# AGENDA

i> Multilayer Perceptron

ii> Backpropagation intuition

iii> ANN {forward propagation information flow}

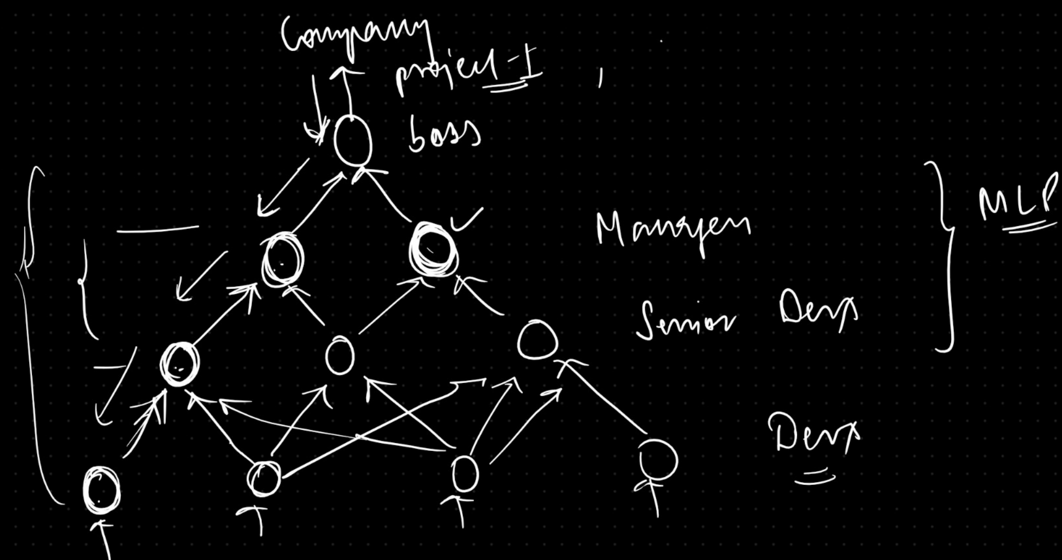
iv> Why we need activation function ? {Proof}

