Deep Learning.

AI - 1960

ML → DL → Human brain

↓

Biologist

↓

Neuron, dendrites ← Computer Scientist

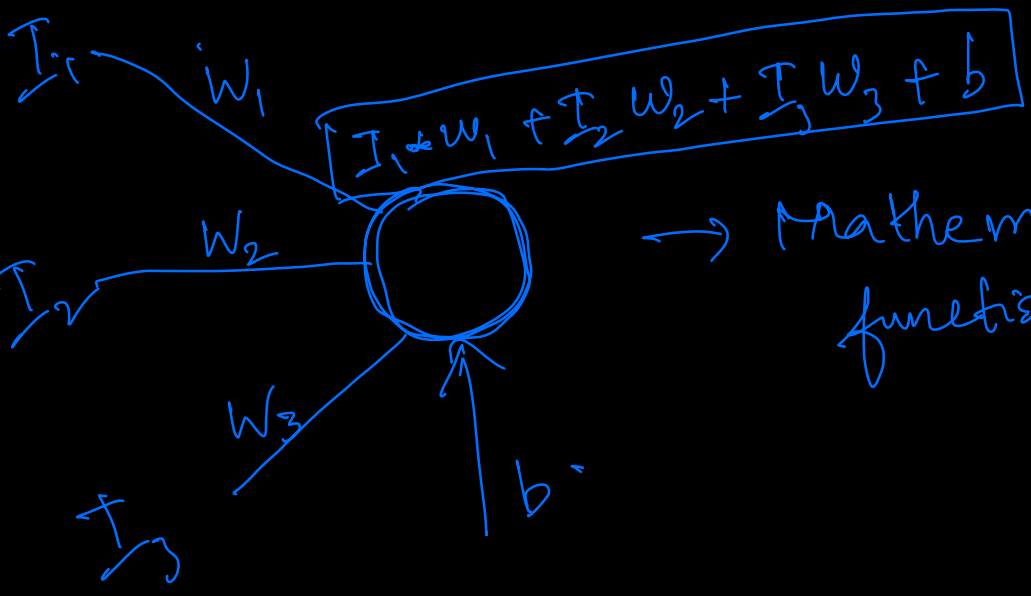
Computer Scientist (Algorithm) + Biologist (Human Brain)

↑

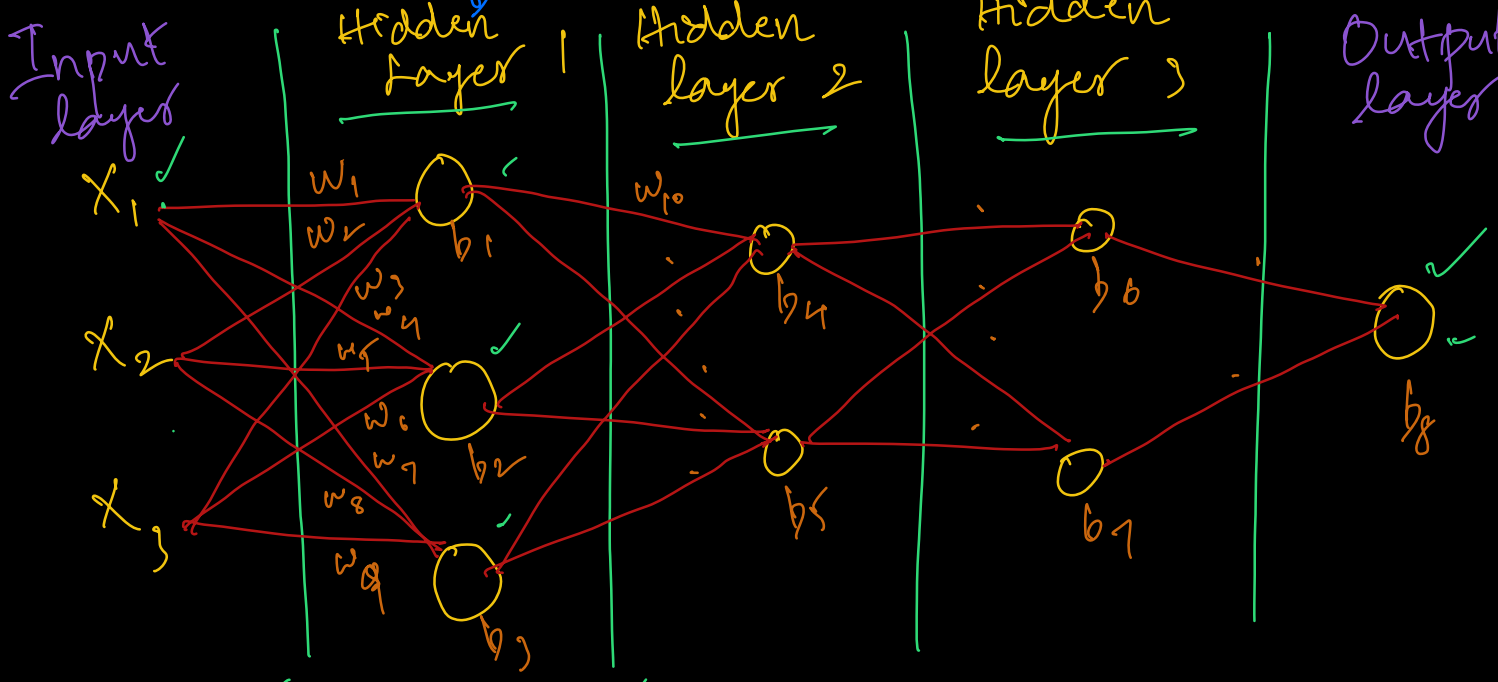
Neural Network —
(Algorithm with Brain functionality)

