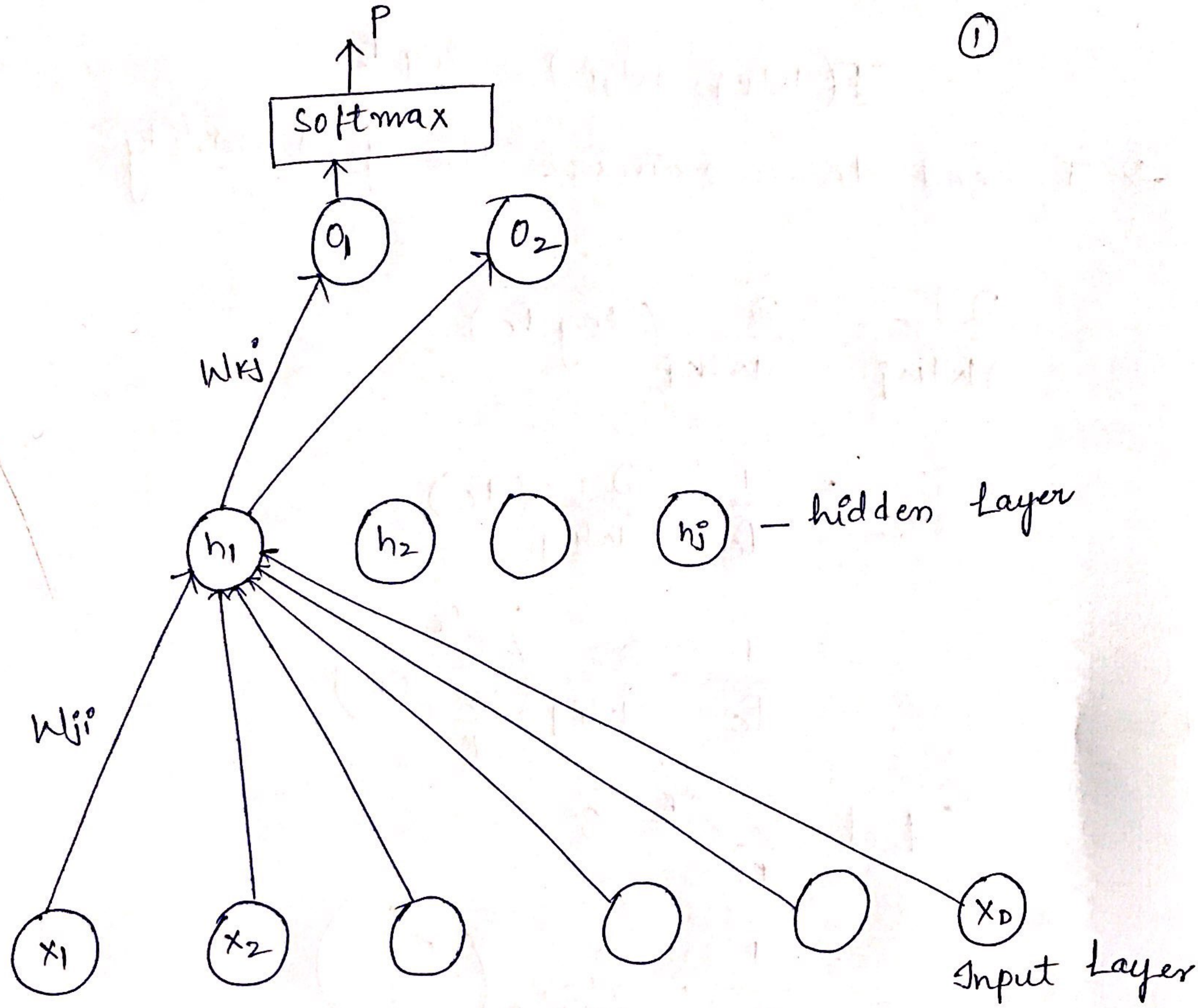
Classification - Multi-Class Using 3-Layer NN



$$\text{Softmax} = \frac{e^{O_1}}{e^{O_1} + e^{O_2} + e^{O_3} + \cdots + e^{O_K}}$$

$O_K \rightarrow$ Number of output Nodes

$$P_c = \frac{e^{O_c}}{e^{O_1} + e^{O_2} + \cdots + e^{O_K}} = \frac{e^{O_c}}{\sum_{i=1}^{K} e^{O_K}}$$

