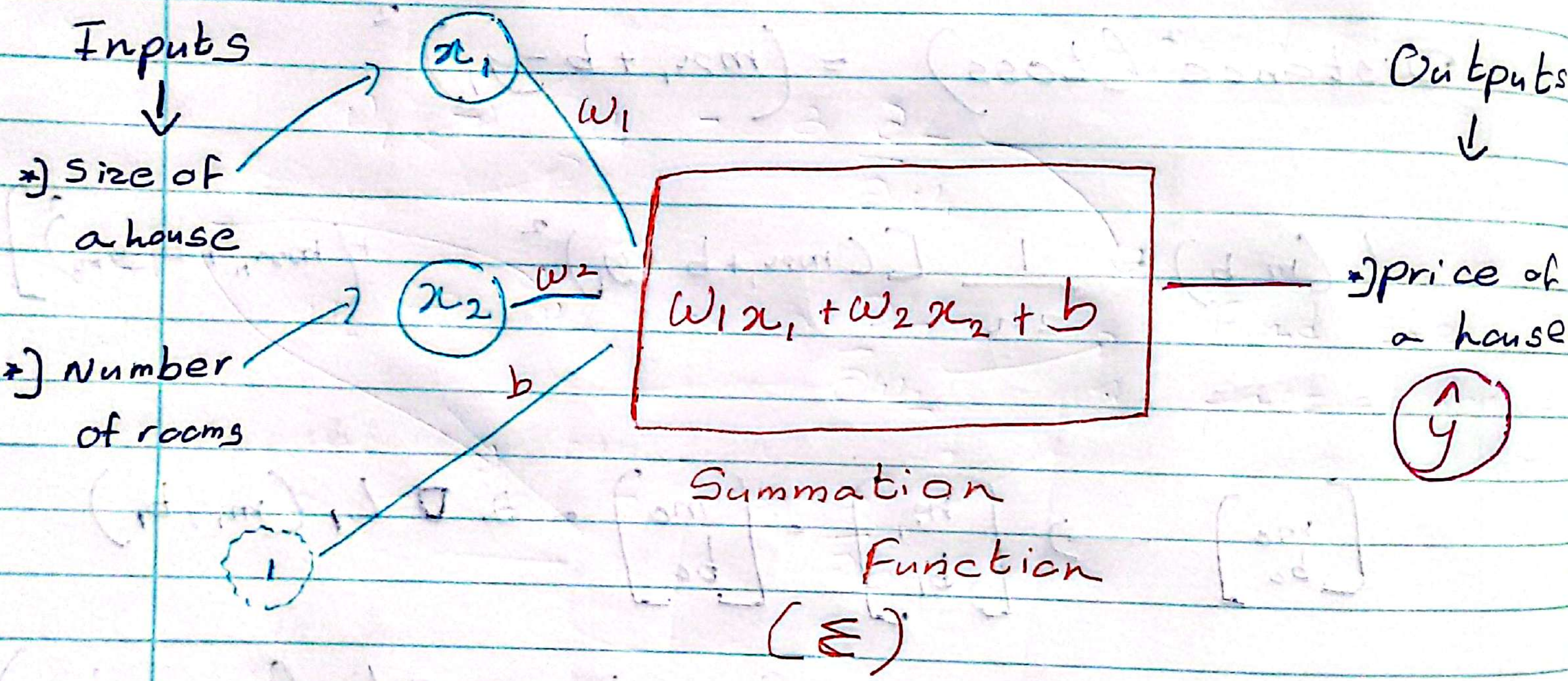
Week 03

1] Regression with a preceptron

Single Layer Neural Network Perceptron



$$\hat{y} = w_1x_1 + w_2x_2 + b$$

Main Goal:

