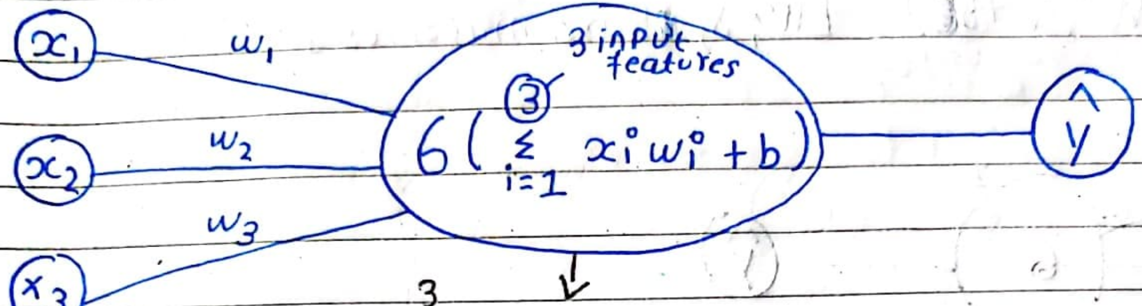
feed-forward in Neural Networks

① Simple Neural Network (perceptron):



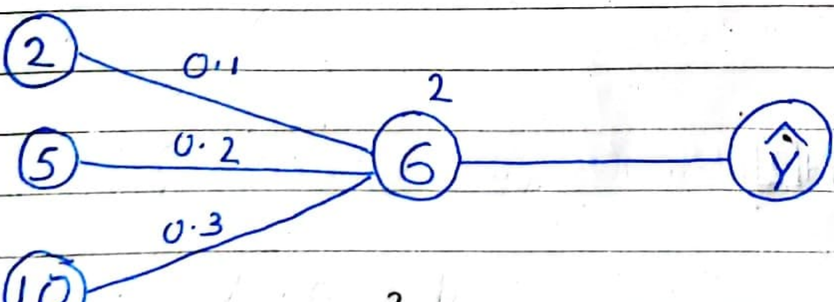
$$\sum_{i=1}^{3} x_i w_i = x_1 w_1 + x_2 w_2 + x_3 w_3$$

Now, working mechanism

CGPA	Iq	level	Placement
2	5	10	0
5	8	20	1

let's random assign weights and bias:

$w_1 = 0.1$ $\quad w_2 = 0.2$ $\quad w_3 = 0.3$ $\quad b = 2$



$$\therefore \hat{y} = \sigma\left(\sum_{i=1}^{3} x_i w_i + b\right) = \sigma(x_1 w_1 + x_2 w_2 + x_3 w_3 + b)$$

$$= \sigma(2 \times 0.1 + 5 \times 0.2 + 10 \times 0.3 + 2)$$

$$= \sigma(6.2)$$

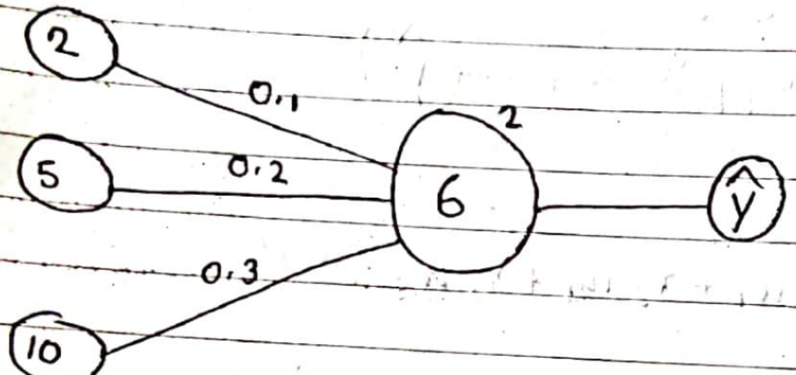
$$\therefore \sigma(x) = f(x) = \frac{1}{1 + e^{-x}} \qquad = \sigma(6.2) = f(6.2) = \frac{1}{1 + e^{-6.2}}$$

$$= 0.99$$

$$= \hat{y} = 0.99$$

We use matrix multiplication way to achieve same result like this which is used in Deep learning libraries also:

$$\hat{y} = \sigma(W^T X + B) \quad [W, X, B \text{ are matrixes}]$$



$$\hat{y} = \sigma(W^T X + B)$$

$$W = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix}, \quad X = \begin{bmatrix} 2 \\ 5 \\ 10 \end{bmatrix} \qquad B = \begin{bmatrix} 2 \\ \\ \end{bmatrix}$$

$$\hat{y} = \sigma\left(\begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ 10 \end{bmatrix} + [\,2\,] \right)$$