ANTS

One neuron

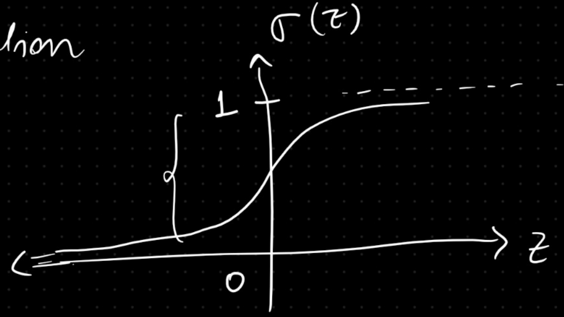
Client

error = 0

Sigmoid activation function

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




domain z value

$(-\infty, \infty)$

i/p range

$[1, 2]$ $(1, 2)$

$[1, 2)$

Range

$(0, 1)$

$z \to \infty \Rightarrow 1$

$z \to -\infty \to 0$

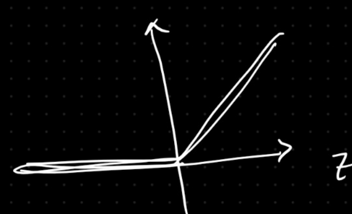
$\infty \to 1$

$-\infty \to 0$

squashing $(0, 1)$

ReLU ⇒ Rectified Linear Unit

$$\text{ReLU}(z) = \begin{cases} 0 & z < 0 \\ z & z \geq 0 \end{cases}$$




domain

$(-\infty, \infty)$

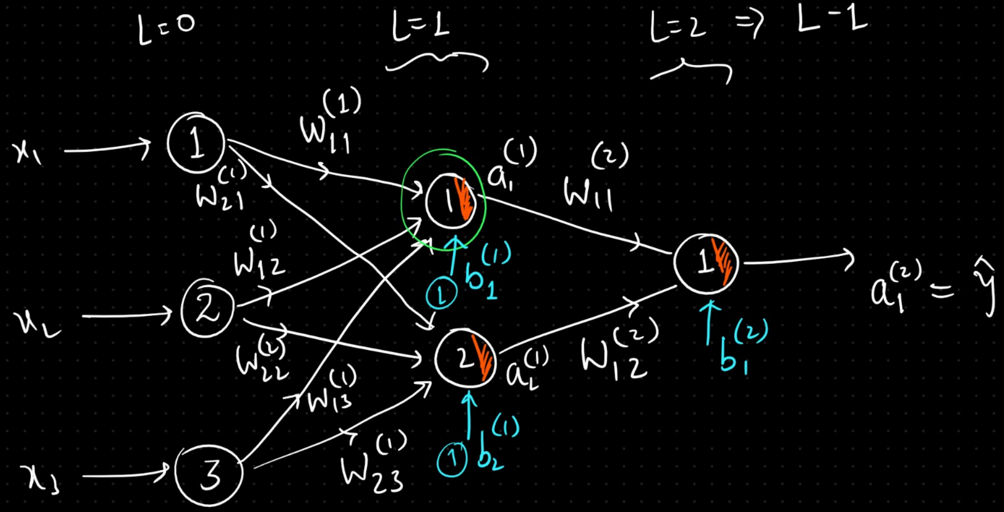
Range

$[0, \infty)$

$\infty \to \infty$

$-\infty \to 0$

$W_{21}^{(N)}$ → Nth layer; 2 → destination; 1 → source




at layer 1.

① $\begin{cases} z_1^{(1)} = W_{11}^{(1)} \cdot x_1 + W_{12}^{(1)} x_2 + W_{13}^{(1)} x_3 + b_1^{(1)} \\ a_1^{(1)} = \sigma(z_1^{(1)}) \end{cases}$

② $\begin{cases} z_2^{(2)} = W_{21}^{(1)} \cdot x_1 + W_{22}^{(1)} x_2 + W_{23}^{(1)} x_3 + b_2^{(1)} \\ a_2^{(1)} = \sigma(z_2^{(1)}) \end{cases}$

final layer {o/p layer}

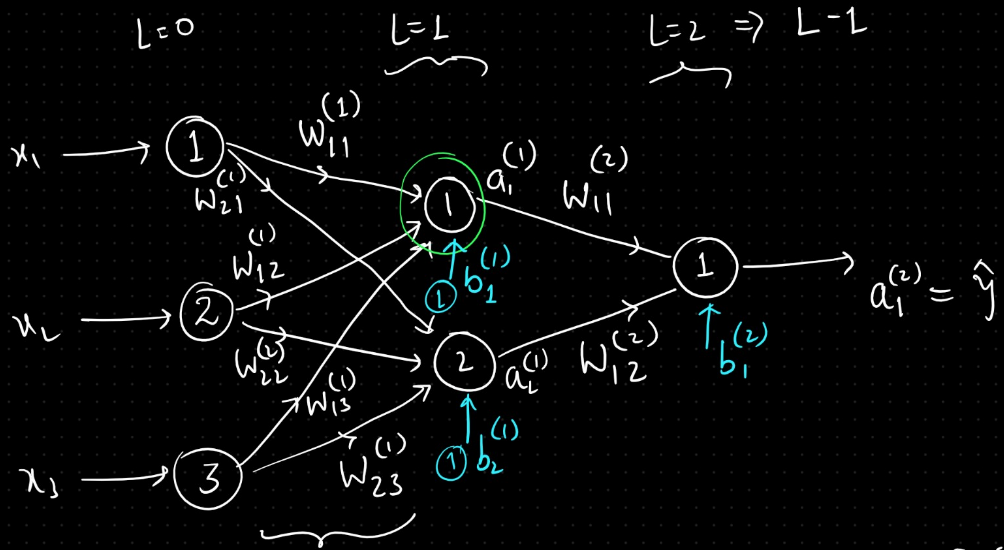
$$z_1^{(2)} = W_{11}^{(2)} a_1^{(1)} + W_{12}^{(2)} a_2^{(1)} + b_1^{(2)}$$

$$a_1^{(2)} = \sigma(z_1^{(2)}) \longrightarrow \hat{y}$$

Weight update rule (General)

$$\begin{cases} W = W + \Delta W \\ \Delta W = -\eta \dfrac{\partial e}{\partial W} \end{cases} \qquad \begin{rcases} b = b + \Delta b \\ \Delta b = -\eta \dfrac{\partial e}{\partial b} \end{rcases}$$

