Back Propagation

Topics to be covered

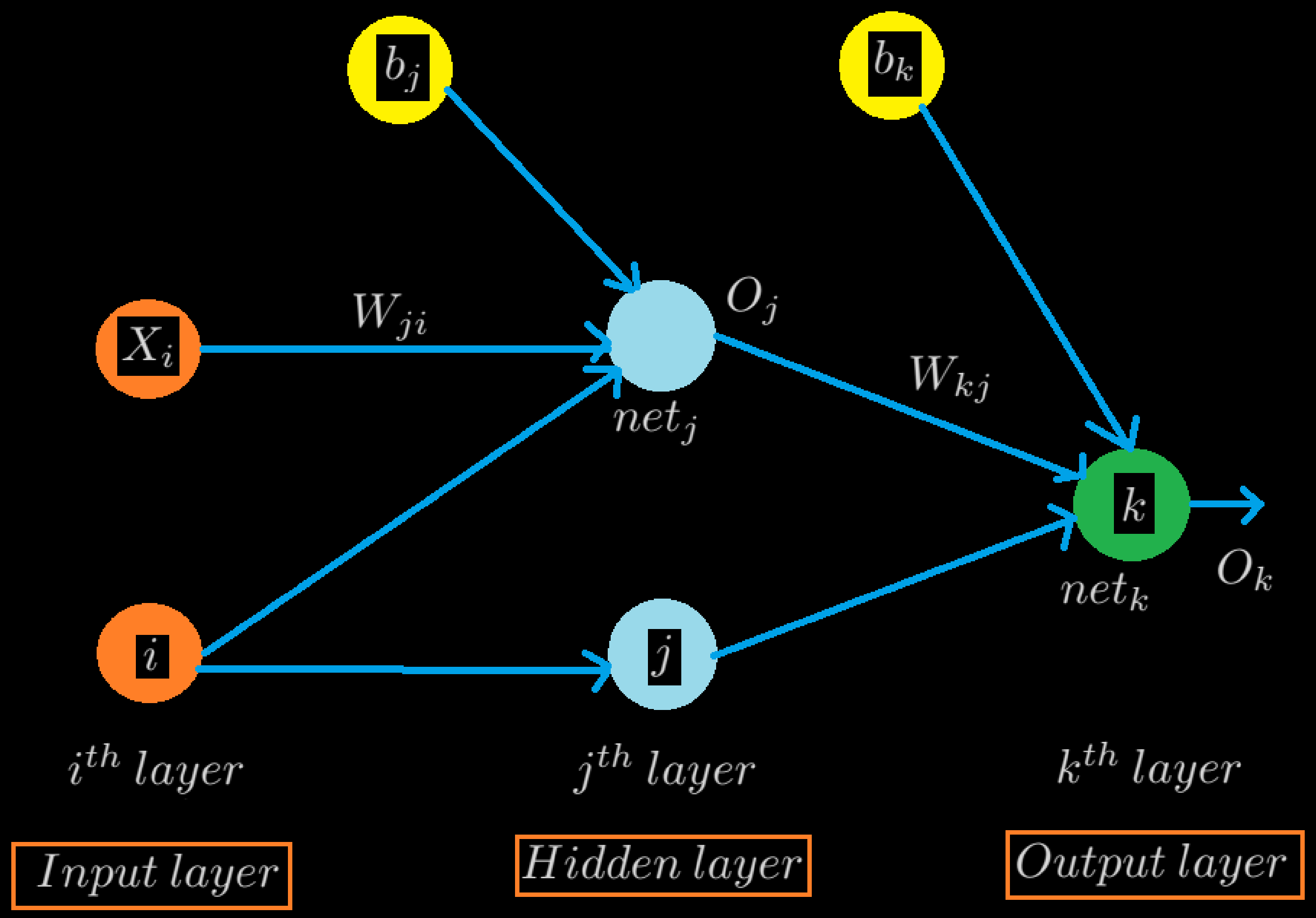
- Forward Propagation
- Backward Propagation

