Deep Learning

(9:05PM) Start.

$$y \sim f(x) + \epsilon$$

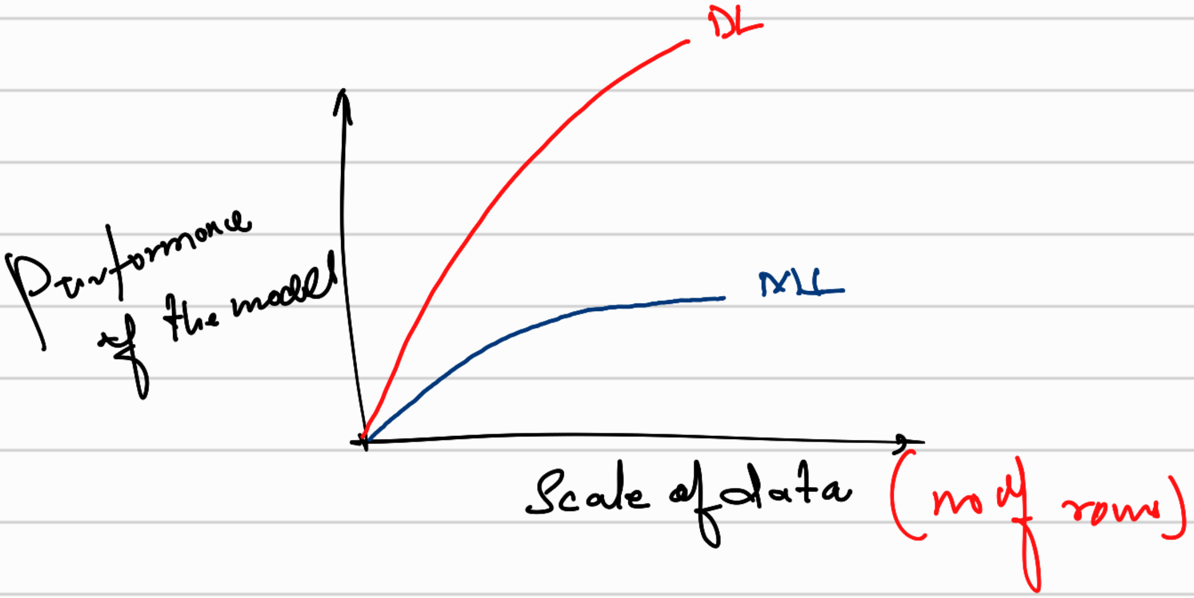
- Reg
- Class

- Parameters
  - Linear
  - Log
  - SVM
- non parameters.

Parametric form $y = f(x)$

Equation

Technique → f



ML →

x_1	x_2	y

DL → automate the feature engineering.

Tabular data

$y \leftarrow f(x)$

x_1	x_2	x_3	y

Artificial neural network

image data

$y \leftarrow f(\text{image})$

→ Convolution neural

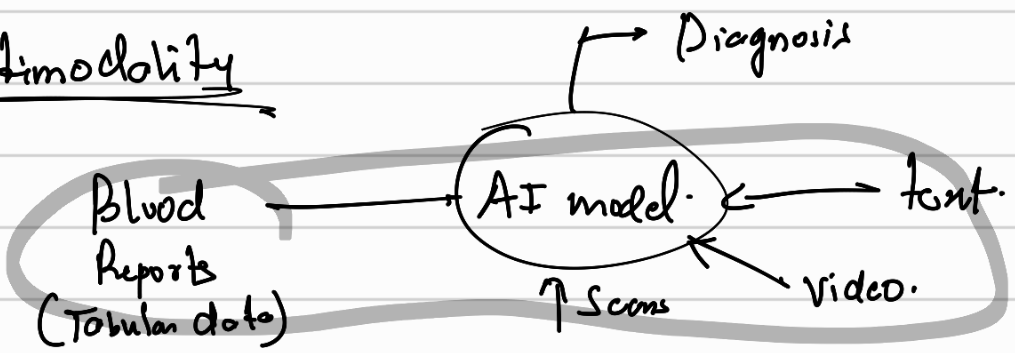
$y \leftarrow f(\text{text})$. → RNN, LSTM, Transformers

$y \leftarrow f(\text{audio})$

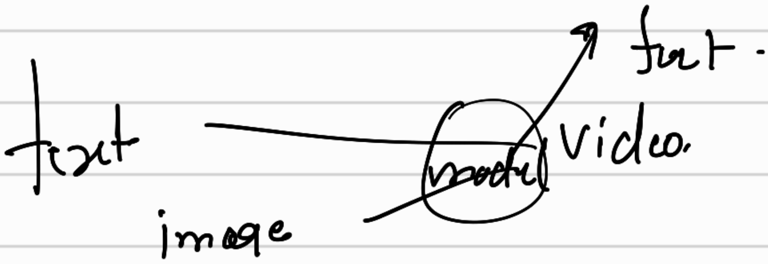
$y \leftarrow f(\text{video})$. → Sequences

$y \leftarrow f(\text{graph})$. → graph neural network

Multimodality

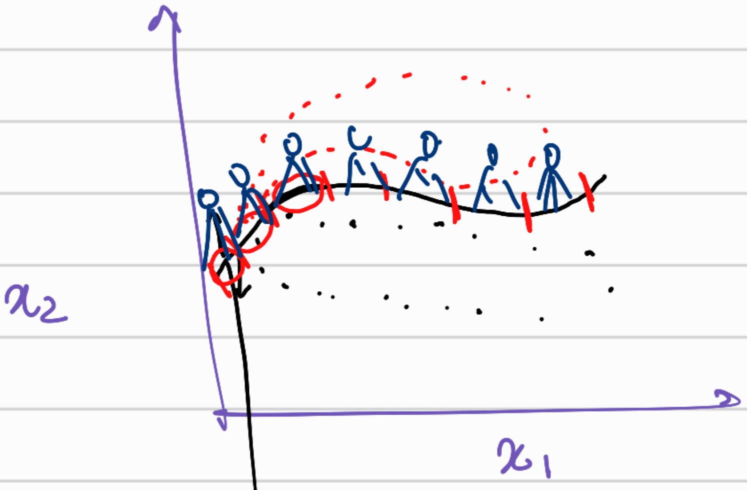


Diagnosis → f(Reports, scans, video, text,).

