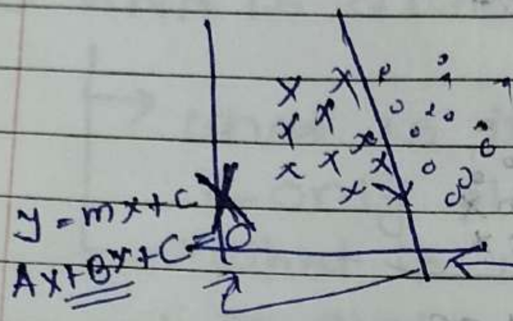
Logistic Regression.

→ Requirement:- Data should be linearly saperable.



← linearly saperable.

← we find the quation of this.

⇒ perceptron Trick:-

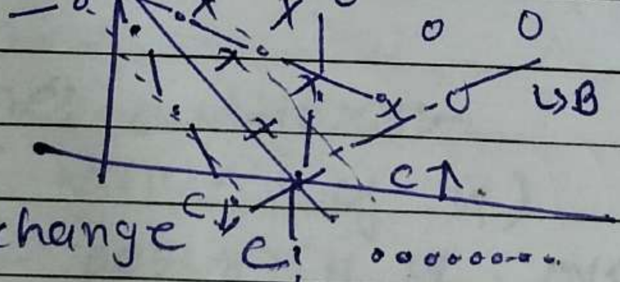
→ How to label region:-

@ desmos

$Ax+By+C>0$ positive region

$Ax+By+C<0$ negative region

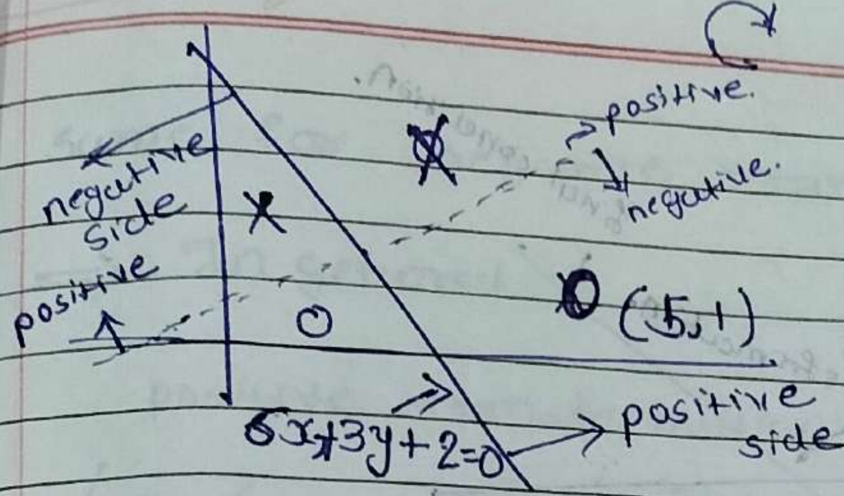
$Ax+By+C=0$ no the line



if I change C ········

if I change A —·—·—·—·

if I change B —$\bar{B}$—$\bar{B}$—$\bar{B}$—



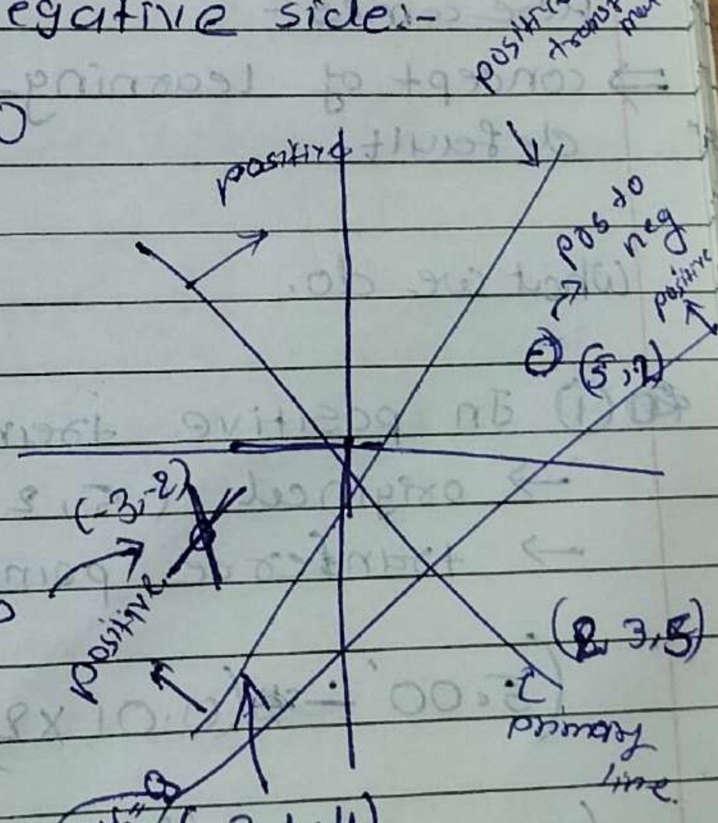
How to apply transformation:-

① transformation on positive side:-

$$(6-5)x + (3-1)y + (2-\underbrace{1}_{\text{by default}}) = 0$$
$$1x + 2y + 1 = 0$$

② transformation on negative side:-

$$(-3+2)x + (3-2)y + 6 = 0$$
$$-1x + 1y + 6 = 0$$

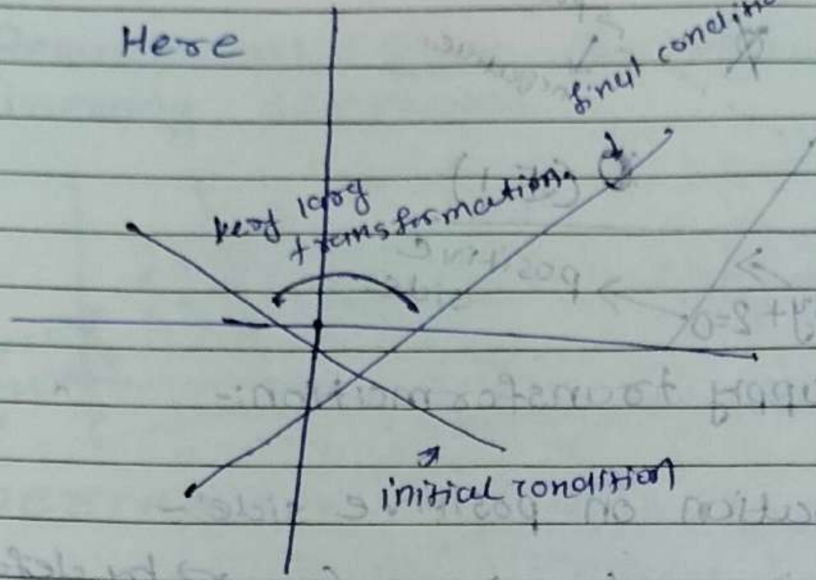


observation:

(negative to positive થવું હોય તો negative transformation)

(positive to negative થવું હોય તો positive transformation)

Here



→ Rule:- We cannot apply every long transformation in M.L.

Here comes:-

