BACKPROPOGATION

① Create a feed-forward neural networks with $n^{[0]}$ input units, $n^{[1]}$ hidden units & $n^{[2]}$ Output units.

② Initialize all network weights to small random values (between -0.5 & 0.5)

③ Until Termination Condition is met
(# of iterations, change in weights almost zero or Total Errors almost same w.r.t. last few iterations)

Do

For Each Training Example, Do #Stochastic Gradient Descent

3.1 Propogate the I/P forward through the network i.e.

→ Compute outputs of neurons/units at layer [1] using inputs of layer [0]

→ Forward propogate the outputs of layer [1] to compute the outputs of layer [2]

3.2 For each network output unit k, calculate error E_k as

$$\delta_k^{[2]} \left(\frac{\partial E}{\partial z_k^{[2]}}\right) = \left(y_k - a_k^{[2]}\right) * a_k^{[2]} * \left(1 - a_k^{[2]}\right)$$

3.3 For each hidden unit j, calculate error E_j as

$$\delta_j^{[1]} \left(\frac{\partial E}{\partial z_j^{[1]}}\right) = a_j^{[1]} * \left(1 - a_j^{[1]}\right) \sum_k w_{jk}^{[2]} \delta_k$$

Update network weights

3.4

For weights associated with Output layer:

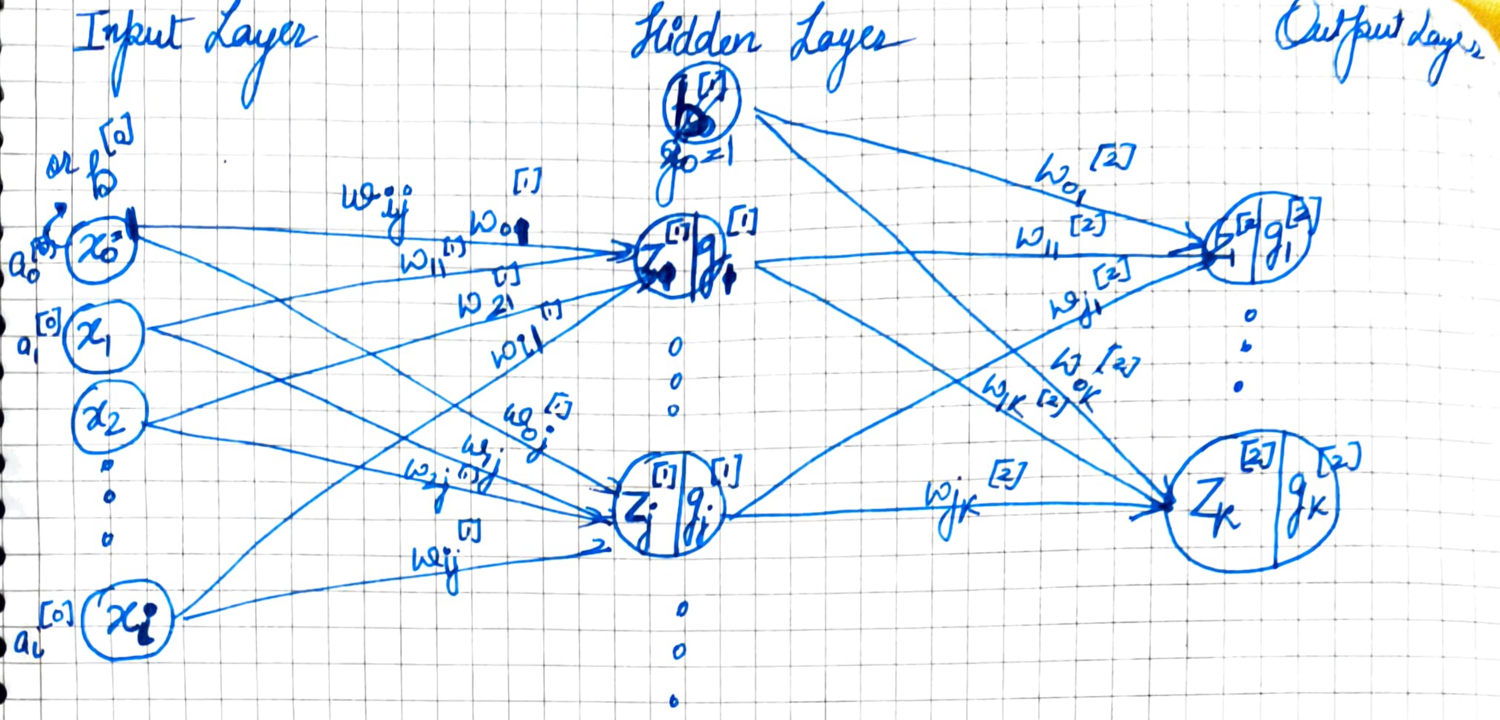
$$w_{jk}^{[2]} = w_{jk}^{[2]} + \underbrace{\alpha\, \delta_k^{[2]} * a_j^{[1]}}_{\Delta w_{jk}^{[2]}}$$

