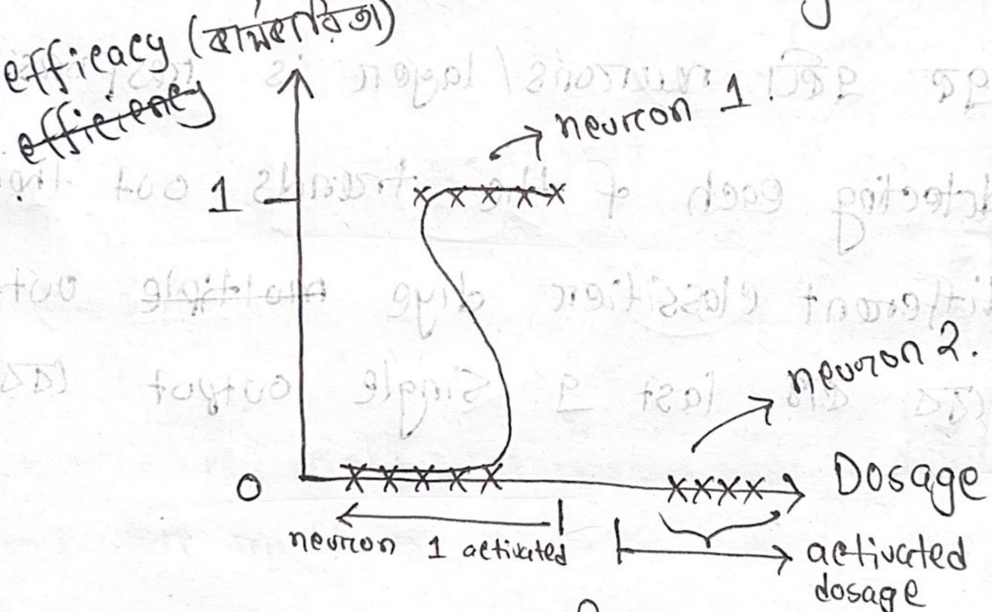
Neural Networks

neural network এর প্রত্যেকটা neuron একটা logistic regresson হিসেবে কাজ করতে পারে। Each neuron is a regression model.



medicine dosage কম হলে efficiency কম। আবার বেশি Dosage হলে উল্টা কাজ হয়।

Napa, 500 mg এর নিচে ~~খেলে~~ খেলে, efficacy 0. যদি, 500-3000 mg এর মধ্যে খাই তাহলে efficacy 1. Again, 3000 mg এর উপরে খেলে efficacy তখন 0.

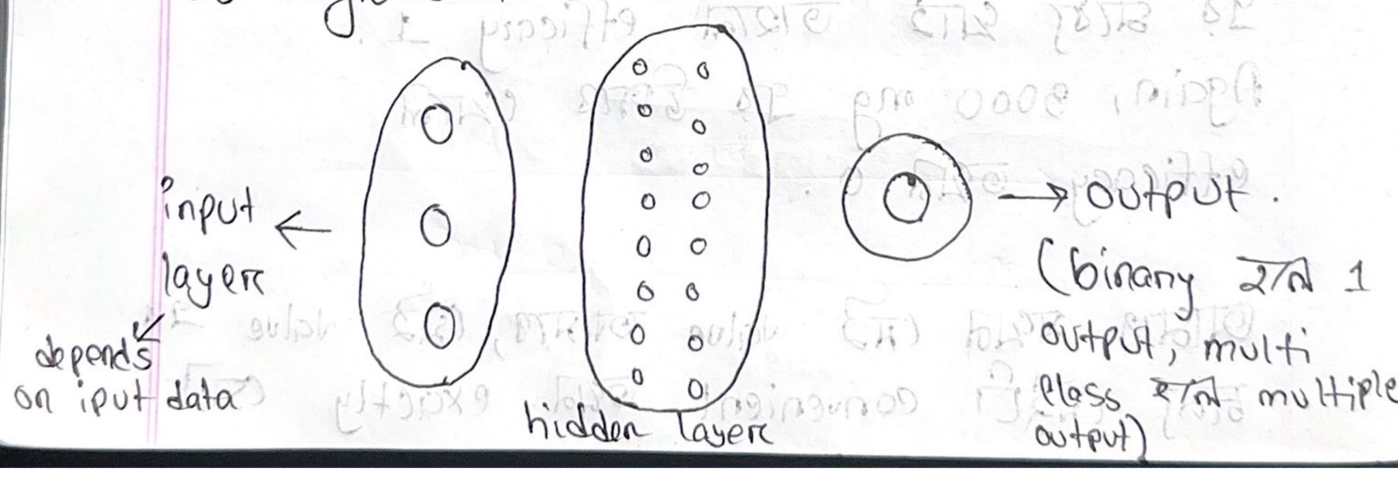
আমার যখন যেই value আসবে, সেই value এর জন্য যেইটা convenient তখন exactly সেই

