PAGE No
DATE / /

Week-8 12^{th} October Monday

Theory Task

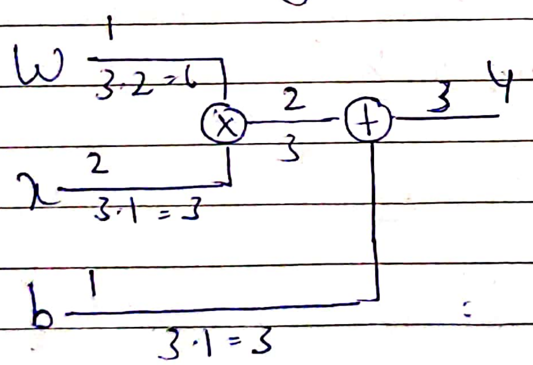
Q. Generate Computation graphs for Linear regression and logistics Regression and compute gradients

Sol:- Consider Linear Regression model

$f(x) = Y = Wx + b = \cancel{f(x,w)}$

Y = predicted value
x = input vector
W = weights to be learned
b = bias

Computation graph   Let $(w, x, b) = (1, 2, 1)$



$\frac{df}{df} = 1$ $\frac{df}{db} = 3.1 = 3$

$\frac{df}{dw} = 6$ $\frac{df}{dx} = 3$

Consider logistics regression model.

$f(x,y,z) = w_0 x + w_1 y + w_2 z$

(x,y,z) = input vector
(w_0, w_1, w_2) = weights to be learned.

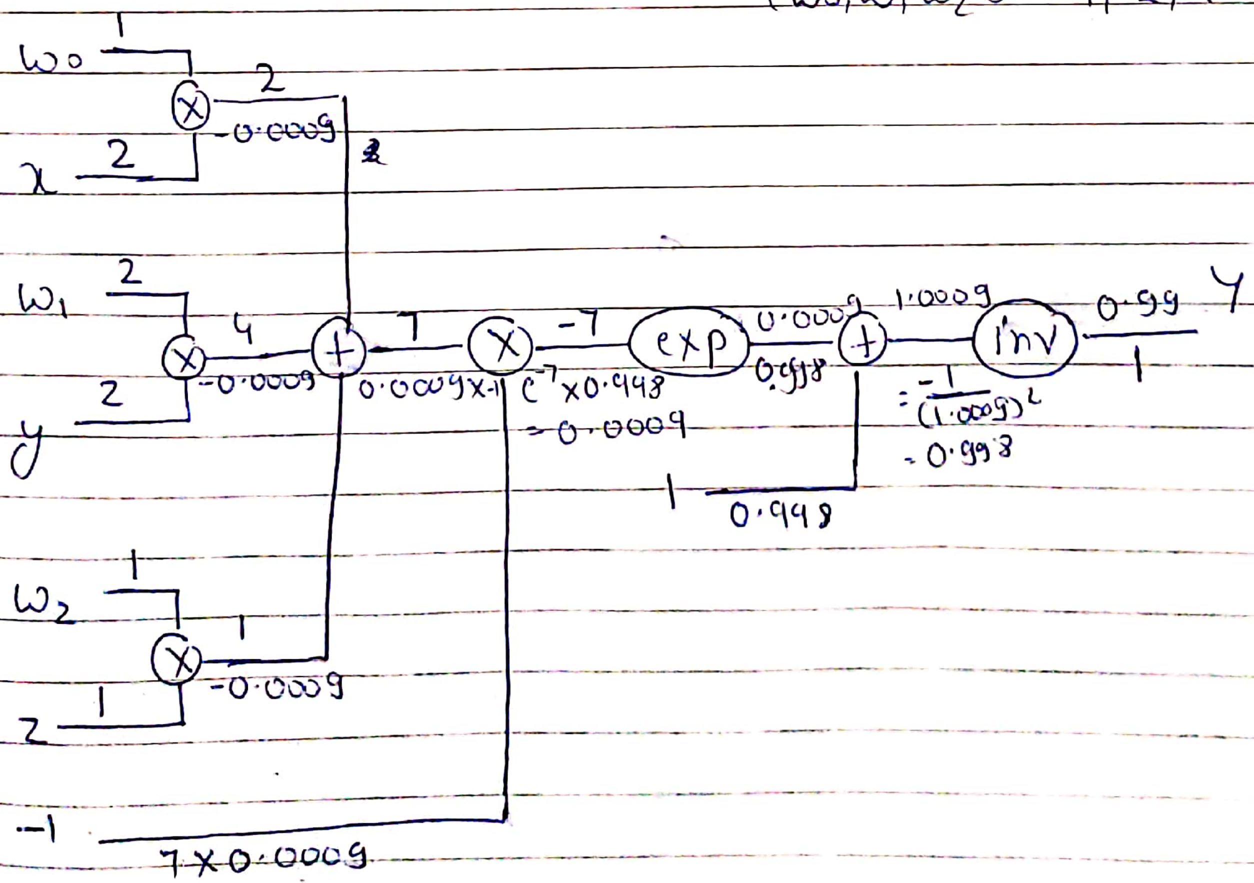
Using sigmoid activation function at output node.