⇒ concept of learning-rate = 0.01 by default.

What we do.

① In positive transformation.

→ original : (5, 2, 1)

→ transferee point : (2, 3)

$$\left(5.00 - (0.01 \times 2)\right)x + (2 - 0.01 \times 3) + (1 - 0.01(1))$$

$$= (5.00 - 0.02)x + (2.00 - 0.03)y + (1.00 - 0.01)$$

$$= 4.98x + 1.97y + 0.99 = 0$$

↳ final line.

same for negative transformation.

$\Rightarrow$ In general.

positive transformation:- (positive region side moving)

$$\left(A-(L.R.\times a)\right)x+\left(B-(L.R.\times b)\right)y+\cancel{C.L.R.C}\ (C-(L.R\times c))=0$$

negative transformation:- (negative region side moving)

$$\left(A+(L.R\times a)\right)x+\left(B+(L.R\times b)\right)+\left(C+(L.R\times c)\right)=0$$

Algorithm:-

x_0	x_1 CGPA	x_2 IQ	Y placed
1	7.5	81	1
1	8.9	109	1
1	7.0	81	1

$$Ax + By + C = 0$$

$$w_0 + w_1 x_1 + w_2 x_2 = 0$$

$w_0 = C$

$w_1 = A$

$w_2 = B$

$$w_0 x_0 + \omega_1 x_1 + \omega_2 x_2 = 0$$

In general, $\sum_{i=0}^{2} \omega_i x_i = 0$ ≥ 0 placed

< 0 not placed

let epoch $= 1000$, $\eta = 0.01$

original.