neuron অথবা সেই layer টা activate হয়না। So, prediction accordingly দিতে পারে।

Main idea এটা।

multiple classifier কে একসাথে জোড়া দেওয়া but এক একটা neurons / layer is responsible for detecting each of the trends out there. So, different classifier diye ~~multiple~~ output ~~বে~~ বের করে last এ single output বের করি।

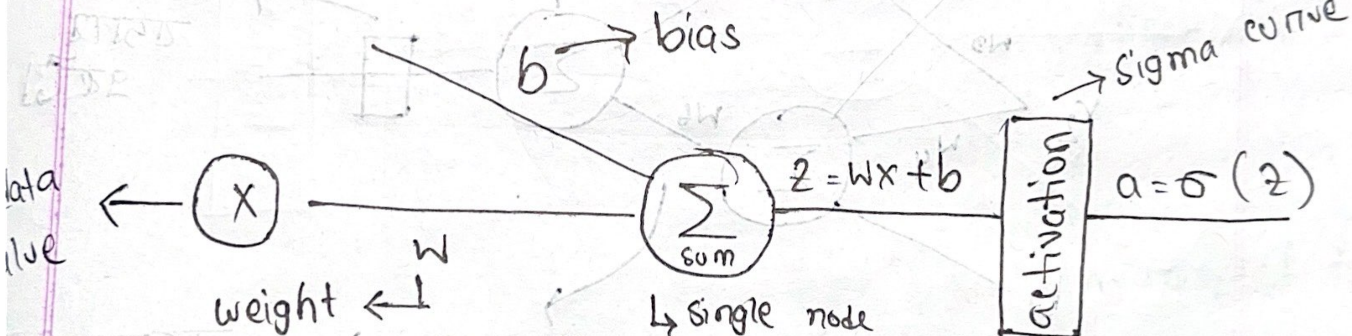
It can work like classification and as well as regression model.



Logistic regression to neural networks

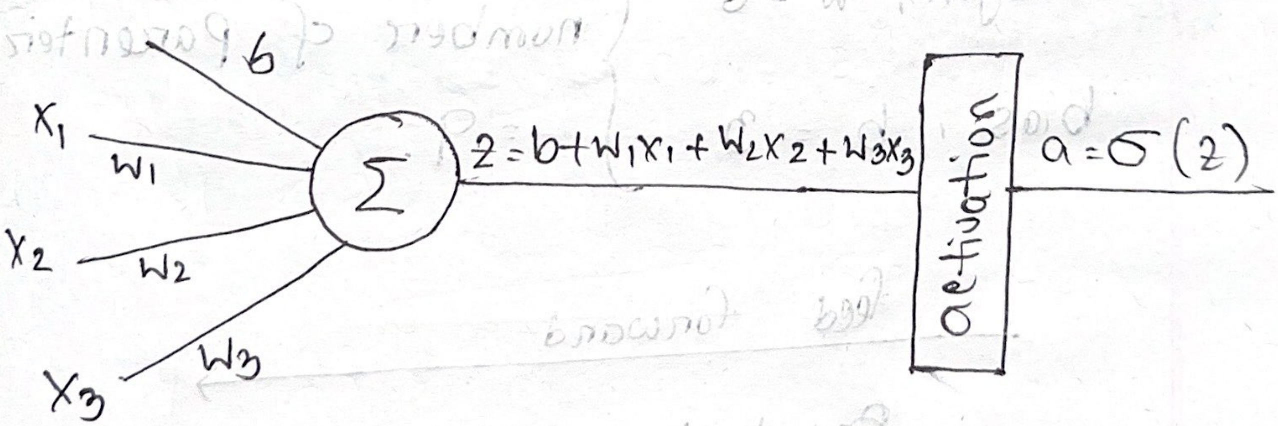
→ logistic regression can be expressed as a single neuron neural network.