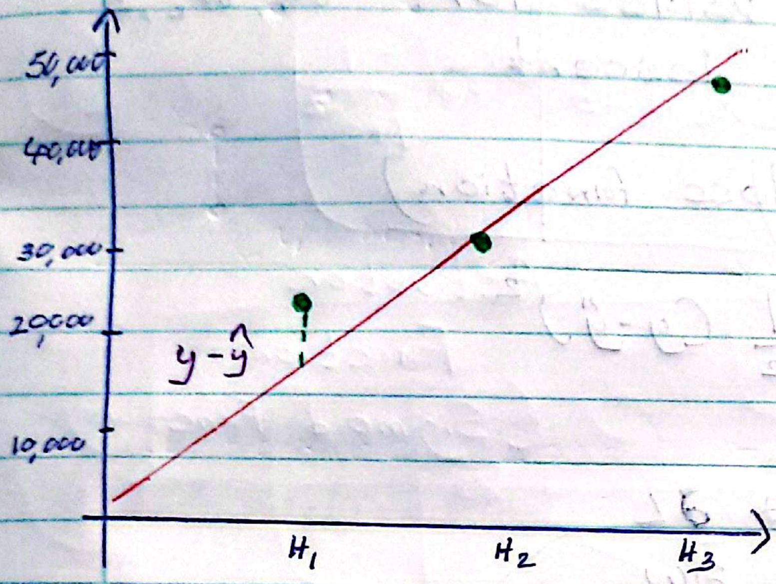
Find weights and bias that will optimize the predictions

(Reduce the errors in the predictions)

$\Downarrow$

The Loss Function

2) The Loss Function (Linear Regression)



	y	$\hat{y}$	$\frac{1}{2}(y-\hat{y})$
$H_1 \rightarrow$	20k $\rightarrow$	15k $\rightarrow$	Error
$H_2 \rightarrow$	30k $\rightarrow$	30k $\rightarrow$	Error
$H_3 \rightarrow$	50k $\rightarrow$	45k $\rightarrow$	Error

Mean Squared Error

*) prediction Function:

$$\hat{y} = w_1 x_1 + w_2 x_2 + b$$

*) Loss Function:

$$L(y, \hat{y}) = \frac{1}{2}(y-\hat{y})^2$$

*) Main Goal:

Find w_1, w_2, b that give $\hat{y}$ with the least error

3] Gradient Descent (Linear Regression)

To find optimal values for: w_1, w_2, b
we need gradient descent

(minimize the loss function)

$$L_{(y,\hat{y})} = \frac{1}{2}(y-\hat{y})^2$$

Some
Initial
Values

$$w_1 \rightarrow w_1 - \alpha \frac{\partial L}{\partial w_1}$$

$$w_2 \rightarrow w_2 - \alpha \frac{\partial L}{\partial w_2}$$

Find the partial derivatives

$$b \rightarrow b - \alpha \frac{\partial L}{\partial b}$$

$$\hat{y} = w_1 x_1 + w_2 x_2 + b$$

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

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = -(y-\hat{y})$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = -(y-\hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = -(y-\hat{y})x_2$$

$$w_1 \rightarrow w_1 - a\left(-x_1\left(y-\hat{y}\right)\right)$$

$$w_2 \rightarrow w_2 - a\left(-x_2\left(y-\hat{y}\right)\right)$$

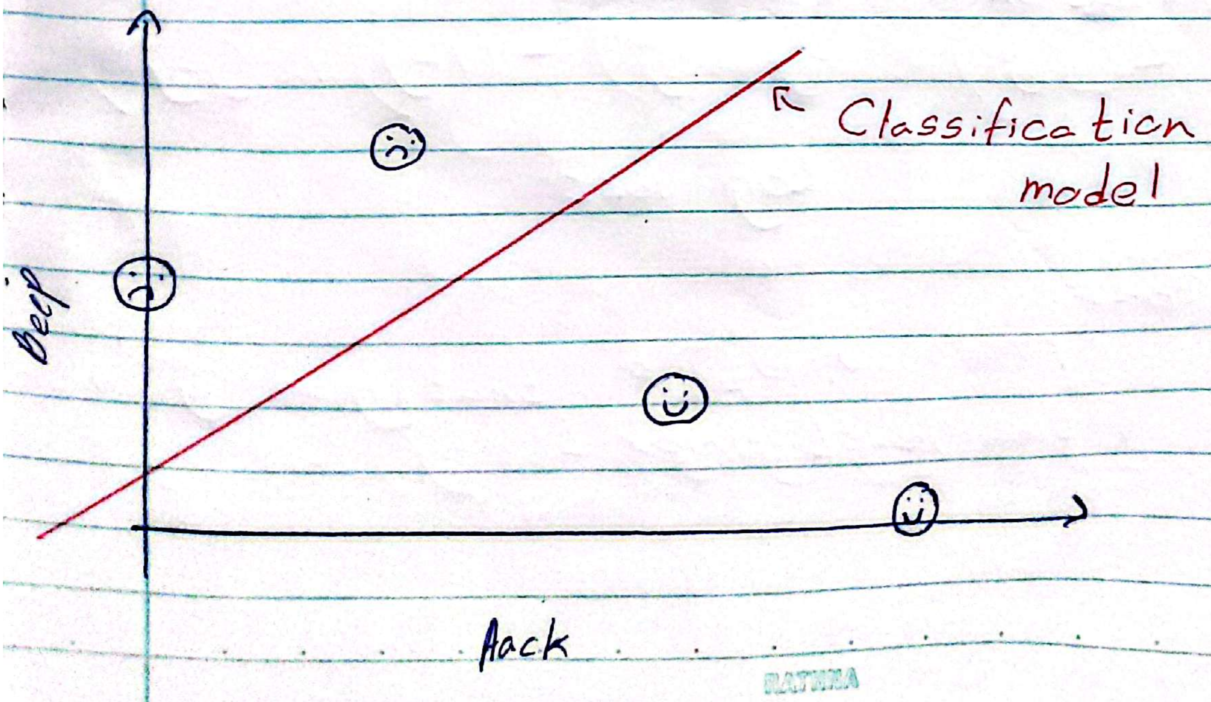
$$b \rightarrow b - a\left(-\left(y-\hat{y}\right)\right)$$