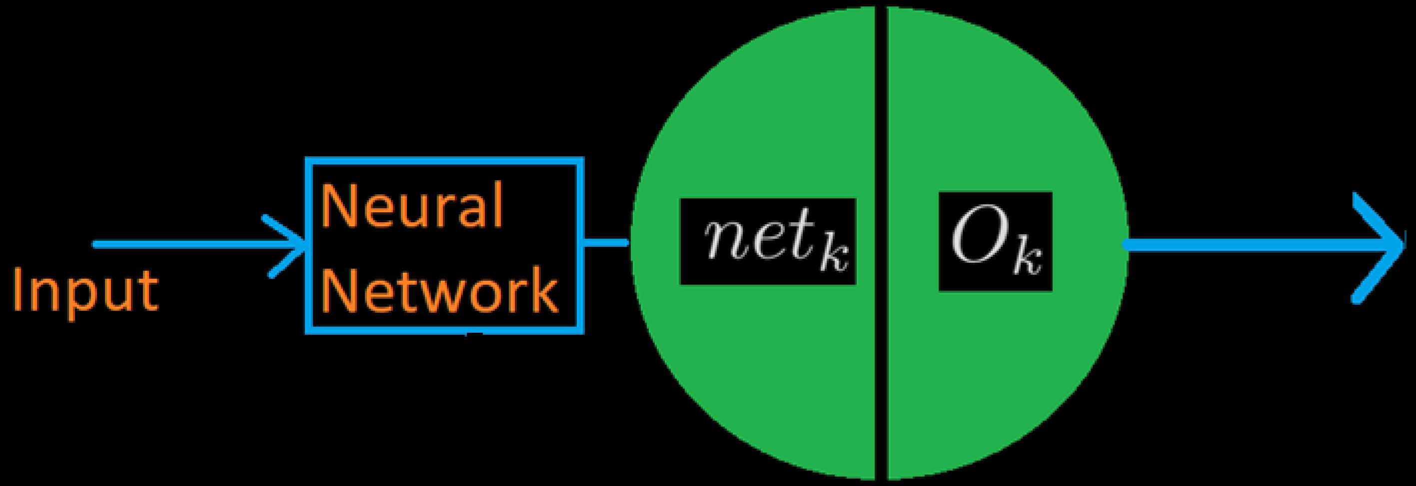
Learning Of Neural Networks

Neural Network learns in two stages / propagations

- Forward Propagation
- Backward Propagation

Standard Neural Network




Input
Neural
Network
net_k
O_k

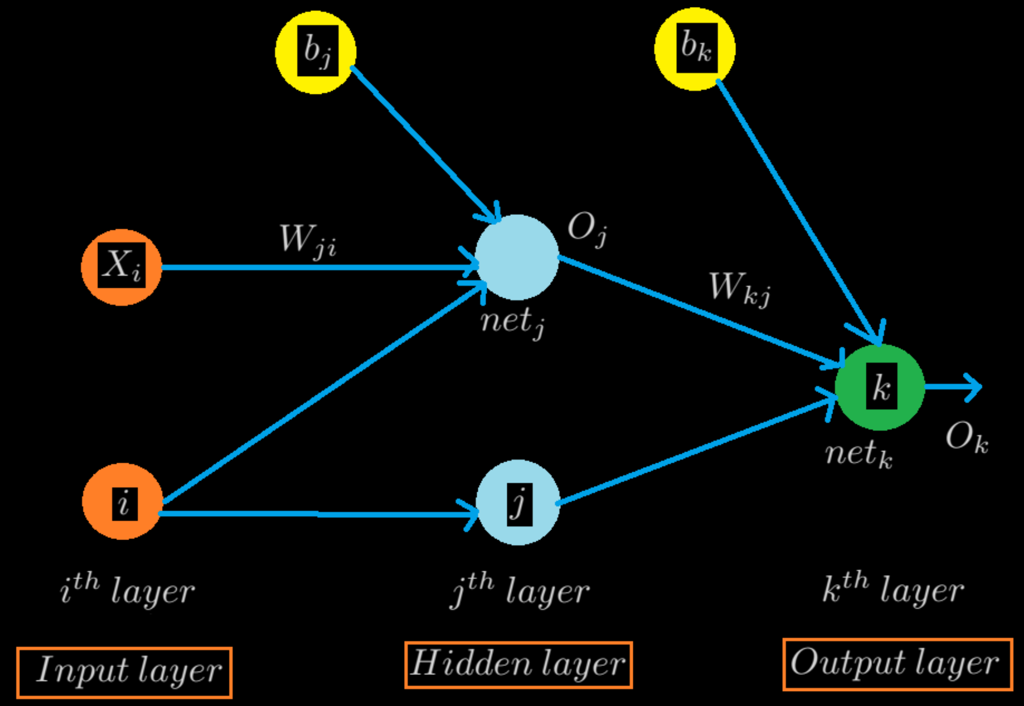
Forward Propagation

Forward path