$$= \sigma([0.1\times 2 + 0.2\times 5 + 0.3\times 10] + [2])$$

$$= \sigma([6.2])$$

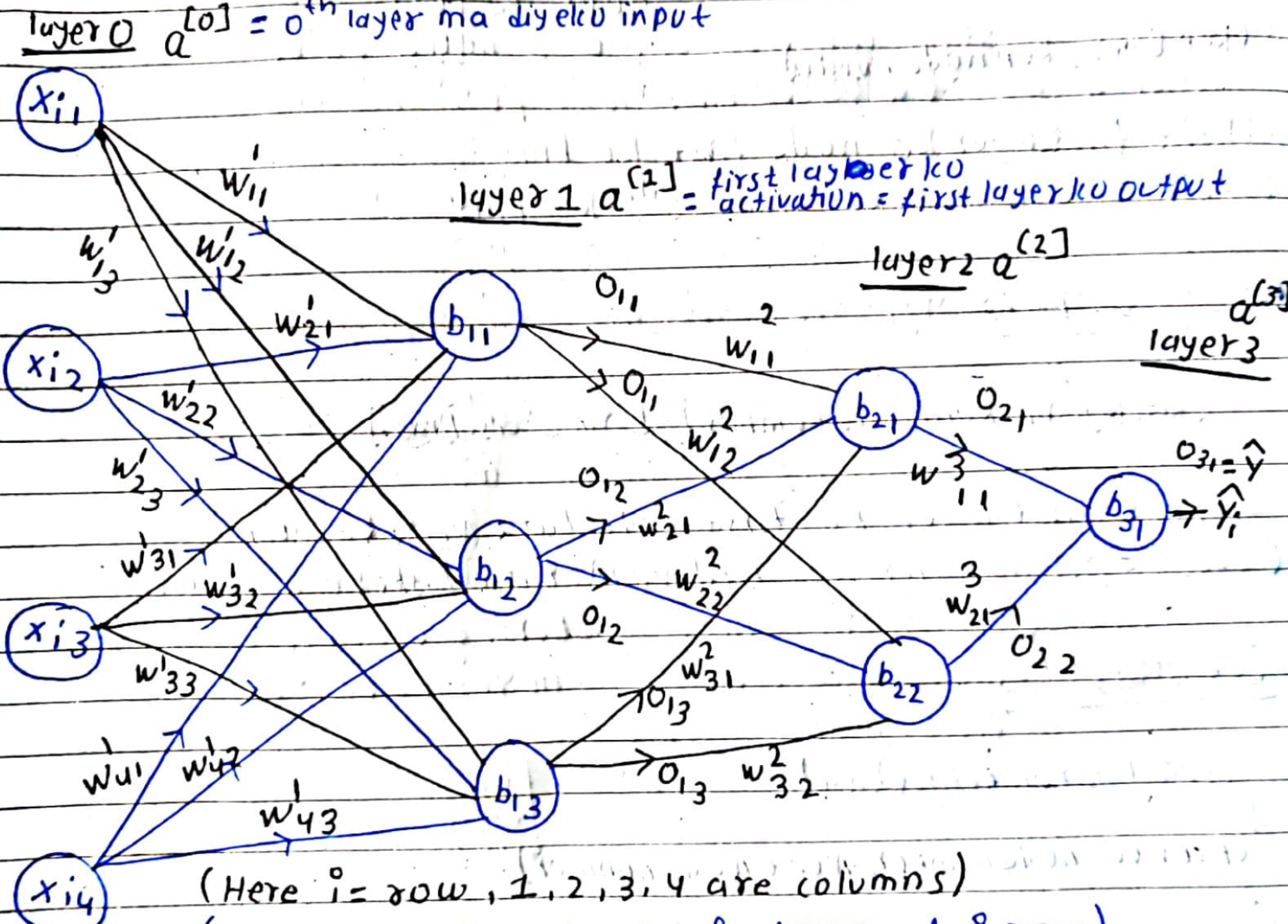
$$= \sigma(x) = f(x) = \frac{1}{1+e^{-x}}$$

$$\therefore \sigma(6.2) = f(6.2) = \frac{1}{1+e^{-6.2}}$$

$$= 0.99$$

② Multi Neuron Neural Network/multi layered perceptron:

let's start with the Notation first:



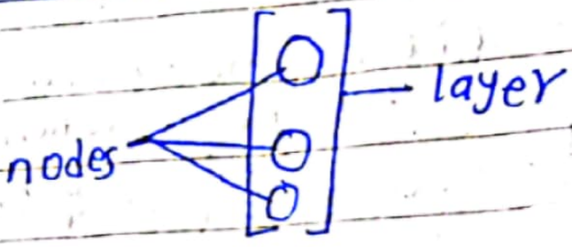
(Here i = row, 1, 2, 3, 4 are columns)
(x_{i1}, x_{i2}, x_{i3}, x_{i4} are features of i row)

let's see total trainable parameters

Input layer = (4x3) weights + 3 bias = 15
hidden layer 1 = (3x2) weights + 2 bias = 8
hidden layer 2 = (2x1) weights + 1 bias = 3
∴ total trainable parameters = 26

How notation is given let's see

for bias



$b_{(\text{kun layer, kun node number})}$

$b(1,1)$ = first layer ko node first ko bias
$b(2,5)$ = Second layer ko node 5^{th} ko bias

for outputs

$O_{(\text{kun layer, kun node number})}$ [Same as bias]

$O_{(1,1)}$ = first layer ko first node bata niskeko output
$O(2,3)$ = Second layer ko Third node bata niskeko output

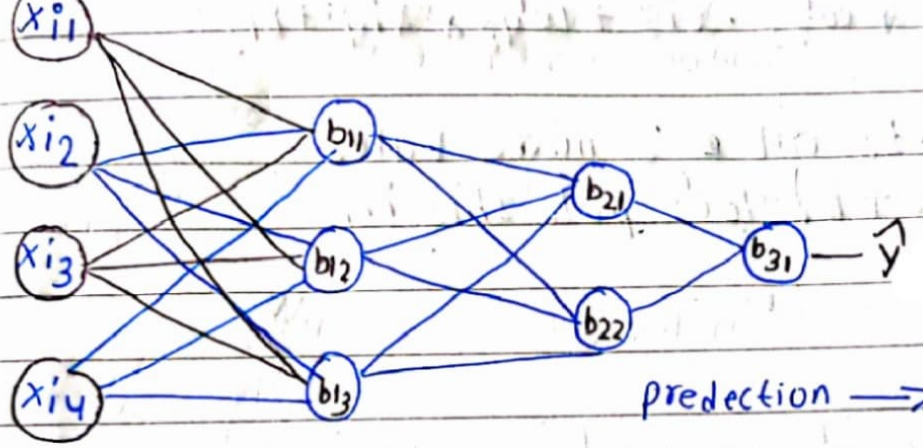
for weights

$W^{(\text{kun layer ma Jadai xa})}_{(\text{previous node number, Next node number})}$

W^{1}_{23} = (first layer ma Jadai xa paila layer ko Second node bata first layer ko 3rd node ma)

W^{2}_{32} = (Second layer ma Jadai xa paila layer ko 3rd node bata Second node ma)

let's see forward propogation



prediction $\longrightarrow$ $\sigma(w^T x + b)$

layer 1

$$= \begin{bmatrix} w'_{11} & w'_{12} & w'_{13} \\ w'_{21} & w'_{22} & w'_{23} \\ w'_{31} & w'_{32} & w'_{33} \\ w'_{41} & w'_{42} & w'_{43} \end{bmatrix}^T \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \end{bmatrix} + \begin{bmatrix} b_{11} \\ b_{12} \\ b_{13} \end{bmatrix}$$

$$= \underset{(3\times 4)}{\begin{bmatrix} w'_{11} & w'_{21} & w'_{31} & w'_{41} \\ w'_{12} & w'_{22} & w'_{32} & w'_{42} \\ w'_{13} & w'_{23} & w'_{33} & w'_{43} \end{bmatrix}} \underset{(4\times 1)}{\begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \end{bmatrix}} + \begin{bmatrix} b_{11} \\ b_{12} \\ b_{13} \end{bmatrix}$$

$= (4 \times 1)$

$= (3 \times 1)$

$$= \begin{bmatrix} w'_{11} \cdot x_{i1} + w'_{21} \cdot x_{i2} + w'_{31} \cdot x_{i3} + w'_{41} \cdot x_{i4} \\ w'_{12} \cdot x_{i1} + w'_{22} + x_{i2} + w'_{32} \cdot x_{i3} + w'_{42} \cdot x_{i4} \\ x'_{13}\, x_{i1} + w'_{23} \cdot x_{i2} + w'_{33} \cdot x_{i3} + w'_{43} \cdot x_{i4} \end{bmatrix} + \begin{bmatrix} b_{11} \\ b_{12} \\ b_{13} \end{bmatrix}$$

$$= \begin{bmatrix} w^1_{11} \cdot x_{i1} + w^1_{21} \cdot x_{i2} + w^1_{31} \cdot x_{i3} + w^1_{41} \cdot x_{i4} + b_{11} \\ w^1_{12}\, x_{i1} + w^1_{22} \cdot x_{i2} + w^1_{32} \cdot x_{i3} + w^1_{42} \times x_{i4} + b_{12} \\ w^1_{13}\, x_{i1} + w^1_{23} \cdot x_{i2} + w^1_{33} \cdot x_{i3} + w^1_{43} \times x_{i4} + b_{13} \end{bmatrix}$$

