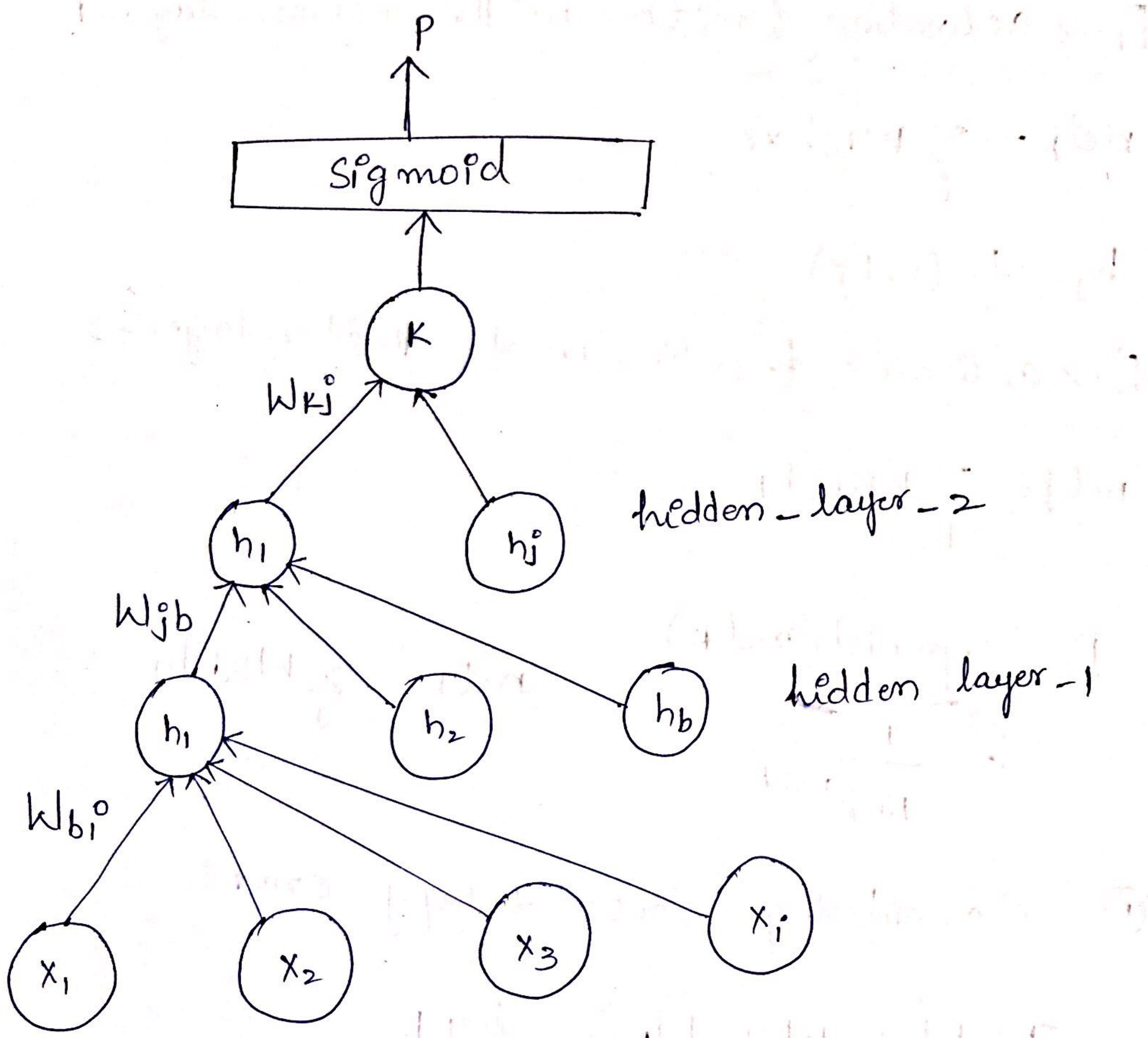
→ Binary Classification Using 4-layer NN :-



W_{bi} → Weights from input layer to hidden_layer-1

W_{jb} → Weights from hidden_layer_1 to hidden_layer-2

W_{kj} → Weights from hidden_layer_2 to output layer

$h_b = f_1(net_b)$

$f_1 \rightarrow$ activation function in the hidden_layer_1

$$net_b = \sum_i W_{bi} \cdot x_i$$

$h_j = f_2(net_j)$

$f_2 \rightarrow$ activation function in the hidden_layer-2

$$net_j = \sum_b W_{jb} \cdot h_b$$

$$P = \text{sigmoid}(net_K) \qquad net_K = \sum_j W_{Kj} \cdot h_j$$

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

① Considering Cross-Entropy error

$$J(W_{Kj}, W_{jb}, W_{bi}) = \log P.$$

$$\frac{\partial J}{\partial W_{Kj}} = \frac{\partial}{\partial W_{Kj}} (\log P)$$

$$= \frac{1}{P} \cdot \frac{\partial}{\partial W_{Kj}} (P)$$

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

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

$$= \frac{1}{P} \cdot \frac{e^{-net_k}}{(1+e^{-net_k})^2} \cdot \frac{\partial}{\partial W_{kj}} (net_k)$$

