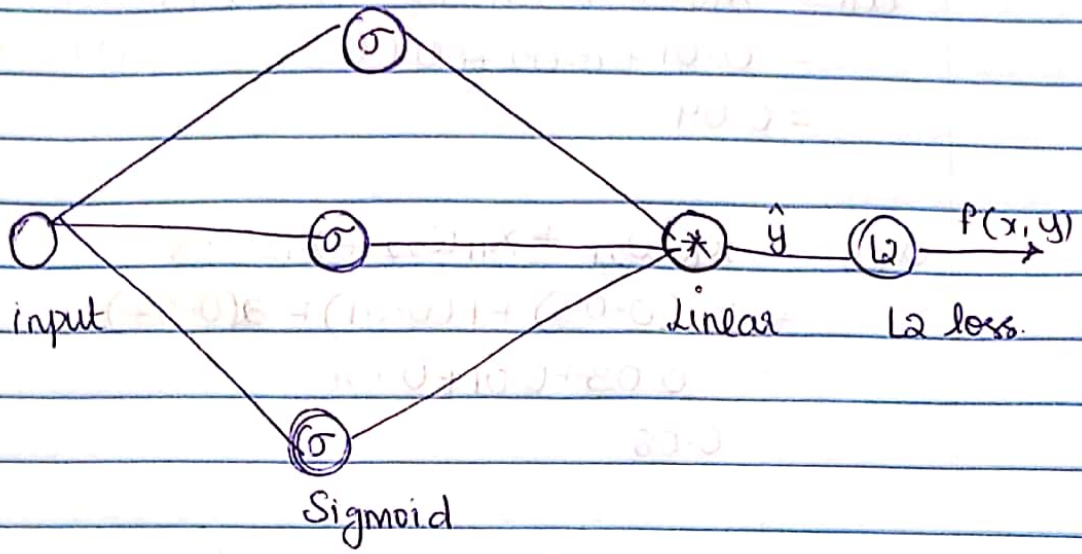
Deep Learning:

1)

a)



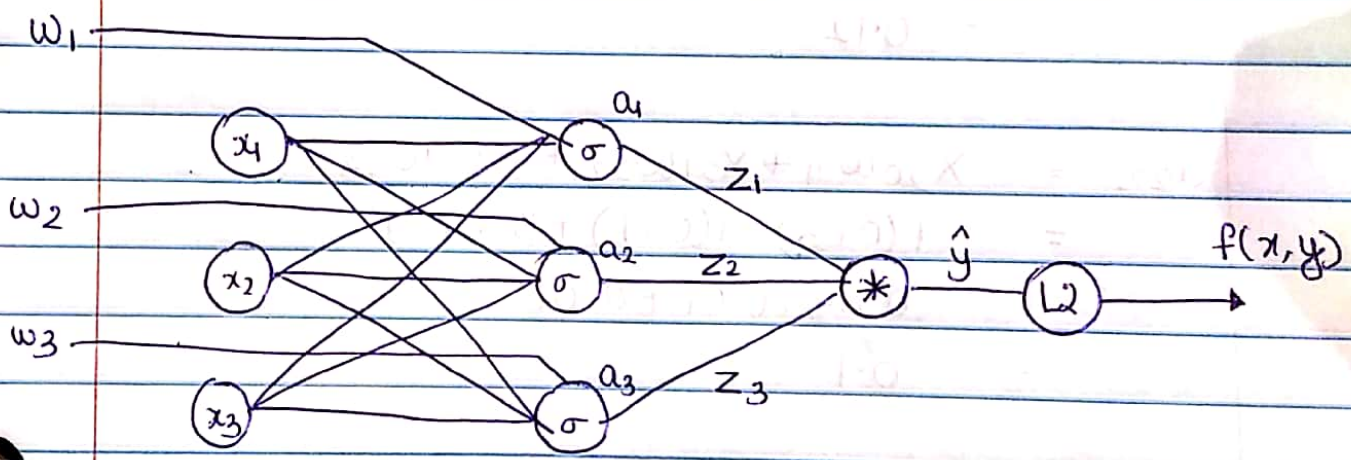
b) $w_1 = (0.01, 0.02, 0.03)$ $w_2 = (0.03, 0.01, 0.02)$