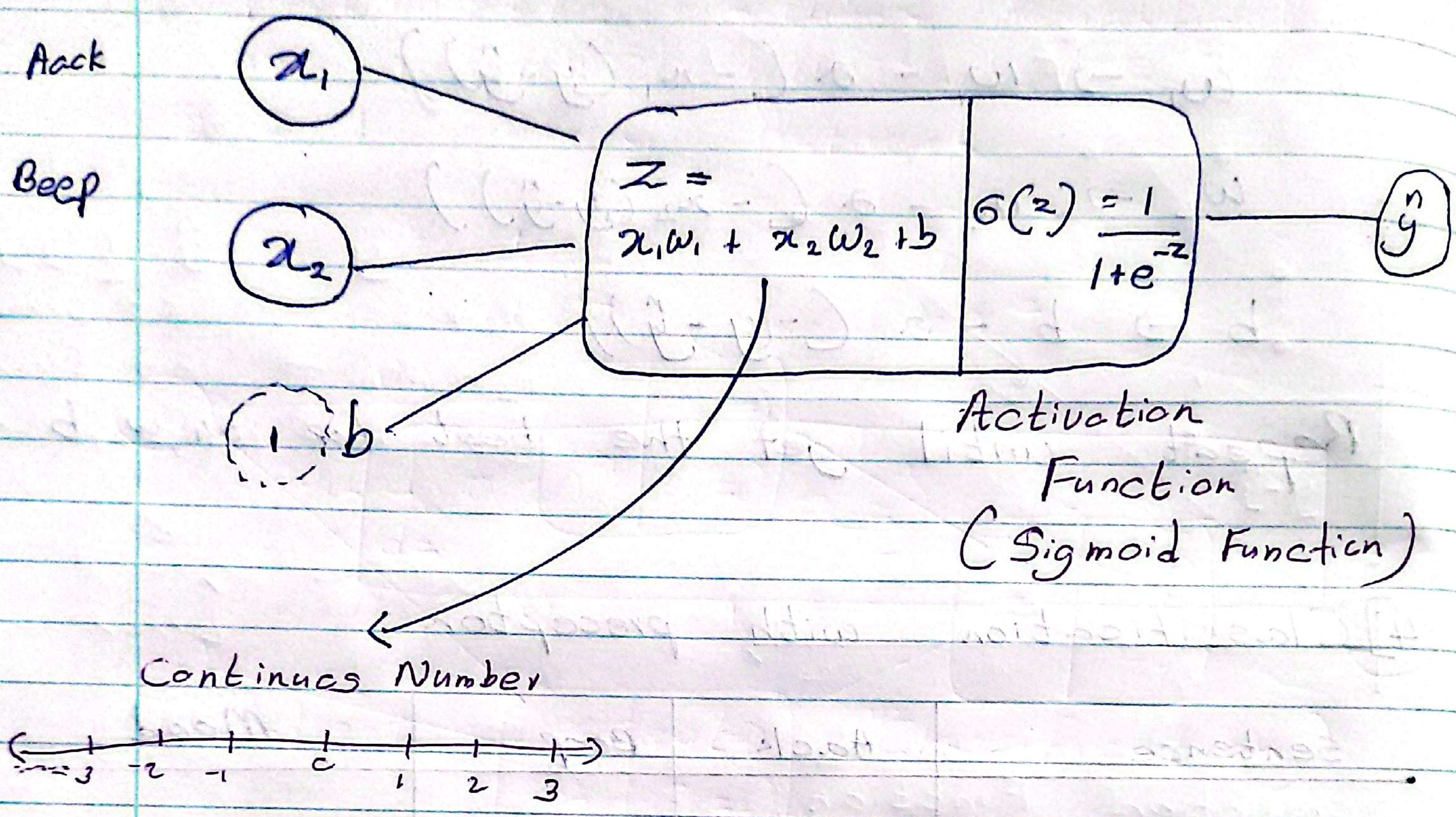
Repeat until get the best for w_1, w_2, b