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

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

$$= \frac{P^{\cancel{2}}}{\cancel{P}} \cdot e^{-net_k} \cdot h_j$$

$$= P \cdot e^{-net_k} \cdot h_j$$

$$\boxed{W_{kj}^{New} = W_{kj}^{old} + \lambda \cdot P \cdot e^{-net_k} \cdot h_j}$$

$$\frac{\partial J}{\partial w_{jb}} = \frac{\partial}{\partial w_{jb}} (\log p)$$

$$= \frac{1}{p} \cdot \frac{\partial}{\partial w_{jb}} (p)$$

$$= \frac{1}{p} \cdot \frac{\partial}{\partial w_{jb}} \left(\frac{1}{1+e^{-net_K}} \right)$$

$$= \frac{1}{p} \cdot \frac{\partial}{\partial net_K} \left(\frac{1}{1+e^{-net_K}} \right) \cdot \frac{\partial net_K}{\partial w_{jb}}$$

$$= \frac{1}{p} \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot \frac{\partial}{\partial w_{jb}} \left(\sum w_{Kj} \cdot h_j \right)$$

$$= \frac{1}{p} \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot \sum_K w_{Kj} \cdot \frac{\partial}{\partial w_{jb}} (h_j)$$

$$= \frac{1}{p} \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot \sum_K \cdot w_{Kj} \cdot \frac{\partial}{\partial w_{jb}} \cdot f_2(net_j)$$

$$= \frac{p^{\cancel{2}}}{\cancel{p}} \cdot \left(\sum_K w_{Kj} \cdot \frac{\partial}{\partial net_j} f(net_j) \cdot \frac{\partial net_j}{\partial w_{jb}} \right) e^{-net_K}$$

$$= p \cdot \left(\sum_K w_{Kj} \cdot f'(net_j) \cdot \frac{\partial}{\partial w_{jb}} \left(\sum_b w_{jb} \cdot h_b \right) \right) e^{-net}$$

$$= p \cdot e^{-net_K} \cdot \left(\sum_K W_{kj} \cdot F'(net_j) \cdot h_b \right)$$

$$= p \cdot e^{-net_K} \left(\sum_K W_{kj} \cdot F'(net_j) \right) \cdot h_b$$

$$\boxed{W_{jb}^{New} = W_{jb}^{old} + \lambda \cdot p \cdot e^{-net_K} \cdot \left(\sum_K W_{kj} \cdot F'(net_j) \right) \cdot h_b}$$

$$\frac{\partial J}{\partial W_{bi}} = \frac{\partial}{\partial W_{bi}} (\log p)$$

$$= \frac{1}{p} \cdot \frac{\partial}{\partial W_{bi}} (p)$$

$$= \frac{1}{p} \cdot \frac{\partial}{\partial W_{bi}} \left(\frac{1}{1+e^{-net_K}} \right)$$

$$= \frac{1}{p} \cdot \frac{\partial}{\partial net_K} \left(\frac{1}{1+e^{-net_K}} \right) \cdot \frac{\partial net_K}{\partial W_{bi}}$$

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

$$= \frac{1}{p} \cdot \frac{e^{-net_K}}{\left(1+e^{-net_K}\right)^2} \sum_{K,j} W_{kj} \cdot \left(\frac{\partial h_j}{\partial W_{bi}} \right)$$

$$= \frac{p^{\cancel{2}}}{\cancel{p}} \cdot e^{-net_k} \cdot \sum_j w_{kj} \cdot \frac{\partial}{\partial w_{bi}} (h_j)$$

$$= p \cdot e^{-net_k} \cdot \sum_{k,j} w_{kj} \cdot \frac{\partial}{\partial w_{bi}} f_2(net_j)$$

$$= p \cdot e^{-net_k} \sum_{k,j} w_{kj} \cdot \frac{\partial}{\partial net_j} f_2(net_j) \cdot \frac{\partial net_j}{\partial w_{bi}}$$

$$= p \cdot e^{-net_k} \cdot \sum_{k,j} w_{kj} \cdot f_2'(net_j) \cdot \frac{\partial}{\partial w_{bi}} \left(\sum_b w_{jb} \cdot h_b \right)$$

$$= p \cdot e^{-net_k} \cdot \sum_{k,j} w_{kj} \cdot f_2'(net_j) \cdot w_{jb} \cdot \frac{\partial}{\partial w_{bi}} (h_b)$$

$$= p \cdot e^{-net_k} \cdot \sum_{k,j} w_{kj} \cdot f_2'(net_j) \cdot w_{jb} \cdot \frac{\partial}{\partial w_{bi}} (f_1(net_b))$$

$$= p \cdot e^{-net_k} \cdot \sum_{k,j} w_{kj} \cdot f_2'(net_j) \cdot w_{jb} \cdot \frac{\partial}{\partial net_b} f_1(net_b) \cdot \frac{\partial net_b}{\partial w_{bi}}$$

$$= p \cdot e^{-net_k} \cdot \sum_{k,j} w_{kj} \cdot f_2'(net_j) \cdot w_{jb} \cdot f_1'(net_b) \cdot \frac{\partial}{\partial w_{bi}} \sum_i w_{bi} \cdot x_i$$

$$= p \cdot e^{-net_k} \sum_{k,j} W_{kj} \cdot f_2'(net_j) \cdot W_{jb} \cdot F_1'(net_b) \cdot x_i$$

$$\boxed{W_{bi}^{New} = W_{bi}^{old} + \lambda\, p \cdot e^{-net_k} \cdot \left(\sum_{k,j} W_{kj} \cdot f_2'(net_j) \cdot W_{jb} \cdot F_1'(net_b) \right) \cdot x_i}$$