→ where, for a single feature and a single neuron, the network looks like:



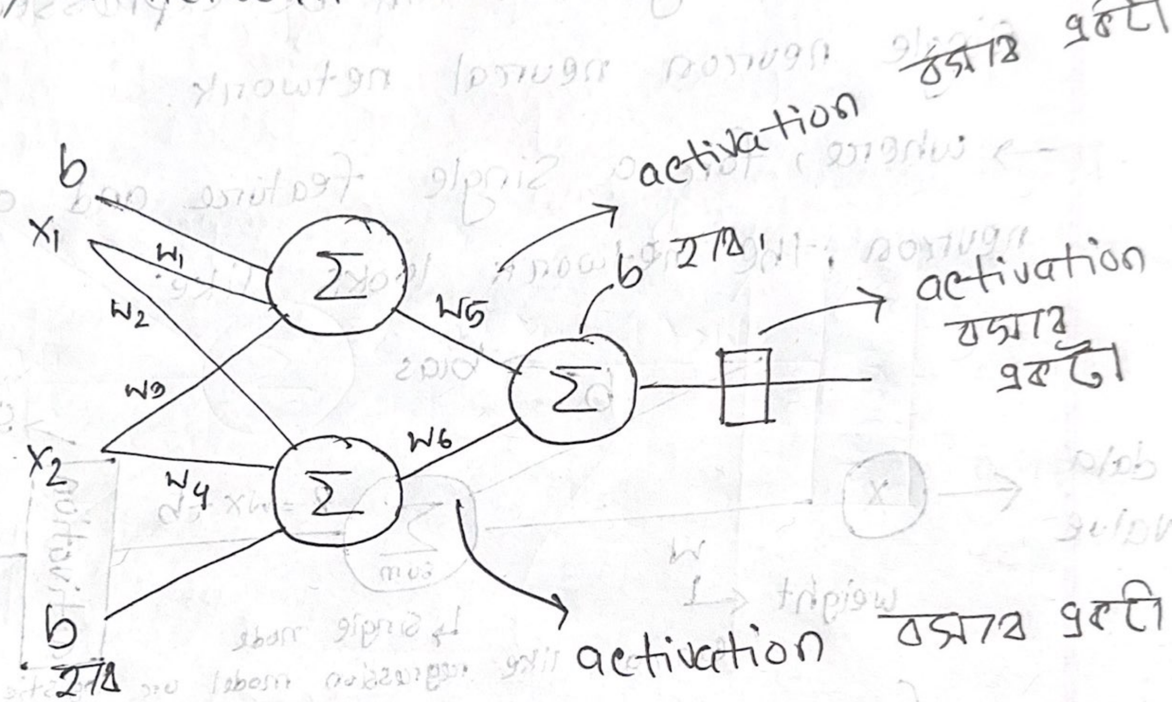
can act like regression model or logistic regression classifier

→ for multiple features on a single neuron,



→ Single layer neural networks are referred to

→ a network can involve multiple neurons



weight, $w = 6$
bias, $b = 3$
} number of parameters, $= 9$.