$x_i \in N$

in prediction $x_i \in P_0$

$x_i \to$ student know about placement.

for i in range (epochs):

if $x_i \in N$ and $\sum_{i=0}^{2} m_i x_i \geq 0$:

L.R. Algo

$\omega_{new} = \omega_{old} - \eta(x_i)$

if $x_i \in P$ and $\sum_{i=0}^{2} m_i x_i < 0$:

$\omega_{new} = \omega_{old} + \eta(x_i)$

Modefication:-

for i in ~~epo~~ range(epochs):

$\omega_n = \omega_0 + \eta(y_i - \hat{y}_i)x_i$

Is this work?

let consider

y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	
0	0	0	case:1
0	1	-1	case:2
1	0	1	case:3
1	1	0	case:4.

from: case: 1 & case: 3.

$\omega_n = \omega_0 + \eta(0-0)x_i$

$\omega_n = \omega_0 + \eta x_i \rightarrow$ no change.

from case: 2

$\omega_n = \omega_0 + \eta(0-1)x_i$

↳ placement નહી હોવા થી
but
model ને બોલા હુવા
positive to negative.
(substract).

$\omega_n = \omega_0 - \eta x_i$ (our first condition).

from case: 3.

$\omega_n = \omega_0 + \eta(1-0)x_i$

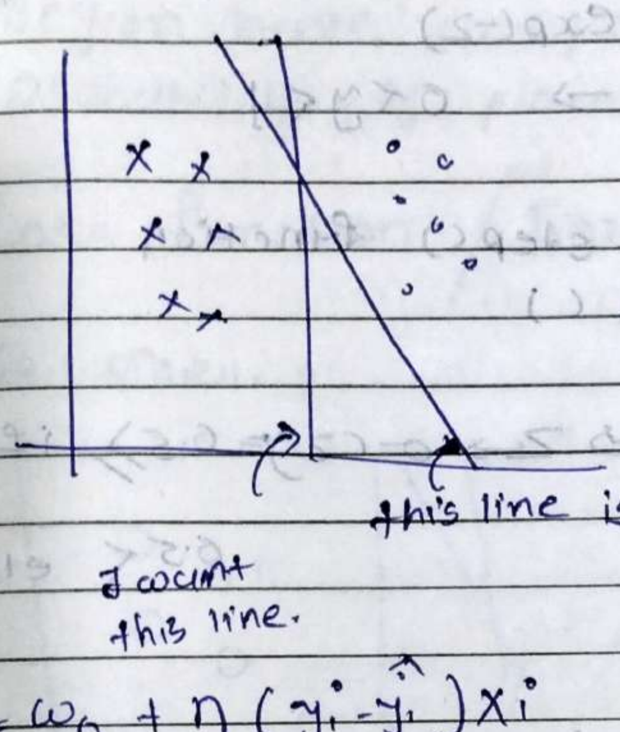
↳ placement હોવા થી
model says
નહી હુવા
(negative to positive)
(Add)

$\omega_0 = \omega_0 + \eta x_i$ (second condition) effectiveness

logistic Regression sklearn Algo > Perseptron Algo

⇒ perseptron की error / accuracy को improve करने some modefied algo.

what I want ?