Issue ↓ (Vanishing and exploding gradient descent)

Neural Network → basic building block → Neuron



I → Input
W → weights
b → bias



X_1	X_2	X_3	Y
1	2	6	10
2	5	10	20

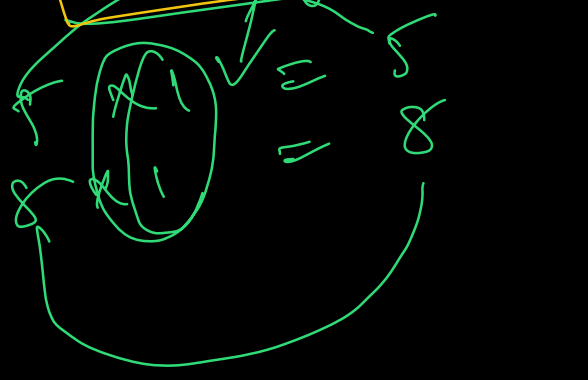
→ Sample Data

Regression → Output layer

one neuron

Activation function

Identity function



Classification

Target → Y or no → Binary

→ 0, 1, 2 → Multiclass

→ 0, 1, 2 ... 9

Y or N ⇒ 2 → 2 neurons

buy or not buy ⇒ 2 → 2 neurons

Y, N, maybe ⇒ 3 → 3 neuron

0, 1, 2, 3, 4 ⇒ 5 → 5 neuron

0, 1, 2, 3, 4, ... 9 ⇒ 10 → 10 neuron

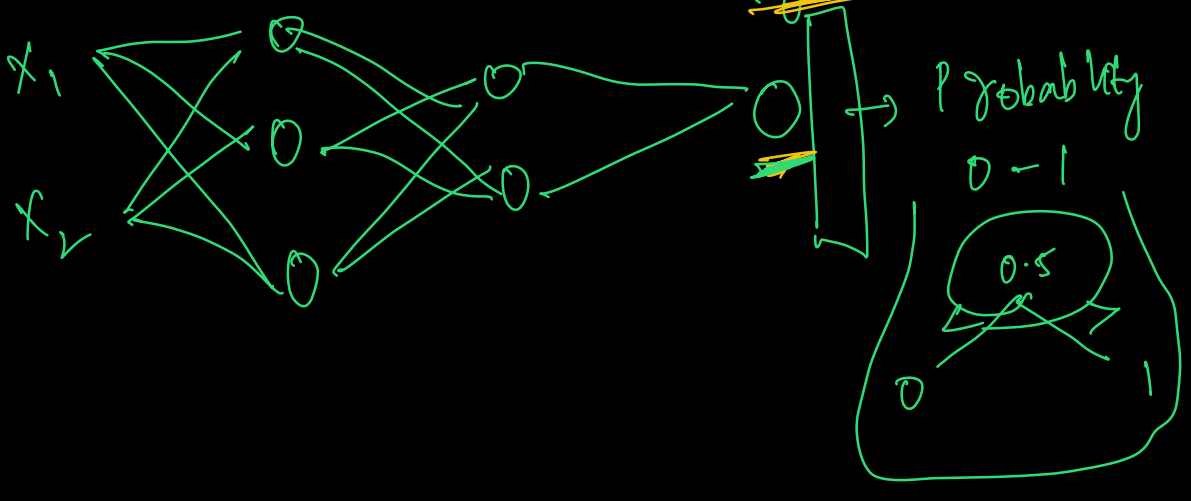
Activation function for classification

Softmax → multi class or binary

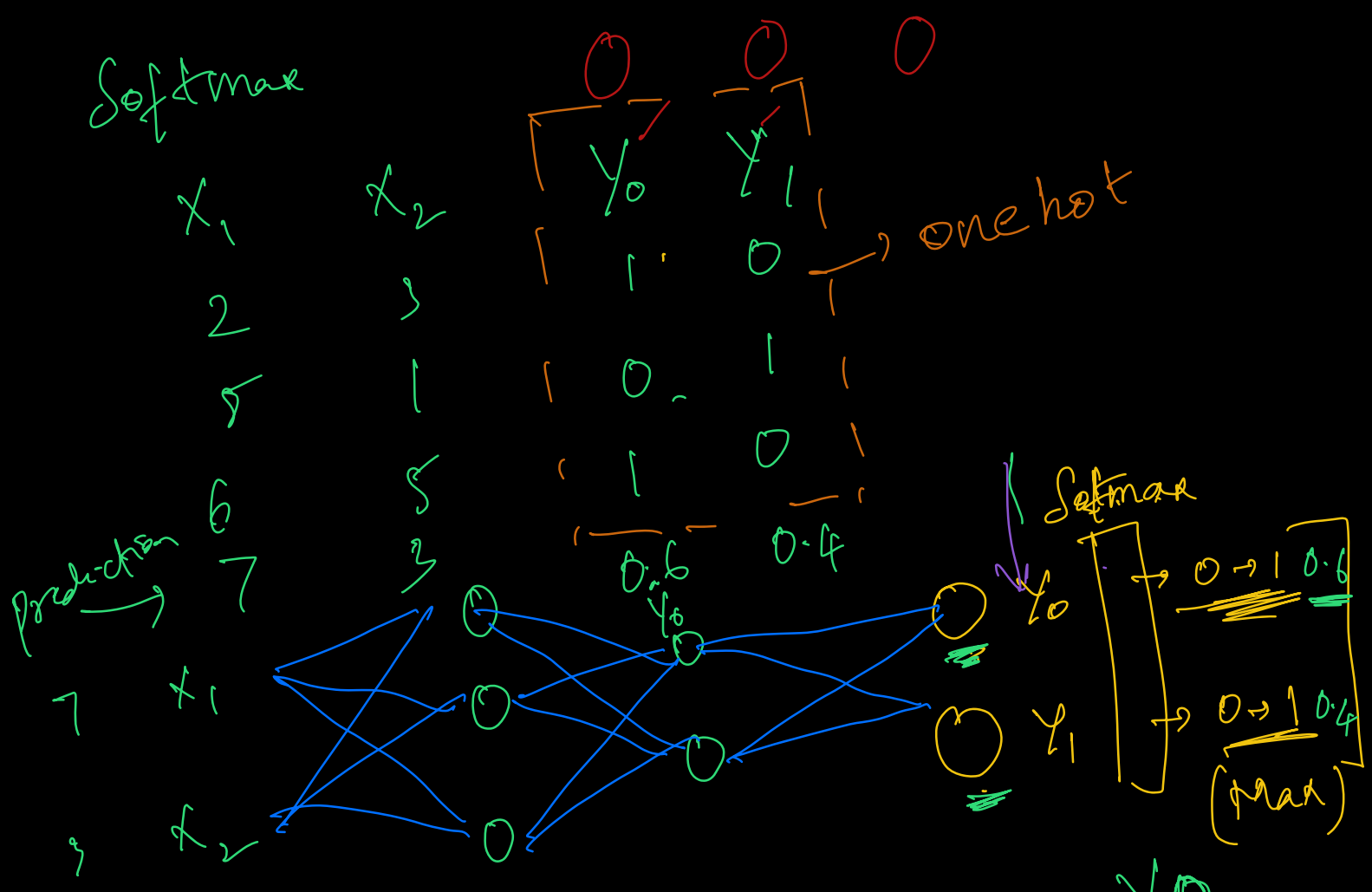
If Sigmoid → one neuron → probability → 0 or 1 → Binary classification with one neuron at output layer

