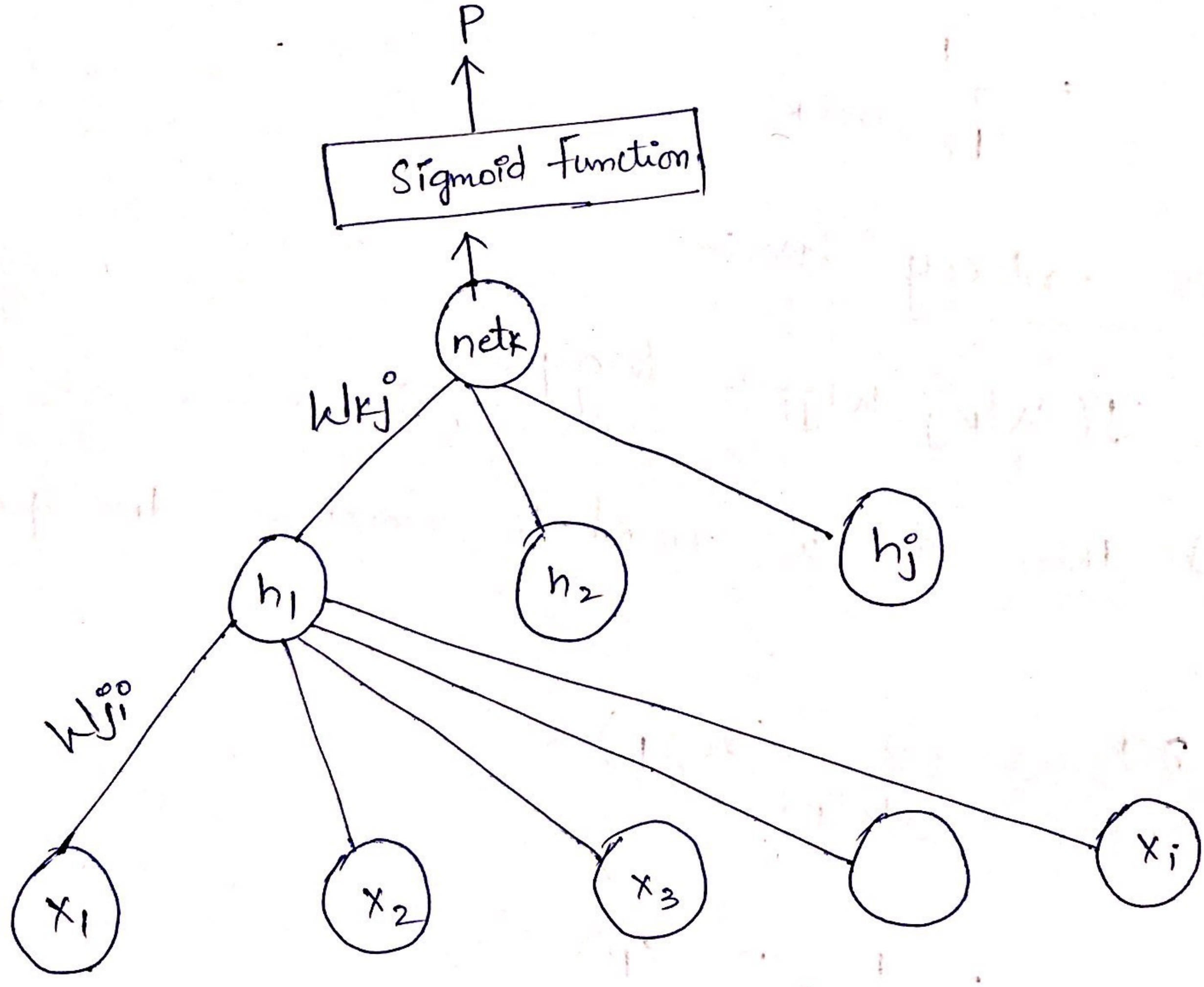
(b) Using Cross Entropy error

→ Cross entropy loss or log loss, measures the performance of a classification model whose output is a probability value between 0 and 1.

→ We need to increase the probability i.e increase the log-like~~li~~hood (log-likelihood)

→ Increase the log-likelihood.



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

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

$$h_j = F(net_j)$$

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

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

$$Z_K = \text{sigmoid}(net_K)$$

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

<u>Cross entropy loss:-</u>

$$J(W_{Kj}, W_{ji}) = \log P$$

$\rightarrow$ In this case we need to maximize the probability.

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

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

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

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

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

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

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

$$W_{Kj}^{New} = W_{Kj}^{old} + \lambda \frac{\partial J}{\partial W_{Kj}}$$

$$W_{Kj}^{New} = W_{Kj}^{old} + \underbrace{\lambda \cdot (1/P) \cdot \frac{e^{-net_K}}{(1+e^{-net_K})^2} \cdot h_j}_{\text{Further Simplification.}}$$

$$\Rightarrow \left(\frac{1}{P}\right) \cdot \frac{e^{-net_K}}{(1/P)^2} \cdot h_j. \qquad \boxed{P = \frac{1}{1+e^{-net_K}}}$$

$$\Rightarrow \frac{P^{\cancel{2}}}{\cancel{P}} \cdot e^{-net_K} \cdot h_j.$$

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

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

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

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

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

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

$$= \frac{p^{\cancel{2}}}{\cancel{p}} \cdot e^{-net_k} \cdot W_{kj} \cdot \frac{\partial (h_j)}{\partial W_{ji}}$$

$$= p \cdot e^{-net_k} \cdot W_{kj} \cdot \frac{\partial}{\partial W_{ji}} (F(net_j))$$

$$= p \cdot e^{-net_k} \cdot W_{kj} \cdot \frac{\partial F(net_j)}{\partial net_j} \cdot \frac{\partial net_j}{\partial W_{ji}}$$

$$= p \cdot e^{-netk} \cdot W_{kj} \cdot F'(netj) \cdot \frac{\partial}{\partial W_{ji}} \left(\sum_{i}^{25} W_{ji} \cdot x_i \right)$$

$$= p \cdot e^{-netk} \cdot W_{kj} \cdot F'(netj) \cdot x_i$$

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

$$\boxed{W_{ji}^{New} = W_{ji}^{old} + \lambda \cdot p \cdot e^{-netk} \cdot W_{kj} \cdot F'(netj) \cdot x_i}$$