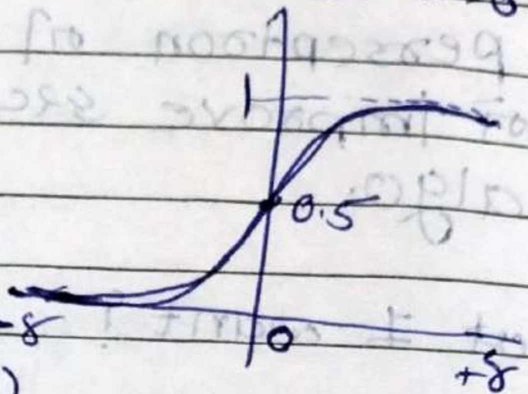
$w_n = w_0 + \eta(y_i - \hat{y}_i)x_i$

I want $y_i - \hat{y}_i \neq 0$ at case : 1 & case : 4

It hold some value.

sol^n

Sigmoid function:-



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

$$= \frac{1}{1+\exp(-z)}$$

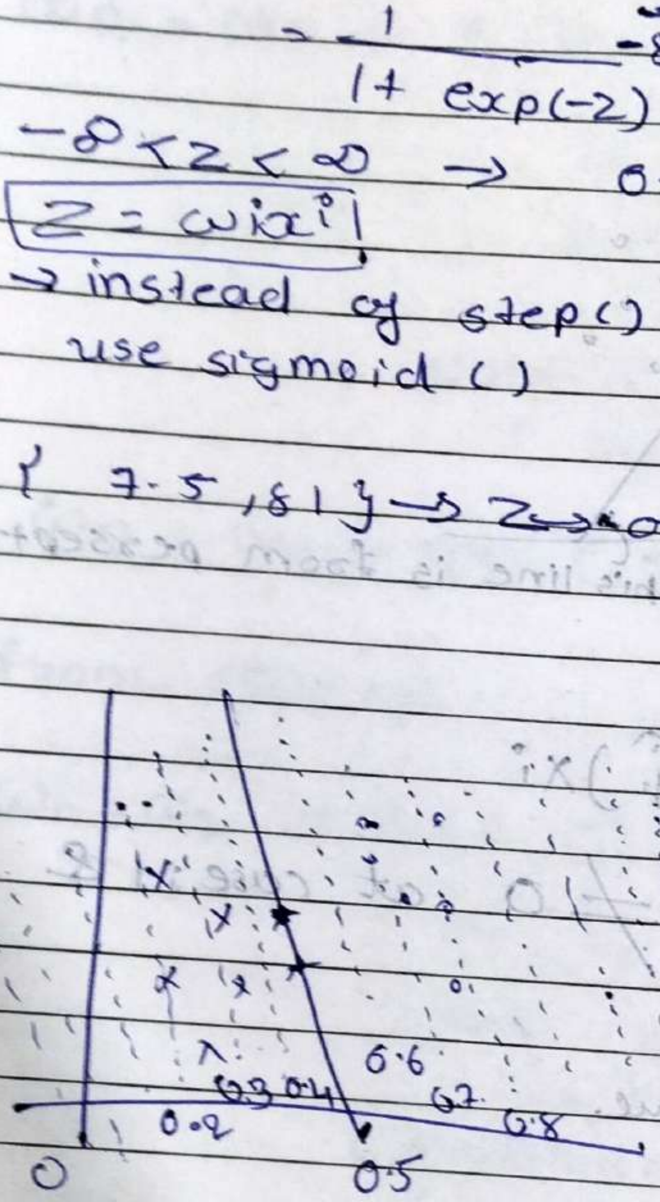
$-\infty < z < \infty \rightarrow 0 < y < 1$

$\boxed{z = w_i x^i}$

→ instead of step() function use sigmoid()