4] Classification with precepton

Sentence	Aack	Beep	Mood
Aack aack aack!	3	0	Happy ☺
Bep Beep!	0	2	Sad ☹
Aack beep beep beep!	1	3	Sad ☹
Aack beep aack	2	1	Happy ☺

$\underbrace{\hspace{6cm}}_{\text{Dataset}}$

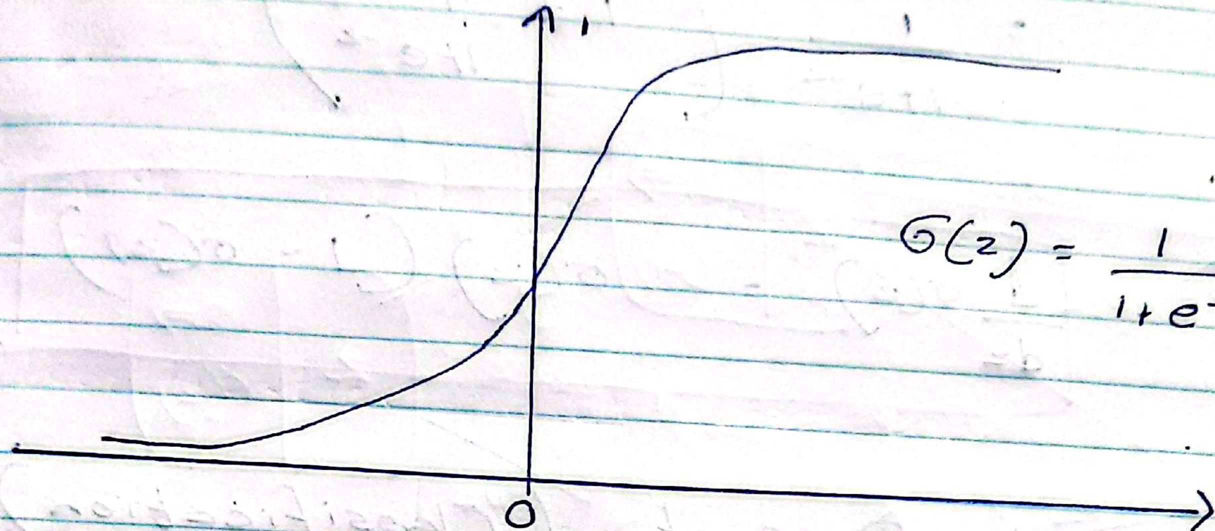




*) Activation function used to convert the all numbers ~~t~~ in the line to 0 - 1

5] Sigmoid function

Input is the number line, output is the interval 0 - 1



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

Derivative of a Sigmoid function

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

$$\frac{d}{dz}\sigma(z) = -1(1+e^{-z})^{-1-1}\left(\frac{d}{dz}(1) + \frac{d}{dz}(e^{-z})\right)$$