→ Using CROSS-ENTROPY ERROR in this case

$$J(W_{kj}, W_{ji}) = \log P_c$$

→ I need to maximize the probability

$$\frac{\partial J}{\partial W_{kj}} = \frac{\partial}{\partial W_{kj}}(\log P_c)$$

$$= \frac{1}{P_c} \cdot \frac{\partial}{\partial W_{kj}}(P_c)$$

$$= \frac{1}{P_c} \cdot \frac{\partial}{\partial W_{kj}}\left(\frac{e^{O_c}}{\sum_K e^{O_K}}\right)$$

Let $\sum_K e^{O_K} = S$

$$= \frac{1}{P_c} \cdot \frac{\partial}{\partial W_{kj}}\left(\frac{e^{O_c}}{S}\right)$$

$O_c = net\,c$ or for simplicity considering it as $net\,K$

$$net\,k = \sum_j W_{kj} \cdot h_j$$

$$= \frac{1}{P_c} \cdot \frac{\partial}{\partial W_{kj}}\left(\frac{e^{O_c}}{S}\right)$$

$$= \frac{1}{P_c} \cdot \frac{\partial}{\partial net_K} \left(\frac{e^{O_c}}{S} \right) \cdot \frac{\partial net_K}{\partial W_{Kj}} \quad (3)$$

$$= \frac{1}{P_c} \cdot \frac{\partial}{\partial net_K} \left(\frac{e^{O_c}}{S} \right) \cdot h_j$$

$$\frac{\partial}{\partial net_K} \left(\frac{e^{O_c}}{S} \right) = \frac{S \cdot \frac{\partial}{\partial net_K} e^{O_c} - e^{O_c} \cdot \frac{\partial}{\partial net_K} (S)}{S^2}$$

When $c == K$

$$= \frac{S \cdot e^{O_c} - e^{O_c} \cdot \frac{\partial}{\partial net_K} (S)}{S^2}$$

$$= \frac{S \cdot e^{O_c} - e^{O_c} \cdot e^{O_K}}{S^2}$$

$$= \frac{e^{O_c} (S - e^{O_K})}{S^2} \quad \text{when } (K == c)$$

When $K \neq C$

$$\frac{S \cdot \frac{\partial}{\partial net_K} e^{O_c} - e^{O_c} \cdot \frac{\partial}{\partial net_K} \left(\sum_K e^{O_K} \right)}{S^2}$$

(4)

$$= \frac{0 - e^{O_c} \cdot e^{O_K}}{S^2}$$

$$= \frac{-e^{O_c} \cdot e^{O_K}}{S^2}$$

By combining both the conditions i.e $K == C$ and $K \neq C$ we can write this as

$$(Y_K - P_K) h_j$$

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

$$\frac{\partial J}{\partial W_{ji}} = \sum_K \frac{\partial}{\partial W_{ji}} (\log P_c)$$

$$= \sum_K \frac{1}{P_c} \cdot \frac{\partial}{\partial W_{ji}} (P_c)$$

$$= \sum_K \frac{1}{P_c} \cdot \frac{\partial}{\partial W_{ji}} \left(\frac{e^{O_c}}{S} \right)$$

$$= \sum_K \frac{1}{P_c} \cdot \frac{\partial}{\partial net_K} \cdot \left(\frac{e^{net_K}}{S} \right) \cdot \frac{\partial net_K}{\partial W_{ji}}$$

$$= \sum_K \frac{1}{P_c} \cdot \frac{\partial}{\partial net_c} \left(\frac{e^{(net_c)}}{S} \right) \cdot W_{Kj} \cdot \frac{\partial}{\partial W_{ji}} (h_j)$$

5

$$= \sum_K \underbrace{(Y_K - P_K)}_{\text{from the previous calculations.}} \cdot W_{Kj} \cdot \frac{\partial}{\partial W_{ji}} (h_j)$$

$$= \sum_K (Y_K - P_K) \cdot W_{Kj} \cdot F'(net_j) \cdot X_i$$

$$\boxed{W_{ji}^{New} = W_{ji}^{old} + \lambda \sum_K (Y_K - P_K) \cdot W_{Kj} \cdot F'(net_j) \cdot X_i}$$