$\frac{\partial y}{\partial x}$ → Change in y w.r.t. small chane in x




at layer 1

$$\begin{bmatrix} W_{11}^{(1)} & W_{12}^{(1)} & W_{13}^{(1)} \\ W_{21}^{(1)} & W_{22}^{(1)} & W_{23}^{(1)} \end{bmatrix}_{2\times3} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{3\times1} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \end{bmatrix}_{2\times1} = \begin{bmatrix} Z_1^{(1)} \\ Z_2^{(1)} \end{bmatrix}_{2\times1} \xrightarrow{\text{act}} \begin{bmatrix} \sigma(z_1) \\ \sigma(z_2) \end{bmatrix} \to \begin{bmatrix} a_1^{(1)} \\ a_2^{(1)} \end{bmatrix}$$

at final layer

$$\begin{bmatrix} W_{11}^{(2)} & W_{12}^{(2)} \end{bmatrix}_{2\times2} \begin{bmatrix} a_1^{(1)} \\ a_2^{(1)} \end{bmatrix}_{2\times1} + \begin{bmatrix} b_1^{(2)} \end{bmatrix}_{1\times1} = \begin{bmatrix} Z_1^{(2)} \end{bmatrix}_{1\times1} \xrightarrow{\text{act}} \begin{bmatrix} \sigma(z_1) \end{bmatrix} \to \begin{bmatrix} a_1^{(2)} \end{bmatrix} \to \hat{y}$$

$$W = W - \eta \cancel{\frac{\partial e}{\partial W}}^{\,0} \qquad b = b - \eta \cancel{\frac{\partial e}{\partial b}}^{\,0}$$

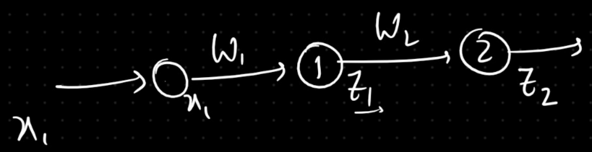
$W = W \Rightarrow$ no weight update

$b = b \Rightarrow$ no bias update

$\nabla \Rightarrow$

Why we need activation function?

$\sigma$ X

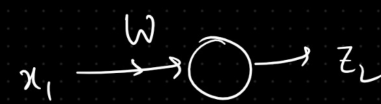



$z_2 = \hat{y}$

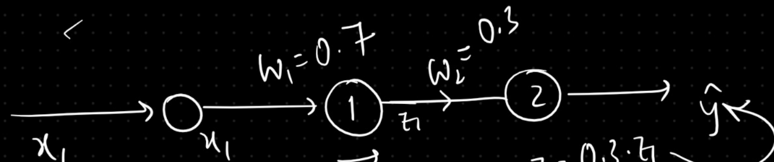
$$z_1 = W_1 x_1 \quad (1) \quad , \quad z_2 = W_2 \cdot z_1$$

$$z_2 = W_2 W_1 x_1$$

$$z_2 = W x_1$$

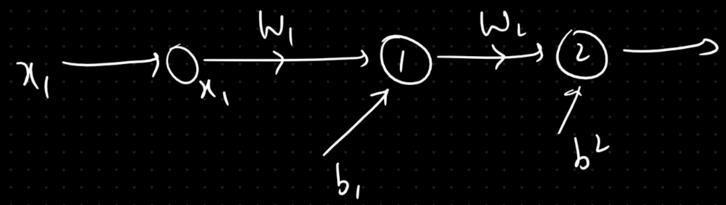



No activation



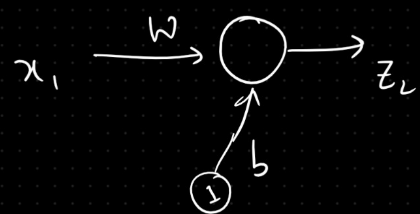

$z_1 = 0.7 \cdot x_1$ $\quad z_2 = 0.3 \cdot z_1$

$$z_2 = 0.3 \cdot z_1$$
$$= 0.3 \times 0.7 \times x_1$$
$$= 0.21 \times x_1$$
$$z_2 = W x_1$$