$\{7.5, 8.1\} \rightarrow z \rightarrow \sigma(z) = 0.5 \geq$ if 1

$0.5 <$ else 0



shows probability:-

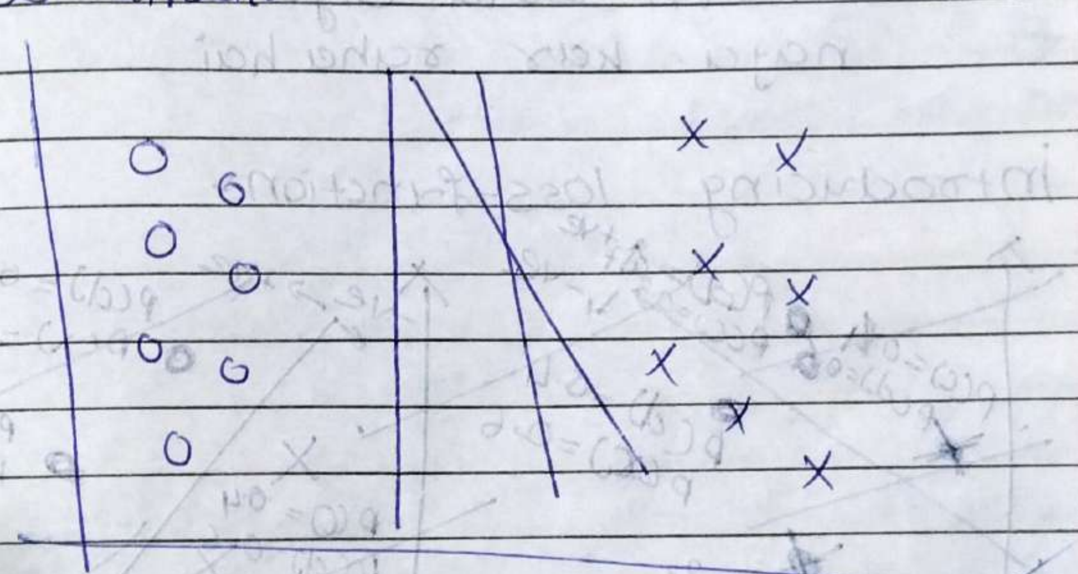
→ students which are on the line have 50 % probability of placed.

→ students which are on positive region have 60 %, 70 %, 80 %, 90 %, 99.99 % probability of placed.

→ student which are on negative region have 40 %, 30 %, 20 %, 10 %, 0.01 % probability of placed.

⇒ Loss function / Error function.

આવું કંઈક હોવાના.



→ perseptron technique graph will improved by appling logistic regression simmiler to sklearn but still some dissimilarities occur.

Why ?

→ sklearn નું logistic regression using sigmoid function થી samples લેતું હતું એના કરતાં આપણું model આપણા sample લે છે.

→ It is normal becuse it totally based on random-ness.

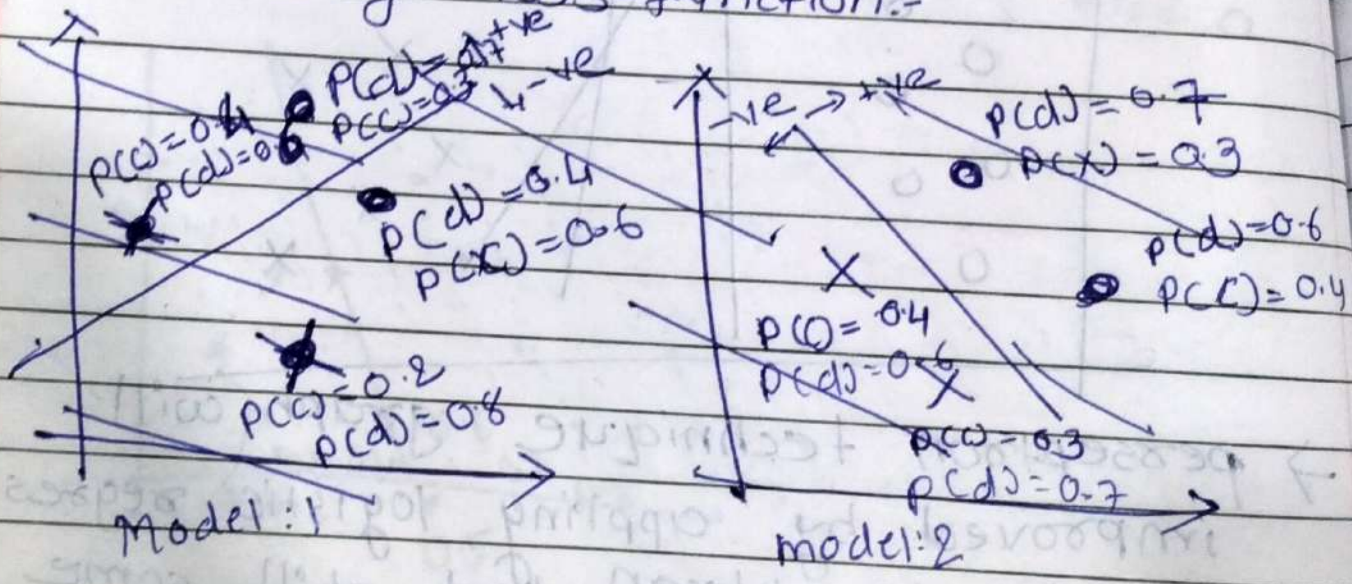
→ ∴ Differe is generated.