$$= -1(1+e^{-z})^{-2}\left(0 + e^{-z}\left(\frac{d}{dz}(-z)\right)\right)$$

$$= (1+e^{-z})^{-2}\, e^{-z}$$

$$= \frac{1}{(1+e^{-z})^{2}} = \frac{e^{-z}+1-1}{(1+e^{-z})^{2}}$$

$$= \frac{1+e^{-z}-1}{(1+e^{-z})^{2}}$$

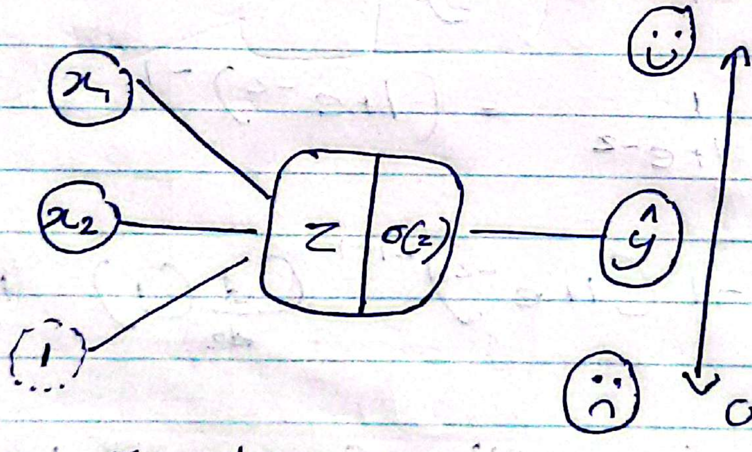
$$= \frac{1}{(1+e^{-z})} - \frac{1}{\ldots}$$

$$\frac{d}{dz}\sigma(z) = \frac{1}{(1+e^{-z})} - \left(\frac{1}{(1+e^{-z})}\right)\left(\frac{1}{1+e^{-z}}\right)$$

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

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

6] Gradient Descent (Classification)



Prediction Function:

$$\hat{y} = \sigma(w_1x_1 + w_2x_2 + b)$$

Loss Function :

$$L(y,\hat{y}) = -y\ln(\hat{y}) - (1-y)\ln(1-\hat{y})$$

Log loss

Main Goal :

Find w_1, w_2, b that give $\hat{y}$ with the least error

To find optimal values for : w_1, w_2, b
we need gradient descent

Initial starting point

$$w_1 \rightarrow w_1 - \alpha \frac{\partial L}{\partial w_1}$$

$$w_2 \rightarrow w_2 - \alpha \frac{\partial L}{\partial w_2}$$

$$b \rightarrow b - \alpha \frac{\partial L}{\partial b}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} = \cancel{\frac{-y-\hat{y}}{\hat{y}(1-\hat{y})}} \; \hat{y}(1-\hat{y})x_1$$

$$-(y-\hat{y})x_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} = -(y-\hat{y})x_2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = -(y-\hat{y})$$

$$w_1 \rightarrow w_1 - \alpha(-x_1(y-\hat{y}))$$

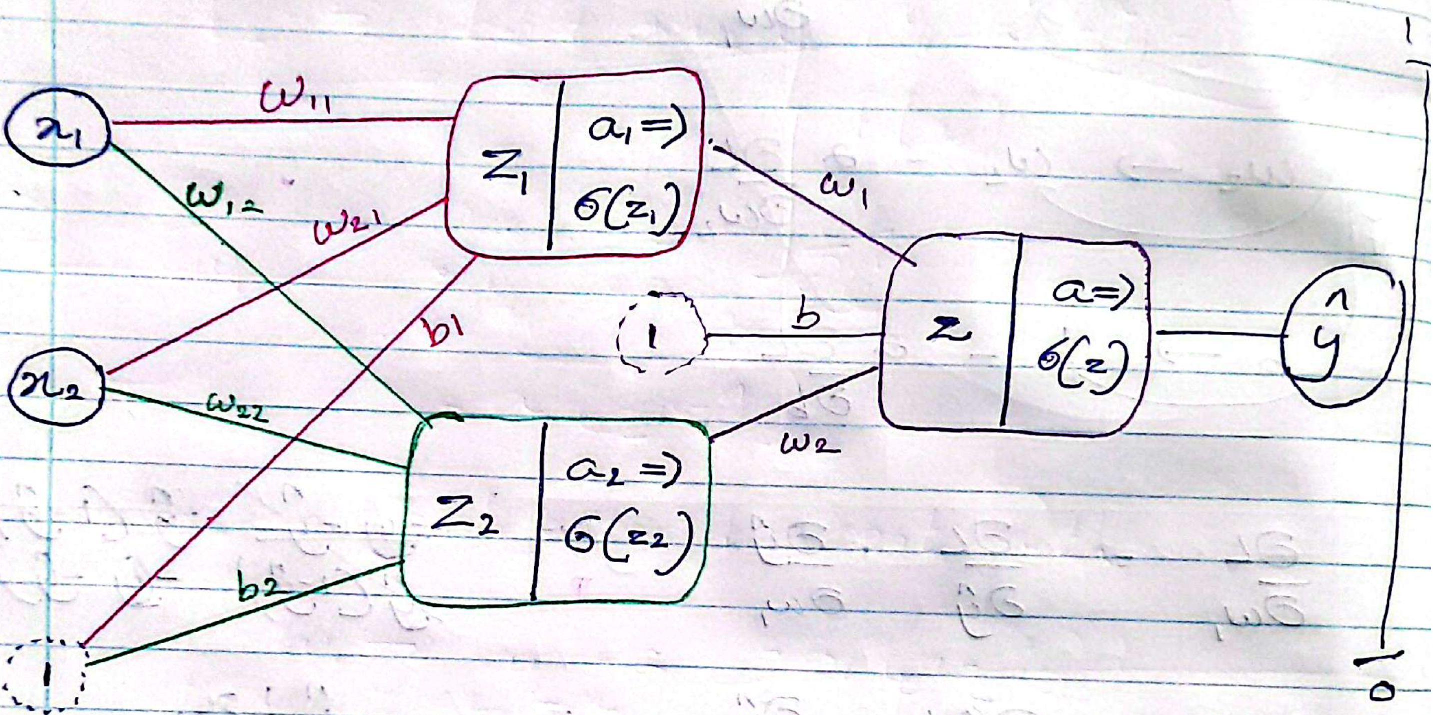
$$w_2 \rightarrow w_2 - \alpha(-x_2(y-\hat{y}))$$