x_1	x_2	Y (target Binary)
2	3	0
4	1	1
6	5	0

Sigmoid / Softmax



Softmax



Binary → 2 neurons output layer | Softmax ✗

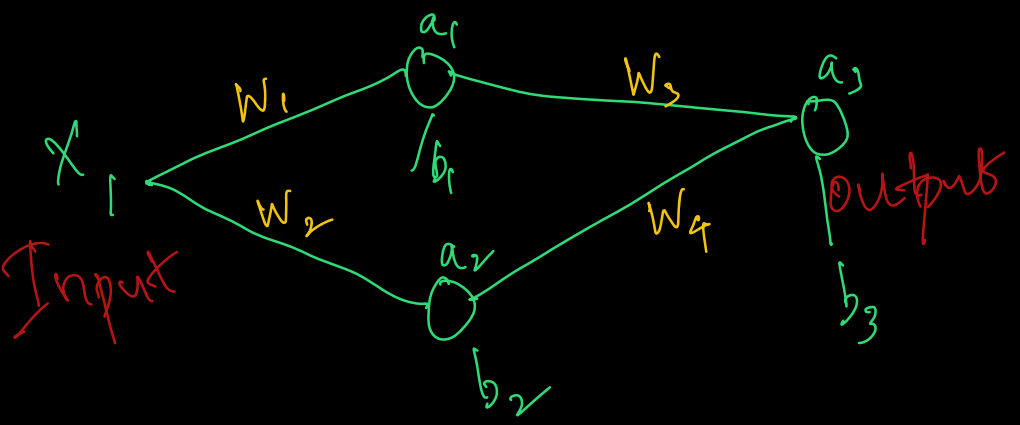
1 neuron output layer | Sigmoid ←

Multi Class → n neurons output layer | Softmax

n → unique number of classes in Target data

No Sigmoid

X → input
y → output → Regression
1 → Hidden layer
2 → neurons in hidden layer



$$a_1 = X_1 W_1 + b_1$$

Chain rule

$$a_2 = X_1 W_2 + b_2$$

$$a_3 = a_1 * W_3 + a_2 * W_4 + b_3$$

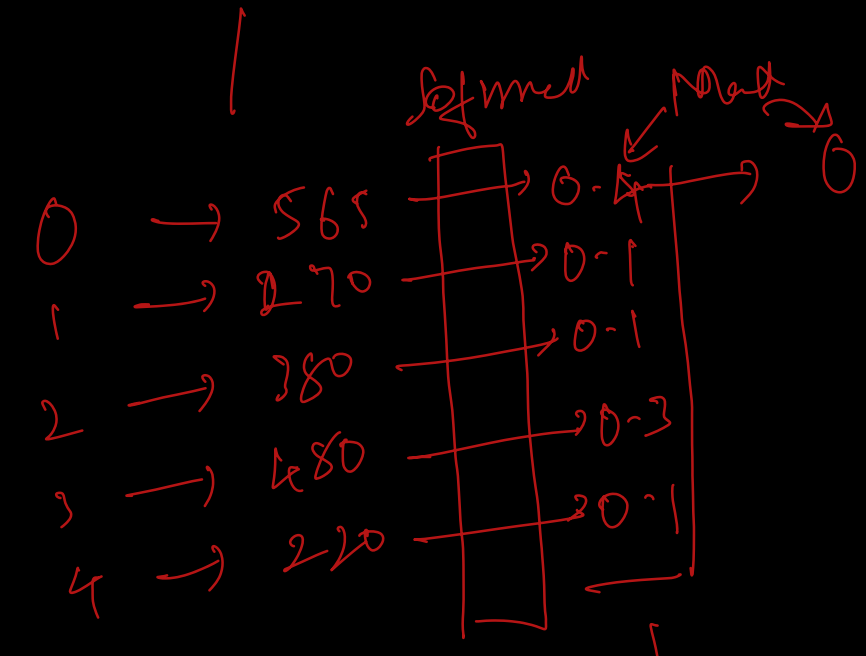
$$a_3 = (X_1 W_1 + b_1) * W_3 + (X_1 W_2 + b_2) * W_4 + b_3$$

