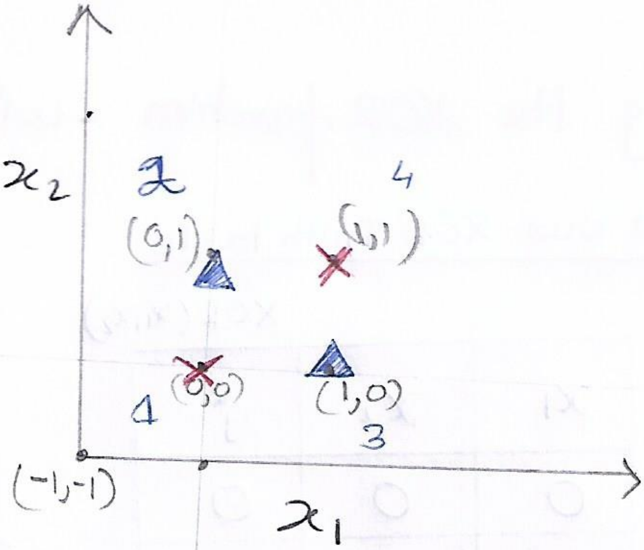
Introduction to Feed Forward Networks.

The XOR operation

x_1	x_2	$XOR(x_1,x_2)$
0	0	0
0	1	1
1	0	1
1	1	0

Let us see how the plot looks!



We have the following input: (x_1, x_2)

1] (0,0)

2] (0,1)

3] (1,0)

4] (1,1).

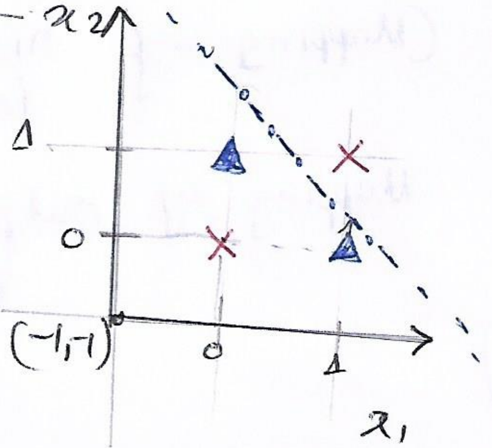
X → output takes 0

▲ → output takes 1

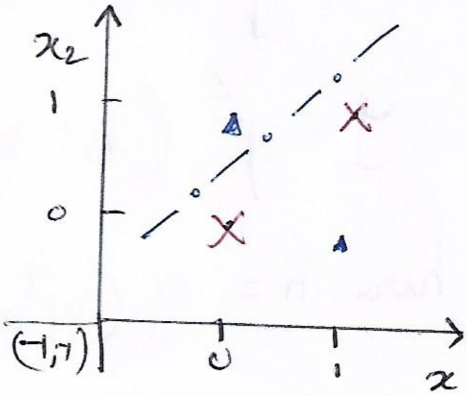
Q] Can we draw a straight line to separate the X from the ▲'s?

Let us try

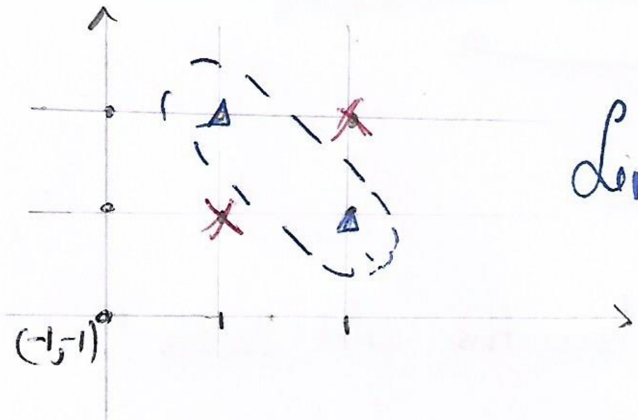
Trial 1



Trial 2



How can we separate the 1's from the 0's.



