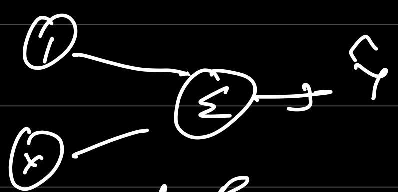
Linear Regression - $\hat{y} = w^T x + w_0$

Logistic Regression

$$\hat{y} = \sigma(z), \quad z = w^T x + w_0$$



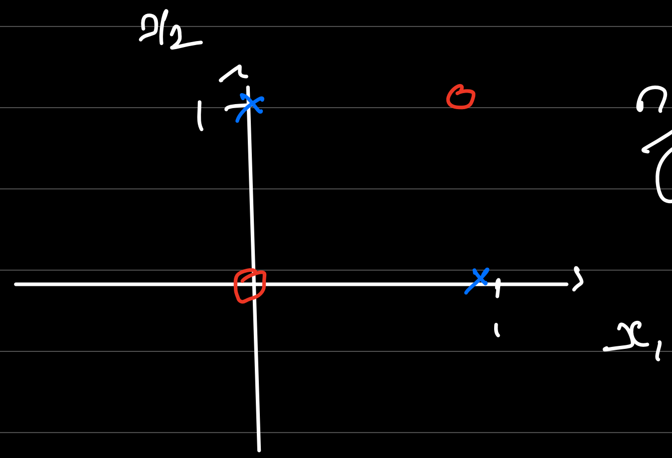
LR

$$\sigma(z) = \frac{1}{1+e^{-z}} = \hat{y}$$

XOR - GATE

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

x_i, 0 or 1



Linearly not separable
Classification

$$\hat{y} = x_1 + x_2 - 2x_1x_2 \rightarrow$$ Model like a regression problem

$$\hat{y} = w_1x_1 + w_2x_2 + w_0 x_3$$

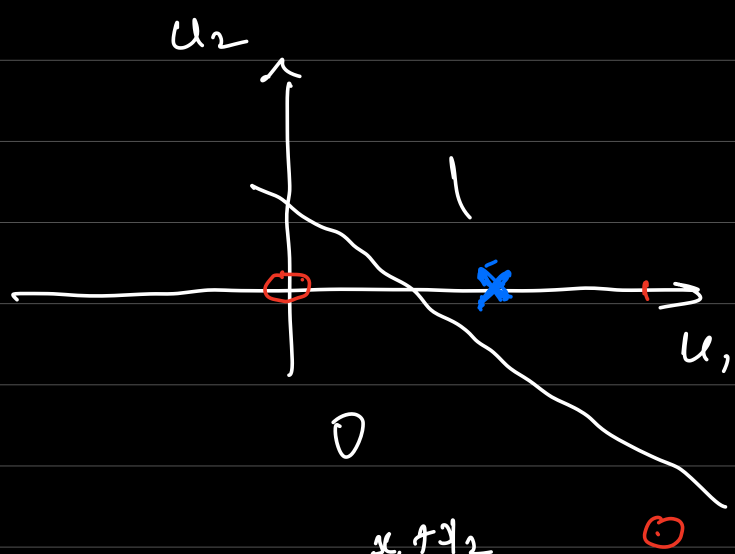
$$\hat{y} = w_1x_1 + w_2x_2 + w_3x_3$$
$\hookrightarrow x_1 x_2$

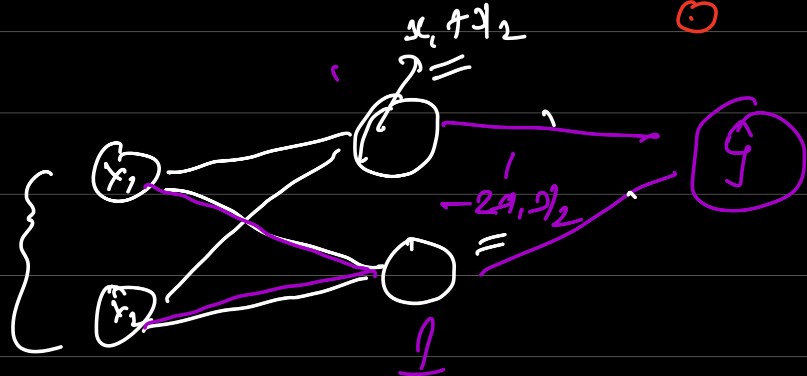
non-linear Combination of input features