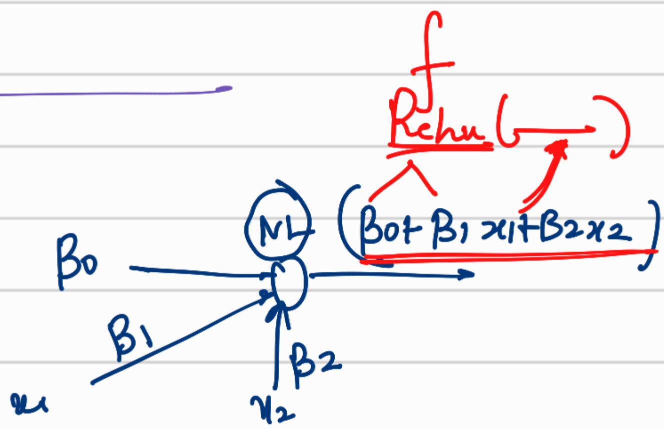


Classification

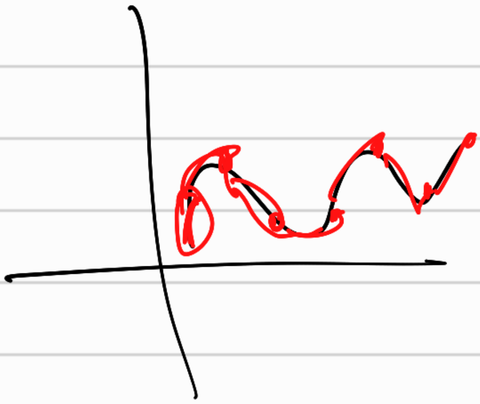
$y = f(x)$

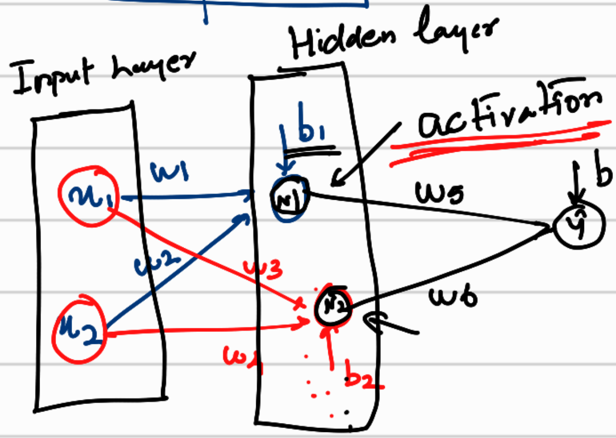


B_0, B_1, Non-linear function



x_1	x_2	Y
✓	✓	✓
✓	✓	✓
✓	✓	✓
✓	✓	✓





$$\hat{y} = w_5 N_1 + w_6 N_2 + b_3$$

$$\Rightarrow N_1 = w_1 x_1 + w_2 x_2 + b_1$$

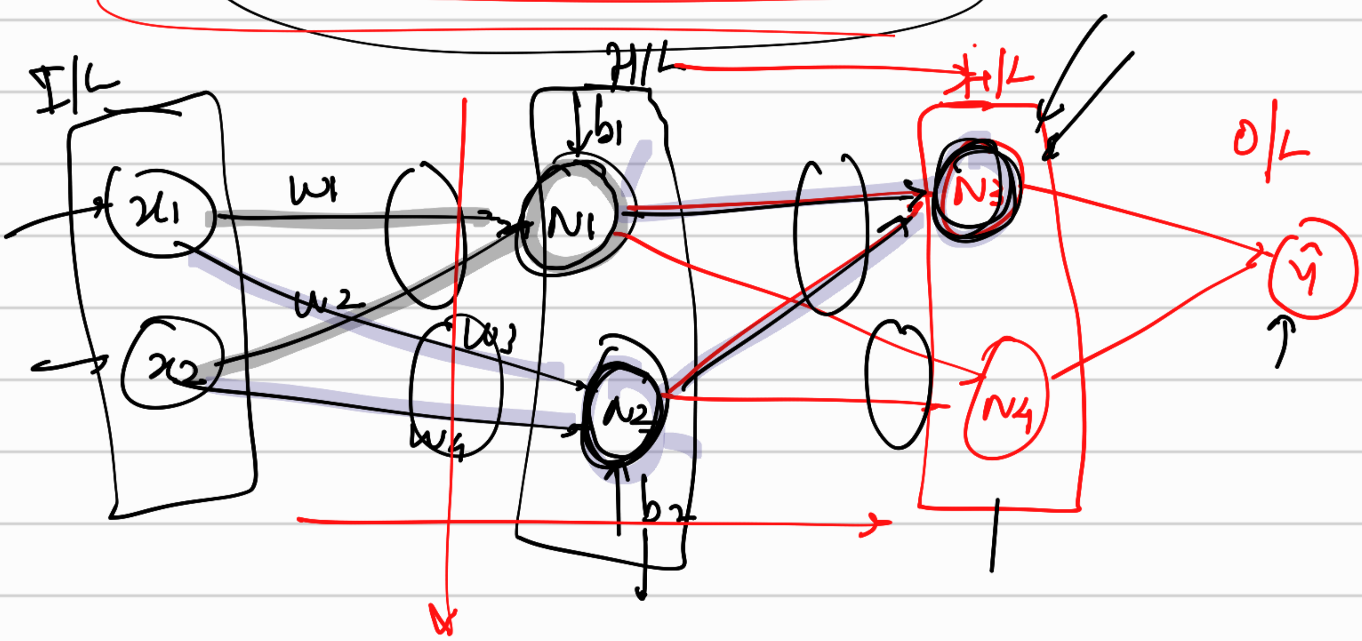
$$N_2 = w_3 x_1 + w_4 x_2 + b_2$$

$$\hat{y} = w_5 (w_1 x_1 + w_2 x_2 + b_1) + w_6 (w_3 x_1 + w_4 x_2 + b_2) + b_3$$

$$\Rightarrow \hat{y} = w_5 w_1 x_1 + w_5 w_2 x_2 + w_5 b_1 + w_6 w_3 x_1 + w_6 w_4 x_2 + w_6 b_2 + b_3$$

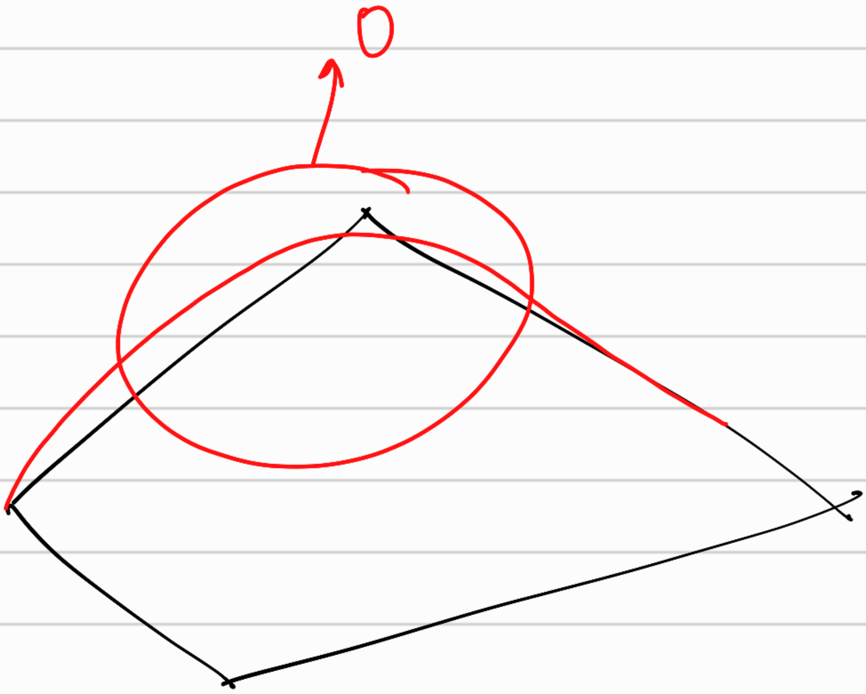
$$\hat{y} = x_1 (w_5 w_1 + w_6 w_3) + x_2 (w_5 w_2 + w_6 w_4) + w_5 b_1 + w_6 b_2 + b_3$$

