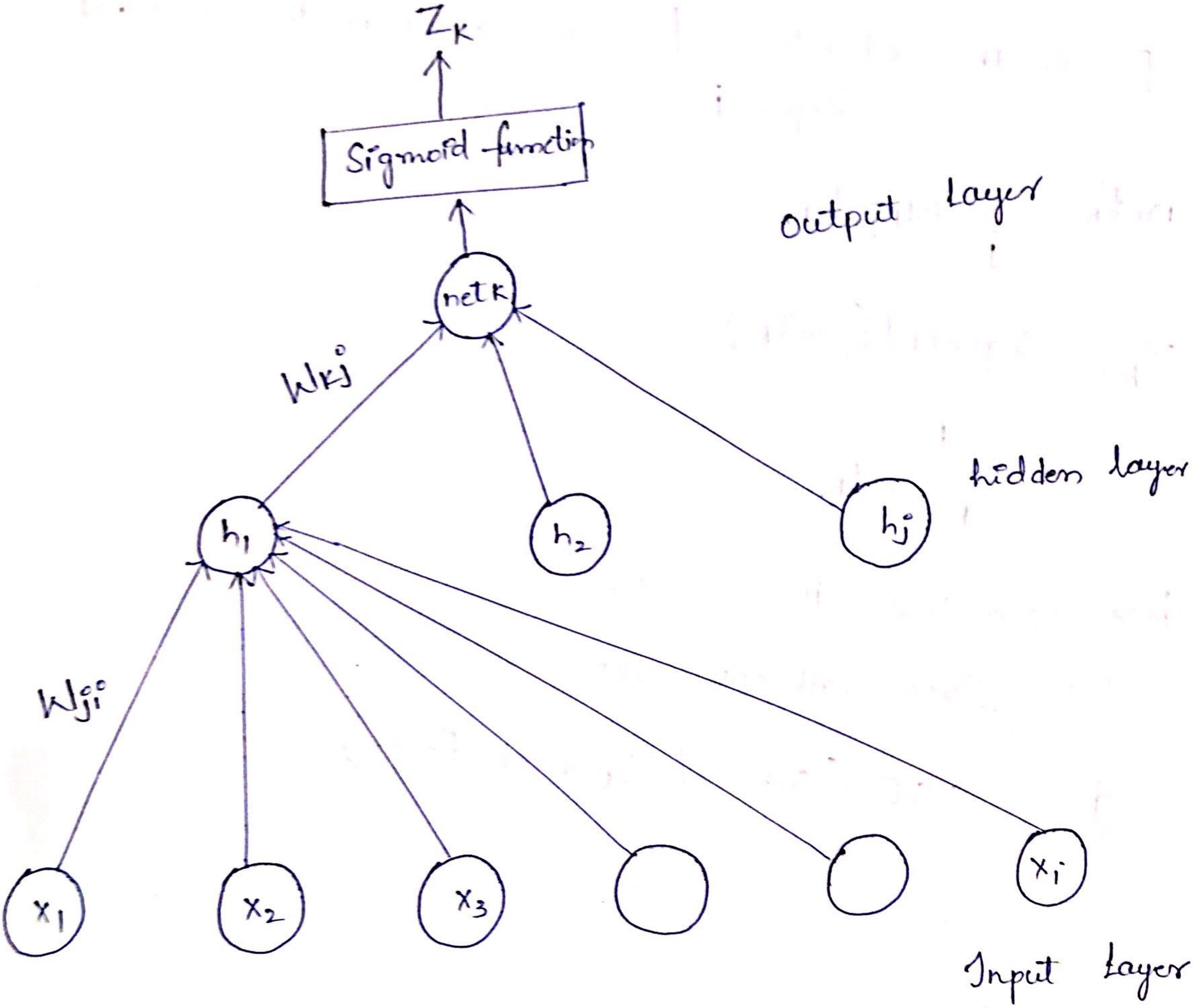
→ Binary Classification (0/1) Using 3-Layer NN:-



W_{ji} = weights from the input layer to the hidden layer.