Linear Separation is not possible!

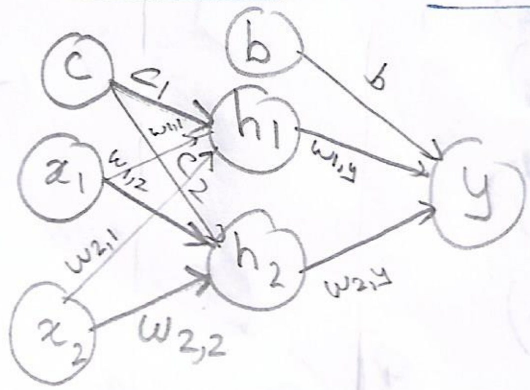
Learning the XOR function with Feed Forward Networks

Revisit our XOR Truth table.

$XOR(x_1, x_2)$

x_1	x_2	y^*
0	0	0
0	1	1
1	0	1
1	1	0

Consider the following Feed fwd network



Consider a mapping as follows

$f(x : w, b) = x^T w + b$

Now $h = f^1(x : w, c)$ (Mapping of hidden to input Layer

$y = f^2(h : w, b)$ Mapping of output to hidden.

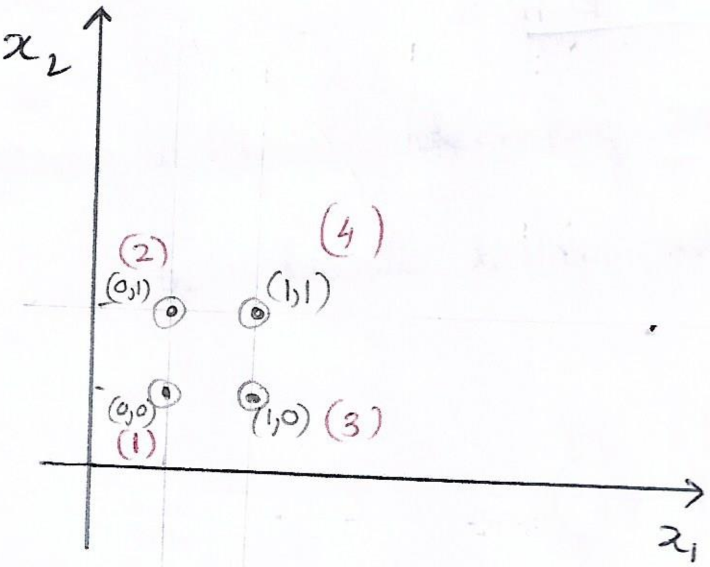
here $h = g(W^T x + c)$

g is the activation function to be used.

Solving XOR by Linear Functions.

Consider original x space as shown below.

x_1	0	0	1	1
x_2	0	1	0	1
y	0	1	1	0



Let us try Fitting Linear Functions with weights w_1, w_2

Thus for the four points in question, we have the following.

$$w_1 \times 0 + w_2 \times 0 + b = 0 \quad — (1)$$

$$w_1 \times 0 + w_2 \times 1 + b = 0 \quad — (2)$$

$$w_1 \times 1 + w_2 \times 0 + b = 0 \quad — (3)$$

$$w_1 \times 1 + w_2 \times 1 + b = 0 \quad — (4)$$

if from (1) we get $b = 0$

and from (2), (3) if $w_1 = 1$ or $w_2 = 1$

then (4) does NOT hold!

Thus a Linear Fit is not directly possible!

Drawbacks of Linear Model.

Vector of weights + Scalar bias parameter

Affine transformation from 'x' to 'h'.

1. Entire vector of bias parameters would be needed!
2. Activation Function to be applied elementwise.

Consider the following.

$$h_i = g(x^T w_{:i} + c_i)$$

ReLU → Rectified Linear Unit.