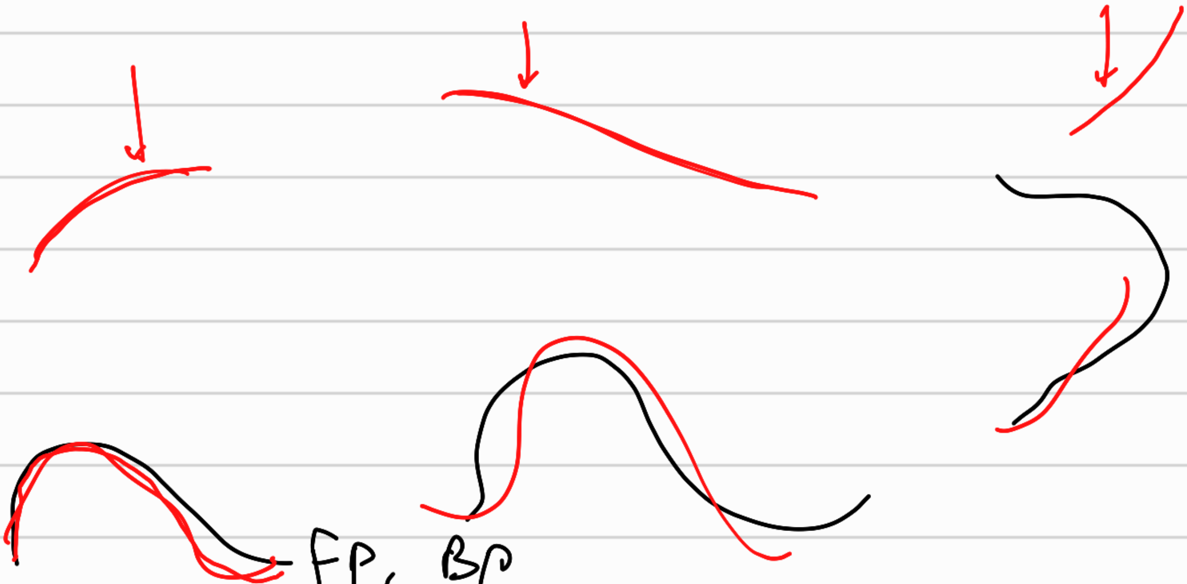
$$\Rightarrow \hat{y} = A x_1 + B x_2 + C$$



Objective of the NN training:

: Minimization of the loss ($f(y, \hat{y})$) by changing weights, and the biases





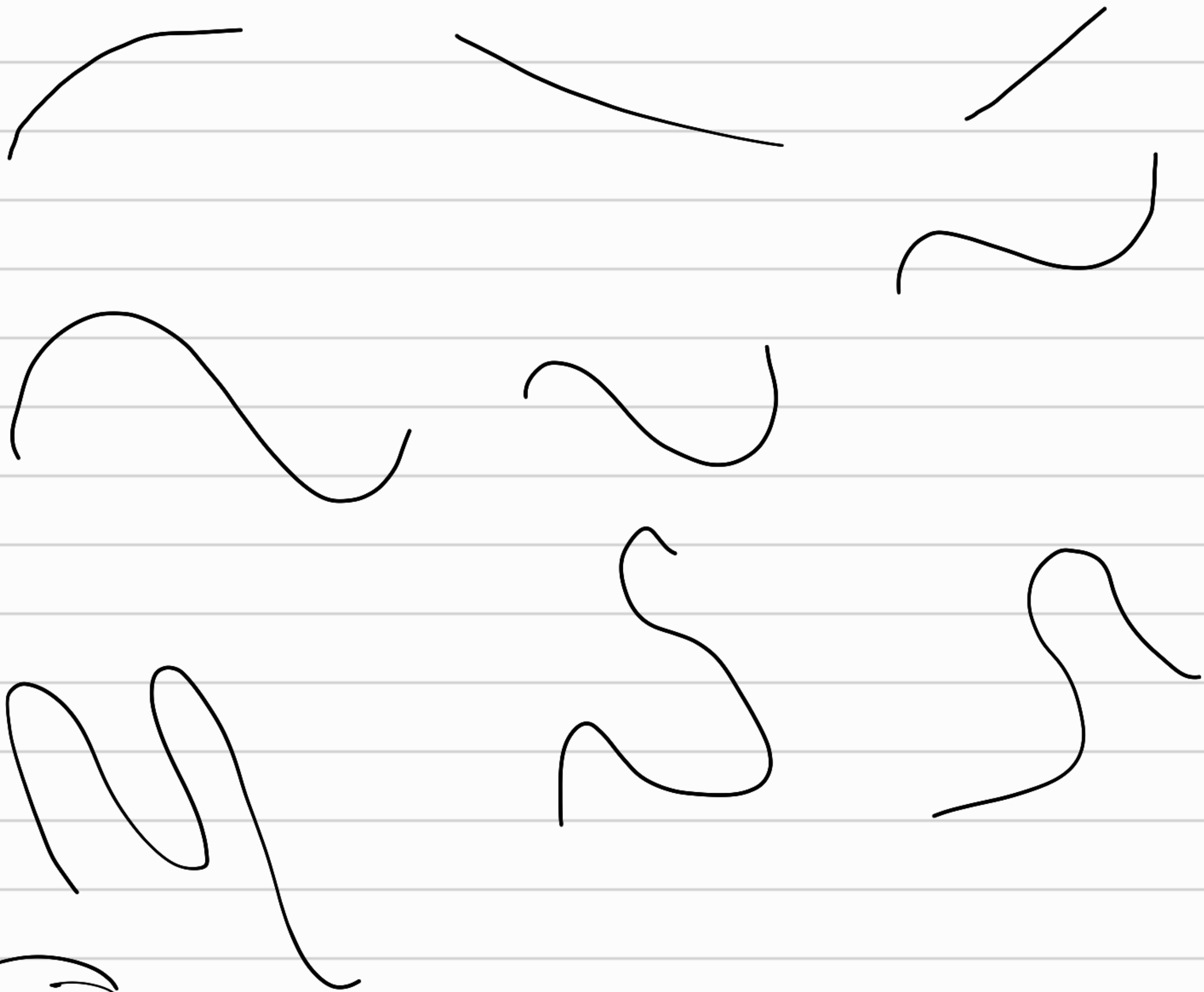
FP, BP

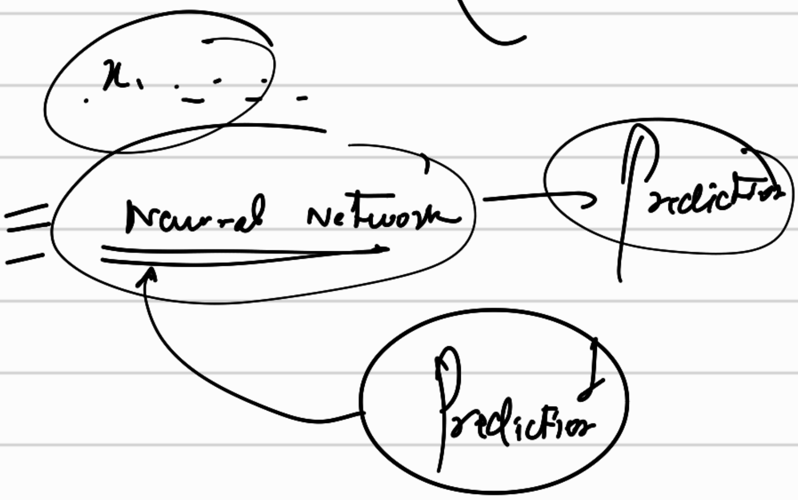
Gradient Descent → FP, BP

Code

tf, Keras

No HL, No Neurons → Hyperparameters



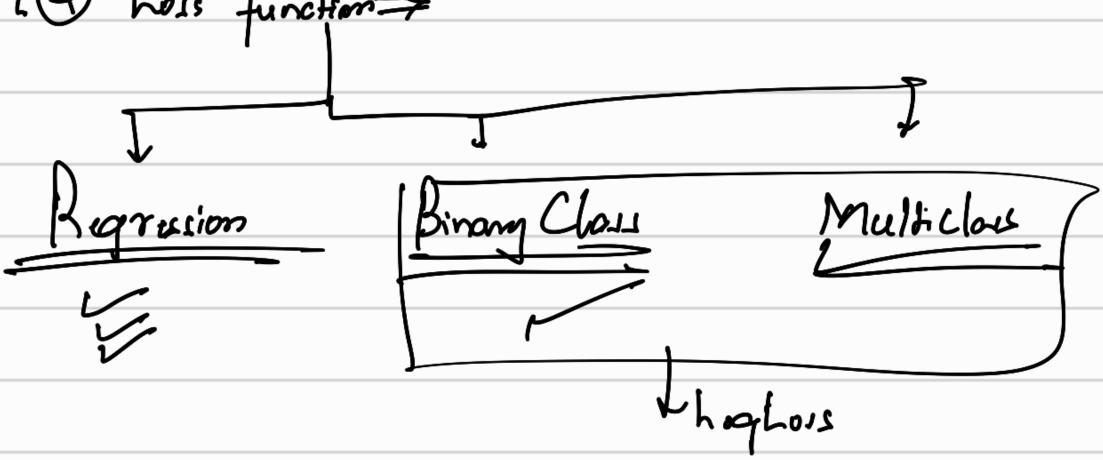


Explainability

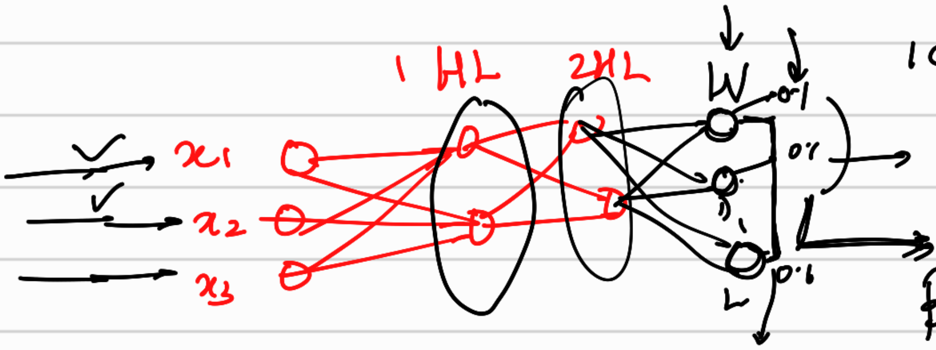
Lecture-2. 2/8/2024

Forward Propagation and Backward Propagation

1. Define the NN architecture (HL, Neurons).
2. Activation functions
3. Initialise the weights and biases. → random
4. Loss function →



5. Activation function at the o/p layer.



1 class o/Rest

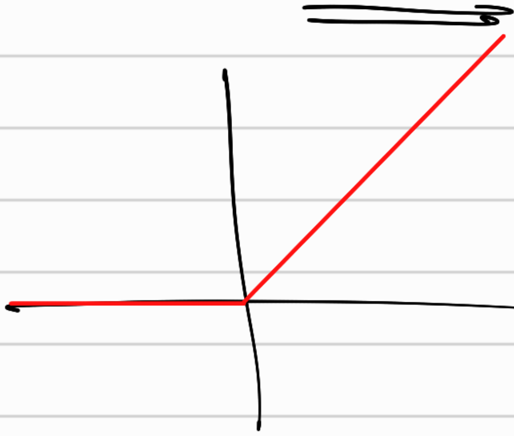
$-\infty, +\infty$

0-1 (Sigmoid).

