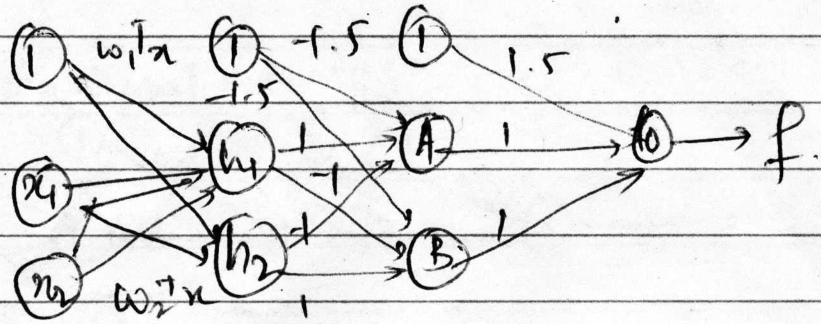
Exercise 7.3



We have, $h_1(x) = \text{sign}(w_1^T x)$ $\quad h_2(x) = \text{sign}(w_2^T x)$ } 1st hidden layer

Now, $A = (1.5)(1) + (1)(h_1(x)) + (-1)(h_2(x))$

$\therefore A(x) = \text{sign}(1.5 + h_1(x) - h_2(x))$

similarily, $B = (-1.5)(1) + (-1)(h_1(x)) + (1)(h_2(x))$

$= -1.5 - h_1(x) + h_2(x)$

$\therefore$ output $B(x) = \text{sign}(-1.5 - h_1(x) + h_2(x))$

finally for output layer:-

$= (1.5)(1) + (1)A(x) + (1)B(x)$

$\therefore f = \text{sign}(1.5 + A(x) + B(x))$

$\therefore$ putting $A(x)$ & $B(x)$,

$f = \text{sign}(1.5 + \text{sign}(1.5 + h_1(x) - h_2(x)) + \text{sign}(h_2(x) - h_1(x) - 1.5))$

$$\boxed{f = \text{sign}\left(\text{sign}\left(h_1(x) - h_2(x) - \frac{3}{2}\right) - \text{sign}\left(h_1(x) - h_2(x) + \frac{3}{2}\right) + \frac{3}{2}\right)}$$

Hence proved, where $h_1(x) = \text{sign}(w_1^T x)$

& $h_2(x) = \text{sign}(w_2^T x)$.

classmate

Date ____
Page ____

Exercise 7.7

$$E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N}(\tanh(w^T x_n) - y_n)^2$$

S.T

$$\nabla E_{in}(w) = \frac{2}{N}\sum_{n=1}^{N}(\tanh(w^T x_n) - y_n)(1-\tanh^2(w^T x_n))x_n.$$

∴ Taking derivative w.r.t w of $E_{in}(w)$.

$$\therefore \frac{\partial E_{in}(w)}{\partial w} = \frac{2}{N}\sum_{n=1}^{N}\frac{\partial(\tanh(w^T x_n) - \cancel{y_n}^{\,0})}{\partial w}.(\tanh(w^T x_n) - y_n)$$

$$= \frac{2}{N}\sum_{n=1}^{N}(1-\tanh^2(w^T x_n))\frac{\partial(w^T x_n)}{\partial w} . \tanh(w^T x_n - y_n)$$

$$= \frac{2}{N}\sum_{n=1}^{N}(1-\tanh^2(w^T x_n))(x_n)(\tanh(w^T x_n) - y_n).$$

$$\therefore = \frac{2}{N}\sum_{n=1}^{N}(\tanh(w^T x_n) - y_n)(1-\tanh^2(w^T x_n) - y_n).x_n$$

Hence proved.