W_{kj} = Weigths from hidden layer to the output layer

$net_j = \sum_i W_{ji} \cdot x_i$

$h_j = f(net_j)$

$f \rightarrow$ is the activation function used in the hidden layer j

$net_K = \sum_j W_{Kj} \cdot h_j$

$Z_K =$ sigmoid (net_K)

$$= \frac{1}{1+e^{-net_K}}$$

We can use two errors :-

(a) Cross entropy error

(b) SSE (Sum of Squared Error)

SSE :-

$J(W_{Kj}, W_{ji}) = \frac{1}{2}\sum_K (Y - Z_K)^2$

In this case $K=1$

$= \frac{1}{2}(Y - Z_K)^2$

1. Considering Loss as the SSE

$$\frac{\partial J}{\partial W_{kj}} = \frac{\partial}{\partial W_{kj}}\left(\frac{1}{2}(Y-Z_k)^2\right)$$

$$= \frac{1}{2} \cdot \frac{\partial}{\partial W_{kj}}(Y-Z_k)^2$$

$$= \left(\frac{1}{\cancel{2}}\right)(\cancel{2})(Y-Z_k) \cdot \frac{\partial}{\partial W_{kj}}(Y-Z_k)$$

$$= (Y-Z_k) \cdot \frac{\partial}{\partial W_{kj}}(-Z_k)$$

$$= -(Y-Z_k) \cdot \frac{\partial}{\partial W_{kj}}(Z_k)$$

$$= -(Y-Z_k) \cdot \frac{\partial}{\partial W_{kj}} \left(\frac{1}{1+e^{-net_k}}\right)$$

$$= -(Y-Z_k) \cdot \frac{\partial}{\partial net_k}\left(\frac{1}{1+e^{-net_k}}\right) \cdot \frac{\partial (net_k)}{\partial W_{kj}}$$

$$\boxed{\frac{\partial}{\partial x}\left(\frac{1}{1+e^{-x}}\right) = \frac{e^{-x}}{(1+e^{-x})^2}}$$

$$= -(Y-Z_K) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot \frac{\partial}{\partial W_{kj}} \left(\sum_j W_{kj} \cdot h_j \right)$$

$$= -(Y-Z_K) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot h_j$$

$$W_{kj}^{New} = W_{kj}^{old} - \lambda \cdot \frac{\partial J}{\partial W_{kj}}$$

$$\boxed{W_{kj}^{New} = W_{kj}^{old} + \lambda (Y-Z_K) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot h_j}$$

$$\frac{\partial J}{\partial W_{ji}} = \frac{\partial}{\partial W_{ji}} \left(\frac{1}{2} (Y-Z_K)^2 \right)$$

$$= \left(\frac{1}{\cancel{2}}\right)(\cancel{2})(Y-Z_K) \cdot \frac{\partial}{\partial W_{ji}} (Y-Z_K)$$

$$= (Y-Z_K) \cdot \frac{\partial}{\partial W_{ji}} (Y-Z_K)$$

$$= -(Y-Z_K) \cdot \frac{\partial}{\partial W_{ji}} (Z_K)$$

$$= -(Y-Z_K) \cdot \frac{\partial}{\partial W_{ji}} \left(\frac{1}{1+e^{-net_K}} \right)$$

$$= -(Y-Z_K) \cdot \frac{\partial}{\partial net_K}\left(\frac{1}{1+e^{-net_K}}\right) \cdot \frac{\partial net_K}{\partial W_{ji}}$$

$$= -(Y-Z_K) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot \frac{\partial (net_K)}{\partial W_{ji}}$$

$$= -(Y-Z_K) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot \frac{\partial}{\partial W_{ji}}\left(\sum_j W_{Kj} \cdot h_j\right)$$

$$= -(Y-Z_K) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot W_{Kj} \cdot \frac{\partial}{\partial W_{ji}}(h_j)$$

$$= -(Y-Z_K) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot W_{Kj} \cdot \frac{\partial}{\partial W_{ji}}(F(net_j))$$

$$= -(Y-Z_K) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot W_{Kj} \cdot \frac{\partial}{\partial net_j} F(net_j) \cdot \frac{\partial net_j}{\partial W_{ji}}$$

$$= -(Y-Z_K) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot W_{Kj} \cdot F'(net_j) \cdot \frac{\partial}{\partial W_{ji}}\left(\sum W_{ji} \cdot X_i\right)$$

$$= -(Y-Z_K) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot W_{Kj} \cdot F'(net_j) \cdot X_i$$

$$W_{ji}^{New} = W_{ji}^{old} - \lambda \frac{\partial J}{\partial W_{ji}}$$

$$\boxed{W_{ji}^{New} = W_{ji}^{old} + \lambda (Y - Z_k) \cdot \frac{e^{-net_k}}{(1+e^{-net_k})^2} \cdot W_{kj} \cdot f'(net_j) \cdot x_i}$$