$w_3 = (0.02, 0.03, 0.01)$ $V = (0.01, 0.02, 0.03, 0.04)$

Training data: $x_1, y_1 = [(1, 2), 8]$

$x_2, y_2 = [(1, 3), 11]$

$x_3, y_3 = [(2, 2), 10]$



$a_{11} = x_{10}w_{11} + x_{11}w_{12} + x_{12}w_{13}$

$= 0.01 + 0.02 + 0.06$

$= 0.09$

$a_{12} = x_{10}w_{21} + x_{11}w_{22} + x_{12}w_{23}$

$= 1(0.03) + 1(0.01) + 2(0.02)$

$= 0.03 + 0.01 + 0.04$

$= 0.08$

$a_{13} = x_{10}w_{31} + x_{11}w_{32} + x_{12}w_{33}$

$= 1(0.02) + 1(0.03) + 2(0.01)$

$= 0.02 + 0.03 + 0.02$

$= 0.07$

$a_{21} = x_{20}w_{11} + x_{21}w_{12} + x_{22}w_{13}$

$= 1(0.01) + 1(0.02) + 3(0.03)$

$= 0.01 + 0.02 + 0.09$

$= 0.12$

$a_{22} = x_{20}w_{21} + x_{21}w_{22} + x_{22}w_{23}$

$= 1(0.03) + 1(0.01) + 3(0.02)$

$= 0.03 + 0.01 + 0.06$

$= 0.1$

$$a_{23} = x_{20}w_{31} + x_{21}w_{32} + x_{22}w_{33}$$
$$= 1(0.02) + 1(0.03) + 3(0.01)$$
$$= 0.02 + 0.03 + 0.03$$
$$= 0.08$$

Applying Sigmoid:

$$z_{11} = \frac{1}{1+e^{-a_{11}}} = \frac{1}{1+e^{-0.09}} = 0.522$$

$$z_{12} = \frac{1}{1+e^{-a_{12}}} = \frac{1}{1+e^{-0.08}} = 0.519$$

$$z_{13} = \frac{1}{1+e^{-a_{13}}} = \frac{1}{1+e^{-0.07}} = 0.517$$

$$z_{21} = \frac{1}{1+e^{-a_{21}}} = \frac{1}{1+e^{-0.21}} = 0.529$$

$$z_{22} = \frac{1}{1+e^{-a_{22}}} = \frac{1}{1+e^{-0.1}} = 0.524$$

$$z_{23} = \frac{1}{1+e^{-a_{23}}} = \frac{1}{1+e^{-0.08}} = 0.519$$

$$Z_{381} = \frac{1}{1+e^{-a_{381}}} = \frac{1}{1+e}$$

$$a_{31} = x_{30}w_{11} + x_{31}w_{12} + x_{32}w_{13}$$
$$= 1(0.01) + 2(0.02) + 2(0.03)$$
$$= 0.01 + 0.04 + 0.06$$
$$= 0.11$$

$$a_{32} = x_{30}w_{21} + x_{31}w_{22} + x_{32}w_{23}$$
$$= 1(0.03) + 2(0.01) + 2(0.02)$$
$$= 0.03 + 0.02 + 0.04 = 0.09$$

$$a_{33} = x_{30}w_{31} + x_{31}w_{32} + x_{32}w_{33}$$
$$= 1(0.02) + 2(0.03) + 2(0.01)$$
$$= 0.02 + 0.06 + 0.02 = 0.1$$

$$Z_{31} = \frac{1}{1+e^{-0.11}} = 0.527$$

$$Z_{32} = \frac{1}{1+e^{-0.09}} = 0.522$$

$$Z_{33} = \frac{1}{1+e^{-0.1}} = 0.524$$

O/p:

V_0 = bias

$\hat{y}_1 = V_0 + V_1 Z_{11} + V_2 Z_{12} + V_3 Z_{13}$

$= 0.01 + (0.02)(0.522) + (0.03)(0.519) + (0.04)(0.517)$

$= 0.0566$

$y_1 = 8$

L2 loss $= (y_1 - \hat{y}_1)^2 = 63.096$

$\hat{y}_2 = V_0 + V_1 Z_{21} + V_2 Z_{22} + V_3 Z_{23}$

$= (0.01) + (0.02)(0.526) + (0.03)(0.524) + (0.04)(0.519)$

$= 0.0570$

$y_2 = 11$

L2 loss $= (y_2 - \hat{y}_2)^2 = 119.74$

$\hat{y}_3 = V_0 + V_1 Z_{31} + V_2 Z_{32} + V_3 Z_{33}$

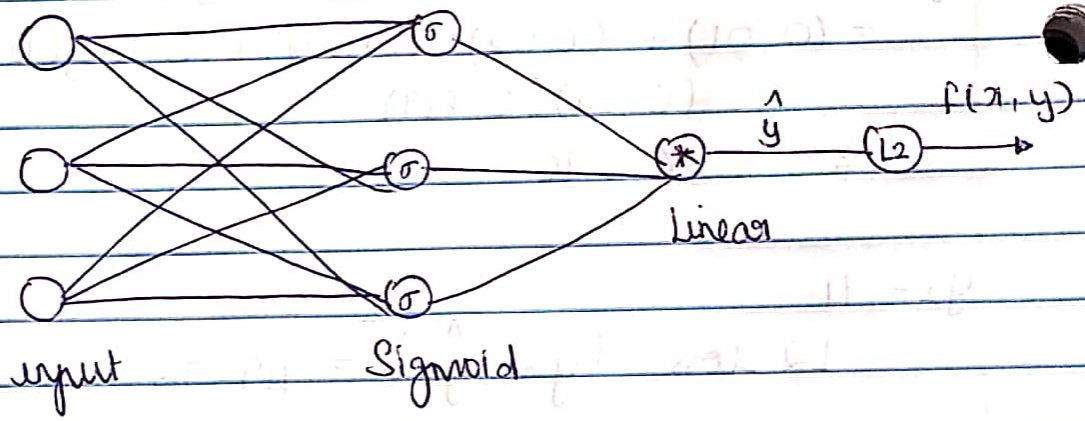
$= 0.01 + 0.02(0.527) + 0.03(0.522) + 0.04(0.524)$

$= 0.0571$

$y_3 = 10$

L2 loss $= (y_3 - \hat{y}_3)^2 = 98.86$

(c) Gradients using backpropagation:



Back propagating from L to the input layers, the gradients are required. They are:

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial L}{\partial \hat{y}_1} \times \frac{\partial \hat{y}_1}{\partial z_{1\text{linear}}} \times \frac{\partial z_1}{\partial w_{11}}$$

=

$$\frac{\partial L}{\partial \hat{y}_1} = \partial(y-\hat{y}_1)^2 = 2(y-\hat{y}_1)(+1)$$

$$= +2(0.0566-8)$$

$$= -15.886.$$

$$\frac{\partial \hat{y}_1}{\partial z_{11}} = v_1 = 0.02. \qquad \frac{\partial z_{11}}{\partial w_{11}} = \frac{-1 e^{-a_{11}}}{(1+e^{-a_{11}})^2} * x_0.$$

$$= 0.249.$$

$$\frac{\partial L}{\partial w_{11}} = 0.02 \times 0.249 \times (-15.886)$$

$$= -0.079.$$

$$\frac{\partial L}{\partial w_{12}} = \frac{\partial L}{\partial \hat{y}_1} * \frac{\partial \hat{y}_1}{\partial z_{11}} * \frac{\partial z_{11}}{\partial w_{12}}$$

$$\frac{\partial L}{\partial \hat{y}_1} = -15.886 \qquad \frac{\partial \hat{y}_1}{\partial z_{11}} = 0.02 \quad \{\text{From prev. sub div}\}$$

