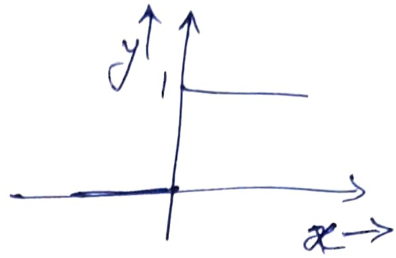
Different Non-linear functions:

① Threshold function:

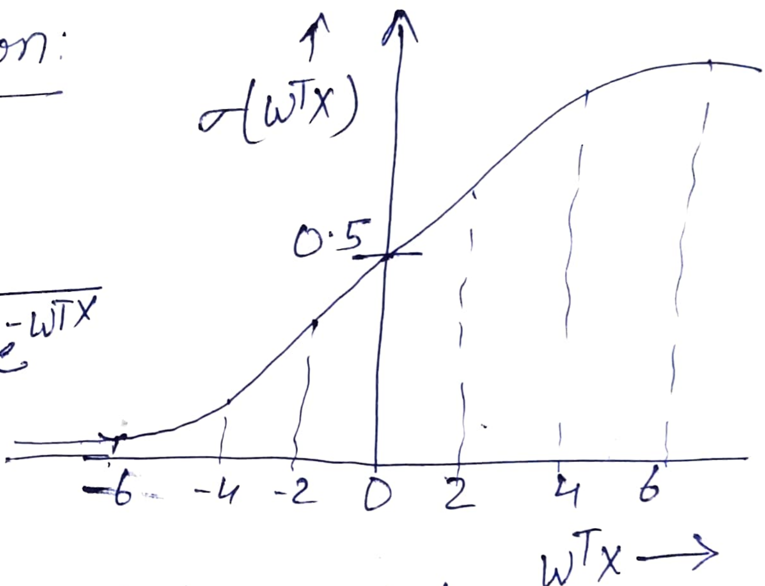
$$Y = \begin{cases} 1 & ; x \geq 0 \\ 0 & ; x < 0 \end{cases}$$



② Logistic Regression:

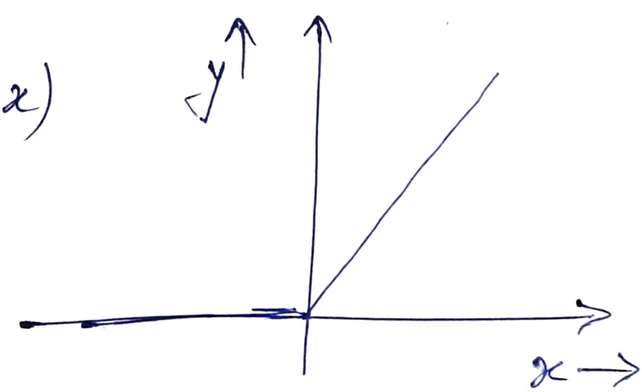
Sigmoid function:

$$\sigma(W^T x) = \frac{1}{1+e^{-W^T X}}$$



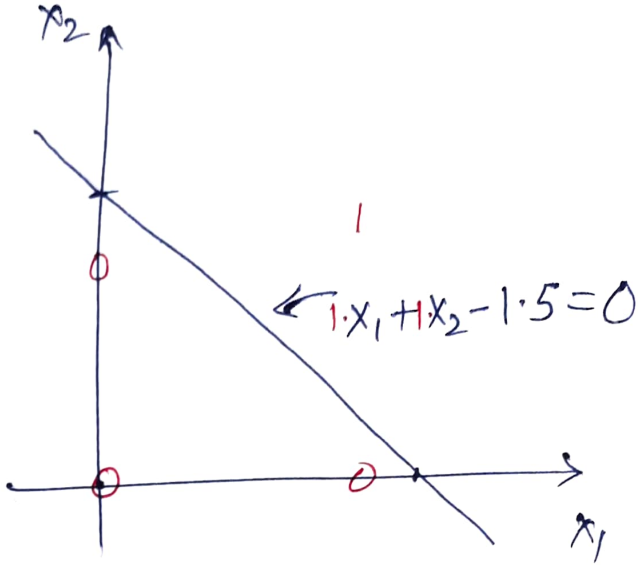
③ ReLU: Rectified Linear Unit:

$$Y = \max(0, x)$$



① AND function:

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



$$X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \quad W = \begin{bmatrix} -1\cdot 5 \\ 1 \\ 1 \end{bmatrix} \Rightarrow X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}_{3\times 4}$$

By considering all feature vectors:

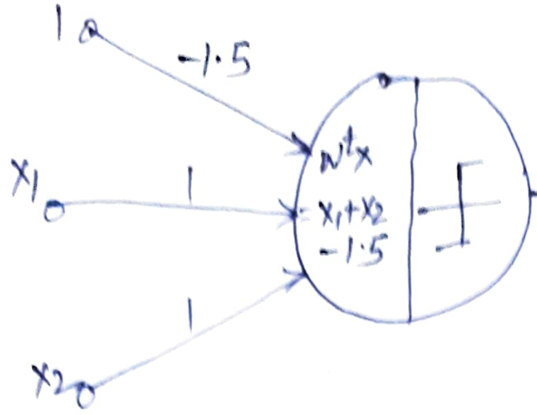
$$X^T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}_{4\times 3} \quad \& \quad W = \begin{bmatrix} -1\cdot 5 \\ 1 \\ 1 \end{bmatrix}_{3\times 1}$$

$$(X^T W) = \begin{bmatrix} -1\cdot 5 \\ -0\cdot 5 \\ -0\cdot 5 \\ 0\cdot 5 \end{bmatrix}_{4\times 1} \quad \underline{\text{OR}} \quad W^T X = \begin{bmatrix} -1\cdot 5 & -0\cdot 5 & -0\cdot 5 & 0\cdot 5 \end{bmatrix}_{1\times 4}.$$

$$y(x^T W) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Non-linear f^n:

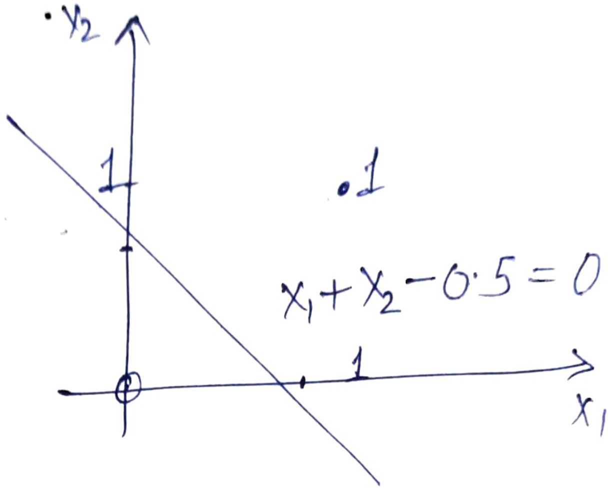
Threshold f^n $y(x) = \begin{cases} 1 ; & x \geq 0 \\ 0 ; & x < 0 \end{cases}$



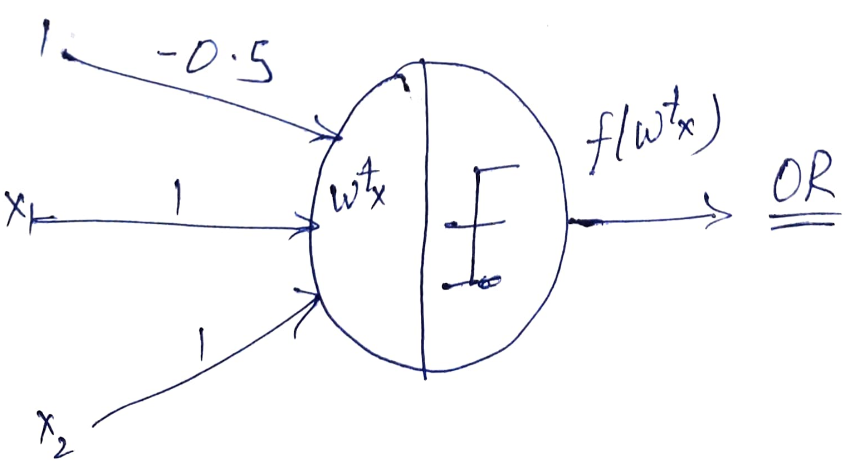
$$y = f(w^t x) = f\left(\sum_{i=0}^{2} w_i x_i\right)$$