$$\hat{y} = \underbrace{x_1 + x_2}_{u_1} \overbrace{- 2x_1 x_2}^{} $$
u_2

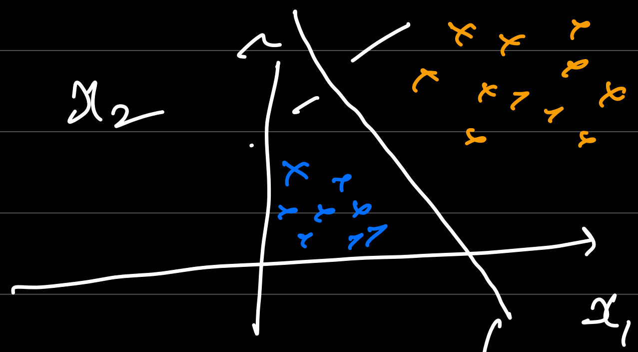
u_1 u_2

x_1	x_2	x_1+x_2	$-2x_1x_2$	y
0	0	0	0	0
0	1	1	0	1
1	0	1	0	1
1	1	2	-2	0





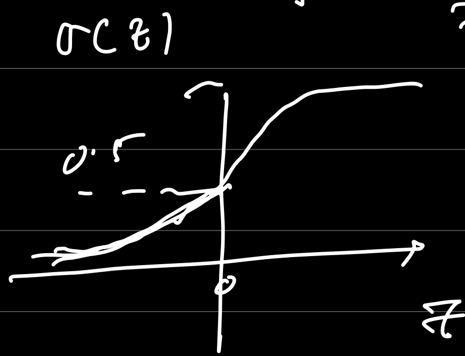
Intermediate Computation
a 'hidden layer'



$z = 0$

$z = w_0 + w_1 x_1 + w_2 x_2$

$\hat{y} = \sigma(z) \sim [0 ; 1)$



$z < 0$

$\Rightarrow \sigma(z) < 0.5$

~ 0

$z > 0$

$\Rightarrow \sigma(z) > 0.5$

$\hookrightarrow 1$

Logic gates

OR GATE

x_1	x_2	Y	$\hat{y}$	z
0	0	0	< 0.5	< 0
0	1	1	> 0.5	> 0
1	0	1	> 0.5	> 0
1	1	1	> 0.5	> 0

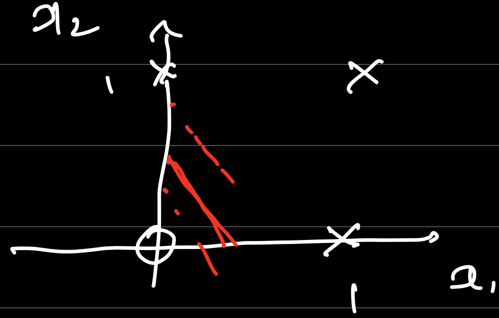
$$z = w_0 + w_1 x_1 + w_2 x_2$$

$x_1 = 0, \quad x_2 = 0$ $\qquad \sigma(z)$

$z = w_0 \quad , \quad -1$

$x_1 = 0 \qquad z = w_0 + w_1 = -1 + w_1$

$\qquad z = w_0 + w_2 = -1 + w_2$



$\sigma(z)$

AND GATE

x_1	x_2	y	$\hat{y}$	z
0	0	0	< 0.5	< 0
0	1	0	< 0.5	< 0
1	0	0	< 0.5	< 0
1	1	1	≥ 0.5	> 0