$$\frac{\partial z_{11}}{\partial w_{12}} = \frac{0.913}{\text{Or } 3.663} * 1 = 0.249.$$

$$\frac{\partial L}{\partial w_{12}} = -15.886 * 0.02 * 0.249.$$

$$\frac{\partial L}{\partial w_{12}} = -0.079.$$

$$\frac{\partial L}{\partial w_{13}} = \frac{\partial L}{\partial \hat{y}_1} * \frac{\partial \hat{y}_1}{\partial z_{1b}} * \frac{\partial z_{1b}}{\partial w_{13}}$$

$$= -15.886 * 0.02 * 0.249 * 2$$

$$= -0.498$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial L}{\partial \hat{y}_2} * \frac{\partial \hat{y}_2}{\partial z_2} * \frac{\partial z_2}{\partial w_{21}}$$

$$\frac{\partial L}{\partial \hat{y}_2} = (y - \hat{y})^2 = 2(y - \hat{y}) \cdot (-1).$$

$$= 2(0.057 - 11)$$

$$= -21.886$$

$$\frac{\partial y_2}{\partial z_2} = V_2 = 0.03$$

$$\frac{\partial z_2}{\partial w_{21}} = \frac{e^{-a_2}}{(1+e^{-a_2})^2} * (-x_0)$$

$$= 0.249.$$

$$\frac{\partial L}{\partial w_{21}} = -21.886 * 0.03 * 0.249$$

$$= -0.1635$$

$$\frac{\partial L}{\partial w_{22}} = \frac{\partial L}{\partial \hat{y}_2} * \frac{\partial \hat{y}_2}{\partial z_2} * \frac{\partial z_2}{\partial w_{22}}$$

$$\frac{\partial L}{\partial \hat{y}_2} = -21.886 \qquad \frac{\partial \hat{y}_2}{\partial z_2} = 0.03.$$

$$\frac{\partial z_2}{\partial w_{22}} = 0.249.$$

$$\frac{\partial L}{\partial w_{22}} = 0.249 * 0.03 * (-21.886)$$

$$\frac{\partial L}{\partial w_{22}} = -0.1635$$

$$\frac{\partial L}{\partial w_{23}} = \frac{\partial L}{\partial \hat{y}_2} * \frac{\partial \hat{y}}{\partial z_2} * \frac{\partial z_2}{\partial w_{23}}$$

$$\frac{\partial L}{\partial \hat{y}_2} = -21.886 \qquad \frac{\partial \hat{y}_2}{\partial z_2} = 0.03.$$

$$\frac{\partial z_2}{\partial w_{23}} = 0.747.$$

$$\frac{\partial L}{\partial w_{23}} = -0.490.$$

$$\frac{\partial L}{\partial w_{31}} = \frac{\partial L}{\partial \hat{y}_3} * \frac{\partial \hat{y}_3}{\partial z_3} * \frac{\partial z_3}{\partial w_{31}}.$$

$$\frac{\partial L}{\partial \hat{y}_3} = -2(10 - 0.0571) = -19.885$$

$$\frac{\partial \hat{y}_3}{\partial z_3} = 0.04$$

$$\frac{\partial z_3}{\partial w_{31}} = \frac{-0.895}{3.594} * -1.$$

$$= 0.2490.$$

$$\frac{\partial L}{\partial w_{31}} = -19.885 * 0.04 * 0.249$$

$$= -0.198$$

$$\frac{\partial L}{\partial w_{32}} \quad \frac{\partial z_3}{\partial w_{32}} = \frac{-0.895}{3.594} * 2.$$

$$= 0.498$$

$$\frac{\partial L}{\partial w_{32}} = -19.885 * 0.04 * 0.498$$

$$= -0.396.$$

$$\frac{\partial L}{\partial w_{33}} = \frac{\partial L}{\partial \hat{y}_3} \times \frac{\partial \hat{y}_3}{\partial z_3} * \frac{\partial z_3}{\partial w_{33}}$$