$\boxed{\text{neuron } j:}$

$$net_j = \sum_i W_{ji} X_i + b_j$$

$$O_j = f(net_j) = \sigma(net_j) = \frac{1}{1 + e^{-net_j}}$$



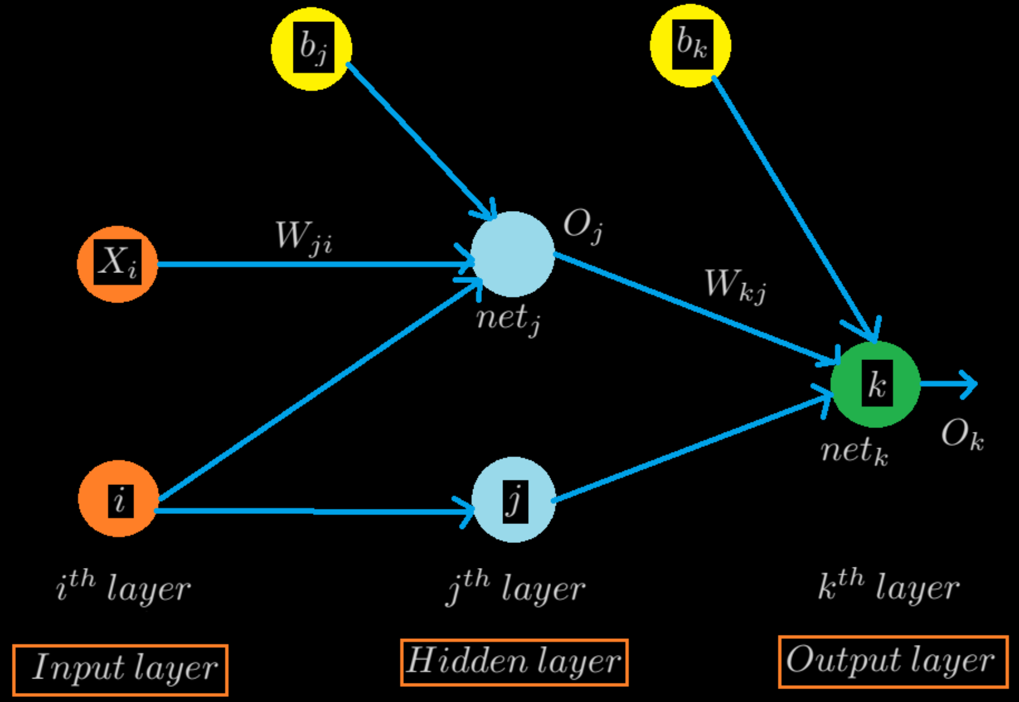
Forward path

$\boxed{\text{neuron } k :}$

Output of neuron k is the output of neural network

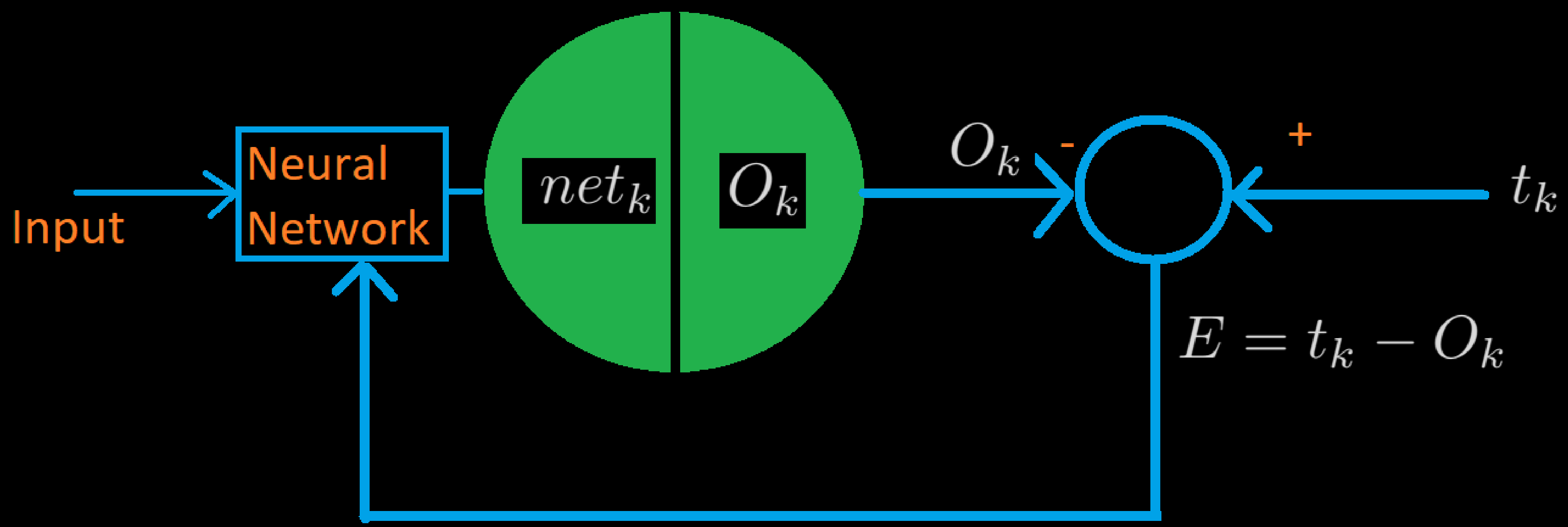
$$net_k = \sum_j W_{kj} O_j + b_k$$

$$\boxed{O_k = f(net_k) = \sigma(net_k) = \frac{1}{1 + e^{-net_k}}}$$



Backward Propagation

Backward Propagation



Error is propagated back through the layers of the Neural Network and the weights of the Neural Network are adapted iteratively.