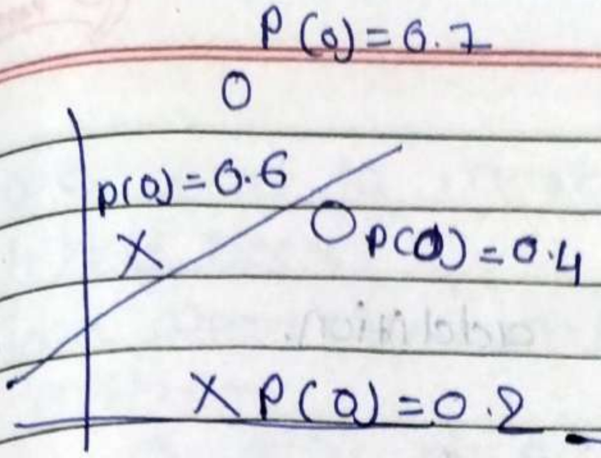
→ But there is a big differcne occurs.

→ what does it mean?

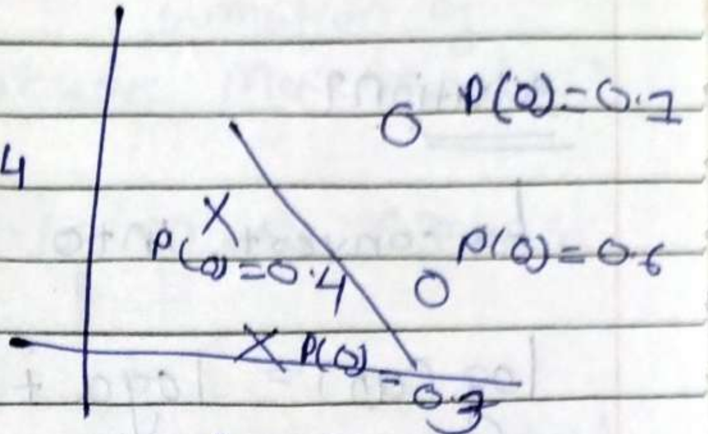
- sklearn wala algo kuchh to naya kar raha hai.

introoducing loss-function:-





Model:1



Model:2

product of poobability:-

For Model:1 (dot) $0.7 \times 0.4 \times 0.8 \times 0.4$

$= 0.28 \times 0.32$

$= 0.0896$ (max likely hood)

For Model:2 : $0.7 \times 0.6 \times 0.6 \times 0.7$

$= 0.1764$ (max likely hood)

~~Ao~~ $0.1764 > 0.0896$

Model 2 > Model 1

$0.3 \times 0.3 = 0.09$

$0.3 \times 0.3 \times 0.3 = 0.027$

$0.3 \times 0.3 \times 0.3 \times 0.3 = 0.0081$

Becomes smaller.

મોટા ડેટાસેટ પર કામ કરતી વખત આ વેલ્યુ બહું નાની થઈ જાય છે. (problem) solution?

↳ convert into addition.

$\log(ab) = \log a + \log b$.

$\log(0.7) + \log(0.4) + \log(0.4) + \log(0.8)$
=

But?
$\log(0.3) = -0.522$
$\log(0.9) = -0.045$
$\log(0.1) = -1$
$\log(0.01) = -2$
$\log(0.99) = -1.04$

conclusion $\log(0.\_\_\_) = -ve$.

∴ solution?

~~log(max)=~~

$-\log(0.7) - \log(0.4) - \log(0.4) - \log(0.8)$

↳ cross entropy

↳ sumation of ~~A~~ of log of maximum likelyhood.