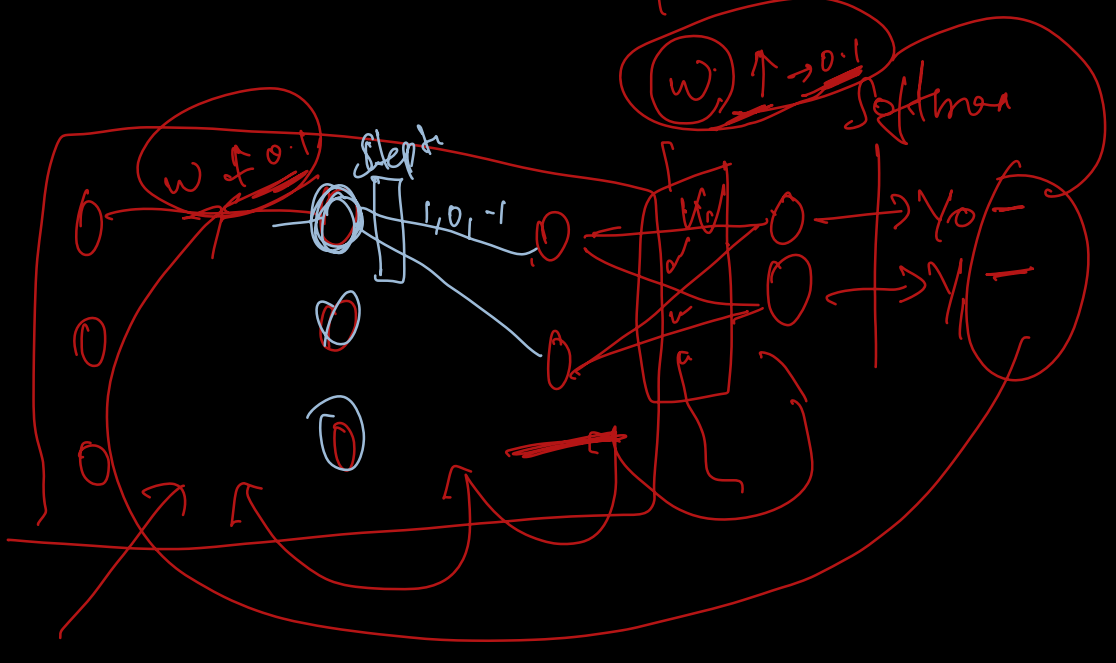
W_1, W_2, W_3, W_4
b_1, b_2, b_3

7 parameters | DL

LR $X_1 m + b$ | 2 parameters | ML

Activation Function





Activation function

Non-Linearity

Not Mandatory

Complex problem → Activation function is needed.