② OR Function :

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

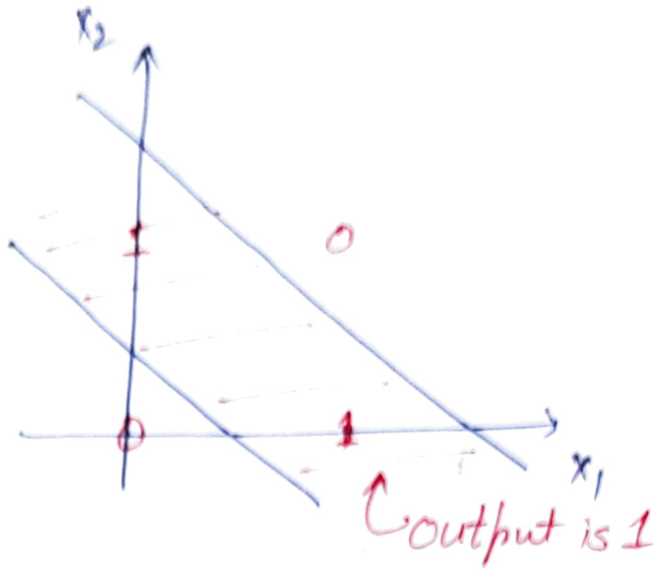


$$x^T w = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -0.5 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.5 \\ 0.5 \\ 1.5 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$



XOR Function:

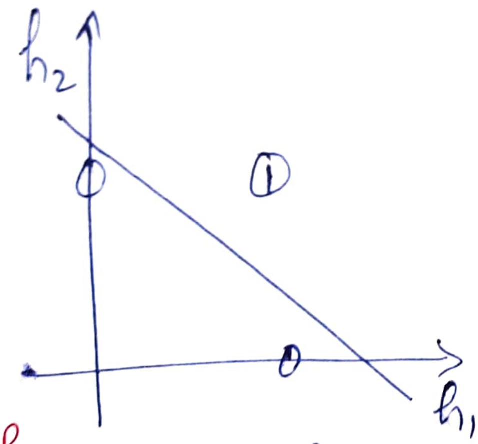
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0



$$x_1 \oplus x_2 = \underbrace{(x_1 + x_2)}_{\text{OR}} \cdot \underbrace{(\bar{x}_1 + \bar{x}_2)}_{\text{NAND}} \qquad \left[\,\overline{x_1 \cdot x_2} = \bar{x}_1 + \bar{x}_2\,\right].$$

AND

x_1	x_2	$h_1 = x_1 + x_2$	$h_2 = \bar{x}_1 + \bar{x}_2$	$h_1 \cdot h_2 = x_1 \oplus x_2$
0	0	0	1	0
0	1	1	1	1
1	0	1	1	1
1	1	1	0	0