$$z = w_0 + w_1 x_1 + w_2 x_2$$

$w_0 = -1.5, \quad w_1 = 1, \quad w_2 = 1$

$w_0 = -3, \quad w_1 = 2, \quad w_2 = 2$

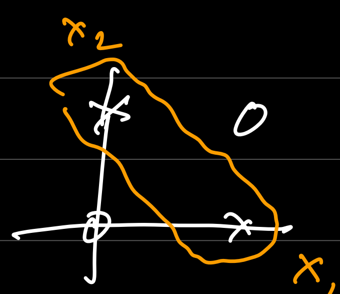
NOR - GATE

x_1	x_2	Y	$\hat{y}$	z
0	0	1	≥ 0.5	≥ 0
0	1	0	< 0.5	< 0
1	0	0	< 0.5	< 0
1	1	0	< 0.5	< 0

$z = \quad w_0 + w_1 x_1 + w_2 x_2$

$\quad 0.5 \quad 1 \quad -2 \quad -2$

XOR GATE



x_1	x_2	Y
0	0	0
0	1	1
1	0	1
1	1	0

$$XOR(x_1, x_2) \rightarrow [x_1 + x_2] - (2x_1 x_2)$$

function

$$= NOR(\underbrace{NOR(x_1, x_2)}_{u_1}, \underbrace{AND(x_1, x_2)}_{u_2})$$

$$= NOR(u_1, u_2)$$

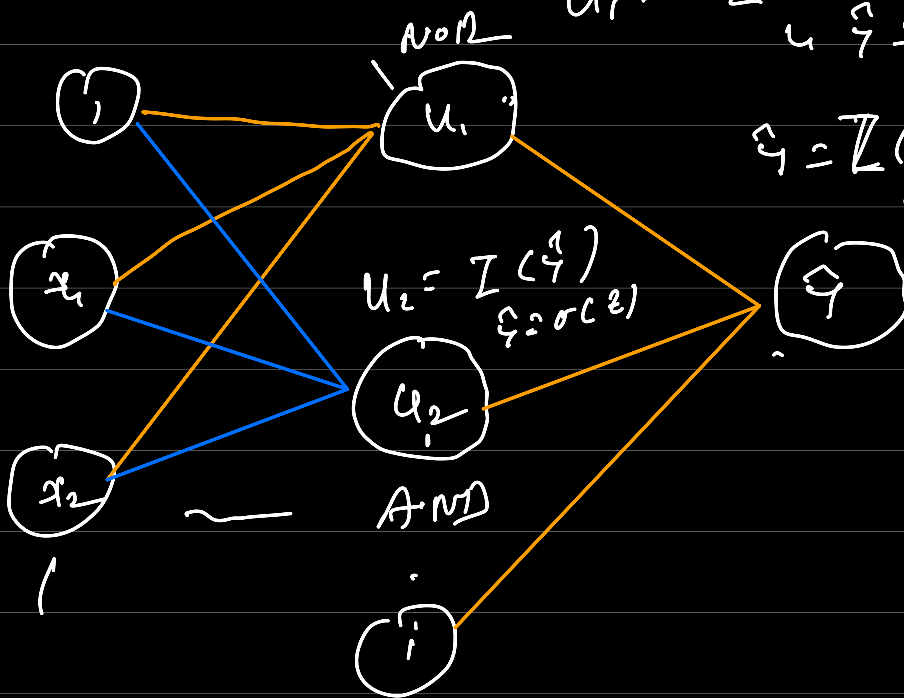
↳ Linearly separable

$u_1 = \mathbb{I}(\hat{y})$

↳ $\hat{y} = \sigma(z)$

$\hat{y} = \mathbb{I}(\sigma(z))$

$z = \sum wx$



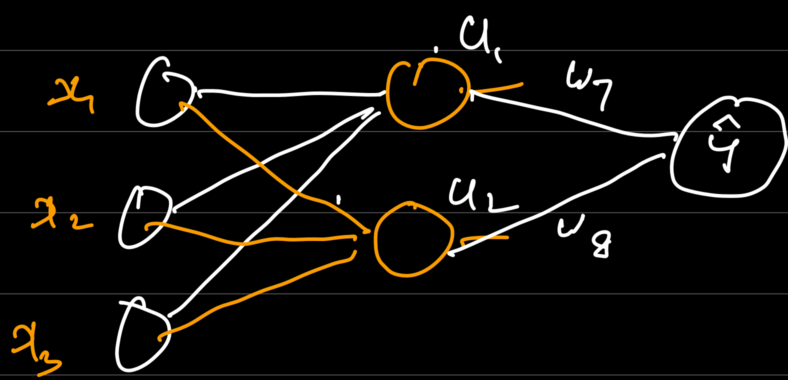
$u_2 = \mathbb{I}(\hat{y})$

$\hat{y} = \sigma(z)$

↳ Intermediate layer of Computation

Pitts - McCullough Neuron

