WHAT YOU WILL STUDY IN TODAY VIDEO ?

- What is MLP?
- Design of a MLP
- Mathematical Working of MLP
- Drawbacks of MLP

Prerequisites
1. Differentiation
2. Gradient Descent
3. Perceptron

Introduction

① MLP ⟶ Multi Layer Perceptron

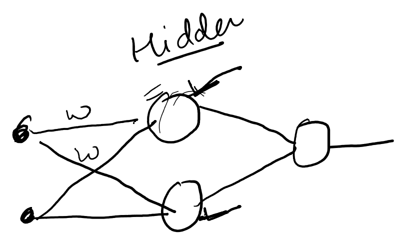
② ANN model
- ⟶ Classification ✓
- ⟶ Regression ✓

③ Why needed?
- ⟶ Solve non linear problems (real / complex)
- ⟶ MLP → foundation block modern neural networks (CNN, RNN, LSTM)

Perceptron $\xrightarrow{\text{foundation block}}$ MLP (ANN) $\xrightarrow{\text{foundation block}}$ CNN, RNN etc.

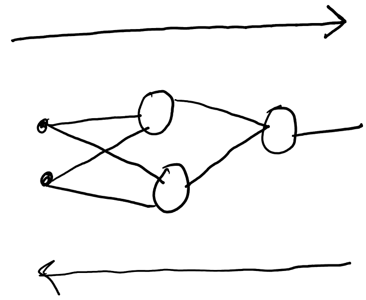
④ Structure.

Input layer ⟶ Hidden layers (1 or more) ⟶ O/P layer



* Each layer has neurons (perceptrons)

* Connected
* Each connection has weights & biases
* Each neuron has activation fn



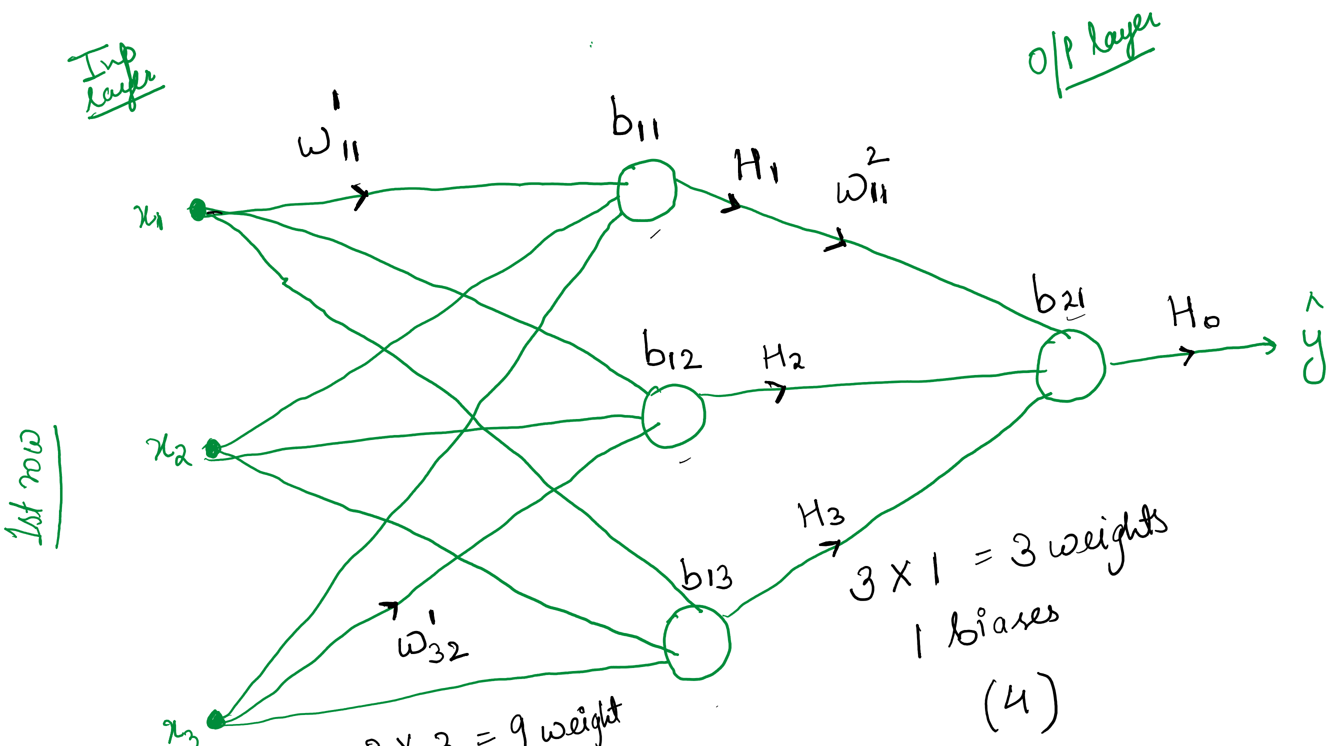
* { Forward Propagation
  Loss
  Back Propagation
  Weights update → Optimizer }