$$= -19.885 * 0.04 * 0.498$$

$$= -0.396.$$

$\frac{\partial L}{\partial v_{10}}$ for $\hat{y_i}$

$$\frac{\partial L}{\partial v_{10}} = \frac{\partial L}{\partial \hat{y_i}} \cdot \frac{\partial \hat{y_i}}{\partial v_{10}}$$

$$= 2(\hat{y_i} - y) * 1$$

$$= -15.886$$

$$\frac{\partial L}{\partial v_{11}} = \frac{\partial L}{\partial \hat{y_i}} \cdot \frac{\partial \hat{y_i}}{\partial v_{11}}$$

$$= 2(\hat{y} - y) * z_{11}$$

$$= -15.886 * 0.522$$

$$= -8.292$$

$$\frac{\partial L}{\partial v_{12}} = \frac{\partial L}{\partial \hat{y_i}} \cdot \frac{\partial \hat{y_i}}{\partial v_{12}}$$

$$= 2(\hat{y} - y) * z_{12}$$

$$= -15.886 * 0.519 = -8.244$$

$$\frac{\partial L}{\partial v_{13}} = \frac{\partial L}{\partial \hat{y_i}} \cdot \frac{\partial \hat{y_i}}{\partial v_{13}} = 2(\hat{y} - y) * z_{13}$$

$$= -8.213$$

For $\hat{y}_2$

$$\frac{\partial L}{\partial v_{10}} = \frac{\partial L}{\partial \hat{y}_2} * \frac{\partial \hat{y}_2}{\partial v_{10}}$$

$$= 2(\hat{y}_2 - y) * 1$$

$$= -21.886$$

$$\frac{\partial L}{\partial v_{11}} = \frac{\partial L}{\partial \hat{y}_2} * \frac{\partial \hat{y}_2}{\partial v_{11}} = -21.886 * Z21(0.529)$$

$$= -11.57.$$

$$\frac{\partial L}{\partial v_{12}} = \frac{\partial L}{\partial \hat{y}_2} * \frac{\partial \hat{y}_2}{\partial v_{12}} = -21.886 * 0.524$$

$$= -11.468$$

$$\frac{\partial L}{\partial v_{13}} = \frac{\partial L}{\partial \hat{y}_2} * \frac{\partial \hat{y}_2}{\partial v_{13}} = -21.886 * 0.579$$

$$= -12.671.$$

For $\hat{y}_3$

$$\frac{\partial L}{\partial v_{10}} = \frac{\partial L}{\partial \hat{y}_3} * \frac{\partial \hat{y}_3}{\partial v_{10}}$$

$$= 2(\hat{y} - y) * 1$$

$$= -19.885$$

$$\frac{\partial L}{\partial v_{11}} = \frac{\partial L}{\partial \hat{y}_3} * \frac{\partial \hat{y}_3}{\partial v_{11}}$$

$$= -19.885 * 0.527 \ (z_{31})$$

$$= -10.479$$

$$\frac{\partial L}{\partial v_{12}} = \frac{\partial L}{\partial \hat{y}_3} * \frac{\partial \hat{y}_3}{\partial v_{12}}$$

$$= -19.885 * 0.522$$

$$= -10.379$$

$$\frac{\partial L}{\partial v_{13}} = \frac{\partial L}{\partial \hat{y}_3} * \frac{\partial \hat{y}_3}{\partial v_{13}}$$

$$= -19.885 * 0.524$$

$$= -10.419$$

(d) $$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}_1} \times \frac{\partial \hat{y}_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_1}$$

Obtaining values from prev sub division

$$\frac{\partial L}{\partial w_1} = \begin{bmatrix} -0.0792 \\ -0.0792 \\ -0.1585 \end{bmatrix}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}_2} \times \frac{\partial \hat{y}_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_2}$$

Obtaining values from prev. sub. div

$$\frac{\partial L}{\partial w_2} = \begin{bmatrix} -0.1635 \\ -0.1635 \\ -0.490 \end{bmatrix}$$

$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial \hat{y}_3} \times \frac{\partial \hat{y}_3}{\partial z_3} \times \frac{\partial z_3}{\partial w_3}$$

Obtaining values from prev. sub. div.

$$\frac{\partial L}{\partial w_3} = \begin{bmatrix} -0.198 \\ -0.396 \\ -0.396 \end{bmatrix}$$

2) $f(x, y) = (2x + 3y)^2$

~~$\nabla f(x,y) = 2(2x+3y)$~~.

