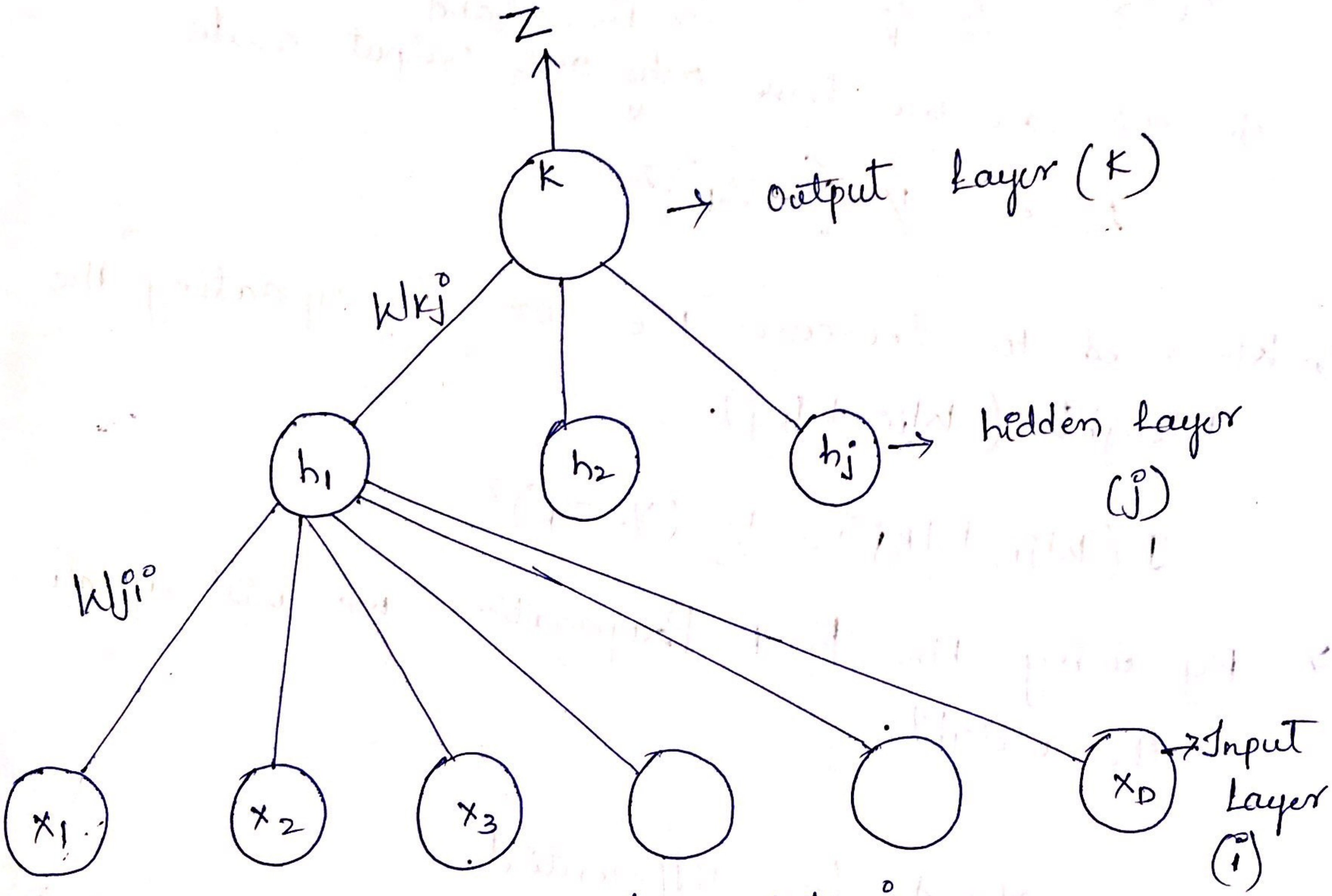
① 3-Layers Neural Network for Regression :-



W_{ji} → Weights from the Layer i to j

W_{kj} → Weights from the layer j to K

$h_j = F(net_j)$ → F → Is the Activation Function that is used in the hidden layer

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

$K = net_K$ → Z → Output i-e predicted output

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

Scanned with CamScanner

2

→ The Loss Function in this case is SSE. (Sum of Squared Errors)

predicted Label

$$\text{Loss} = \frac{1}{2} \sum_{K} (Y_K - Z_K)^2$$

True Label

In this case we have only one output node

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

→ We need to decrease the loss by updating the weights (W_{ji}, W_{Kj}).

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

→ By using the Back Propagation we will update the weights.

∂ → stands for differential.

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

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

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

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

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

$$= Z_K = net_K$$

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

$$= -(Y - Z_K) \cdot \frac{\partial}{\partial W_{Kj}} \left(\sum_j W_{Kj} \cdot h_j \right)$$

$$= -(Y - Z_K) \cdot h_j$$

By using the Gradient Descent Update Rule:-

$$W_{Kj}^{New} = W_{Kj}^{old} - \lambda \left(\frac{\partial J}{\partial W_{Kj}} \right)$$

$\lambda \rightarrow$ Learning rate

$$W_{Kj}^{New} = W_{Kj}^{old} - \lambda \left(-(Y - Z_K) \cdot h_j \right)$$

$$\boxed{W_{Kj}^{New} = W_{Kj}^{old} + \lambda (Y - Z_K) \cdot h_j}$$

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

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

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

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

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

$$= -(Y - Z_K) \cdot \frac{\partial}{\partial w_{ji}} \left(\sum_{j} w_{Kj} \cdot h_j\right)$$

$$= -(Y - Z_K) \cdot w_{Kj} \cdot \frac{\partial}{\partial w_{ji}} (h_j)$$

$$= -(Y - Z_K) \cdot w_{Kj} \cdot \frac{\partial}{\partial w_{ji}} \cdot F(net_j)$$

$$\boxed{\frac{\partial}{\partial x} F(g(x)) = \frac{\partial F(g(x))}{\partial g(x)} \cdot \frac{\partial g(x)}{\partial (x)}}$$

Scanned with CamScanner

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

$$= -(Y-Z_K) \cdot W_{kj} \cdot F'(net_j) \cdot \frac{\partial \left(\sum_i W_{ji} \cdot x_i\right)}{\partial W_{ji}}$$

$$= -(Y-Z_K) \cdot W_{kj} \cdot F'(net_j) \cdot \frac{\partial}{\partial W_{ji}} \left(\sum_i W_{ji} \cdot x_i\right)$$

$$= -(Y-Z_K) \cdot W_{kj} \cdot F'(net_j) \cdot x_i$$

$F'(\ )$ → differential of the activation function used in the hidden layer.

$$W_{ji}^{New} = W_{ji}^{old} - \lambda \left(\frac{\partial J}{\partial W_{ji}}\right)$$

$$\boxed{W_{ji}^{New} = W_{ji}^{old} + \lambda (Y-Z_K) \cdot W_{kj} \cdot F'(net_j) \cdot x_i}$$