Design

Data:

	x_1	x_2	x_3	y
1st row	—	—	—	—
2nd row	—	—	—	—
3rd row	—	—	—	—

Model



$3 \times 3 = 9$ weight
3 biases
(12)

$3 \times 1 = 3$ weights
1 biases
(4)

① Trainable Params $\longrightarrow$ (weights & biases)

$= 16$ params

② ij ⟶ i ⟶ layer no
j ⟶ node no.

(layer notation)

③ weights notation ⟶ w_{ij} ⟶ layer (weight is coming)
⟶ node from which it is coming
⟶ node going

Mathematical Working

Data:

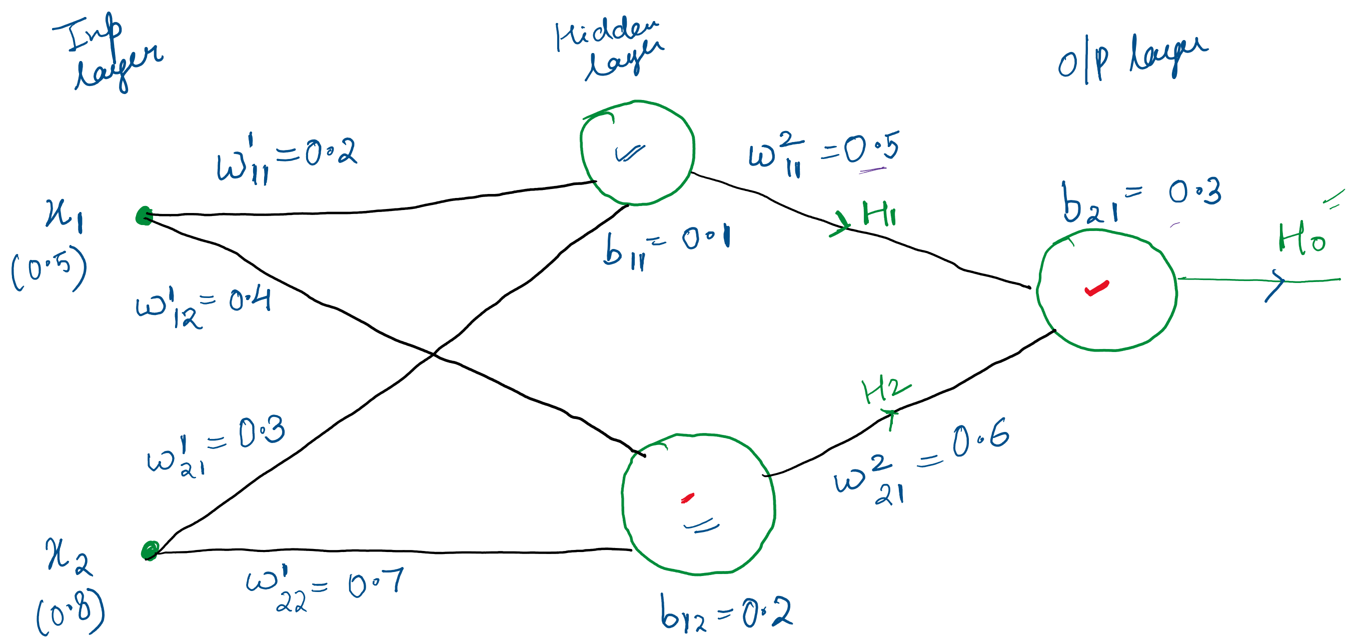
x_1	x_2	y
0.5	0.8	1

Model → Input layer → Hidden layer → O/P layer

① Initialization

* Weights & biases
* Choose a activation fn (Sigmoid)
* Choose Loss fn (Binary Cross entropy)
* Choose optimizer (Gradient Descent)
* Choose learning rate ($\alpha = 0.1$)

In next videos
- Activation fn
- Loss fn
- Optimizers



② Forward Propagation

* Hidden layer (Neuron 1) :-

$$Z_1 = \omega_1 x_1 + \omega_2 x_2 + b$$
$$= (0.2 \times 0.5) + (0.3 \times 0.8) + 0.1$$
$$= 0.44.$$

Apply activation fn

$$\sigma(0.44) = \frac{1}{1+e^{-z}}$$
$$= \frac{1}{1+e^{-0.44}}$$