feed forward →

~~feed~~ backward propagation ←

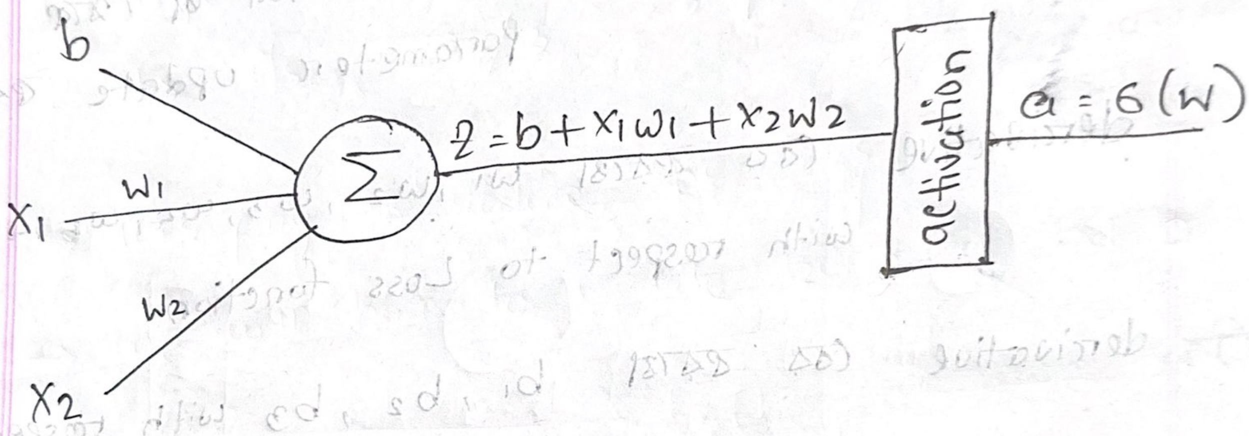
feed forward → starts with assumptions. then last এ একটা loss function পাবো, তাই অনুসারে loss function এর থেকে parameter update করবো।

derivative বের করবো $w_1, w_2, w_3, w_4, w_5, w_6$ with respect to Loss function.

derivative বের করবো b_1, b_2, b_3 with respect to loss function.

Chain rule করে Derivation করলে বিশাল বড় হবে, তাই matrix এ নিয়ে calculation করা হয়।

Neuron Perceptron is nothing but a logistic regression.



2 feature = x_1, x_2

parameters = b, w_1, w_2

output = a

Example:

$W = .7$

$b = .1$

Dataset

x	y
.1	0
.2	0
.3	1
.4	?

$z = .7x + .1$

$$L = -0\log\frac{1}{1+e^{-.7*.1-.1}} - (1-0)\log\left(1-\frac{1}{1+e^{-.7*.1-.1}}\right)$$

$$-0\log\frac{1}{1+e^{-.7\times.2-.1}} - (1-0)\log\left(1-\frac{1}{1+e^{-.7*.2-.1}}\right)$$

$$-1\log\frac{1}{1+e^{-.7\times.3-.1}} - (1-1)\log\left(1-\frac{1}{1+e^{-.7\times.3-.1}}\right)$$

$$= -\log\left(1-\frac{1}{1+e^{-.7\times.1-.1}}\right) - \log\left(1-\frac{1}{1+e^{-.7\times.2-.1}}\right) - \log\frac{1}{1+e^{-.7\times.3-.1}}$$

$$= -\log(1-.54) - \log(1-.55) - \log .57$$

$$= -\log .46 - \log .45 - \log .57$$

$= \cdot 337 + \cdot 347 + \cdot 244$

$= \cdot 928$.

Average loss value accross the 3 data points is $\cdot 928/3 = \cdot 309$.

→ lets try to reduce the loss value.