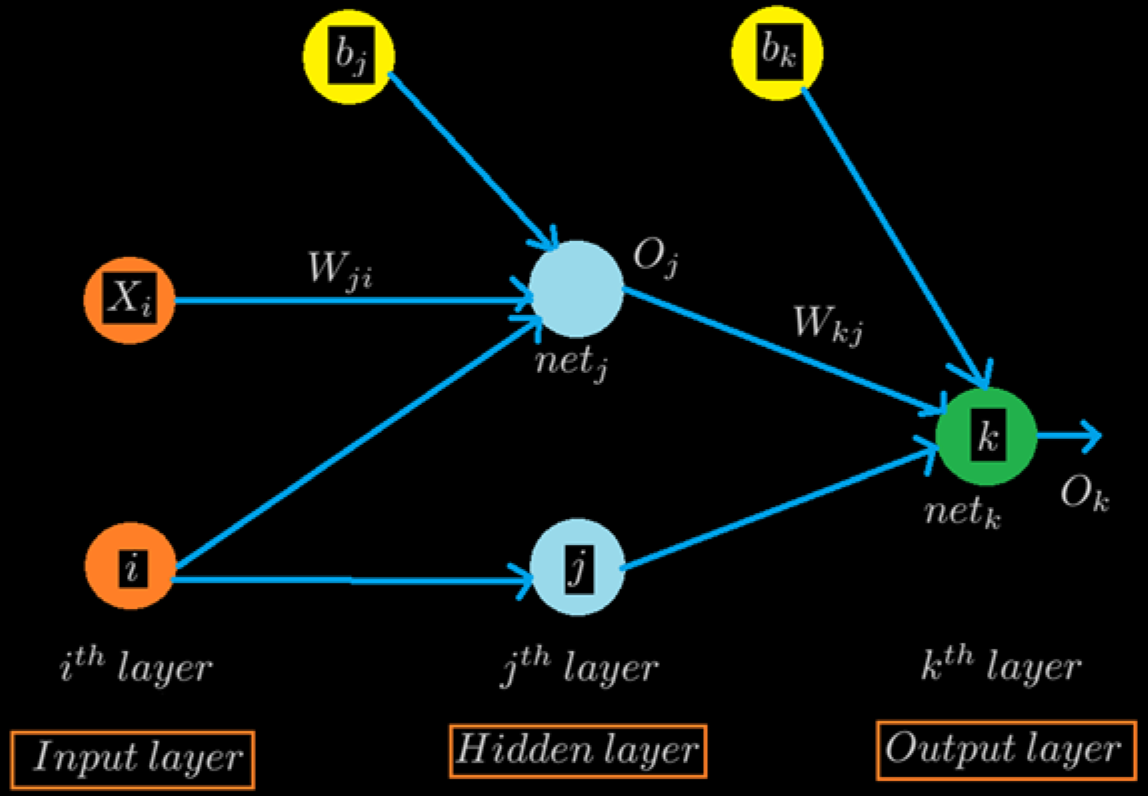
Back Propagation

From output k to hidden j

$$\frac{\partial E}{\partial W_{kj}} = \frac{\partial E}{\partial O_k}\frac{\partial O_k}{\partial net_k}\frac{\partial net_k}{\partial W_{kj}}$$

$$\frac{\partial E}{\partial W_{kj}} = \underbrace{\frac{\partial E}{\partial O_k}\frac{\partial O_k}{\partial net_k}}_{\delta_k}\left(\frac{\partial net_k}{\partial W_{kj}}\right)$$

where δ_k = error signal between layers k and j.



$$\boxed{\text{Finding the value of } \frac{\partial E}{\partial O_k}}$$

$$E = t_k - O_k$$

$$E = \frac{1}{2}(t_k - O_k)^2$$

$$\frac{\partial E}{\partial O_k} = 2 \times \frac{1}{2}(t_k - O_k)(-1)$$

$$\boxed{\frac{\partial E}{\partial O_k} = -(t_k - O_k)}$$

Finding the value of $\dfrac{\partial O_k}{\partial net_k}$

Now

$$O_k = \frac{1}{1 + e^{-net_k}}$$

$$\frac{\partial O_k}{\partial net_k} = \frac{\partial}{\partial net_k}\left[\frac{1}{1 + e^{-net_k}}\right]$$

$$\frac{\partial O_k}{\partial net_k} = \frac{\partial}{\partial net_k}(1 + e^{-net_k})^{-1}$$

$$= -(1 + e^{-net_k})^{-2}(0 + e^{-net_k})(-1)$$

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

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

$$= O_k \left[\frac{1 + e^{-net_k} - 1}{(1 + e^{-net_k})} \right]$$

Breaking the L.C.M

$$= O_k \left[\frac{1 + e^{-net_k}}{1 + e^{-net_k}} - \frac{1}{1 + e^{-net_k}} \right]$$

$$\boxed{\frac{\partial O_k}{\partial net_k} = O_k(1 - O_k)}$$

$$\boxed{\text{Finally,} \quad \frac{\partial net_k}{\partial W_{kj}}}$$

$$\boxed{net_k = \sum_j W_{kj} O_j + b_k}$$

$$\frac{\partial net_k}{\partial W_{kj}} = \frac{\partial}{\partial W_{kj}} \sum W_{kj} O_j + b_k$$

$$\boxed{\frac{\partial net_k}{\partial W_{kj}} = O_j}$$

Thus,

$$\frac{\partial E}{\partial W_{kj}} = \frac{\partial E}{\partial O_k}\frac{\partial O_k}{\partial net_k}\frac{\partial net_k}{\partial W_{kj}}$$

$$\frac{\partial E}{\partial W_{kj}} = -\underbrace{(t_k - O_k)O_k(1 - O_k)}_{\delta_k}O_j$$