→ we want to increase summation of Maximum likely hood.

→ we want to minimise cross entropy

→ $0.5 > 0.3$ ✓

- Summation of maximum likely hood greater.

$$-\log(0.5) \qquad -\log(0.3)$$

$$0.301 < 0.522$$

- It means $\log(mn)$ is has to be less

∴ Conclusion:- Cross entropy minimum लीचाना try करना. For best model

Time to Generallise:-

$$L = \log(ab) = \cancel{\sum_{i=1}^{n} -\log(y_i)}$$

$$L = \log(ab) = \cancel{\sum_{i=1}^{n} -\log(1-y_i)}$$

But :-

$$L = \log(mn) = \sum_{i=\phi}^{n} -y_i \log(\hat{y_i}) - (1-y_i)\log(1-\hat{y_i})$$

Average crossentropy

$$\frac{L}{n} = \frac{\log(mn)}{n} = \frac{-1}{n}\left(\sum_{i=1}^{\infty} y_i \log(\hat{y_i}) + (1-y_i)\log(1-\hat{y_i})\right)$$

→ log loss error
→ binary cross entropy function.

→ Use gradient Decent to calculate above equation.

Rows : m , column : n
(m students) (n parameters)

	1	2	3	...	n
1	X_{11}	X_{12}	X_{13}	...	X_{1n}
2	X_{21}	X_{22}	X_{23}	...	X_{2n}
3	X_{31}	X_{32}	X_{33}	...	X_{3n}
⋮					
m	X_{m1}	X_{m2}	X_{m3}	...	X_{mn}

$$\hat{y}_1 = \sigma\,(\omega_1 x_{11} + \omega_2 x_{12} + \omega_3 x_{13} \cdots + \omega_n x_{1n} + \omega_0)$$
$$\hat{y}_2 = \sigma\,(\omega_1 x_{21} + \omega_2 x_{22} + \omega_3 x_{23} \cdots + \omega_n x_{2n} + \omega_0)$$
$$\hat{y}_3 = \sigma\,(\omega_1 x_{31} + \omega_2 x_{32} + \omega_3 x_{33} \cdots + \omega_n x_{3n} + \omega_0)$$
$$\vdots$$
$$\hat{y}_m = \sigma\,(\omega_1 x_{m1} + \omega_2 x_{m2} + \omega_3 x_{m3} \cdots + \omega_n x_{mn} + \omega_0)$$

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_m \end{bmatrix}$$

$$= \begin{bmatrix} \sigma(\omega_1 x_{11} + \omega_2 x_{12} + \omega_3 x_{13} \cdots + \omega_n x_{1n} + \omega_0) \\ \sigma(\omega_1 x_{21} + \omega_2 x_{22} + \omega_3 x_{23} \cdots + \omega_n x_{2n} + \omega_0) \\ \sigma(\omega_1 x_{31} + \omega_2 x_{32} + \omega_3 x_{33} \cdots + \omega_n x_{3n} + \omega_0) \\ \cdots \\ \sigma(\omega_1 x_{m1} + \omega_2 x_{m2} + \omega_3 x_{m3} \cdots + \omega_n x_{mn} + \omega_0) \end{bmatrix}$$

$$\hat{y} = \sigma\left(\begin{bmatrix} 1 & x_{11} & x_{12} \cdots x_{1n} \\ 1 & x_{21} & x_{22} \cdots x_{2n} \\ 1 & x_{31} & x_{32} \cdots x_{3n} \\ \cdots & & \\ 1 & x_{m1} & x_{m2} \cdots x_{mn} \end{bmatrix} \begin{bmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_m \end{bmatrix} \right)$$

$$\hat{y} = \sigma(XW)$$
$\hookrightarrow$ in matrix form:

$$L = -\frac{1}{m} \sum_{i=1}^{m} \left[\underbrace{y_i \log(\hat{y_i})} + \sum_{i=1}^{m} (1-y_i)\log(1-\hat{y_i}) \right]$$

$$\sum_{i=1}^{m} y_i \log \hat{y_i}$$

$$= y_1 \log \hat{y_1} + y_2 \log \hat{y_2} + y_3 \log \hat{y_3} + \dots + y_m \log \hat{y_m}$$