$$b \rightarrow b - \alpha(-(y-\hat{y}))$$

*) Repeat this step until you get a preetty good model.

7] Classification with a Neural network

2, 2, 1 Neural Network



*] $a_1 = \sigma(z_1)$

$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$

*] $a_2 = \sigma(z_2)$

$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$

*] $\hat{y} = \sigma(z)$

$z = a_1 w_1 + a_2 w_2 + b$

$$L(y, \hat{y}) = -y \log(\hat{y}) - (1-y) \log(1-\hat{y})$$

Minimizing log - loss

$$z_1 = x_1 \omega_{11} + x_2 \omega_{21} + b_1 \quad \} \quad a_1 = \sigma(z_1)$$

$$z_2 = x_1 \omega_{12} + x_2 \omega_{22} + b_2 \quad \} \quad a_2 = \sigma(z_2)$$

$$z = a_1 \omega_1 + a_2 \omega_2 + b \quad \} \quad \hat{y} = \sigma(z)$$

*) $$\frac{\partial L}{\partial \omega_{11}} = \frac{\partial z_1}{\partial \omega_{11}} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$= -x_1 \omega_1 a_1 (1-a_1)(y-\hat{y})$$

perform gradient descent with,

$$\omega_{11} \rightarrow \omega_{11} - a \frac{\partial L}{\partial \omega_{11}}$$ to find optimal value of ω_{11}

*) $$\frac{\partial L}{\partial b_1} = \frac{\partial z_1}{\partial b_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z}{\partial a_1} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial L}{\partial \hat{y}}$$

$$= -\omega_1 a_1 (1-a_1)(y-\hat{y})$$

perform gradient descent with

$$b_1 \rightarrow b_1 - a(-\omega_1 a_1 (1-a_1)(y-\hat{y}))$$

to find the optimal value of b_1

*) Simillarly ω_{22}

Date

$$\boxed{\begin{aligned} w_{21} &\rightarrow w_{21} + \alpha x_1 w_2 a_2(1-a_2)(y-\hat{y}) \\ w_{22} &\rightarrow w_{22} + \alpha x_2 w_2 a_2(1-a_2)(y-\hat{y}) \\ b_2 &\rightarrow b_2 + \alpha w_2 a_2(1-a_2)(y-\hat{y}) \end{aligned}}$$

*) $\dfrac{\partial L}{\partial w_1} = \dfrac{\partial z}{\partial w_1} \cdot \dfrac{\partial \hat{y}}{\partial z} \cdot \dfrac{\partial L}{\partial \hat{y}}$

$= -a_1(y-\hat{y})$

perform gradient descent with,

$w_1 \rightarrow w_1 - \alpha(-a_1(y-\hat{y}))$

to find optimal value of w_1

*) Same to w_2, b

$w_1 \rightarrow w_1 + \alpha a_1(y-\hat{y}))$

$w_2 \rightarrow w_2 + \alpha a_2(y-\hat{y}))$

$b \rightarrow b + \alpha(y-\hat{y}))$