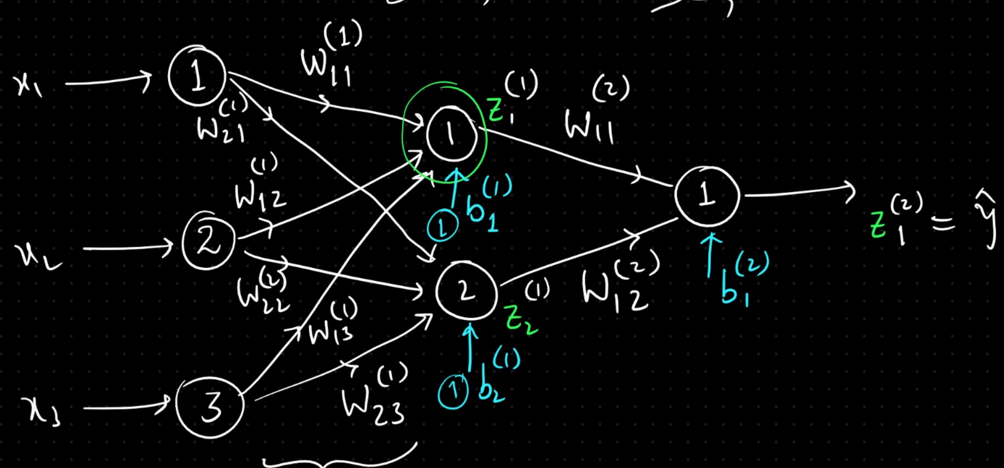



$z_1 = w_1 x_1 + b_1$ $\qquad$ $z_2 = \underbrace{w_2 z_1 + b_2}$

$$z_2 = w_2(w_1 x_1 + b_1) + b_2$$
$$= \underbrace{w_2 \cdot w_1}_{\Downarrow} x_1 + \underbrace{w_2 b_1 + b_2}$$
$$\underbrace{z_2 = w \cdot x_1 + b}$$




$L=0$ $\qquad$ $\underbrace{L=1}$ $\qquad$ $\underbrace{L=2} \Rightarrow L-1$



at layer 1.

$$① \left\{ z_1^{(1)} = W_{11}^{(1)} \cdot x_1 + W_{12}^{(1)} x_2 + W_{13}^{(1)} x_3 + b_1^{(1)} \right.$$

$$② \left\{ z_2^{(2)} = W_{21}^{(1)} \cdot x_1 + W_{22}^{(1)} x_2 + W_{23}^{(1)} x_3 + b_2^{(1)} \right.$$

final layer {o/p layer}

$$z_1^{(2)} = W_{11}^{(2)} z_1^{(1)} + W_{12}^{(2)} z_2^{(1)} + b_1^{(2)}$$

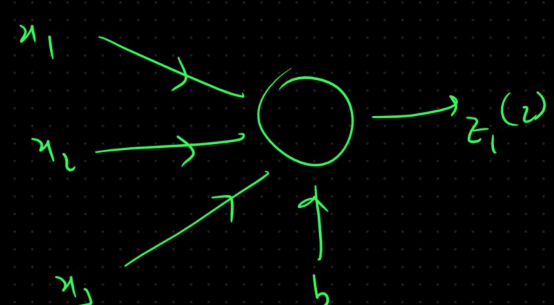
$$z_1^{(2)} \longrightarrow \hat{y}$$

$$= W_{11}^{(2)} \left\{ W_{11}^{(1)} \cdot x_1 + W_{12}^{(1)} \cdot x_2 + W_{13}^{(1)} \cdot x_3 + b_1^{(1)} \right\}$$
$$+ W_{12}^{(2)} \left\{ W_{21}^{(1)} \cdot x_1 + W_{22}^{(1)} x_2 + W_{23}^{(1)} \cdot x_3 + b_2^{(1)} \right\}$$
$$+ b_1^{(2)}$$

$$= Ax_1 + Bx_2 + Cx_3 + D_1$$
$$+ Px_1 + Qx_2 + Rx_3 + D_2$$
$$+ b_1^{(2)}$$

$$= \underbrace{(A+P)}x_1 + \underbrace{(B+Q)}x_2 + \underbrace{(C+R)}x_3 + \underbrace{D_1 + D_2 + b_1^{(2)}}$$

$$z_1^{(2)} = W_1 x_1 + W_2 x_2 + W_3 x_3 + b$$



1. Beginner → Python Machine Learning (ML+DL)
   → Sebastian Raschka
2. Advance → Hands on Machine Learning (ML+DL)
   → Aurelien Geron

CNN → LeNet, AlexNet

CUSTOM → Benchmark Datasets: CIFAR 10, Imagenet

$z_1 = \uparrow\uparrow \rightarrow (0,1)$  $z_1 = \uparrow\uparrow\uparrow \rightarrow y$

$z_2 = \uparrow\uparrow\uparrow$

—O— 97%

—O— 03%

—O— 00%