$$\boxed{\frac{\partial E}{\partial W_{kj}} = -\delta_k O_j}$$

$\boxed{\text{Update the weight } W_{kj}}$

$$W_{kj}(new) = W_{kj}(old) + \eta \frac{\partial E}{\partial W_{kj}}$$

$$W_{kj}(new) = W_{kj}(old) + \eta(-\delta_k O_j)$$

$$\boxed{W_{kj}(new) = W_{kj}(old) - \eta \delta_k O_j}$$

where η = Learning rate

From hidden j to input i

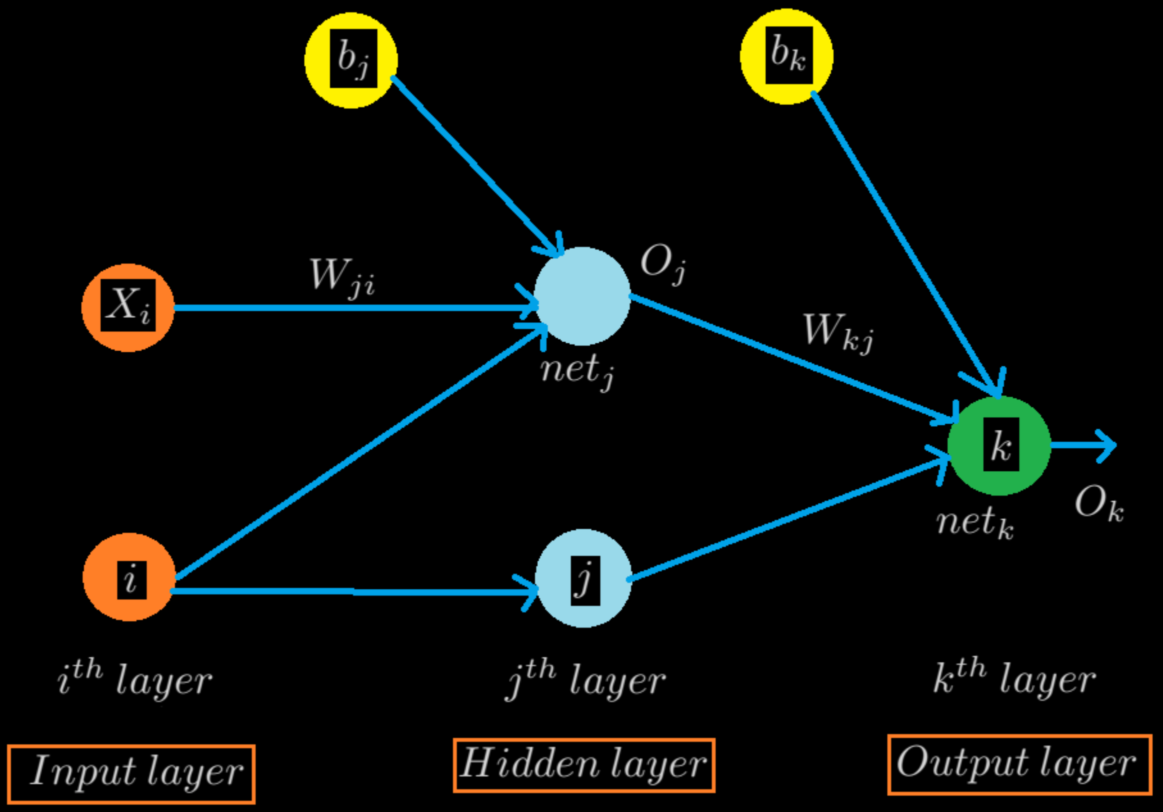
$$\frac{\partial E}{\partial W_{ji}} = \frac{\partial E}{\partial O_k}\frac{\partial O_k}{\partial net_k}\frac{\partial net_k}{\partial O_j}\frac{\partial O_j}{\partial net_j}\frac{\partial net_j}{\partial W_{ji}}$$

$$\frac{\partial net_k}{\partial O_j} = \frac{\partial}{\partial O_j}\sum W_{kj}O_j + b_k$$

$$\boxed{\frac{\partial net_k}{\partial O_j} = W_{kj}}$$

$$\frac{\partial net_j}{\partial W_{ji}} = \frac{\partial}{\partial W_{ji}}\sum W_{ji}X_i + b_j$$

$$\boxed{\frac{\partial net_j}{\partial W_{ji}} = X_i}$$



$$\frac{\partial E}{\partial W_{ji}} = -(t_k - O_k)O_k(1 - O_k)W_{kj}f^{'}(net_j)X_i$$

$$\frac{\partial E}{\partial W_{ji}} = -\underbrace{(t_k - O_k)O_k(1 - O_k)W_{kj}f^{'}(net_j)}_{\delta_j}X_i$$

$$\frac{\partial E}{\partial W_{ji}} = -\underbrace{\overbrace{(t_k - O_k)O_k(1 - O_k)}^{\delta_k}W_{kj}f^{'}(net_j)}_{\delta_j}X_i$$

$$\boxed{\frac{\partial E}{\partial W_{ji}} = -\ \delta_j X_i}$$

$\boxed{\text{Update the weight } W_{ji}}$

$$W_{ji}(new) = W_{ji}(old) + \eta \frac{\partial E}{\partial W_{ji}}$$

$$\boxed{W_{ji}(new) = W_{ji}(old) - \eta \delta_j X_i}$$