Also, if $w \to \infty$. $\Rightarrow \lim_{w \to \infty} \nabla E_{in}(w)$

$$\therefore \lim_{w \to \infty} \nabla E_{in}(w) = \lim \frac{2}{N}\sum_{n=1}^{N}(\tanh(w^T x_n) - y_n)(1-\tanh^2(w^T x_n) - y_n)x_n$$

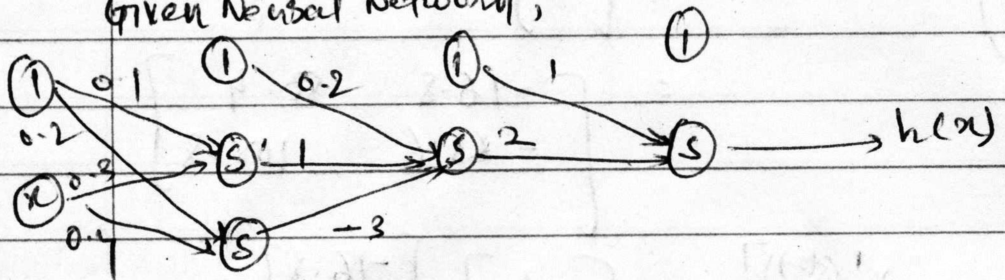
when $w \to \infty$, $\tanh^2(w^T x_n) \longrightarrow 1$.

∴ The quantity $(1-\tanh^2(w^T x_n)) = 1 - 1 = 0$.

Hence gradient = 0 no matter what & it will take long long time to converge, or might go in an infinite loop, as well as if this happens, ~~pocket~~ perceptron will have difficult time to classify the data properly & will make things difficult to optimize.

Exercise 7.8

Given Neural Network,



we have asked to compute $s^{(l)}$, $x^{(l)}$, $\delta^{(l)}$ & $\partial e / \partial w^{(l)}$

$\therefore$ $W^{(1)} = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix}$ $\quad W^{(2)} = \begin{bmatrix} 0.2 \\ 1 \\ -3 \end{bmatrix}$ $\quad W^{(3)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$x = 2, y = 1$

$\therefore$

$X(0)$	$S(1)$	$X^{(1)}$	$S^{(2)}$	$X^{(2)}$	$S^{(3)}$	$X^{(3)}$
$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 0.7 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0.7 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -2.1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ -2.1 \end{bmatrix}$	$\begin{bmatrix} -3.2 \end{bmatrix}$	$\begin{bmatrix} -3.2 \end{bmatrix}$

$\therefore$ $\delta^{(3)} = 2(x^{(2)} - 1)(1)$ $\quad$ As $\theta'(s) = 1$

$= 2(-3.2 - 1)$

$= (2)(-4.2) = \boxed{[-8.4]}$ —— (1)

$\delta^{(2)} = \theta'(s^{(2)}) \otimes [w^{(3)} \delta^{(3)}]_1^{d^{(2)}}$

$= 1 \otimes [2 \times (-8.4)] = [-16.8]$ —— (2)

$\delta^{(1)} = \theta'(s^{(1)}) \otimes [w^{(2)} \delta^{(2)}]_1^{d^{(1)}} = 1 \otimes \left[\begin{bmatrix} 1 \\ -3 \end{bmatrix} [-16.8] \right]$

$= \begin{bmatrix} -16.8 \\ 50.4 \end{bmatrix}$ —— (3)

$$\frac{\partial(e)}{\partial(w^{(1)})} = x^{0}(\delta^{(1)})^{T} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} -16.8 & 50.4 \end{bmatrix}$$

$$= \begin{bmatrix} -16.8 & 50.4 \\ -33.6 & 100.8 \end{bmatrix}$$

$$\frac{\partial(e)}{\partial(w^{(2)})} = X^{1}(\delta^{(2)})^{T} = \begin{bmatrix} 1 \\ 0.7 \\ 1 \end{bmatrix} \begin{bmatrix} -16.8 \end{bmatrix}$$

$$= \begin{bmatrix} -16.8 \\ -11.76 \\ -16.8 \end{bmatrix}$$

$$\frac{\partial(e)}{\partial(w^{(3)})} = x^{(2)}(\delta^{(3)})^{T} = \begin{bmatrix} 1 \\ -2.1 \end{bmatrix} \begin{bmatrix} -8.4 \end{bmatrix}$$

$$= \begin{bmatrix} -8.4 \\ 17.64 \end{bmatrix}.$$

Thus, all calculations done!