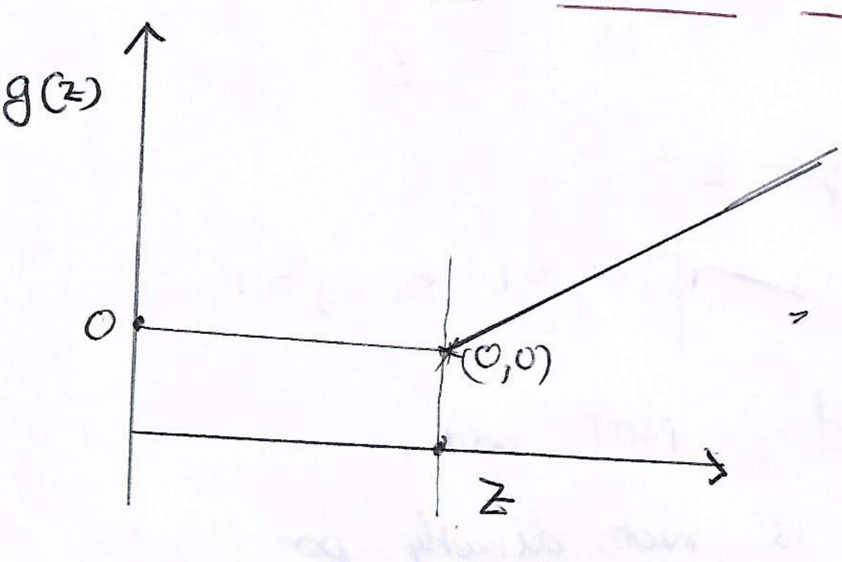
↳ how do we model this?

$$g(z) = \max(0, z)$$

Thus our final output maps to input in the following way:

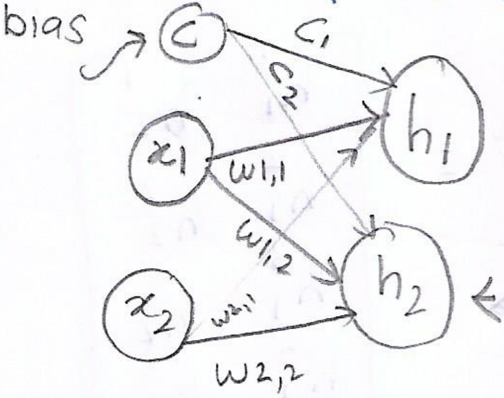
$$f(x : W, c, w, b) = w^T \max\{0, (W^T x + c)\} + b$$

RELU → Activation Function



Let us try Linear Activation fn. with Feed Forward Network.

Consider input to hidden Layer diagram;



h_1 ← hidden Layer Node 1

h_2 ← hidden Layer Node 2

Thus we have: for the 1st hidden node

$$0 \times w_{1,1} + 0 \times w_{2,1} + C_1 = h_1^1$$

$$0 \times w_{1,1} + 1 \times w_{2\times 1} + C_1 = h_1^2$$

$$1 \times w_{1,1} + 0 \times w_{2\times 1} + C_2 = h_1^3$$

$$1 \times w_{1,1} + 1 \times w_{2,1} + C_2 = h_1^4$$

[X] MATRIX FOR[M]

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \overset{[W]}{\begin{bmatrix} w_{1,1} \\ w_{2,1} \end{bmatrix}} + \begin{bmatrix} C_1 \end{bmatrix} = \begin{bmatrix} h_1^1 \\ h_1^2 \\ h_1^3 \\ h_1^4 \end{bmatrix}$$

and similarly for the 2nd hidden node

$$0 \times w_{1,2} + 0 \times w_{2,2} + C_2 = h_2^1$$

$$0 \times w_{1,2} + 1 \times w_{2\times 2} + C_2 = h_2^2$$

$$1 \times w_{1,2} + 0 \times w_{2\times 2} + C_2 = h_2^3.$$

$$1 \times w_{1,2} + 0 \times w_{2\times 2} + C_2 = h_3^4$$

[X] MATRIX FORM

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} w_{1,2} \\ w_{2,2} \end{bmatrix} + \begin{bmatrix} C_2 \end{bmatrix} = \begin{bmatrix} h_2^1 \\ h_2^2 \\ h_2^3 \\ h_2^4 \end{bmatrix}$$