$$\Delta w_{jk}^{[2]} \left(\alpha \frac{\partial E}{\partial w_{jk}^{[2]}}\right) = \frac{\partial E}{\partial z_k^{[2]}} * \frac{\partial z_k^{[2]}}{\partial w_{jk}^{[2]}}$$

For weights associated with hidden layer:

$$w_{ij}^{[1]} = w_{ij}^{[1]} + \underbrace{\alpha\, \delta_j^{[1]} * x_i^{[0]}}_{\Delta w_{ij}^{[1]}}$$

$$\Delta w_{ij}^{[1]} = \left(\alpha \frac{\partial E}{\partial w_{ij}^{[1]}}\right) = \frac{\partial E}{\partial z_j^{[1]}} * \frac{\partial z_j^{[1]}}{\partial w_{ij}^{[1]}}$$



Forward Propogation

1. Computation of Outputs of Hidden Layer

⇒ $z_j^{[1]} = x_0 w_{0j}^{[1]} + x_1 w_{1j}^{[1]} + x_2 w_{2j}^{[1]} + \dots \; x_i w_{ij}^{[1]}$

$= \sum_i x_i w_{ij}^{[1]}$ | $a_j^{[1]} = g(z_j)^{[1]} = g_j^{[1]} = \dfrac{1}{1+e^{-z_j^{[1]}}}$

2. Computation of Outputs of Output Layer

⇒ $z_k^{[2]} = a_0^{[1]} w_{0k}^{[2]} + a_1^{[1]} w_{1k}^{[2]} + \dots \; a_j^{[1]} w_{jk}^{[2]}$

$z_k^{[2]} = \sum_j a_j^{[1]} w_{jk}^{[2]}$ | $a_k^{[2]} = g(z_k)^{[2]} = g_k^{[2]} = \dfrac{1}{1+e^{-z_k^{[2]}}}$

In general

$$\boxed{z_i^{[l]} = \sum_k w_{ki}^{[l]} a_k^{[l-1]} \quad \text{and} \quad a_i^{[l]} = \frac{1}{1+e^{-z_i^{[l]}}}}$$

$$E = \frac{1}{2} \sum_k \left(y_k^{[2]} - a_k^{[2]}\right)^2$$

Backward Propogation algo based on Stochastic Gradient Descent

Error

$$z_j^{[1]} = \sum_i x_i w_{ij}^{[1]} \quad \Big| \quad a_j^{[1]} = \frac{1}{1+e^{-z_j}} \rightarrow z_k^{[2]} = \sum_j a_j^{[1]} w_{jk}^{[2]} \quad \Big| \quad a_k^{[2]} = \frac{1}{1+e^{-z_k^{[2]}}} \rightarrow E = \frac{1}{2}\sum_k \left(y_k - a_k^{[2]}\right)^2$$

E_1, E_2, E_3, E_K

$$a = \frac{1}{(1+e^{-z})}$$

$$\frac{\partial a}{\partial z} = -1 * \frac{e^{-z}}{(1+e^{-z})^2} * -1$$

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

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

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

$$= a(1-a)$$

Aim: To find $\Delta w_{jk} \left(\frac{\partial E}{\partial w_{jk}}\right)$ and $\Delta w_{ij} \left(\frac{\partial E}{\partial w_{ij}}\right)$

$$\frac{\partial E}{\partial w_{jk}^{[2]}} = \frac{\partial E}{\partial a_k^{[2]}} * \frac{\partial a_k^{[2]}}{\partial z_k^{[2]}} * \frac{\partial z_k^{[2]}}{\partial w_{jk}^{[2]}}$$

$$\Delta w_{jk}^{[2]} = \frac{2}{2} \cancel{\sum_k} \underbrace{\left(y_k - a_k^{[2]}\right)} * (-1) * a_k^{[2]} * \left(1 - a_k^{[2]}\right) * \cancel{\sum_j} a_j^{[1]}$$

$$w_{jk} = w_{jk} - \alpha \Delta w_{jk}$$

$$= w_{jk} + \alpha \underbrace{\left(y_k - a_k^{[2]}\right) * a_k^{[2]}\left(1 - a_k^{[2]}\right)}_{E_k} * \cancel{\sum_j} a_j^{[1]}$$