$\nabla f(x, y) \Rightarrow \frac{\partial f(x,y)}{\partial x} = 2(2x+3y) \cdot 2$.

$\frac{\partial f(x,y)}{\partial y} = 2(2x+3y)3$.

b) $F(x,y) = \begin{bmatrix} x^2+2y \\ 3x+4y^2 \end{bmatrix}$

$$DF(x,y) = \begin{bmatrix} \frac{\partial F_1(x,y)}{\partial x} & \frac{\partial F_1(x,y)}{\partial y} \\ \frac{\partial F_2(x,y)}{\partial x} & \frac{\partial F_2(x,y)}{\partial y} \end{bmatrix}$$

$$= \begin{bmatrix} 2x & 2 \\ 3 & 8y \end{bmatrix}$$

$DF(1,2) \Rightarrow x=1,\ y=2$

$$\Rightarrow \begin{bmatrix} 2(1) & 2 \\ 3 & 8(2) \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 3 & 16 \end{bmatrix}$$

c) $G(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad F(x,y) = \begin{bmatrix} x^2+2y \\ 3x+4y^2 \end{bmatrix}$

$D(F \cdot G)(2)$ with chain rule!

$$D(F \circ G)(x) = DF(G(x)) \cdot D(G(x))$$

$$D(G(x)) = \begin{bmatrix} \partial x/\partial x \\ \partial x^2/\partial x \end{bmatrix} = \begin{bmatrix} 1 \\ 2x \end{bmatrix}$$

$$DF(G(x)) = \begin{matrix} 1 \\ 3 \end{matrix} \begin{bmatrix} 1^2 + 2y \\ 6x^2 + 4y^2 \end{bmatrix}$$

$$D(F \cdot G)x = \begin{bmatrix} 1 \\ 12x \end{bmatrix}$$

$$D(F \cdot G)(2) = \begin{bmatrix} 1 \\ 24 \end{bmatrix}$$

Without chain rule:

$$D(F \circ G)(x) = DF(G(x))$$

So

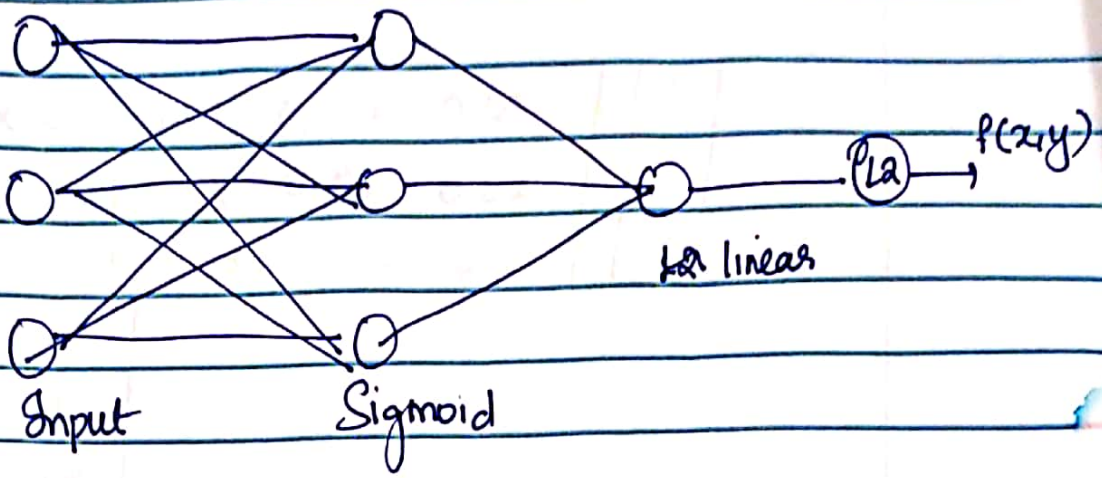
$$(F \circ G)(x) = \begin{bmatrix} x^2 + 2x^2 \\ 3x + 4(x^2)^2 \end{bmatrix}$$