$$\boxed{H_1 = 0.608} = 0.608$$

* Hidden layer (Neuron 2) =

$$Z_2 = (0.4 \times 0.5) + (0.7 \times 0.8) + 0.2$$
$$= 0.96.$$

Apply activation fn

$$= \frac{1}{1+e^{-0.96}} = 0.723$$

$$\boxed{H_2 = 0.723}$$

Output layer $\Rightarrow$

$$Z_0 = \omega_1 H_1 + \omega_2 H_2 + b$$
$$= (0.5 \times 0.608) + (0.6 \times 0.723) + 0.3$$
$$= 1.038.$$

Apply activation fn

$$= \frac{1}{1+e^{-1.038}} = 0.738$$

$$\boxed{H_0 = 0.738}$$

Forward Propagation Done

③ Loss Calculation

Recap

Actual value $\rightarrow y \rightarrow 1$

Predicted value $\rightarrow \hat{y} \rightarrow H_0$ (0.738)

$$\boxed{L = -\left[y \log \hat{y} + (1-y)\log(1-\hat{y})\right]}$$

$$L = -\left[1 \log 0.738 + (1-1)\log(1-0.738)\right]$$

$$= -\log 0.738$$

$$= 0.303$$

$$\boxed{L = 0.303}$$

4 Back Propagation

Recap.

$H_1 = 0.608$ } Hidden layer
$H_2 = 0.723$

$H_0 = O = 0.738$ } o/p layer

$L = 0.303$

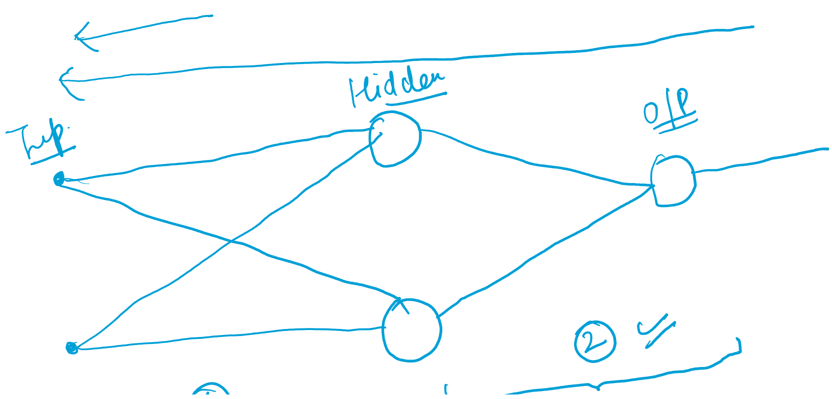
Loss fn

$$\boxed{L = -\left[y \log O + (1-y) \log (1-O) \right]}$$

Note:

1. Calculate derivates wrt Loss (all) $\frac{dy}{dx}$ $\frac{dL}{d}$

2. Three derivatives Loss
   - → Output $\frac{dL}{dH}$
   - → weighted sum(z) $\frac{dL}{dZ}$
   - → weight & biases $\frac{dL}{dw}$, $\frac{dL}{db}$



2

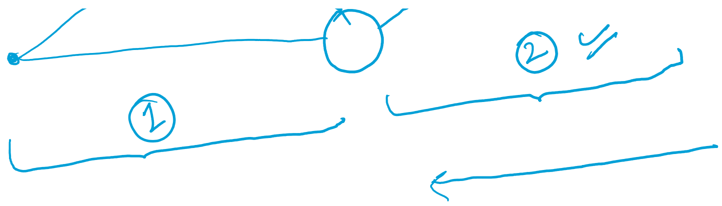

1
2

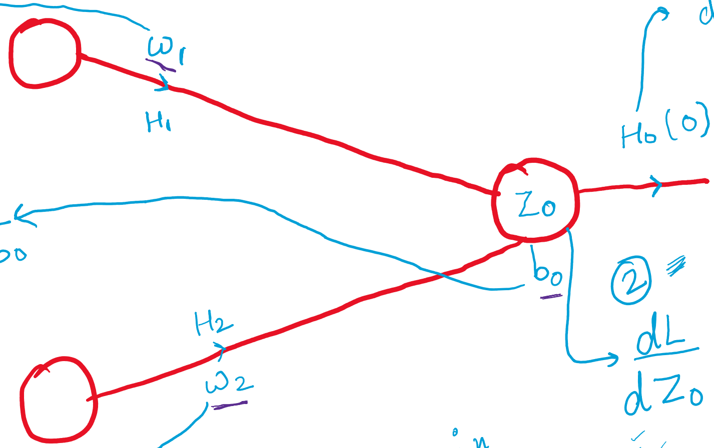
Diagram for Section 2

Hidden

O/P

① $\frac{dL}{dO}$ (calculate directly)

$$\frac{dL}{dw_1} = \frac{dL}{dZ_0} \times \frac{dZ_0}{dw_1}$$



$$\frac{dL}{db_0} = \frac{dL}{dZ_0} \times \frac{dZ_0}{db_0}$$

③

② $$\frac{dL}{dZ_0} = \frac{dL}{dO} \times \frac{dO}{dZ_0}$$

$$\frac{dL}{dw_2} = \frac{dL}{dZ_0} \times \frac{dZ_0}{dw_2}$$

Chain Rule

Why we need Gradient? → for weight updation

Mathematical (Section 2)

① Loss wrt Output

$$\frac{dL}{dO} = -\left[\frac{y}{O} - \frac{1-y}{1-O}\right]$$

$\therefore y = 1$ $\quad \frac{dL}{dO} = \frac{-1}{O} = \frac{-1}{0.738}$

$$\boxed{\frac{dL}{dO} = -0.262}$$

② Loss wrt to weighted sum (z)

$$O = \sigma(z_o)$$

$$= \frac{1}{1+e^{-z_o}}$$

Diff sigmoid fn.

$a(1-a)$

$\therefore H_o = O$

$$\frac{d\sigma}{dz_o} = \sigma(z_o)\,(1-\sigma(z_o))$$

$$\frac{dO}{dz_o} = O\,(1-O)$$

$$= 0.738\,(1-0.738)$$

$$= 0.193$$

$$\boxed{\frac{dO}{dz_o} = 0.193}$$

$$\frac{dL}{dz_o} = \frac{dL}{dO} \times \frac{dO}{dz_o}$$

$$= -0.262 \times 0.193$$

$$= -0.051$$

$$\boxed{\frac{dL}{dz_o} = -0.051}$$

$= -0.051$

③ Loss wrt weights & biases

$$Z_0 = w_1 H_1 + w_2 H_2 + b_0$$

$$\frac{dZ_0}{dw_1} = H_1 \qquad \frac{dZ_0}{dw_2} = H_2 \qquad \frac{dZ_0}{db_0} = 1$$