$$= \begin{bmatrix} y_1 & y_2 & y_3 & \dots & y_m \end{bmatrix}_{1\times m} \begin{bmatrix} \log \hat{y_1} \\ \log \hat{y_2} \\ \log \hat{y_3} \\ \dots \\ \log \hat{y_m} \end{bmatrix}_{m\times 1}$$

$$= \begin{bmatrix} y_1 & y_2 & y_3 & \dots & y_m \end{bmatrix}_{1\times m} \log \left[\begin{bmatrix} \cancel{\log} \hat{y_1} \\ \cancel{\log} \hat{y_2} \\ \cancel{\log} \hat{y_3} \\ \dots \\ \cancel{\log} \hat{y_m} \end{bmatrix}_{m\times 1} \right]$$

Date
Page

$$= \overbrace{[y_1\ y_2\ y_3 \ldots y_m]}^{y} \log(\sigma(\cancel{w}x))$$

$$\cancel{L = -\frac{1}{m} \sum_{i=1}^{m} [}$$

$$L = -\frac{1}{m} \left[y \log(\sigma(XW)) + (1-y) \log(1-\sigma(WX)) \right]$$

here. $y \log \hat{y} = y \log(\sigma(XW))$

↳ w can we find using gradient decent

How to apply gradient decent?

for i in epochs:

$$W = W - \eta \frac{\Delta L}{\Delta W} \quad -(9)$$

↳ learning rate.

$$\frac{\Delta L}{\partial w} = \left[\frac{\partial L}{\partial w_0}, \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \frac{\partial L}{\partial w_n} \right]$$

↳ n+1 elements.

$$L = -\frac{1}{m}\left[y\log\hat{y} + (1-y)\log(1-\hat{y})\right]$$

$$\frac{dL}{d\omega} = \frac{d}{\omega} y\log\hat{y} = y\frac{d}{d\omega}\log(\hat{y})$$

$$= y\frac{1}{\hat{y}}\frac{d}{d\omega}\hat{y}$$

$y = \sigma(x)$

$\frac{dy}{dx} = \sigma(x)\cdot(1-\sigma(x))$

$$= \frac{y}{\hat{y}}\frac{d}{d\omega}\sigma(\omega x)$$

$$= \frac{y}{\hat{y}}\sigma(\omega x)\left[1-\sigma(\omega x)\right]\frac{d(\omega x)}{d\omega}$$

$$= \frac{y}{\cancel{\hat{y}}}\cancel{\hat{y}}\left[1-\hat{y}\right]x$$

$$= y(1-\hat{y})x \quad (1)$$

$$\frac{d}{d\omega}(1-y)\log(1-\hat{y})$$

$$= (1-y)\frac{d}{d\omega}(\log(1-\hat{y}))$$

$$= (1-y)\times\frac{1}{(1-\hat{y})}\frac{d}{d\omega}[1-\hat{y}]$$

$$= -(1-y) \times \frac{1}{(1-\hat{y})} \frac{d}{dw}(\sigma(wx))$$

$$= \frac{-(1-y)}{(1-\hat{y})} \sigma(wx) \times (1-\sigma(wx)) \frac{d}{dw}(wx)$$

$$= \frac{-(1-y)}{\cancel{(1-\hat{y})}} \hat{y} \cancel{(1-\hat{y})} X$$

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

$$\frac{\partial L}{\partial w} = \frac{-1}{m}\left[(1) + (2)\right]$$

$$= \frac{-1}{m}\left[y(1-\hat{y})X - \hat{y}(1-y)X\right]$$

$$= \frac{-1X}{m}\left[y(1-\hat{y}) - \hat{y}(1-y)\right]$$

$$= \frac{-1X}{m}\left[y - \cancel{y\hat{y}} - \hat{y} + \cancel{y\hat{y}}\right]$$

$$\boxed{\frac{\partial L}{\partial w} = \frac{-1}{m}\left[y-\hat{y}\right]X}$$

put in a

$$w = w + \eta \frac{1}{m} (y - \hat{y}) X$$