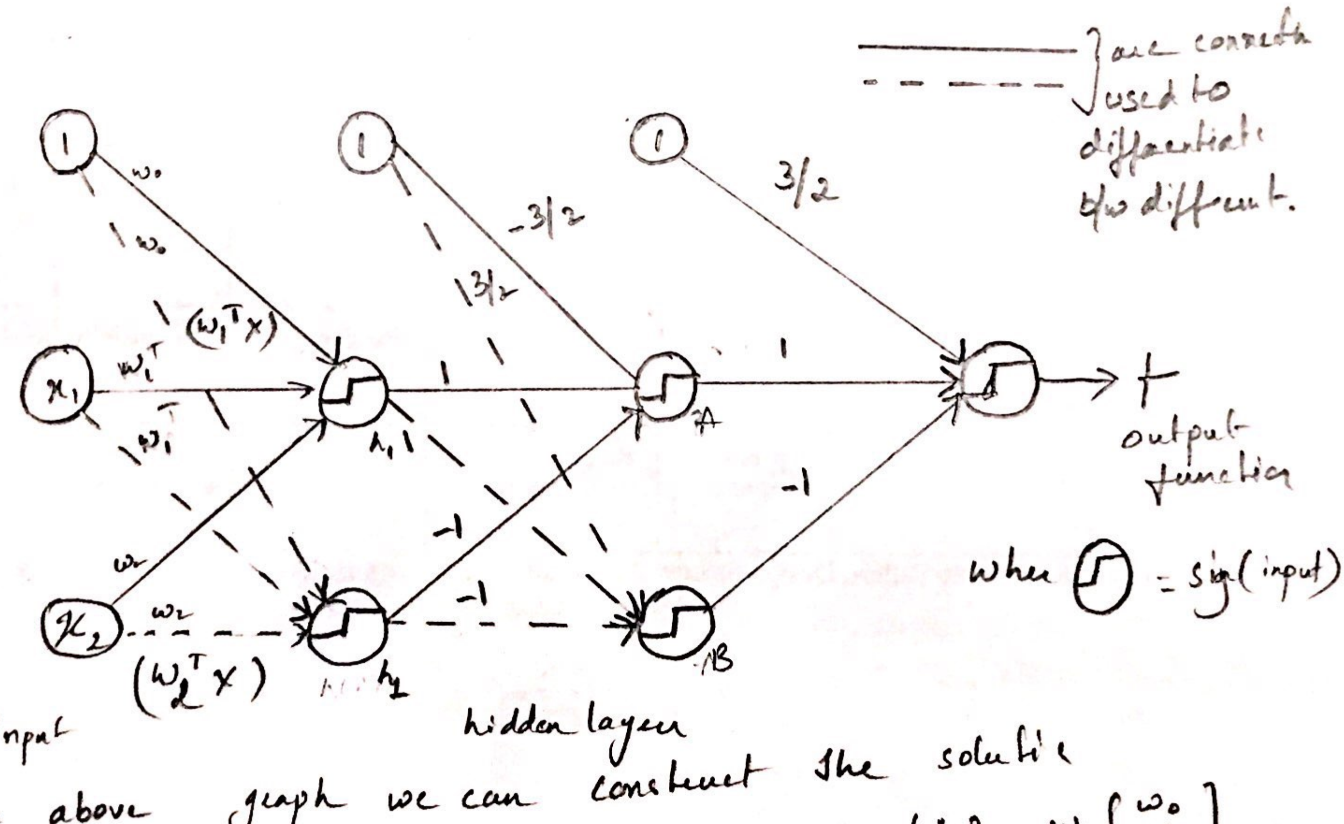
Ex CMPE 257 HW-4 SAMPATH LAKKARAJU 011818781

Ex:- 7.3

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

when $h_1(x) = \text{sign}(w_1^T x)$ & $h_2(x) = \text{sig}(w_2^T x)$

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



from the above graph we can construct the solution

$l_1 = 1$; $h_1(x) = \text{sign}(w_1^T x)$, $h_2 = \text{sign}(w_2^T x)$ when $x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$, $W_i = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$

