Backpropogation Algorithm

Sigmoid function = $\frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right) = \sigma(x)$

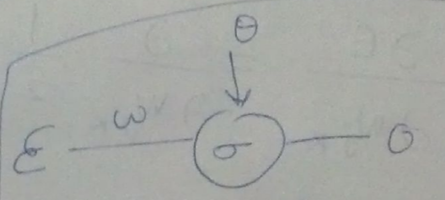
$$= \frac{e^{x}}{(1-e^{x})^2}$$

$$= \frac{(1+e^{-x})-1}{(1+e^{-x})^2}$$

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

$$\sigma' = \sigma(x) - \sigma(x)^2$$

$$\boxed{\sigma' = \sigma(1-\sigma)}$$



$$O = \sigma(\Sigma w + \theta)$$

x_j^l input to node j of layer l

w_{ij}^l weight from layer l-1 node i to l node j

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

θ_j^l = Bias of node j for layer l

O_j^l : output of node j in layer l

t_j : Target value of node j of output layer

Error calculation

$$E = \frac{1}{2} \sum_{k \in K} (O_k - t_k)^2$$

we want to calculate

$$\frac{\partial E}{\partial W_{jk}^{l}}$$

Two cases

→ node is an output node

→ It is a hidden layer

Output layer

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial}{\partial w_{jk}} \frac{1}{2} \sum_{k \in K} (O_k - t_k)^2$$

$$= (O_k - t_k) \, O_k (1 - O_k) \, O_j$$

let define δ_k as $(O_k - t_k) O_k (1-$

so

$$\boxed{\frac{\partial E}{\partial w_{jk}} = O_j \, \delta_k}$$

~~w~~ $\delta_k =$

→ weigth change if it is output layer

Hidden layer node

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \frac{1}{2} \sum_{k \in K} (O_k - t_k)^2$$

$$= \sum_{k \in K} (O_k - t_k) \frac{\partial}{\partial w_{ij}} O_k$$

$$= \sum_{k \in K} (O_k - t_k) \sigma(x_k)(1 - \sigma(x_k)) \frac{\partial x_k}{\partial w_{ij}}$$

$$= \sum_{k \in K} (O_k - t_k) O_k (1 - O_k) \frac{\partial x_k}{\partial O_j} \cdot \underbrace{\frac{\partial O_j}{\partial w_{ij}}}_{\text{no } k \text{ involved so}}$$

$$\frac{\partial E}{\partial w_{ij}} = O_j (1 - O_j) O_i \sum_{k \in K} (O_k - t_k) O_k (1 - O_k) w_{jk}$$

with δ_k

$$\frac{\partial E}{\partial w_{ij}} = O_i O_j (1 - O_j) \sum_{k \in K} \delta^k W_{jk}$$

Simular before we define all terms besides O_i to be δ_j so we have

$$\frac{\partial E}{\partial w_{ij}} = O_i \delta_j$$

for an output layer node $k \in K$

$$\frac{\partial E}{\partial W_{jk}} = O_j \, \delta_k$$

where,

$$\delta_k = O_k (1 - O_k)(O_k - t_k)$$

For an hidden layer node $j \in J$

$$\frac{\partial E}{\partial W_{ij}} = O_i \, \delta_j$$

where,

$$\delta_j = O_j (1 - O_j) \sum_{k \in K} \delta_k W_{jk}$$

bias term

$$\frac{\partial O}{\partial \theta} = 1$$

back propagation algorithm

1. Run network forward with our input data to get the network output
2. for each output node compute

$$\delta_k = O_k (1 - O_k)(O_k - t_k)$$

Scanned by CamScanner

③ for each hidden layer calculate

$$\delta_j = O_j (1 - O_j) \sum_{k \in K} \delta_k W_{jk}$$

④ Update weights and bias

$$\Delta w = -\eta \delta_l O_{l-1}$$

$$\Delta \theta = -\eta \delta_l$$

apply

$$W + \Delta W \rightarrow W$$

$$\theta + \Delta \theta \rightarrow \theta$$