Now, this matrix consists all 3 node value calculated by ($w^T x + b$) let's put this in Sigmoid

$$\sigma\left(\begin{bmatrix} w^1_{11} \cdot x_{i1} + w^1_{21} \cdot x_{i2} + w^1_{31}\, x_{i3} + w^1_{41} \cdot x_{i4} + b_{11} \\ w^1_{12} \cdot x_{i1} + w^1_{22} \cdot x_{i2} + w^1_{32} \cdot x_{i3} + w^1_{42} \cdot x_{i4} + b_{12} \\ w^1_{13} \cdot x_{i1} + w^1_{23} \cdot x_{i2} + w^1_{33} \cdot x_{i3} + w^1_{43} \cdot x_{i4} + b_{13} \end{bmatrix} \right)$$

$$= \begin{bmatrix} O_{11} \\ O_{12} \\ O_{13} \end{bmatrix}$$

[This is the three numbers O_{11} O_{12} O_{13} which is given by 3 nodes of the first hidden layer]

$\therefore$ It is also called $a^{[1]}$ $\therefore$ $a^{[1]} = \begin{bmatrix} O_{11} \\ O_{12} \\ O_{13} \end{bmatrix}$

layer 2

$$\begin{bmatrix} w^2_{11} & w^2_{12} \\ w^2_{21} & w^2_{22} \\ w^2_{31} & w^2_{32} \end{bmatrix}^T \begin{bmatrix} O_{11} \\ O_{12} \\ O_{13} \end{bmatrix} + \begin{bmatrix} b_{21} \\ b_{22} \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix}}_{(2\times 3)} \underbrace{\begin{bmatrix} O_{11} \\ O_{12} \\ O_{13} \end{bmatrix}}_{(3\times 1)} + \underbrace{\begin{bmatrix} b_{21} \\ b_{22} \end{bmatrix}}_{(3\times 1)}$$

(2×1)

$$= \sigma\left(\begin{bmatrix} w^2_{11}\cdot O_{11} + w^2_{21}\cdot O_{12} + w^2_{31}\cdot O_{13} + b_{21} \\ w^2_{12}\cdot O_{11} + w^2_{22}\cdot O_{12} + w^2_{32}\cdot O_{13} + b_{22} \end{bmatrix}\right)$$

$$= \begin{bmatrix} O_{21} \\ O_{22} \end{bmatrix}$$

$\therefore$ It is also called $a^{[2]} = \begin{bmatrix} O_{21} \\ O_{22} \end{bmatrix}$

<u>layer 3</u>

$$\begin{bmatrix} w^3_{11} \\ w^3_{21} \end{bmatrix}^T \begin{bmatrix} O_{21} \\ O_{22} \end{bmatrix} + \begin{bmatrix} b_{31} \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} w^3_{11} & w^3_{21} \end{bmatrix}}_{(1\times 2)} \underbrace{\begin{bmatrix} O_{21} \\ O_{22} \end{bmatrix}}_{(2\times 1)} + \begin{bmatrix} b_{31} \end{bmatrix}$$

1×1

$$= \sigma\left(\begin{bmatrix} w^3_{11}\cdot O_{21} + w^3_{21}\cdot O_{22} + b_{31} \end{bmatrix}\right)$$

$= \begin{bmatrix} O_{31} \end{bmatrix}$ which is $\hat{y}$

$\therefore$ It is also called $a^{[3]} = \begin{bmatrix} O_{31} \end{bmatrix}$ #

So, simply we are doing this

$$a^{[1]} = \sigma\left(a^{[0]} \cdot [W^{[1]}]^T + B^{[1]}\right)$$

$$a^{[2]} = \sigma\left(a^{[1]} \cdot [W^{[2]}]^T + B^{[2]}\right)$$

Here this notation means

$W^{[1]}$ = layer 1 ma gayeko weights

$B^{[1]}$ = layer 1 ko Bais

$\therefore$ W and B are matrices

Collectively we can write

$$\underbrace{\sigma\left(\underbrace{\sigma\left(\underbrace{\sigma\left(a^{[0]} \cdot [W^{[1]}]^T + B^{[1]}\right)}_{a^{[1]}} \cdot [W^{[2]}]^T + B^{[2]}\right)}_{a^{[2]}} \cdot [W^{[3]}]^T + B^{[3]}\right)}_{a^{[3]}}$$