$$D(F \circ G)x = \begin{bmatrix} \frac{\partial(x^2 + 2x^2)}{\partial x} \\ \frac{\partial(3x + 4x^4)}{\partial x} \end{bmatrix} = \begin{bmatrix} 2x + 4x \\ 3 + 16x^3 \end{bmatrix}$$

$$= \begin{bmatrix} 6x \\ 16x^3 + 3 \end{bmatrix}$$

$$D(F \circ G)(2) = \begin{bmatrix} 6 \times 2 \\ 16(8) + 3 \end{bmatrix} = \begin{bmatrix} 12 \\ 131 \end{bmatrix}$$

d)



x_1

$a = W^T \cdot x$

$$= \begin{bmatrix} 0.01 & 0.02 & 0.03 \\ 0.03 & 0.01 & 0.02 \\ 0.02 & 0.03 & 0.01 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.09 \\ 0.08 \\ 0.07 \end{bmatrix}$$

$$z = \sigma(a) = \sigma \begin{bmatrix} 0.09 \\ 0.08 \\ 0.07 \end{bmatrix} = \begin{bmatrix} 0.5224 \\ 0.5199 \\ 0.5174 \end{bmatrix}$$

$$\hat{y} = V^T \cdot z = \begin{bmatrix} 0.01 & 0.02 & 0.03 & 0.04 \end{bmatrix} \begin{bmatrix} 1 \\ 0.5224 \\ 0.5199 \\ 0.5174 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0561 \end{bmatrix}$$

L2 loss $= \hat{y} - y$

$= (0.0567 - 8)^2$

$= 63.095$

x_2

$$a = w^T x = \begin{bmatrix} 0.01 & 0.02 & 0.03 \\ 0.03 & 0.01 & 0.02 \\ 0.02 & 0.03 & 0.01 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 0.12 \\ 0.1 \\ 0.08 \end{bmatrix}$$

$$z = \sigma(a) = \begin{bmatrix} 0.522 \\ 0.519 \\ 0.517 \end{bmatrix}$$

$$\hat{y} = v^T z$$

$$= \begin{bmatrix} 0.01 & 0.02 & 0.03 & 0.04 \end{bmatrix} \begin{bmatrix} 1 \\ 0.522 \\ 0.519 \\ 0.517 \end{bmatrix}$$

$$\hat{y} = \begin{bmatrix} 0.0570 \end{bmatrix}$$

$$\text{L2 loss} = (\hat{y} - y)^2 = (0.0570 - 11)^2$$

$$119.74$$

x_3

$$a = w^T x = \begin{bmatrix} 0.01 & 0.02 & 0.03 \\ 0.03 & 0.01 & 0.02 \\ 0.02 & 0.03 & 0.01 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.11 \\ 0.09 \\ 0.1 \end{bmatrix}$$