many kinds of Activation

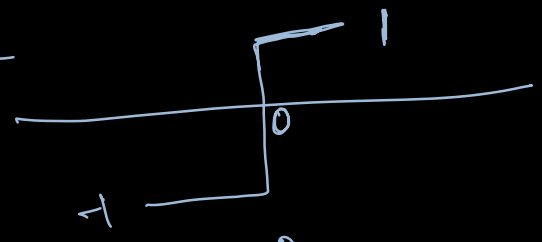
Yes

Vanishing and Exploding Gradient descent

→ One such Activation function mitigated

Significance of Activation function has become a core

1. Step function ✓
2. tanh function
3. Sigmoid function
4. Relu
5. elu

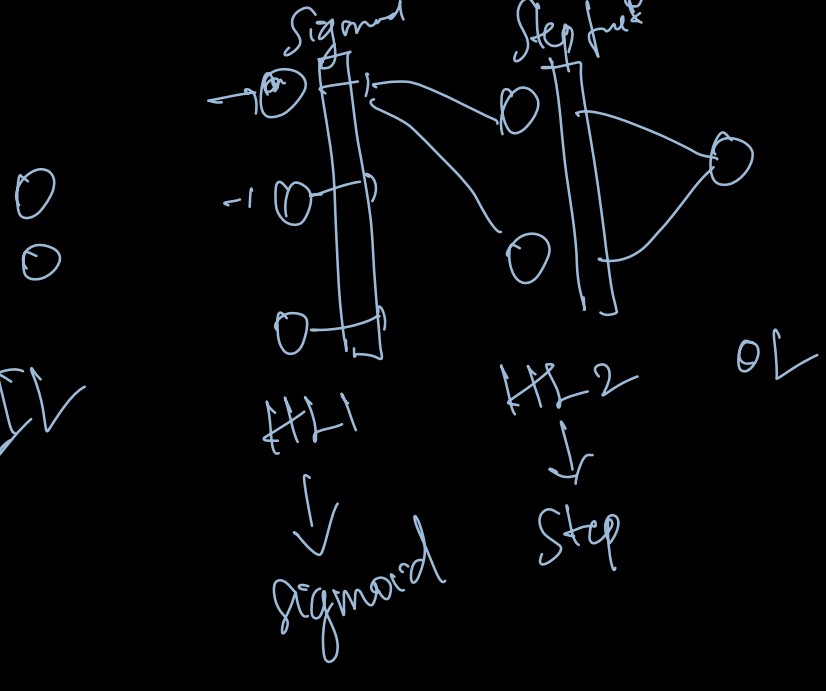


$$\phi(z) = \begin{cases} -1 & z < 0 \\ 0 & z = 0 \\ 1 & z > 0 \end{cases}$$

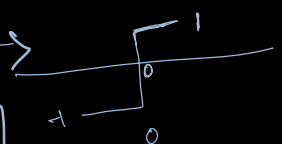
Activation function usage

applied every neuron

on each layer basis activation is applied



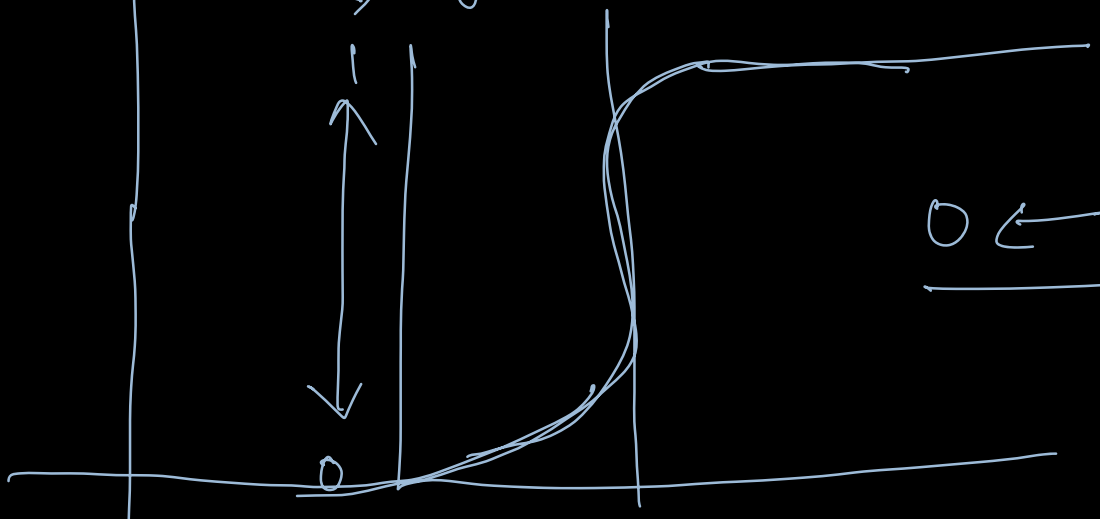
1. Step function
2. Sigmoid function
3. tanh function
4. Relu
5. elu
Selu