Now Combining together:

COMBINED MATRIX FORM

$$\underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}}_{[X]} \underbrace{\begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix}}_{[W]} + \begin{bmatrix} c_1 & c_2 \\ c_1 & c_2 \\ c_1 & c_2 \\ c_1 & c_2 \end{bmatrix} = \begin{bmatrix} h_1^1 & h_2^1 \\ h_1^2 & h_2^2 \\ h_1^3 & h_2^3 \\ h_1^4 & h_2^4 \end{bmatrix}$$

Where we have

$$x = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ x_1 & 0 & 0 & 1 & 1 \\ x_2 & 0 & 1 & 0 & 1 \end{array}$$

$$X = x^T = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

Parameters Assumed for solution are as follows:

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$ ← Weights matrix (input → hidden)

$$c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$ ← bias Vector (input → hidden)

$$w = \begin{bmatrix} w_{1,y} \\ w_{2,y} \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$ ← Weights vector (hidden → output)

$b = [0]$ ← bias (hidden → output)

Computing the Solution with Assumed Parameters.

Step 1: Multiply input matrix with 1st Layer weights

$$[X][W] = \underset{[4\times2]}{\overset{X}{\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}}} \underset{[2\times2]}{\overset{W}{\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}}} = \underset{[4\times2]}{\overset{XW}{\begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 2 & 2 \end{bmatrix}}}$$

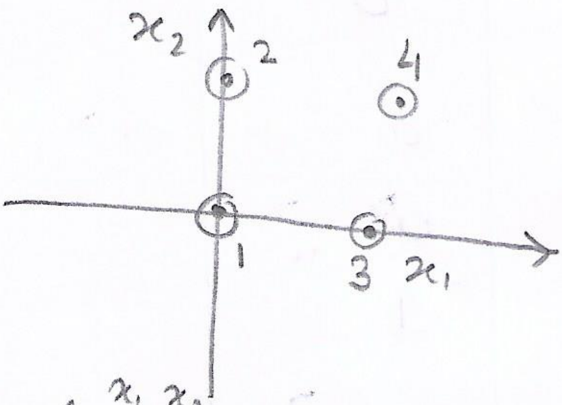
Step 2: Add the bias vector

$[X][W] + C$

Transformed Space.

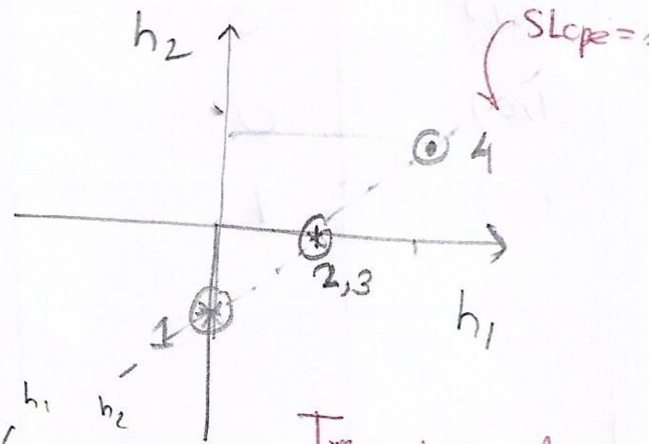
$$\begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 0 & -1 \\ 0 & -1 \\ 0 & -1 \end{bmatrix} = \begin{array}{c} \begin{array}{cc} h_1 & h_2 \end{array} \\ \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ 1 & 0 \\ 2 & 1 \end{bmatrix} \end{array}$$