$l = 2$; $A = \text{sign}\left(h_1(x) - h_2(x) + \frac{3}{2}\right)$; $B = \text{sign}\left(h_1(x) - h_2(x) - \frac{3}{2}\right)$

$l_3 = 3$; $f = \text{sign}\left[\left[\text{sign}\left(\text{sig}(w_1^T x) - \text{sign}(w_2^T x) - \frac{3}{2}\right)\right] - \left[\text{sign}\left(\text{sign}(w_1^T x) + \text{sign}(w_2^T x) + \frac{3}{2}\right)\right] + \frac{3}{2}\right]$

011818781

Ex 7.7

Ans Given $h(x) = \tanh(W^T x)$, and the insample error

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

From Page 85 of LFD the Representation of Ein is in the matrix analog of ordinary differentiation of quadratic and linear functions. To get the gradient of Ein we take instantaneous gradient of each entity in the Representation.

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

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

$$= \frac{1}{N}\sum_{n=1}^{N} 2(\tanh(w^T x_n) - y_n)\left[\frac{d}{dw}\tanh(w^T x_n) - \frac{d}{dw}y_n\right]$$

$$= \frac{2}{N}\sum_{n=1}^{N}(\tanh(w^T x_n) - y_n)\left[(1-\tanh^2(w^T x_n))\cdot\frac{d}{dw}w^T x_n + 0\right]$$

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

0481875|

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

This is the gradient

we know that

$\tanh(\infty) = 1$

$\tanh(0) = 0$

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

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

$= \frac{2}{N}\sum_{n=1}^{N}(1 - y_n)(0)x_n \quad [\text{if } x_n \neq 0]$

$\nabla E_{in}(w) = 0$ which implies the gradient becomes zero, as the weights point is on a flat point of the function.

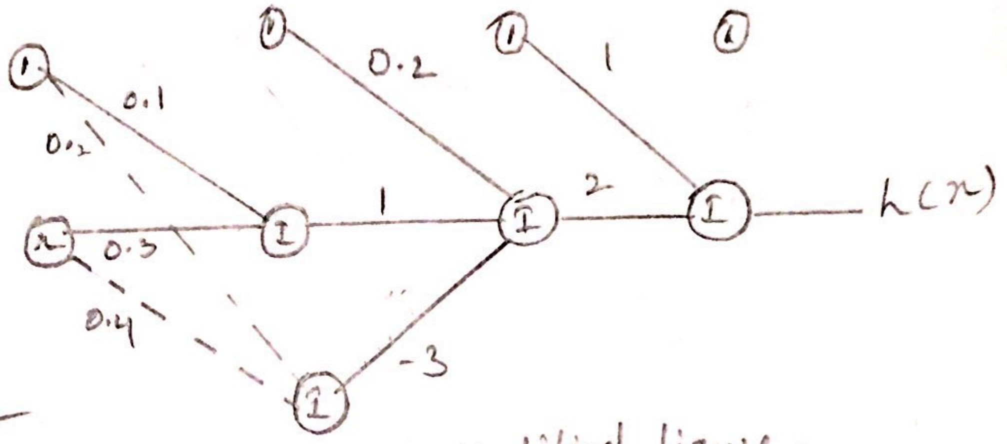
if the gradient is zero, the weight cannot be change according to it to get a minimal posistion in the plane, This provides an issue in optimal solution for perceptron.

0118187081

Ex 7.8:-

Example 7.1 :-

Except
Output transformal
Is identity.



Modified figure.

Let us Represent it
using I when $\theta(s) = s$

From the example we have $x = 2$, $y = 1$ & the weight matrix are

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

Since the output transformation is Identity we can Think that $\theta(x) = x$, $\theta'(x) = 1$ and the following as need to be followed.

The steps Need to be followed.