For

$$\frac{\partial E}{\partial w_{ij}} = \sum_k \frac{\partial E}{\partial z_k} * \frac{\partial z_k^{[2]}}{\partial a_j^{[1]}} * \underbrace{\frac{\partial a_j^{[1]}}{\partial z_j^{[1]}}} * \underbrace{\frac{\partial z_j^{[1]}}{\partial w_{ij}^{[1]}}}$$

$$= -1 * \left(y_k - a_k^{[2]}\right) * a_k^{[2]} * \left(1 - a_k^{[2]}\right)$$

$$* \underbrace{\sum_k w_{jk}^{[2]} \Delta w_{jk}} * \underbrace{a_j^{[1]} * \left(1 - a_j^{[1]}\right)} * \underbrace{x_i}$$

So, for each neuron in O/P layer

$$E_k^{[2]} = \left(y_k - a_k^{[2]}\right) * a_k^{[2]} * \left(1 - a_k^{[2]}\right) \text{ or } \left(y_k - \hat{y}_k\right) * \hat{y}_k * \left(1 - \hat{y}_k\right)$$

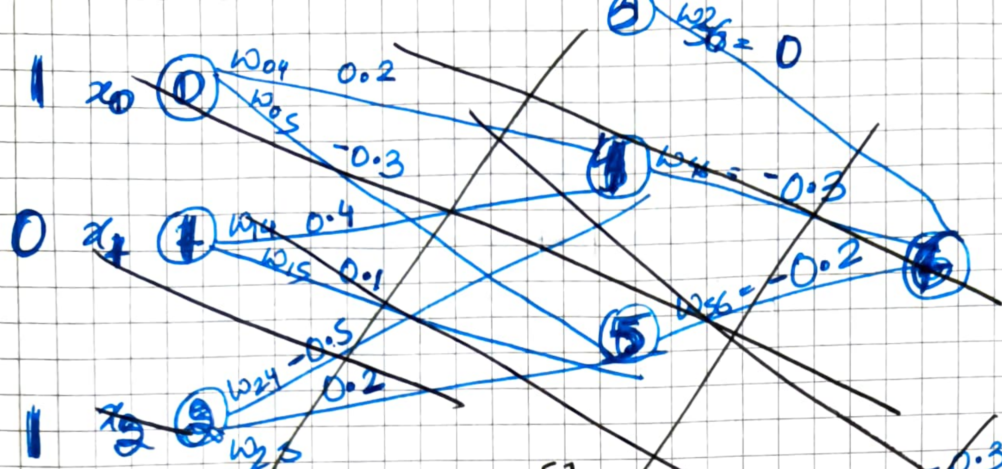
$$w_{jk} = w_{jk} + \alpha E_k * a_j^{[1]}$$

For each neuron in hidden layer

$$E_j^{[1]} = \cancel{E_j^{[2]} *} a_j^{[1]} * \left(1 - a_j^{[1]}\right) * \sum_k w_{jk}^{[2]} E_k$$