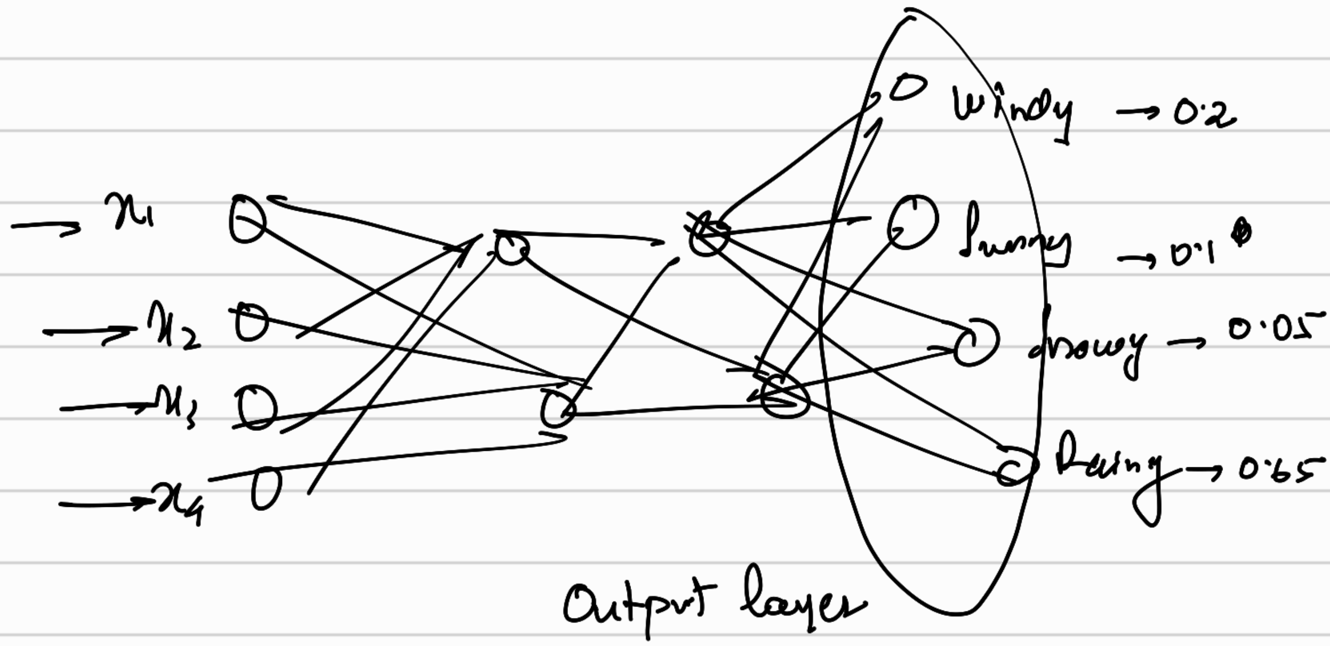
Binary

n classes 1

Multiclassification



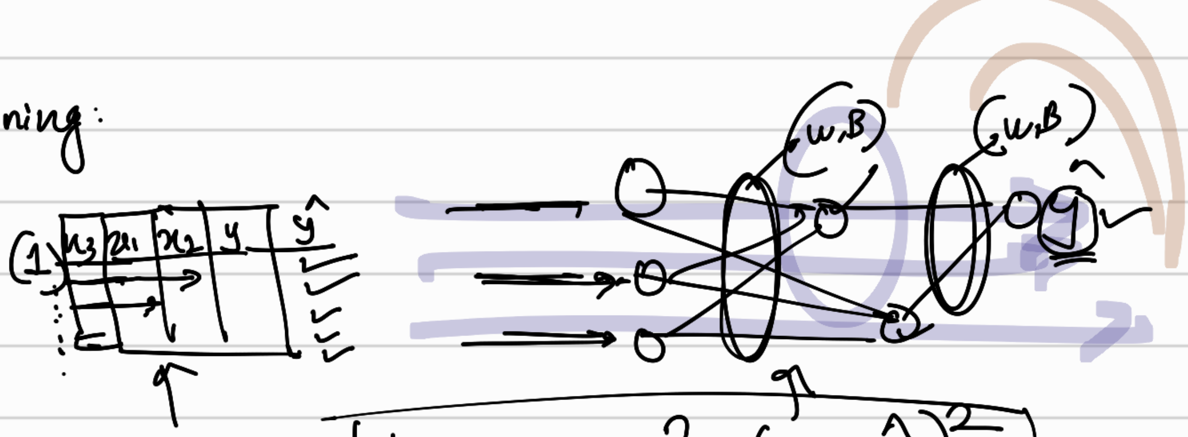
W	D	L
60%	20%	20%



Output layer

Type Problem	No of neurons	activation
Regr	1	linear
BC	1	Sigmoid 0-1
MC	nclasses	Softmax

Training:



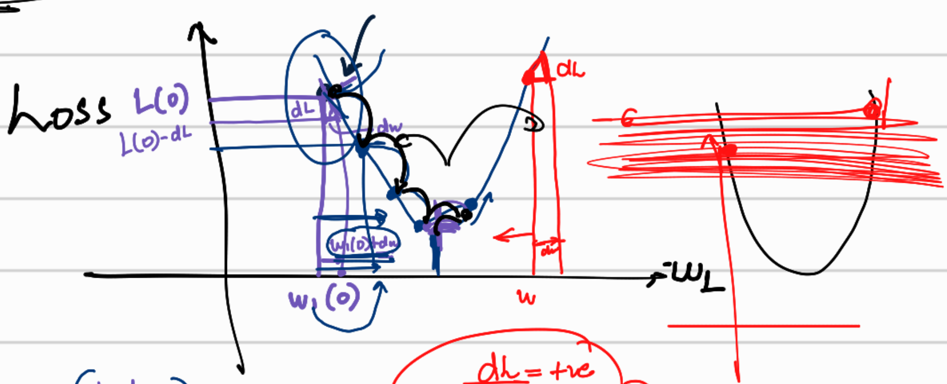
$$\boxed{Loss = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$$

Forward Propagation

Backward Propagation : Reduce the loss by updating the weights and biases.

$$\text{Loss} = f(y_i, w, B, x_i)$$

Gradient Descent:→

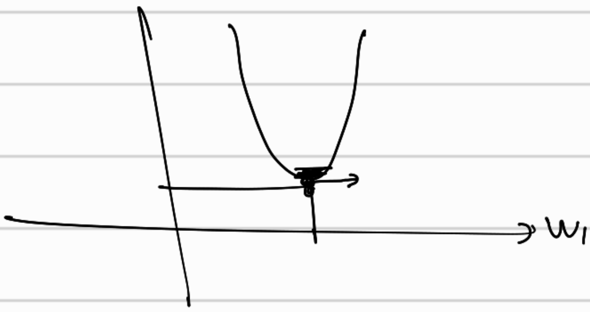


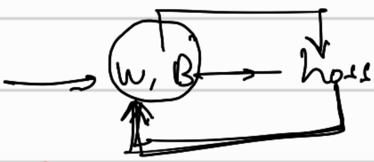
$(dL/dw) = -ve$

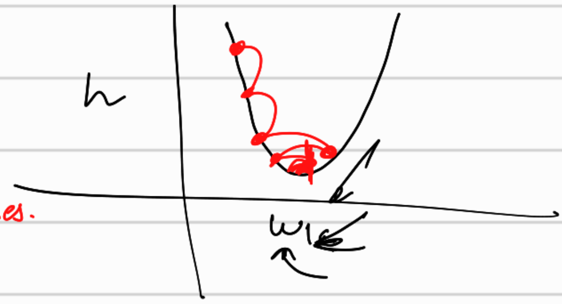
$\frac{dL}{dw} = +ve$

$$\boxed{w_2(1) = w_2(0) - \alpha \frac{dL}{dw_2(0)}}$$

$$w_1(1) = w_1(0) - \alpha \frac{dL}{dw_1(0)}$$







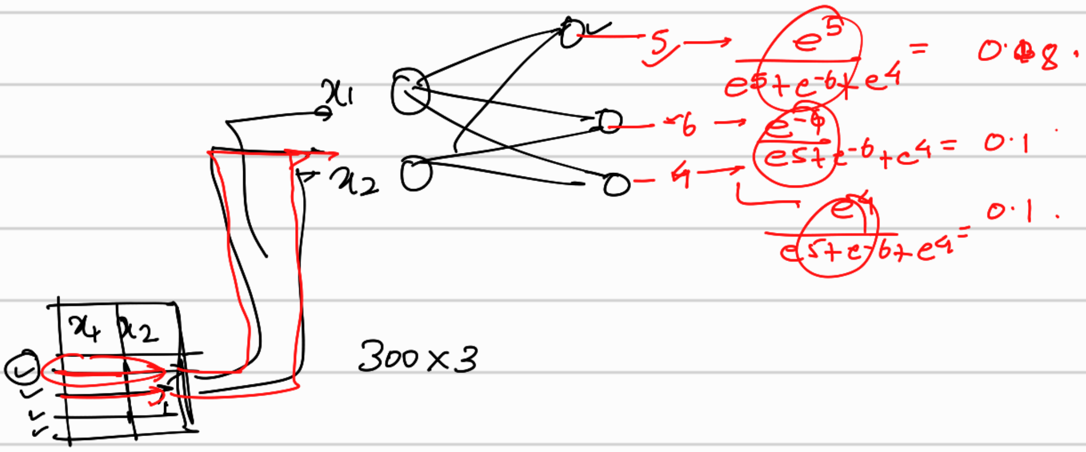
for all weights and biases.