* $\frac{dL}{dw_1} = \frac{dL}{dZ_0} \times \frac{dZ_0}{dw_1} = -0.051 \times 0.608 = -0.031$

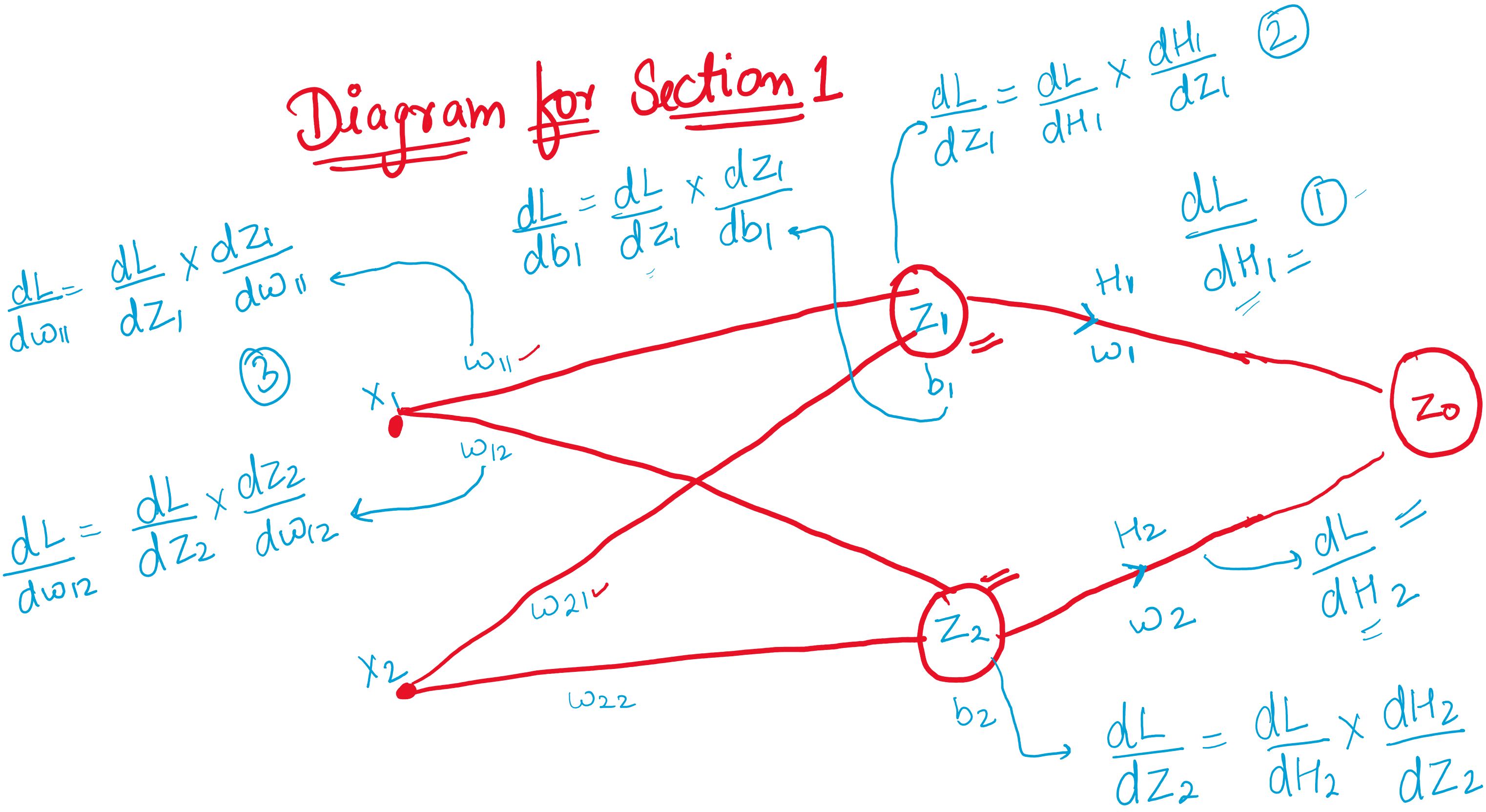
* $\frac{dL}{dw_2} = \frac{dL}{dZ_0} \times \frac{dZ_0}{dw_2} = -0.051 \times 0.723 = -0.037$

* $\frac{dL}{db_0} = \frac{dL}{dZ_0} \times \frac{dZ_0}{db_0} = -0.051 \times 1 = -0.051$

Imp for next step

Done.