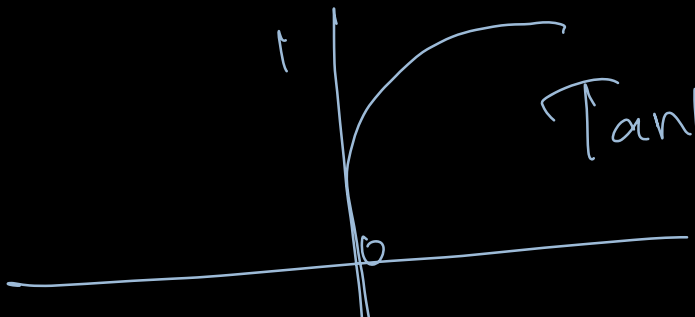
unit step function

$$\phi(z) = \begin{cases} -1 & z<0 \\ 0 & z=0 \\ 1 & z>0 \end{cases}$$

Sigmoid function



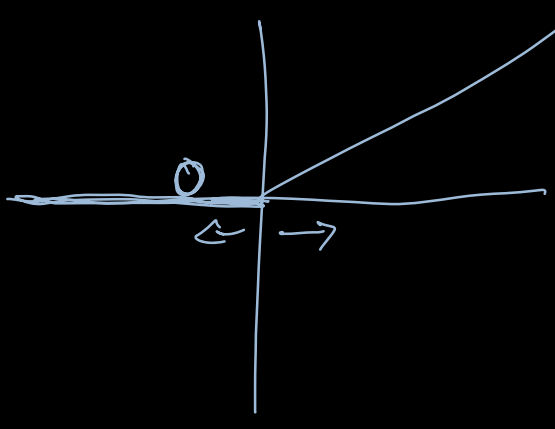
0 ⟷ 1



Tanh

-1 ⟷ +1

Rectified → linear unit

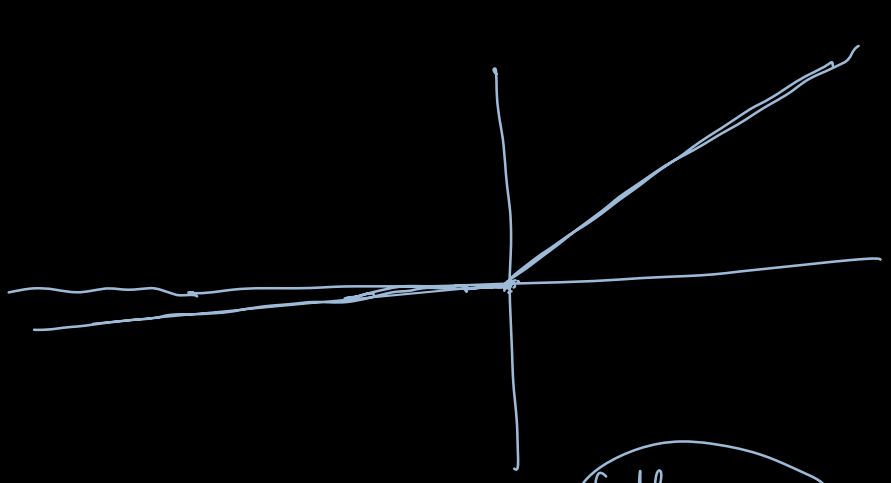


Relu

0

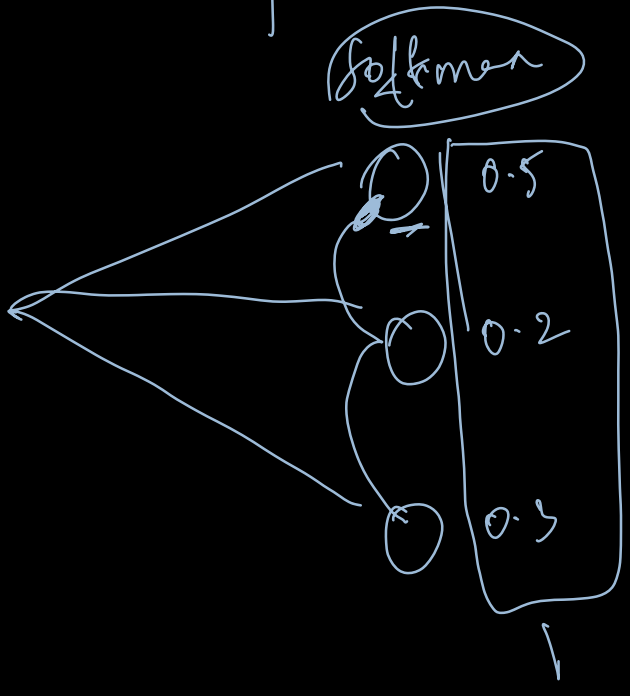
0 ⟷ ∞

elu → exponential Linear Unit

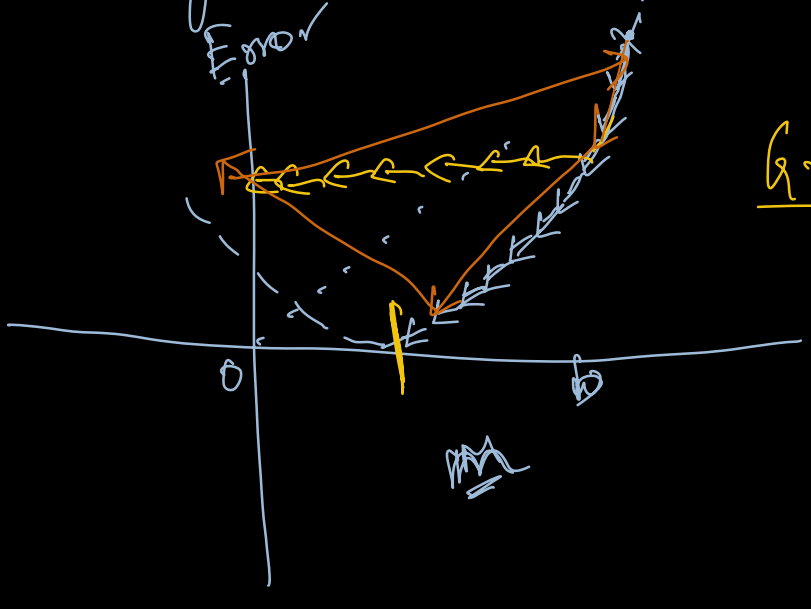


<100

(-2.6)