→ we will have to update both w and b using.

$$\frac{dL}{dw} = \left(\sum_{i=1}^{n}\left(\frac{1}{1+e^{-z_i}} - y_i\right) \times x_i\right)/n$$

$$w = w - \alpha \times \frac{dL}{dw}$$

$$\frac{dL}{db} = \left(\sum_{i=1}^{n}\left(\frac{1}{1+e^{-z_i}} - y_i\right)\right)/n$$

$$b = b - \underbrace{\alpha}_{\text{learning rate}} \times \frac{dL}{db}$$

$$\frac{dL}{dw} = \left(\left(\frac{1}{1+e^{.7\times .1-.1}} - 0\right)\times .1 + \left(\frac{1}{1+e^{.7\times .2-.1}} - 0\right)\times .2\right.$$

$$\left. + \left(\frac{1}{1+e^{.7\times .3-.1}} - 1\right)\times .3\right)/3$$

$$= (.054 + .112 - .127)/3$$

$$= .039/3$$

$$= .013$$

$$\frac{dL}{db} = \left(\left(\frac{1}{1+e^{.7\times .1-.1}} - 0\right) + \left(\frac{1}{1+e^{.7\times .2-.1}} - 0\right) + \right.$$

$$\left.\left(\frac{1}{1+e^{.7\times .3-.1}} - 1\right)\right)/3$$

$$= (.54 + .56 - .42)/3$$

$$= .68/3$$

$$= .23$$

Thus, if $\alpha = 1$, then updated w and b are:

$\rightarrow w = .7 - 1^{*} .013 = .687$

$\rightarrow B = .1 - 1^{*} .23 = -.13$

Recalculating the sum of loss:

$$l = -0\log\frac{1}{1+e^{-.687\times.1+.13}} - (1-0)\log\left(1-\frac{1}{1+e^{-.687\times.1+.13}}\right)$$

$$-0\log\frac{1}{1+e^{-.687\times.2+.13}} - (1-0)\log\left(1-\frac{1}{1+e^{-.687\times.2+.13}}\right)$$

$$-1\log\frac{1}{1+e^{-.687\times.3+.13}} - (1-1)\log\left(1-\frac{1}{1+e^{-.687\times.3+.13}}\right)$$

$$= -\log\left(1-\frac{1}{1+e^{-.687\times.1+.13}}\right) - \log\left(1-\frac{1}{1+e^{-.687\times.2+.13}}\right) - \log\frac{1}{1+e^{-.687\times.3+.13}}$$

$$= -\log(1-.48) - \log(1-.5) - \log .52$$

$$= -\log .52 - \log .50 - \log .52$$

$$= .28 + .3 + .28$$

$$= .86$$

average $= .86/3 = .287$ { slightly better than before }.

x

Given.