1 FP + 1BP → 1 epoch

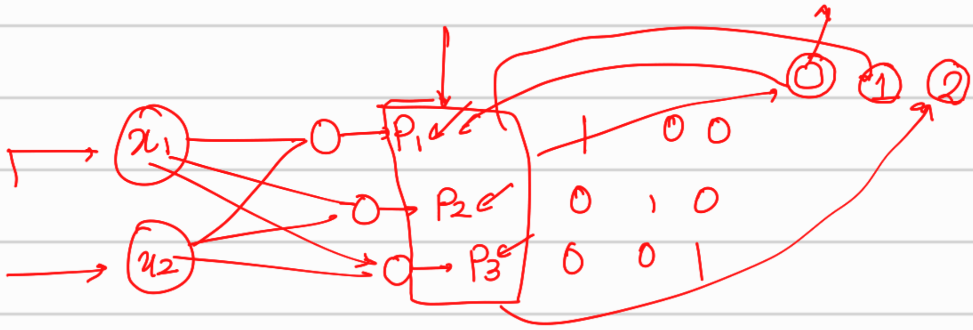
$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \longrightarrow \frac{dL}{d\hat{y}} \text{ easy}$$

$\boxed{\text{Chain Rule of diff}}$

$\frac{dL}{dw_2}$ $\frac{dL}{dw_3}$ $\frac{dL}{dw_1} =$ $\frac{dL}{d\hat{y}}$



$$\text{Log Loss} = \sum_{i=1}^{n} \sum_{j=1}^{K} y_{ij} \log \hat{y}_{ij}$$



1st row

$$-\sum_{j=1}^{3} y_{ij} \times \log \hat{y}_{ij}$$

$$-\left[1 \times \log 0.2 + 0 \times \log 0.3 + 0 \log 0.5\right]$$

$$= -\log 0.2 \Rightarrow \log P[Y]$$

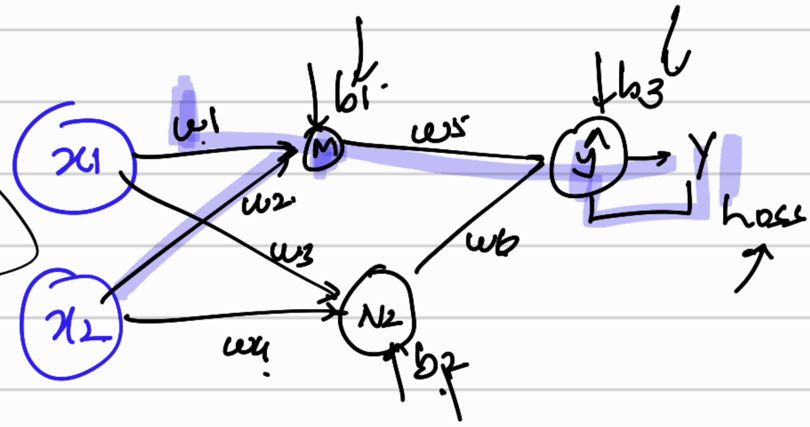
2nd row

$$-\log P[1] = -\log 0.3$$

Lecture-3 7/8/2024

for all ks, biases

$$w_k(t+1) = w_k(t) - \alpha \frac{\partial L}{\partial w_k(t)}$$



$$N_1 = w_1 x_1 + w_2 x_2 + b_1$$

$$N_2 = w_3 x_1 + w_4 x_2 + b_2$$

$$\hat{y} = w_5 \times N_1 + w_6 N_2 + b_3$$

$$L = \sum_{i=1}^{n} \frac{1}{n} \left(y_i - \hat{y}_i\right)^2$$

$$\frac{\partial L}{\partial w_5} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_5} \rightarrow N_1$$

$$\frac{\partial L}{\partial w_6} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_6} \Rightarrow \sum_{i=1}^{n} \frac{1}{n} 2 \left(y_i - \hat{y}_i\right) \times -1 \quad N_2 \; x_i$$

$$y = x^2$$

$$m = y^2$$

$$z = m^2 - 1$$

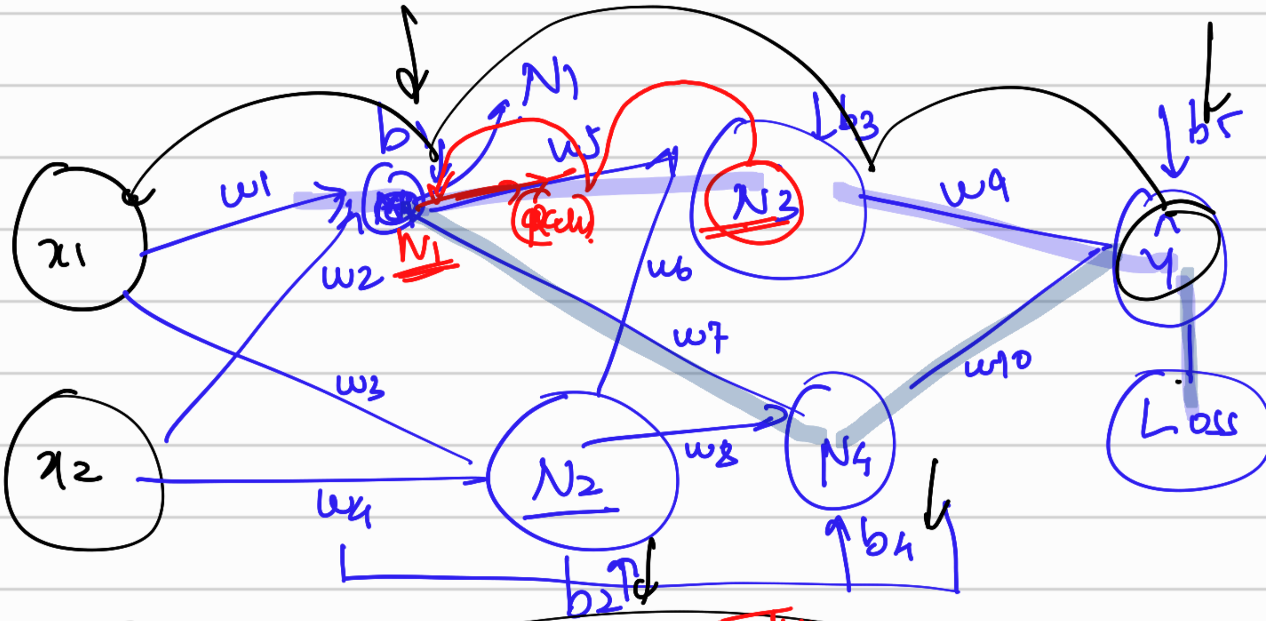
$$\frac{\partial z}{\partial m} = 2m$$

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial m} \times \frac{\partial m}{\partial y} \times \frac{\partial y}{\partial x}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial N_1} \times \frac{\partial N_1}{\partial w_1} \ldots$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial N_1} \times \frac{\partial N_1}{\partial w_2}$$

$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial N_2} \times \frac{\partial N_2}{\partial w_3}$$