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

$z = f(x,y,z) = w_0 x + w_1 y + w_2 z$

Computation Graph

Let $(x,y,z) = 2, 2, 1$
$(w_0, w_1, w_2) = 1, 2, 1$



$\frac{\partial f}{\partial f} = 1$

$\frac{\partial f}{\partial x} = -0.0009$

$\frac{\partial f}{\partial w_0} = 2x - 0.0009 = -0.0018$

$\frac{\partial f}{\partial y} = -0.0018$

$\frac{\partial f}{\partial w_1} = -0.0018$

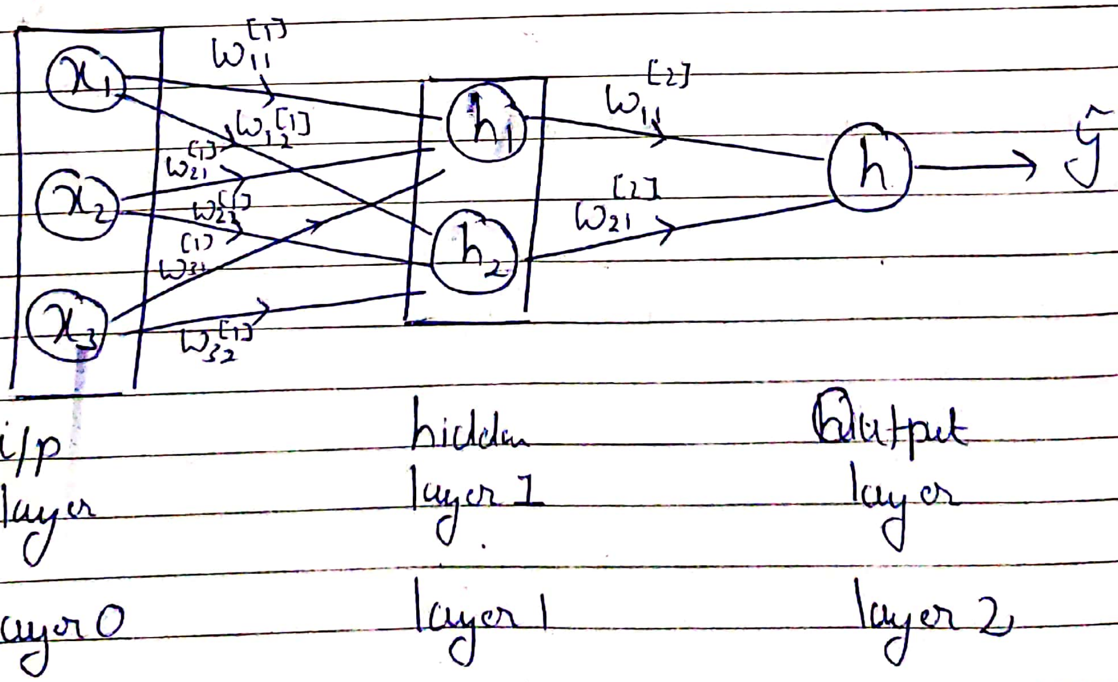
$\frac{\partial f}{\partial z} = -0.0009$

$\frac{\partial f}{\partial w_2} = -0.0009$

Week 8 Wednesday 15th October 2020

Q. Given a neural network, perform feed forward and back propogation. Compute gradient $\frac{\partial L}{\partial W^{[2]}}$ & $\frac{\partial L}{\partial W^{[1]}}$. Here L is loss function.

Sol:-



PAGE No.
DATE / /

Assumption :- Here I am using MSE (Mean Squared Error) function

$$E = \frac{1}{2N} \sum_i (y - \hat{y}_i)^2$$

Here N = 1, because we have only 1 node at output.

$$E = \frac{1}{2} \sum (y - \hat{y})^2$$

y = actual | True label

$\hat{y}$ = predicted value

Notation Meaning

$w_{ij}^{[k]}$ = means weight connecting unit i to unit j at layer k.

input vector = $(x_1, x_2, x_3)^T$

$W^{[1]}$ = weight matrix for layer 1

$$W^{[1]} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix}$$

Weight matrix for layer 2

$$W^{[2]} = \begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix}$$

Feed forward Pass

$$h_1 = w_{11}^{[1]} x_1 + w_{21}^{[1]} x_2 + w_{31}^{[1]} x_3$$

$$h_2 = w_{12}^{[1]} x_1 + w_{22}^{[1]} x_2 + w_{32}^{[1]} x_3$$

$$[h_1 \ h_2] = [x_1 \ x_2 \ x_3] \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix}$$

$$h = h_1 w_{11}^{[2]} + h_2 w_{21}^{[2]}.$$

$$\hat{y} = h$$

$$\text{Error } (E) = \frac{1}{2}(y - \hat{y})^2$$

$$= \frac{1}{2}\left(y - \left(h_1 w_{11}^{[2]} + h_2 w_{21}^{[2]}\right)\right)^2$$

1) $\dfrac{\partial E}{\partial w^{[2]}} = ?$

$$\frac{\partial E}{\partial W^{[2]}} = \frac{\partial E}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial W_2^{[2]}}$$ ((Chain rule))

PAGE No.
DATE / /

$$\frac{\partial E}{\partial \hat{y}} = -(y-\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w^{[2]}} = h_i \quad i=1,2.$$

$$\frac{\partial E}{\partial w^{[2]}} = -(y-\hat{y})\, h_i$$

2) $\frac{\partial E}{\partial w^{[1]}} = ?$

$$\frac{\partial E}{\partial w^{[1]}} = \frac{\partial E}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial h_i} \times \frac{\partial h_i}{\partial w^{[1]}} \qquad \text{(Chain rule)}$$

$$\frac{\partial \hat{y}}{\partial h_i} = w_{ij}^{[k]} \qquad i=1,2 \quad j=1 \quad k=2$$

$$\frac{\partial h_i}{\partial w^{[1]}} = x_i \qquad i=1,2,3.$$

$$\therefore \frac{\partial E}{\partial w^{[1]}} = -(y-\hat{y})\, w_{ij}^{[k]}\, x_i$$