$\omega =$

$b =$

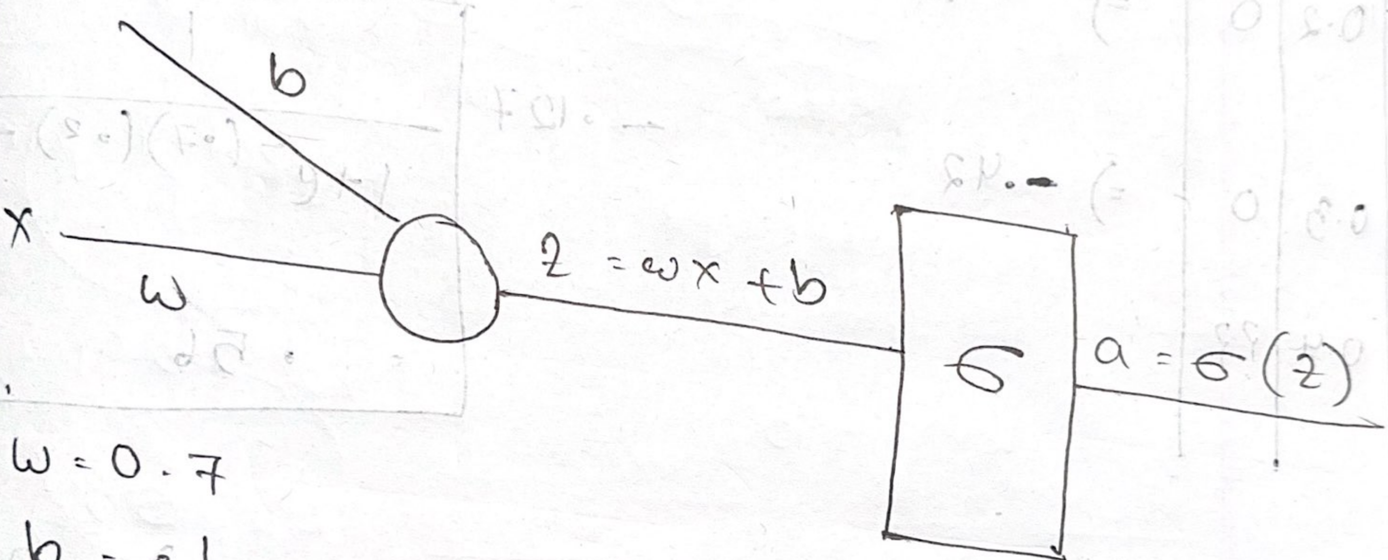
$\alpha =$

$a =$

$\Rightarrow$ (a-

$= .54$

All these calculations just for a neuron
why we keep more side by side it gets
tricky.



$w = 0.7$

$b = .1$

$\alpha = 1$

$a = \dfrac{1}{1+e^{-wx-b}}$

X	Y
.1	0
.2	0
.3	1
.4	?

$\Rightarrow (a-y) \qquad (a-y)x$

$= .54 \qquad (.54)(.1)$

X	Y	
0.1	0	$\Rightarrow .54$
0.2	0	$\Rightarrow .56$
0.3	0	$\Rightarrow -.42$
0.4	??	

$.054$

$(.56) \times (0.2)$ or 0.112

$-.127$

$$\boxed{\frac{1}{1+e^{-(.7)(.2)-.1}} = .56}$$

$$\frac{dL}{dw} = \frac{1}{3}\left(.054 + .112 - .127\right) = .013$$

$$w = w - \alpha \frac{dL}{dw} = (.7) - 1(.013) = .687$$

$$\frac{dL}{db} = \frac{1}{3}\left(0.54 + .56 - .42\right) = .227$$

$$b = b - \alpha \frac{dL}{db} = .1 - (1)(.227)$$

$= -.127$

$$a = \frac{1}{1+e^{-(.647)(.4)-(-.127)}}$$

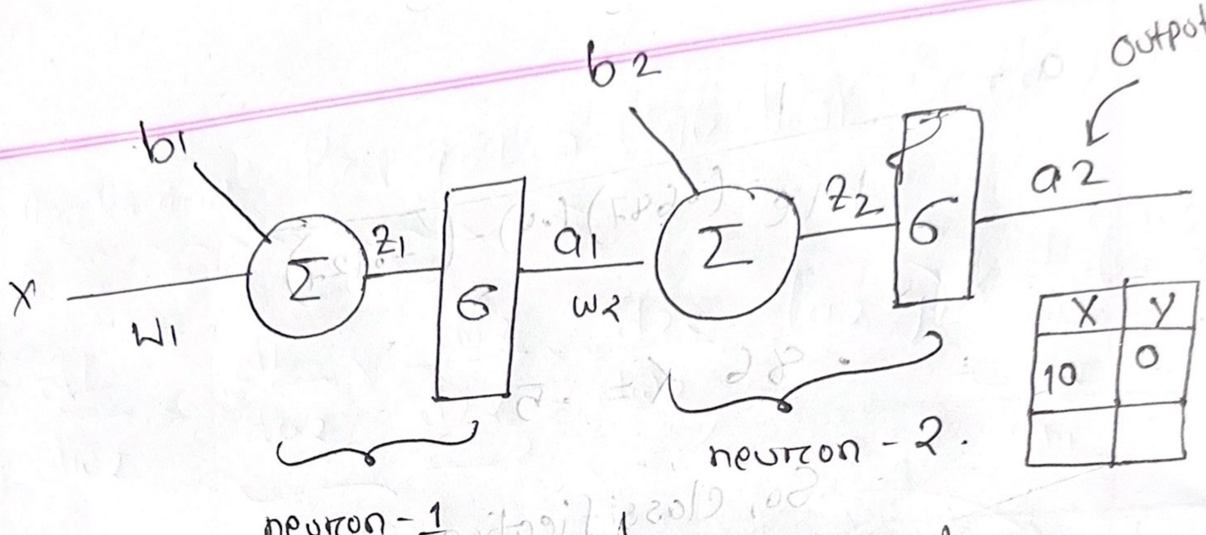
$$= .86 <= .5$$

So, classification আসবে 1.