Plot Original space and Transformed space



	x_1, x_2
1	(0, 0)
2	(0, 1)
3	(1, 0)
4	(1, 1)

ORIGINAL INPUT SPACE



	h_1, h_2
1	(0, −1)
2	(1, 0)
3	(1, 0)
4	(2, 1)

Transformed SPACE

Step 3] Apply ReLU Activation

$g(z) = \max\{0, z) = \max(0, XW + c)$

∮ Consider 1st node of hidden Layer. (h_1)

Original input	h_1	ReLU	$g(h_1)$
(0,0)	0	$g(h_1^1)$	0
(0,1)	1	$g(h_1^2)$	1
(1,0)	1	$g(h_1^3)$	1.
(1,1)	2	$g(h_1^4)$	2

∮ Consider for 2nd Node of hidden Layer (h_2)

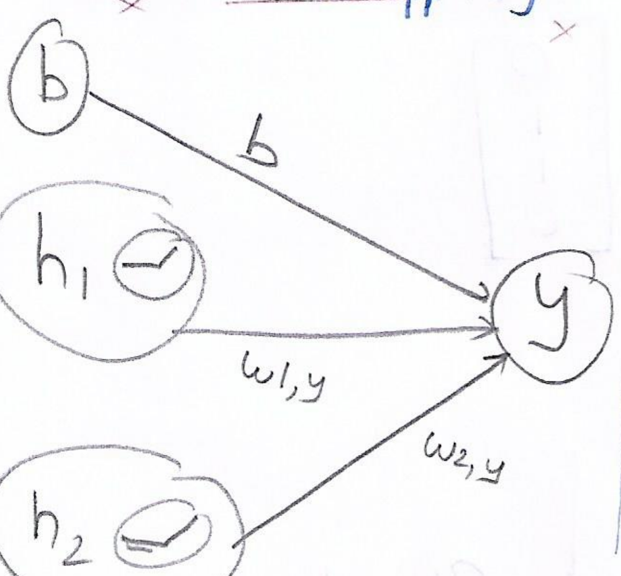
Original input	h_2	ReLU	$g(h_2)$
(0,0)	-1	$g(h_2)^1$	0
(0,1)	0	$g(h_2)^2$	0
(1,0)	0	$g(h_2)^3$	0
(1,1)	1	$g(h_2)^4$	1

Thus post ReLU we have the input transformed as.

$$\begin{array}{cc} x_1 & x_2 \end{array} \qquad \qquad \qquad \qquad \begin{array}{cc} h_1' & h_2' \end{array}$$

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \rightarrow \text{Post ReLU} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 2 & 1 \end{bmatrix}$$

Now consider mapping from hidden to output.



We have.

$$b + 0 \times w_{1,y} + 0 \times w_{2,y} = y^1$$
$$b + 1 \times w_{1,y} + 0 \times w_{2,y} = y^2$$
$$b + 1 \times w_{1,y} + 0 \times w_{2,y} = y^3$$
$$b + 2 \times w_{2,y} + 1 \times w_{2,y} = y^4$$

Putting it all together.

XOR Table

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

hidden Layer to output

MATRIX FORM

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} w_{1,y} \\ w_{2,y} \end{bmatrix} + \begin{bmatrix} b \end{bmatrix} = \begin{bmatrix} y^1 \\ y^2 \\ y^3 \\ y^4 \end{bmatrix}$$

Now Consider Solution parameters as proposed.

$$w = \begin{bmatrix} w_{1,y} \\ w_{2,y} \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

$$b = [0]$$

Thus we have.

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} + [0] = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

Output

Now Compare!

XOR

x_1	x_2	$XOR(x_1, x_2)$
0	0	0
0	1	1
1	0	1
1	1	0

Output From FFN

$$\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

Food For thought?

How do we get the bias and weights?