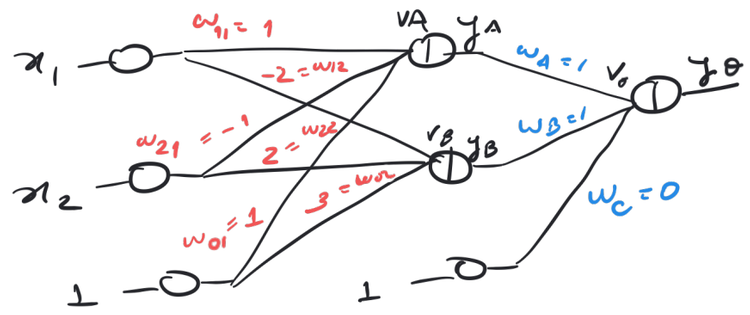
DATA: $x = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ LABEL: $t = 1$



$W_{input \to hidden} = W_{iH} = \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 1 & 3 \end{bmatrix}$ includes bias $(2+1) \times 2$

$W_{hidden \to output} = W_{H0} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ includes bias $(2+1) \times 1$

ACTIVATION FCT.:

$\phi(x) = x$

$\phi'(x) = 1$

PREDICTIONS:

$$y_H = \phi(W_{iH}^T \cdot x_{input}) = \phi\left(\begin{bmatrix} 1 & -1 & 1 \\ -2 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}\right) = \phi\left(\begin{bmatrix} -1 \\ 7 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 7 \end{bmatrix} = \begin{bmatrix} y_A \\ y_B \end{bmatrix}$$

$$y_0 = \phi(W_{H0}^T \cdot y_H) = \phi\left(\begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 7 \\ 1 \end{bmatrix}\right) = \phi(6) = 6 \neq 1$$

Updating the weights directly connected to output:

Let $J(w) = \frac{1}{2} \sum_{i=1}^{N} e_i^2$ with online learning

$$\frac{\partial J(w)}{\partial w_A} = (1-6) \times (-1) \times 1 \times y_A = -5$$

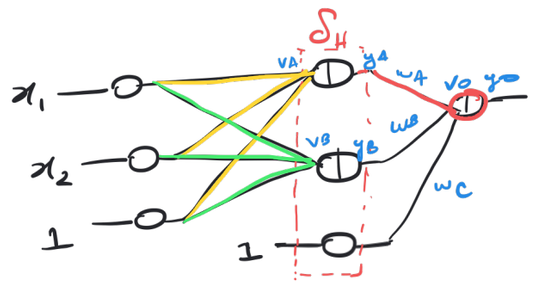
$$\frac{\partial J(w)}{\partial w_B} = (1-6) \times (-1) \times 1 \times y_B = 3.5$$

$$\frac{\partial J(w)}{\partial w_C} = (1-6) \times (-1) \times 1 \times 1 = 5$$

$$\therefore w_{HO} = w_{HO} - \eta \times \begin{bmatrix} -5 \\ 3.5 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} -0.05 \\ 0.35 \\ 0.05 \end{bmatrix} = \begin{bmatrix} 1.05 \\ 0.65 \\ -0.05 \end{bmatrix}$$

Updating weights in the Hidden layer:

Let the local gradient for hidden layer be δ_H:



$$\delta_{H1} = -\left[\sum_{o=1}^{1} E_o \cdot (-1) \cdot \phi'(v_o) \cdot w_A\right] \cdot \phi'(v_A) \quad , \quad E_o = t_o - y_o = 1 - 6$$

$$= -\left[(1-6)(-1) \cdot 1 \cdot 1\right] \cdot 1 = -5$$

Similarly, $\delta_{H2} = -5$ and $\delta_{H3} = 0$.

So, $\delta_H = \begin{bmatrix} -5 \\ -5 \\ 0 \end{bmatrix}$

Updating weights:

$$w_{i1} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + \eta \cdot \delta_{H1} \cdot \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - 0.05 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1.1 \\ 0.95 \end{bmatrix}$$

$$w_{i2} = \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix} + \eta \cdot \delta_{H2} \cdot \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix} - 0.05 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 1.9 \\ 2.95 \end{bmatrix}$$