Diagram for Section 1



Mathematical Section-1.

① Loss wrt Output. (H_1, H_2)

$$\frac{dL}{dH_1} = \frac{dL}{dZ_0} \times \omega_1 = -0.051 \times 0.5$$

$$\boxed{\frac{dL}{dH_1} = -0.0255}$$

$$\frac{dL}{dH_2} = \frac{dL}{dZ_0} \times \omega_2 = -0.051 \times 0.6$$

$$\boxed{\frac{dL}{dH_2} = -0.0306}$$

② Loss wrt weighted Sum(Z)

$H_1 = \sigma(Z_1)$ $\qquad$ $H_2 = \sigma(Z_2)$

Derivative of sigmoid

$$\frac{d\sigma}{dz} = \sigma(z)(1-\sigma(z))$$

$$\frac{dH_1}{dZ_1} = H_1(1-H_1)$$
$$= 0.608(1-0.608)$$

$$\frac{dH_2}{dZ_2} = H_2(1-H_2)$$
$$= 0.723(1-0.723)$$

$$\frac{dH_1}{dZ_1} = 0.238$$

$$\frac{dH_2}{dZ_2} = 0.200$$

Chain Rule.

$$\frac{dL}{dZ_1} = \frac{dL}{dH_1} \times \frac{dH_1}{dZ_1}$$
$$= -0.0255 \times 0.238$$
$$= -0.00607$$

$$\frac{dL}{dZ_2} = \frac{dL}{dH_2} \times \frac{dH_2}{dZ_2}$$
$$= -0.00612$$

Ans to this step

③ Loss wrt weights & biases

Neuron 1.

$$Z_1 = \omega_{11}X_1 + \omega_{21}X_2 + b_1$$

$$\frac{dZ_1}{d\omega_{11}} = X_1 \qquad \frac{dZ_1}{d\omega_{21}} = X_2 \qquad \frac{dZ_1}{db_1} = 1$$

$$\ldots \times dZ_1 \ldots = -0.00607 \times 0.5 = -0.00303$$ for weight ...lation

$$\frac{dL}{dw_{11}} = \frac{dL}{dz_1} \times \frac{dz_1}{dw_{11}} = -0.00607 \times 0.5 = -0.00303$$

$$\frac{dL}{dw_{21}} = \frac{dL}{dz_1} \times \frac{dz_1}{dw_{21}} = -0.00607 \times 0.8 = -0.00486$$

$$\frac{dL}{dw_{21}} = \frac{dL}{dz_1} \times \frac{dz_1}{dw_{21}} = -0.00607 \times 1 = -0.00607$$

$$\frac{dL}{db_1} = \frac{dL}{dz_1} \times \frac{dz_1}{db_1}$$

for weight updation needed (Neuron 1.)

Neuron 2

$$Z_2 = w_{12}X_1 + w_{22}X_2 + b_2$$

$$\frac{dZ_2}{dw_{12}} = X_1 \qquad \frac{dZ_2}{dw_{22}} = X_2 \qquad \frac{dZ_2}{db_2} = 1$$

$$\frac{dL}{dw_{12}} = \frac{dL}{dZ_2} \times \frac{dZ_2}{dw_{12}} = -0.00612 \times 0.5 = -0.00306$$

$$\frac{dL}{dw_{22}} = \frac{dL}{dZ_2} \times \frac{dZ_2}{dw_{22}} = -0.00490$$

$$\frac{dL}{db_2} = \frac{dL}{dZ_2} \times \frac{dZ_2}{db_2} = -0.00612$$

weight updation for Neuron 2.

