)

$$\nabla_w L = \begin{bmatrix} \partial L/\partial w_1 \\ \partial L/\partial w_2 \\ \partial L/\partial w_3 \\ \vdots \end{bmatrix}$$

# Training an MLP:-

Notation




$$W^{k}_{ij}$$

K → layer

i → from

j → to

$O_{ij}$ → i → layer No.

j → input (Neuron)

⇒ So we need to optimised $W^1_{4\times3}, W^2_{3\times2}, W^3_{2\times1}$ = 20 weights.

Step-1

Loss function - $L = \sum_{i=1}^{n} (y-\hat{y})^2 + Reg.$ , Sq. Loss $= L_i = (y_i - \hat{y}_i)^2$

$\boxed{\text{Aim} \rightarrow \text{min Loss}}$ w.r.t. $w_{ij}^{K}$ -

Step-2

(a) initialize - $w_{ij}^{K}$ - randomly / many methods

(b) $$\boxed{(w_{ij}^{K})_{new} = (w_{ij}^{K})_{old} - \eta \left(\frac{\partial L}{\partial w_{ij}^{K}}\right)_{old}}$$

Aim $\rightarrow \left(\frac{\partial L}{\partial w_{ij}^{K}}\right)_{w \to old} \rightarrow$ find it.

(c) perform update untill the convergence ( $w_{new} \simeq w_{old}$ )

To calculate $\left(\frac{dL}{\partial w_{ij}^{k}}\right)$

$\rightarrow$ For $w^3$ :-

$$\boxed{\frac{dL}{dw_{11}^{3}} = \frac{dL}{dO_{31}} \times \frac{dO_{31}}{dw_{11}^{3}}} \longrightarrow \text{chain rule.}$$

$\rightarrow$ For $w^2$ :-

$$\boxed{\frac{dL}{dw_{11}^{2}} = \frac{dL}{dO_{31}} \times \frac{dO_{31}}{dO_{21}} \times \frac{dO_{21}}{dw_{11}^{2}}} \longrightarrow \text{chain rule}$$

$$\boxed{\frac{dL}{dw_{21}^{2}} = \frac{dL}{dO_{31}} \times \frac{dO_{31}}{dO_{21}} \times \frac{dO_{21}}{dw_{21}^{2}}}$$

→ For $w^1$

→ let a Case →

$$\begin{tikzcd} & f \arrow[dr] & & \\ x \arrow[ur] \arrow[dr] & & h \arrow[r] & K \\ & g \arrow[ur] & & \end{tikzcd}$$

$$\Rightarrow \left(\frac{dK}{dh} \times \frac{dh}{dx}\right)$$

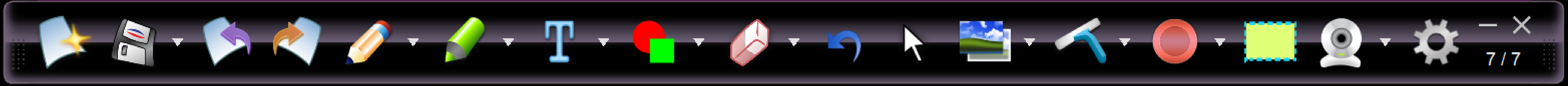
$$\Rightarrow \frac{dh}{dx} = \left(\frac{dh}{df} \times \frac{df}{dx} + \frac{dh}{dg} \times \frac{dg}{dx}\right)$$

$$\Rightarrow \frac{dL}{dw^1_{11}} = \left(\frac{dL}{dO_{31}} \times \frac{dO_{31}}{dw^1_{11}}\right)$$

$$\hookrightarrow \frac{dO_{31}}{dw^1_{11}} = \left(\frac{dO_{31}}{dO_{21}} \times \frac{dO_{21}}{dO_{11}} \times \frac{dO_{11}}{dw^1_{11}} + \frac{dO_{31}}{dO_{22}} \times \frac{dO_{22}}{dO_{11}} \times \frac{dO_{11}}{dw^1_{11}}\right)$$

$$\Rightarrow \frac{dL}{dw^1_{11}} = \frac{dL}{dO_{31}} \left(\quad\right)$$

⇒ we can do the same for other $w^k_{ij}$'s.


7 / 7