$$z = \sigma(a) = \begin{bmatrix} 0.527 \\ 0.522 \\ 0.524 \end{bmatrix}$$

$\hat{y} = 0.0571$ L2 loss $= (\hat{y} - y)^2 = 98.86$

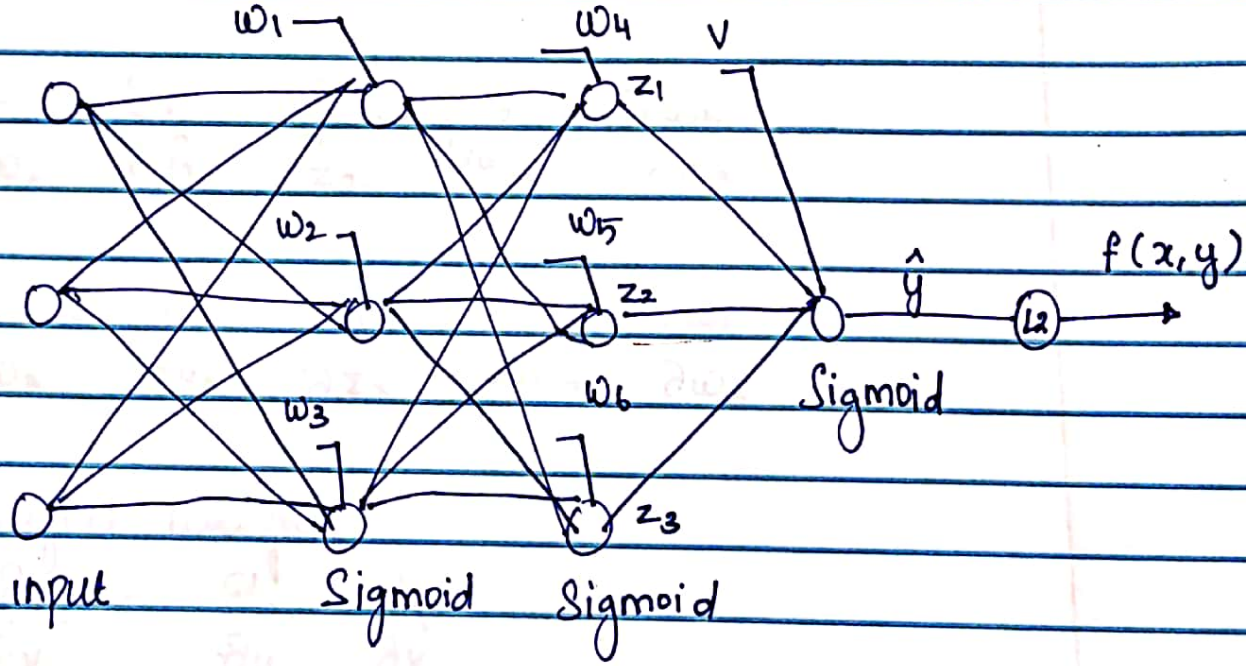
Backward Prop:-

$$\frac{\partial L}{\partial w_1} = \begin{bmatrix} -0.079 \\ -0.079 \\ -0.498 \end{bmatrix}$$ {with respect to prev. answers.

$$\frac{\partial L}{\partial w_2} = \begin{bmatrix} -0.1635 \\ -0.1635 \\ -0.490 \end{bmatrix}$$

$$\frac{\partial L}{\partial w_3} = \begin{bmatrix} -0.198 \\ -0.396 \\ -0.396 \end{bmatrix}$$

(6 3)



Input vector: $x^i \in R^i$

$y^i \in \{0, 1\}$

Gradients of Backpropagation:

$$\frac{\partial L}{\partial w_4} = \frac{\partial L}{\partial \hat{y}_1} \times \frac{\partial \hat{y}_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_4}$$

$$\frac{\partial L}{\partial w_5} = \frac{\partial L}{\partial \hat{y}_2} \times \frac{\partial \hat{y}_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_5}$$

$$\frac{\partial L}{\partial w_6} = \frac{\partial L}{\partial \hat{y}_3} \times \frac{\partial \hat{y}_3}{\partial z_3} \times \frac{\partial z_3}{\partial w_6}$$

Gradients for prev. layers:

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}_1} \times \frac{\partial \hat{y}_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_4} \times \frac{\partial w_4}{\partial w_1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}_2} \times \frac{\partial \hat{y}_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_5} \times \frac{\partial w_5}{\partial w_2}$$

$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial \hat{y}_3} \times \frac{\partial \hat{y}_3}{\partial z_3} \times \frac{\partial z_3}{\partial w_6} \times \frac{\partial w_6}{\partial w_3}$$

Output weights:

$$\frac{\partial L}{\partial v} = \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial v}$$