Back Propagation Done. → Aim = (Get gradients for each weight & biases) → **weight updation**

5 Weight Updation

→ Optimizer = (Gradient we)

→ Learning rate

formula weight updation

$$\boxed{w_{new} = w_{old} - \alpha \times \frac{dL}{dw}}$$

α → learning rate

$\frac{dL}{dw}$ → respective gradient

Weight updation for O/P layer

$$w_{1\,new} = 0.5 - (0.1 \times -0.031)$$

$$\boxed{w_{1\,new} = 0.503}$$

$$w_{2\,new} = 0.6 - (0.1 \times -0.037)$$

$$\boxed{w_{2\,new} = 0.604}$$

$(\cdot \times -0.051)$

$$b_{0_{new}} = 0.3 - (0.1 \times -0.051)$$

$\boxed{b_{0_{new}} = \_\_\_\_ \quad \text{Ans}}$

Do same with Hidden layer (Neuro1, Neuron2.) → Do yourself.

Weight updation Done.

⑥ After Weight Updation

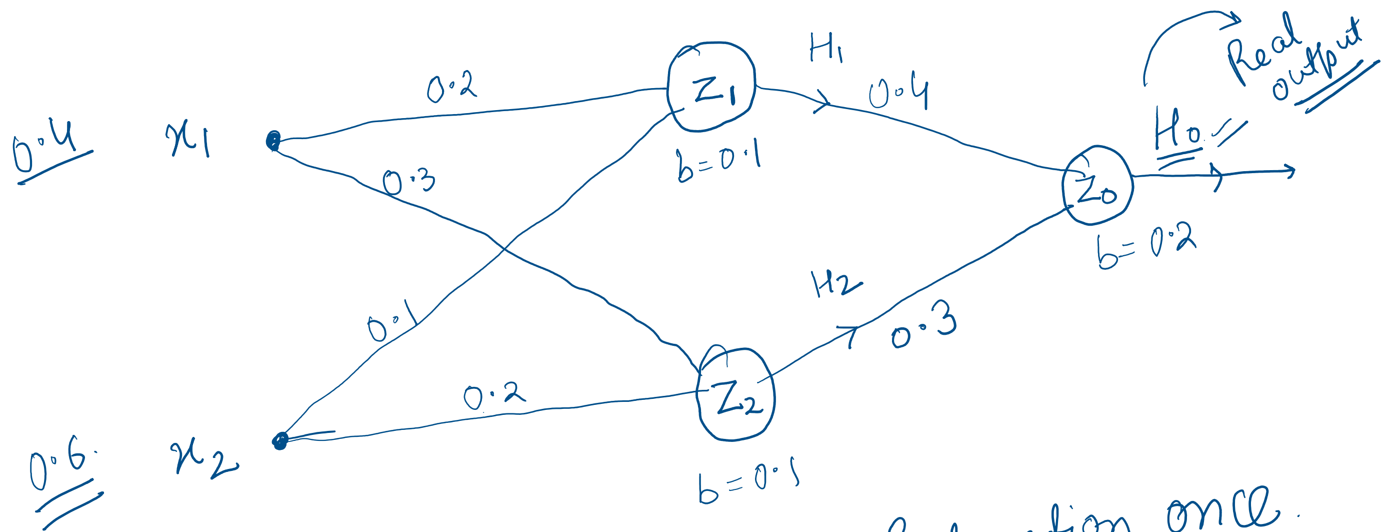
- You have got new weight & biases (Replace with old ones)
- Perform forward Propagation
- Calculate Loss
- Backpropagation again
- Weights update

Loop

Till you get minimum loss → More accurate model.

{ Training of Model is Done ✓ Completed

Prediction Phase



① Perform forward Propagation once.

Drawbacks.

① fail in image data. $\xrightarrow{\text{fix}}$ CNN.

② fails in textual / sequential data $\xrightarrow{\text{fix}}$ RNN LSTM