$\therefore$

X	Y
.4	1

(Ans).



X	y
10	0

Hidden layer

$$z_1 = w_1 x_1 + b$$

$$a_1 = \sigma(z_1) = \frac{1}{1+e^{-z_1}}$$

Output Layer

$$z_2 = w_2 a_1 + b_2$$

$$a_2 = \sigma(z_2) = \frac{1}{1+e^{-z_2}}$$

$$\left.\frac{dL}{dw_2}, \frac{dL}{db_2}, \frac{dL}{dw}, \frac{dL}{db_1}\right\}$$ 4 derivative.

2

output

z_2 σ a_2

X	Y
10	0

-tion - 2.

-ut Layer

$-_2 a_1 + b_2$

$(z_2) = \frac{1}{1+e^{-z_2}}$

-tive.

Forward Propagation

Backward propagation $\longrightarrow$ gradient decent actually

assume, $w = .1$ $b_1 = .1$ $w_2 = .2$ $b_2 = 0.2$

F.P

$z_1 = w_1 x + b_1 = 11/10$

$a_1 = \frac{1}{1+e^{-z_1}} = \frac{1}{1+e^{-1.1}} = .75$

$z_2 = w_2 a_1 + b_2 = (0.2)(.75) + (0.2)$

$= 0.35$

$a_2 = \frac{1}{1+e^{-z_2}} = \frac{1}{1+e^{-(.35)}} = .587$

feed forward

Forward propagation

এটা 0.5
এর দিকে
আমরা চাচ্ছি
ছিল কারণ

X	Y
10	0 - So,

loss function, $a_2 = L = -\frac{1}{n}\sum_{i=1}^{n}\left[y\log a_2 + (1-y)(1-a_2)\right]$

Backward Propagation (Gradient Decent)

$$\frac{dL}{dw_2} = \underbrace{\frac{dL}{da_2} \times \frac{da_2}{dz_2}} \times \underbrace{\frac{dz_2}{dw_2}} = (a_2 - y) \times \underbrace{\frac{d}{dw_2}(w_2a_1 + b_2)}_{a_1}$$

$$= (a_2 - y)\,a_1 \rightarrow w_2$$

$$\frac{dL}{db_2} = \frac{dL}{da_2} \times \frac{da_2}{dz_2} \times \frac{dz_2}{db_2} = (a_2 - y) \rightarrow b_2$$

$$\frac{dL}{dw_1} = \underbrace{\frac{dL}{da_2} \times \frac{da_2}{dz_2}} \times \underbrace{\frac{dz_2}{da_1} \times \frac{da_1}{dz_1}} \times \frac{dz_1}{dw_1}$$

$$= (a_2 - y) \times \frac{d}{da_1}(w_2a_1 + b_2) \times \underbrace{\frac{d}{dz_1}\left(1+e^{-z_1}\right)^{-1}} \times$$

$$\frac{d}{dw_1}(w_1 x + b_1)$$

$$= (a_2 - y) \times w_2 \times a_1 (1 - a_1) \times x$$

$\hookrightarrow \underline{\underline{w_1}}$

$$(-1)(1+e^{-z})^{-2} \times (-e^{-z_1})$$

$$(e^{-z_1}) \times \frac{1}{(1+e^{-z_1})^2}$$

$$\frac{(1-a_1)}{\cancel{a_1}} \times \cancel{a_1^2}\, a_1$$

$$\frac{dL}{db_1} = (a_2 - y) \times w_2 \times a_1 (1 - a_1) \times 1$$

$\hookrightarrow \underline{\underline{b_1}}$

$$\frac{1}{1+e^{-z_1}} = a_1$$

$$\frac{1}{a_1} = 1 + e^{-z}$$

$$e^{-z} = \frac{1-a_1}{a_1}$$