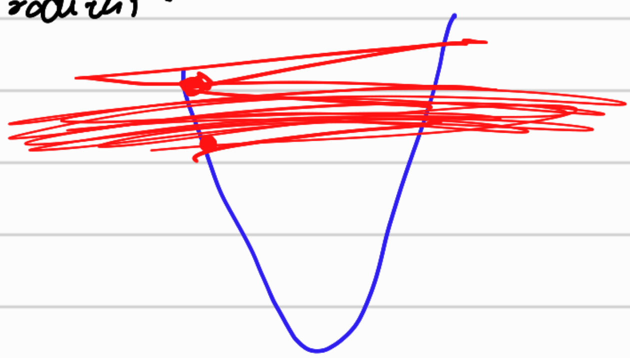
HL
+2

100

activation

$$\frac{\partial L}{\partial w_1} = \left(\frac{\partial L}{\partial \hat{y}} \times \frac{d\hat{y}}{dN_3} \times \frac{dN_3}{\partial N_1} \times \frac{dN_1}{\partial w_1}\right) + \left(\frac{\partial L}{\partial \hat{y}} \times \frac{d\hat{y}}{dN_4} \times \frac{\partial N_4}{\partial N_1} \times \frac{dN_1}{\partial w_1}\right).$$

Vanishing (vs) Exploding gradient.



Loss =

		$\hat{y}$			y	
x_1	x_2	P_1	P_2	P_3	OHE	LE
		0.2	0.1	0.7	010	1
		0.3	0.4	0.3	001	2
		0.5	0.2	0.3	100	0

LE | OHE

LE (0, 1, 2) Indexing

$\log p[y] = \log 0.1$

$y \to 1$

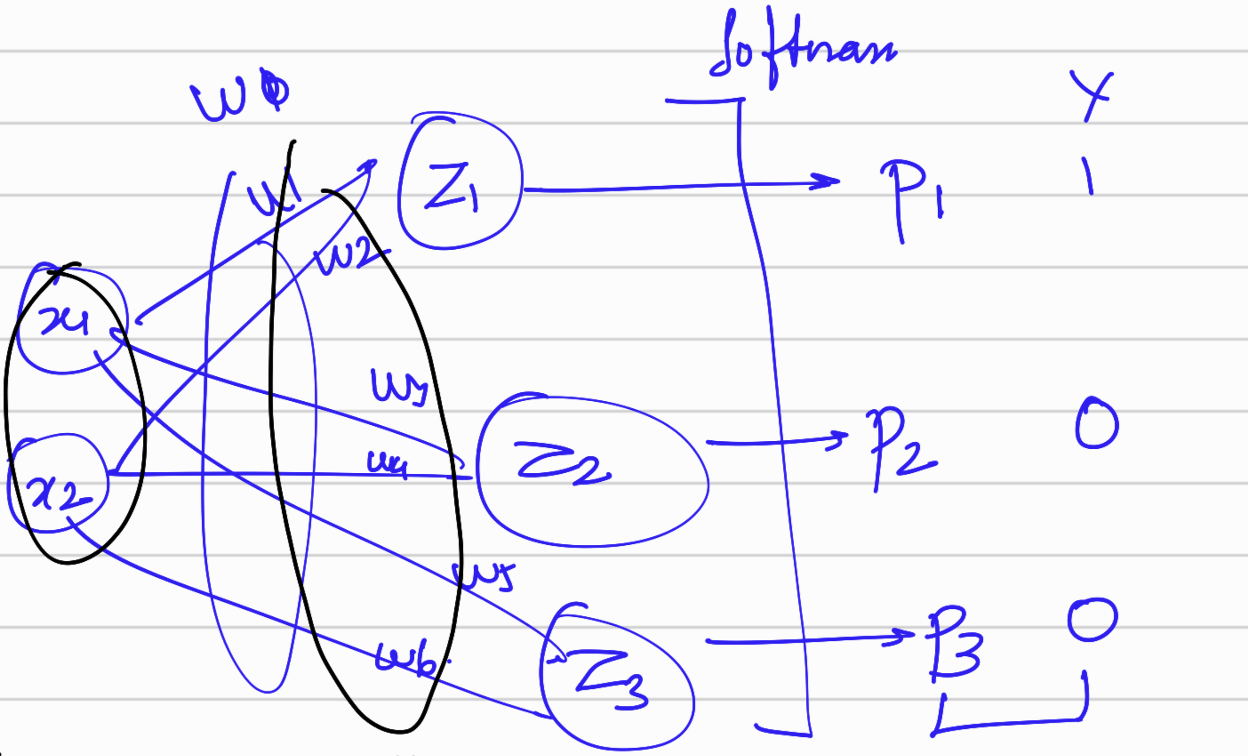
OHE

Multiplication

010, 100, 001

Logloss $= (0 \times \log 0.2 + 1 \times \log 0.1 + 0 \times \log 0.7)$.

$0 \times \log 0.3 + 0 \times \log 0.4 + 1 \times \log 0.5$



$Z = wx + b$

$\frac{dz}{dw} = x$.