earliest version of a neuron



$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} w_1 & w_4 \\ w_2 & w_5 \\ w_3 & w_6 \end{bmatrix} = [u_1 \; u_2$$

1×3 $\qquad$ 3×2

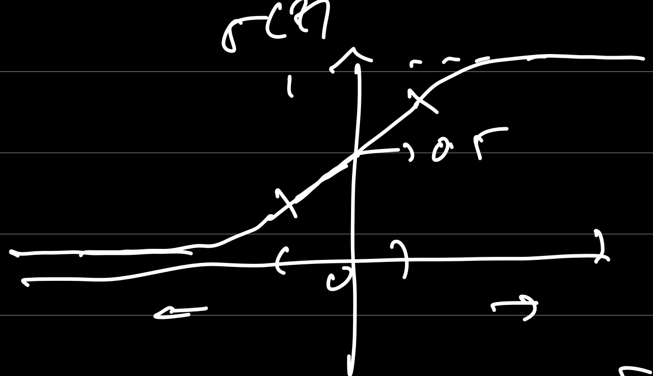
$$\hat{y} = w_7 u_1 + w_8 u_2$$

$$= \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} w_7 \\ w_8 \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} w_1 & w_4 \\ w_2 & w_5 \\ w_3 & w_6 \end{bmatrix} \begin{bmatrix} w_7 \\ w_8 \end{bmatrix}$$

1×3 $\qquad$ 3×2 $\qquad$ 2×1

$$\sigma(z) = \left[\frac{1}{1+e^{-z}}\right]$$



d

$$\left(\frac{d}{dz}\sigma(z)\right)$$

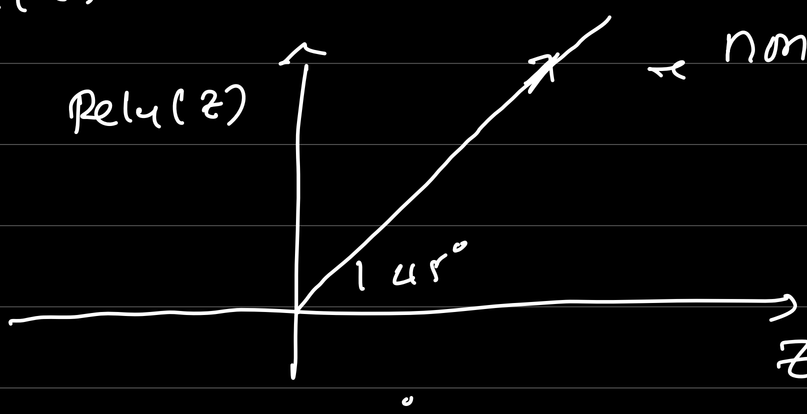
$\downarrow$

$z \to \infty$

$\sigma(z) \to [0, 1]$

Relu(z) → Rectified linear unit

$$\frac{d}{dz} \text{Relu}(z) = 1$$



← non-linearity

$$\text{Relu}(z) = \begin{cases} z, & z \geq 0 \\ 0 & z \leq 0 \end{cases}$$

$$\mathrm{Relu}(z) = \max(0, z)$$