Vanishing and Exploding Gradient descent



Gradient = 0

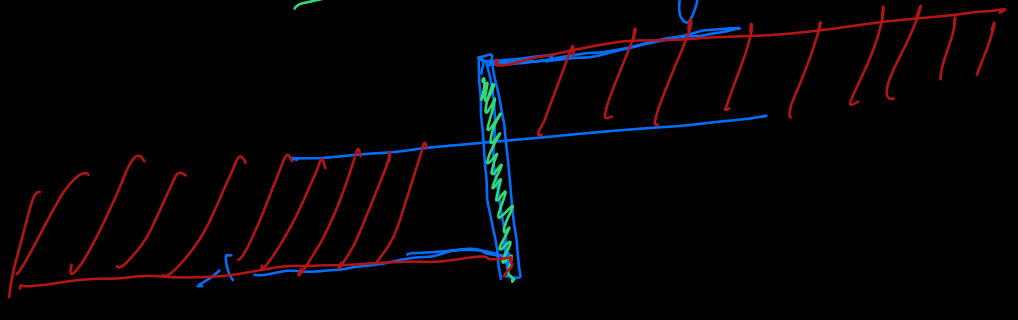
Iteration	w_1	b_1	Vanishing error	exploding
1	0	0	100	100
2	0.1	0.2	80	80
3	0.11	0.21	75	75
4	0.11	0.13	0.21	0.23	75	85
5	0.11	0.10	0.21	0.20	75	70
6	0.11	0.14	0.21	0.18	75	80

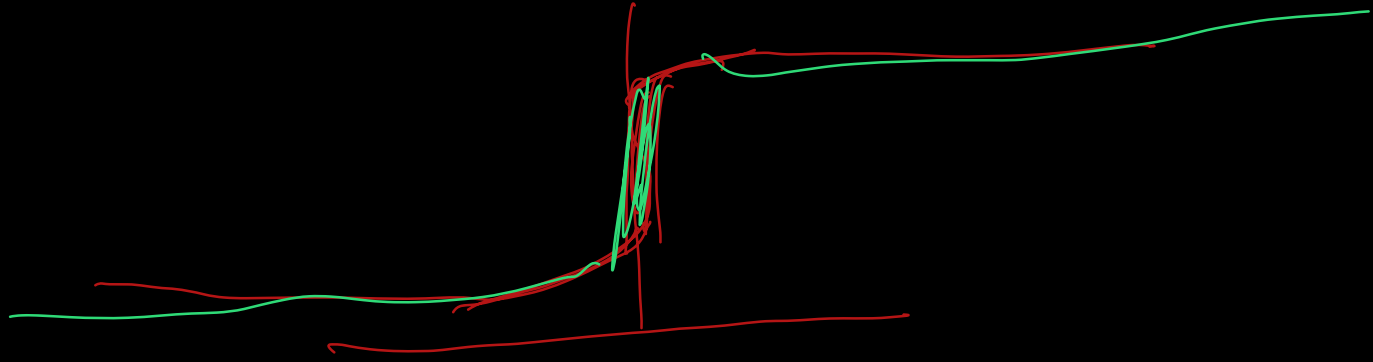
Reason for Vanishing & exploding

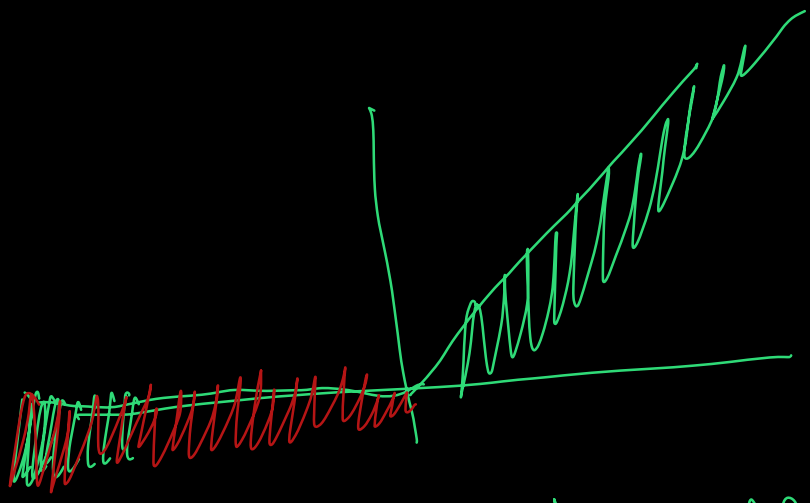
1. Explosion of parametes
2. Dependency on parameters
3. Not choosing the right Activation function

Mitigate the Vanishing and exploding gradient dosent

1. Choose the Activation function Non-Saturating



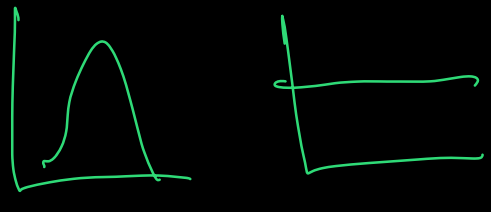




2. Proper Initialization Weights

Normally distributed (mean = 0)

Uniform distribution (−x to +x)



3. Batch Normalization