Forward propagation

(i) $x^{(0)} \leftarrow x$ [Initialization]

(ii) for $l = 1$ to L compute

$$s^{(l)} \leftarrow (W^{(l)})^T x^{(l-1)}$$

$$x^{(l)} \leftarrow \begin{bmatrix} 1 \\ \theta(s^{(l)}) \end{bmatrix}$$

(iii) $h(x) = x^{(L)}$ [output]

Backpropagation to compute Sensitivities $\delta^{(l)}$

Input a data point (x, y); Run forward propagation to get $s^{(l)}, x^{(l)}$,

(i) $\delta^{(L)} \leftarrow 2(x^{(L)} - y)\,\theta'(s^{(L)})$ Initializ

$\theta'(s^{(L)}) = 1 \quad \because \theta(s) = s.$

(ii) for $l = L-1$ to 1 compute

Let $\theta'(s^{(l)}) = [U]_1^{d(l)}$ U is a unit vector

(iii) Compute $\delta^{(l)}$ from $\delta^{(l+1)}$

$$\delta^{(l)} \leftarrow \theta'(s^{(l)}) \otimes \left[W^{(l+1)} \delta^{(l+1)}\right]_1^{d^{(l)}}$$

011818781

From Example we have $x = 2$, $y = 1$ (data point)

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

$\theta(s) = s$, $\theta'(s) = 1$, $\theta'(s^{(L)}) = U$

Computing x^0, x^1, x^2, x^3; s^1, s^2, s^3.

$$x^0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} ; \quad s^1 = \underbrace{\begin{bmatrix} 0.1 & 0.3 \\ 0.2 & 0.4 \end{bmatrix}}_{(W^1)^T} \underbrace{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}_{x} = \begin{bmatrix} 0.1 + 0.6 \\ 0.2 + 0.8 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 1 \end{bmatrix}$$

$$x^1 = \begin{bmatrix} 1 \\ 0.7 \\ 1 \end{bmatrix} ; \quad s^2 = \begin{bmatrix} 0.2 & 1 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ 0.7 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.2 + 0.7 - 3 \end{bmatrix}$$
$$= \begin{bmatrix} -2.1 \end{bmatrix}$$

output Transformation is identity

$$x^2 = \begin{bmatrix} 1 \\ -2.1 \end{bmatrix}, \quad s^3 = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -2.1 \end{bmatrix} = \begin{bmatrix} 1 - 4.2 \end{bmatrix} = \begin{bmatrix} -3.2 \end{bmatrix}$$

$$x^3 = -3.2$$

For computing Sensitivity we need to initialize

$$\delta^L = \delta^3 \leftarrow 2\,(x^{(L)} - y)\,\theta'(s^{(L)}) = 2(-3.2 - 1)(1)$$

$$\delta^3 = \begin{bmatrix} -8.4 \end{bmatrix}$$

01181878)

$$\delta^{(l)} \leftarrow \theta'(s^{(l)}) \otimes [W^{(l+1)} \delta^{(l+1)}]_1^{d(l)}$$ [∵ the bias is not necessary for calc sensitivities

$$\delta^2 = U \otimes [W^3 \delta^3]_1^d = [1] \left[[2] [-8.4] \right]$$

$$\delta^2 = [-16.8]$$

$$\delta^1 = U \otimes [W^2 \delta^2] = U \left[\begin{bmatrix} 1 \\ -3 \end{bmatrix}_{2\times1} [-16.8]_{1\times1} \right] = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} -16.8 \\ -50.4 \end{bmatrix}_{2\times1}$$

$$= \begin{bmatrix} -16.8 \\ 50.4 \end{bmatrix}$$

for the gradients we use.

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

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

$$\frac{\partial e}{\partial w^2} = (x^1)(\delta^2)^T = \begin{bmatrix} 1 \\ 0.7 \\ 1 \end{bmatrix} [-16.8] = \begin{bmatrix} -16.8 \\ -11.76 \\ -16.8 \end{bmatrix}$$

$$\frac{\partial e}{\partial w^3} = (x^2)(\delta^3)^T = \begin{bmatrix} 1 \\ -2.1 \end{bmatrix} [-8.4] = \begin{bmatrix} -8.4 \\ 17.64 \end{bmatrix}$$