$$\frac{dL}{dw} = \frac{dL}{dp} \times \frac{dp}{dz} \times \frac{dz}{dw}$$

$[P_1, P_2, P_3] - [1\ 0\ 0]$

$$\frac{dL}{dw} = [p-y] \times x$$

$z = wx + b$

$\frac{dz}{db} = 1$

$$\frac{dL}{db} = \frac{dL}{dp} \times \frac{dp}{dz} \times \frac{dz}{db}$$

$$\frac{dL}{db} = [p-y]$$

$p-y$

$[p_1 \ p_2 \ p_3] - [1 \ 0 \ 0] = [p_1 - 1 \ \ p_2 \ \ p_3]$

$[p_1 \ p_2 \ p_3] - 0$

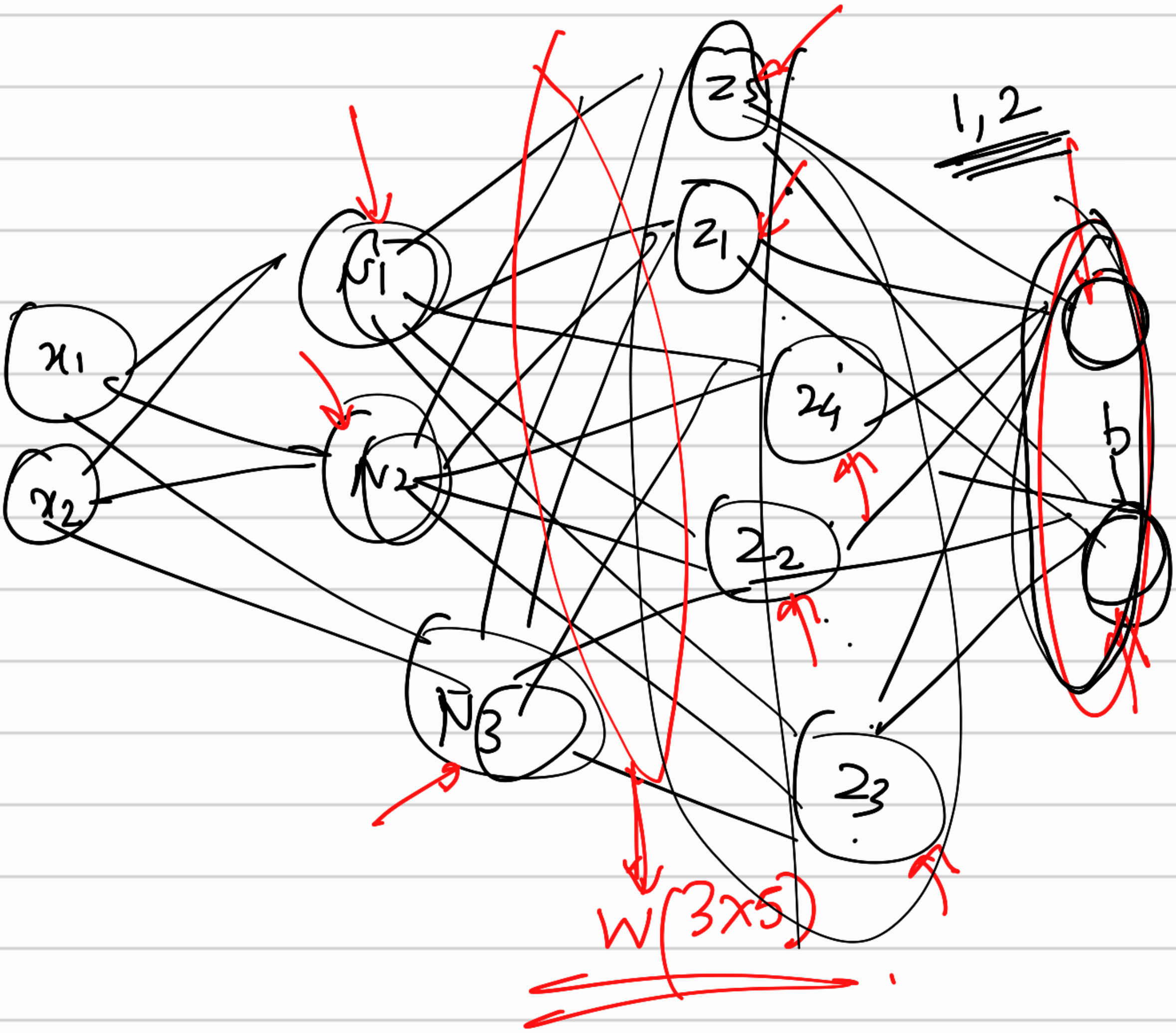
1 $\quad p_1 - 1 \quad p_2 \quad p_3$

2 $\quad p_1 \quad p_2 - 1 \quad p_3$

$\quad p_1 \quad p_2 \quad p_3 - 1$

$$p_1 = \frac{e^{z_1}}{e^{z_1} + e^{z_2} + e^{z_3}}$$

$$\frac{dp_1}{dz_1} =$$



300x3 + 1x3 (Broadcasted manner).

$Z = Wx + B$

$Z = 2\times3$

$X\,W$

$(300\times2) \times (2\times3)$

$Z = 300\times3 \quad + [b_1\ b_2\ b_3]$

300 $\qquad$ 1x3

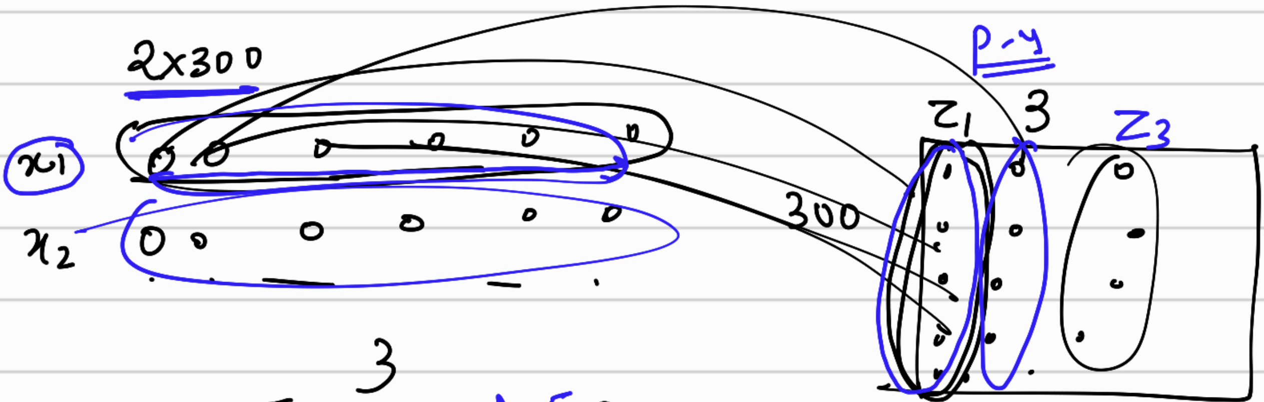
$$1 \quad 0 \quad 0$$

$$0.2 \quad 0.3 \quad 0.5$$

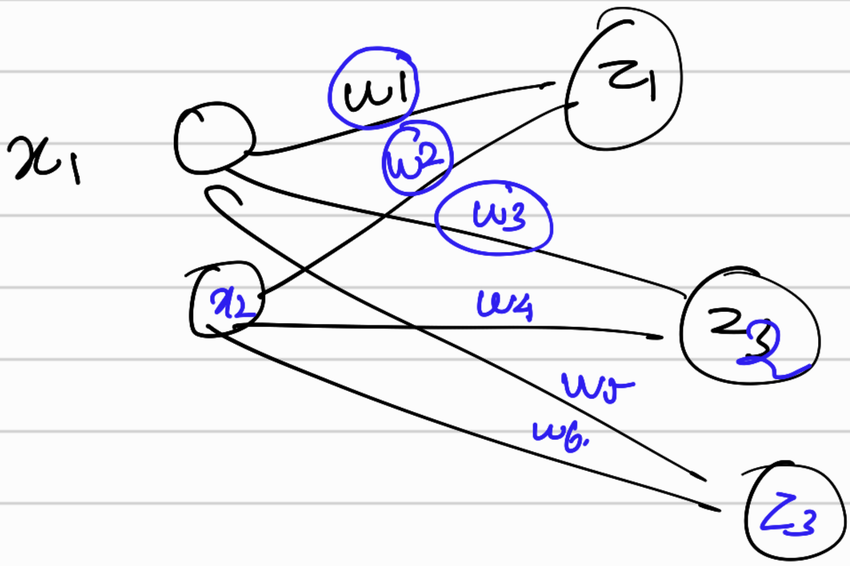
$$\begin{bmatrix} -0.8 & 0.3 & 0.5 \end{bmatrix}$$

$$\underset{4}{\frac{dL}{dz}} = \begin{bmatrix} \boxed{-0.8} & 0.3 & 0.5 \end{bmatrix}$$

$$dw = x^T \, dz$$



$$\begin{array}{cc} & 3 \\ 2 & \begin{bmatrix} \overset{\partial w_1}{0} & \overset{\partial w_3}{0} & \overset{\partial w_5}{0} \\ \overset{\partial w_2}{0} & \underset{\partial w_4}{0} & \underset{\partial w_6}{0} \end{bmatrix} \end{array}$$


x1
w1
w2
w3
x2
w4
w5
w6
z1
z2
z3