$$w_{ij} = w_{ij} + \alpha E_j^{[1]} * x_i^{[0]}$$

Implied

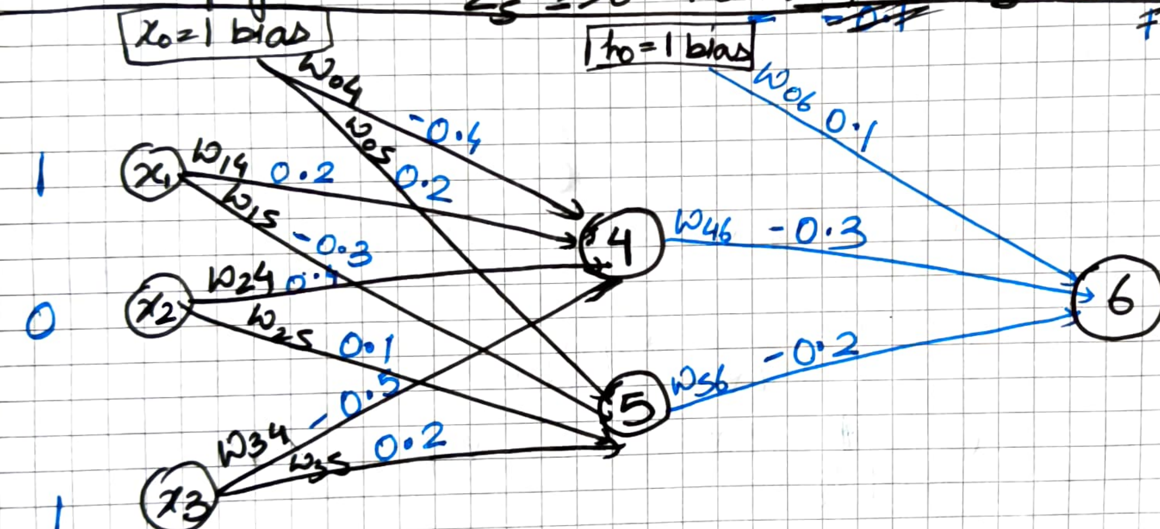


Solved Example.

~~Forward Propagation~~

~~$z_4^{[1]} = 0.2 + 0 + 0 - 0.5 = -0.3 \quad a_4^{[1]} = \frac{1}{1+e^{0.3}}$~~

~~$z_5^{[1]} = -0.3 + 0 + 0.2 = -0.1 \quad a_5^{[1]} = \frac{1}{F}$~~



Forward Propagation

$$z_4^{[1]} = -0.4 + 0.2 - 0.5 = -0.7 \Rightarrow a_4^{[1]} = \frac{1}{1+e^{0.7}} = 0.332$$

$$z_5^{[1]} = 0.2 - 0.3 + 0.2 = 0.1 \Rightarrow a_5^{[1]} = \frac{1}{1+e^{-0.1}} = 0.525$$

$$z_6^{[1]} = 0.1 * 1 + (-0.3)(0.332) + (-0.2)(0.525)$$
$$= -0.105$$

$$a_6^{[1]} = \frac{1}{1+e^{+0.105}} = 0.474$$

Error for neuron at Output layer

$\delta_6 = (y_6 - a_6^{[2]}) * a_6^{[2]} * (1 - a_6^{[2]})$

$= (1 - 0.474) * 0.474 * (1 - 0.474) = 0.1311$

$\Delta w_{06} = 0.1311 * 1$ | $w_{06} = 0.1 + 0.9\, \Delta w_{06} = 0.218$

$\Delta w_{46} = 0.1311 * 0.332$ | $w_{46} = -0.3 + 0.9\, \Delta w_{46} = -0.261$

$\Delta w_{56} = 0.1311 * 0.526$ | $w_{56} = -0.2 + 0.9\, \Delta w_{56} = -0.130$

Error for neuron at hidden layer

$\delta_4 = a_4^{[1]} * (1 - a_4^{[1]}) * \sum_k w_{4k}\, \delta_k$ $w_{46}\, \delta_6$

$= 0.332 * (1 - 0.332) * (-0.3)(0.1311)$

$= -0.0087$

Purane weights

$\delta_5 = a_5^{[1]} * (1 - a_5^{[1]}) * \sum_k w_{4k}\, \delta_k$

$= 0.525 * (1 - 0.53) * (-0.3)(0.1311) = -0.0065$

$\Delta w_{04} = $ $w_{04}^{w_{ij}} = -0.4 + 0.9 * \delta_4 x_0$

$= -0.4 * 0.9 * (-0.0087) * 1 = -0.408$

$\Delta w_{05} = $ $\longrightarrow w_{05} = 0.2 + 0.9 * (-0.0065)^{\delta_5 x_0} * 1$

$= 0.194$

$w_{14} = 0.2 + 0.9 * \delta_4 x_1 = 0.2 + 0.9 * (-0.0087) * 1 = 0.192$

$w_{15} = -0.3 + 0.9 * \delta_5 x_1 = 0.3 + 0.9 * (0.0065) * 1$

$= -0.306$

⋮