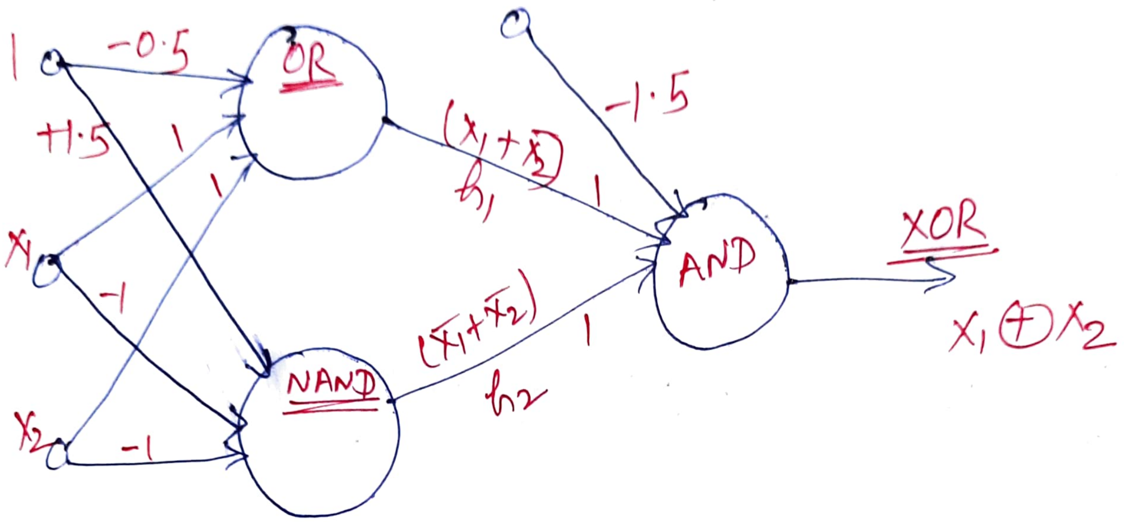


$$\begin{array}{l} \text{OR} \\ \\ \text{NAND weight} \end{array} \underbrace{\begin{bmatrix} -0.5 & 1 & 1 \\ 1.5 & -1 & -1 \end{bmatrix}}_{W_1^t} \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}}_{X} = \begin{bmatrix} -0.5 & 0.5 & 0.5 & 1.5 \\ 1.5 & 0.5 & 0.5 & -0.5 \end{bmatrix} \Rightarrow \int \Rightarrow$$

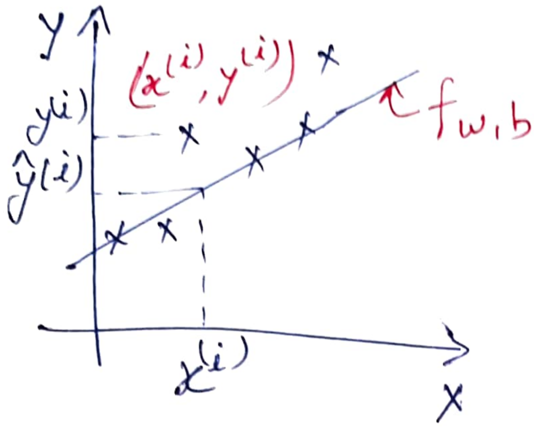
$$\underbrace{\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}}_{h} \begin{array}{l} \Leftarrow h_1 \\ \Leftarrow h_2 \end{array}$$

$$h' W_2 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} -1.5 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.5 \\ 0.5 \\ -0.5 \end{bmatrix} \Rightarrow f \Rightarrow \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$X_1 \oplus X_2$



Cost Function



$(x^{(i)}, y^{(i)})$ = i^{th} training instance with ~~data~~ value label $y^{(i)}$.

m = # training instances

$$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$$

$$f_{w,b}(x^{(i)}) = wx^{(i)} + b$$

cost function: Mean Squared Error (MSE)

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

objective: Find w, b such that $\hat{y}^{(i)}$ is close to $y^{(i)}$ $\forall$ $(x^{(i)}, y^{(i)})$.

Model: $f_{w,b}(x) = wx + b$

Parameters: w, b

cost function: $J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right)^2$

Goal: $\underset{w,b}{\text{minimize}}\ J(w,b)$

Gradient descent algorithm:

$$w = w - \alpha \frac{\partial}{\partial w} J(w,b)$$

Here: α = learning rate

$\frac{\partial}{\partial w} J(w,b)$ = derivative of cost f^n w.r.to w.

$$b = b - \alpha \frac{\partial}{\partial b} J(w,b)$$

NOTE: Update w & b simultaneously based on old w & b value.

Correct Simultaneous update:

$$temp\_w = w - \alpha \frac{\partial J(w,b)}{\partial w}$$

$$temp\_b = b - \alpha \frac{\partial}{\partial b} J(w,b)$$

$$w = temp\_w$$

$$b = temp\_b$$

Incorrect way:

$$temp\_w = w - \alpha \frac{\partial}{\partial w} J(w,b)$$

$$w = temp\_w$$

$$temp\_b = b - \alpha \frac{\partial}{\partial b} J(w,b)$$

$$b = temp\_b.$$