Loss Function OR Binary Cross Entropy....

Gradient Descent

Rows = m, Columns = n

1	2	3	......	n	Y	$\hat{Y}$
X_{11}	X_{12}	X_{13}	......	X_{1n}	Y_1	$\hat{Y}_1$
X_{21}	X_{22}	X_{23}	......	X_{2n}	Y_2	$\hat{Y}_2$
⋮						
X_{m1}	X_{m2}	X_{m3}	......	X_{mn}	Y_m	

The Coefficients required are:-
$\{W_1, W_2, W_3 \cdots W_n\}$
W_0
$(n+1)$

To Calculate the $\hat{Y}_1 \Rightarrow \sigma(W_1 * X_{11} + W_2 * X_{12} + W_3 * X_{13} \cdots W_n * X_{1n} + W_0)$
$= \hat{Y}_1$

Similarly $\hat{Y}_2 \Rightarrow \sigma(W_1 * X_{21} + W_2 * X_{22} + W_3 * X_{23} \cdots W_n * X_{2n} + W_0) = \hat{Y}_2$

It means :-

$$\hat{Y} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \hat{Y}_3 \\ \vdots \\ \hat{Y}_n \end{bmatrix} \rightarrow \begin{bmatrix} \sigma(W_1 * X_{11} + W_2 * X_{12} + W_3 * X_{13} \cdots W_n * X_{1n} + W_0) \\ \sigma(W_1 * X_{21} + W_2 * X_{22} + W_3 * X_{23} \cdots W_n * X_{2n} + W_0) \\ \sigma(W_1 * X_{31} + W_2 * X_{32} + W_3 * X_{33} \cdots W_n * X_{3n} + W_0) \\ \vdots \\ \sigma(W_1 * X_{m1} + W_2 * X_{m2} + W_3 * X_{m3} \cdots W_n * X_{mn} + W_0) \end{bmatrix}$$

$$\hat{Y} = \sigma\left(\begin{bmatrix} (W_1 * X_{11} + W_2 * X_{12} + W_3 * X_{13} \cdots W_n * X_{1n} + W_0) \\ (W_1 * X_{21} + W_2 * X_{22} + W_3 * X_{23} \cdots W_n * X_{2n} + W_0) \\ \vdots \\ (W_1 * X_{m1} + W_2 * X_{m2} + W_3 * X_{m3} \cdots W_n * X_{mn} + W_0) \end{bmatrix}\right)$$

Spiral

Date..........

$\boxed{(1-Y)\log(1-\hat{Y})}$

$$\frac{d}{dW}(1-Y)\log(1-\hat{Y}) \Rightarrow (1-Y)\frac{d}{dW}\log(1-\hat{Y})$$

$$\Rightarrow \frac{(1-Y)}{(1-\hat{Y})}\frac{d}{dW}[1-\hat{Y}] \Rightarrow \frac{(1-Y)}{(1-\hat{Y})}\frac{d}{dW}[1-\hat{Y}]$$

Derivative of '1' is '0'

$$\Rightarrow -\frac{(1-Y)}{(1-\hat{Y})}\frac{d}{dW}\sigma(WX) \Rightarrow \frac{-(1-Y)}{1-\hat{Y}}\sigma(WX) * [1-\sigma(WX)]\frac{d}{dL}(WX)$$

$$\Rightarrow -\frac{(1-Y)}{\cancel{(1-\hat{Y})}}\hat{Y}\cancel{(1-\hat{Y})}X$$

$$\Rightarrow -(1-Y)\hat{Y}X \text{ or } \cancel{-\hat{Y}\frac{(1-Y)}{(1-\hat{Y})}} \quad \boxed{-\hat{Y}(1-Y)X} \; (2)$$

Combining (1) and (2)

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

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

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

$$\boxed{\frac{\Delta L}{\Delta W} \Rightarrow \frac{-1}{m}(Y-\hat{Y})X}$$

Gradient Descent Formula will be:-

$$W = W - \eta\frac{\Delta L}{\Delta W} \Rightarrow \boxed{W = W + \eta\frac{1}{m}(Y-\hat{Y})X}$$

"Putting the derivation value in place of $\frac{\Delta L}{\Delta W}$"

Shape=(n+1, 1)

$$W = \begin{bmatrix} W_0 \\ W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} \quad X = \begin{bmatrix} X_1 & X_2 & X_3 & \cdots & X_{1n} \\ X_{m1} & X_{m2} & X_{m3} & \cdots & X_{mn} \end{bmatrix} \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}$$

Shape=(m, n+1)

Shape=(m, 1)

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

Shape=(m, 1)

Spiral

Date.....................

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

where $\hat{Y} = \sigma(XW)$

Loss Function in Matrix Form

Loss Function in Matrixterm

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

Gradient Descent

$W = [\quad]$

for i in epochs:

$$W = W - \eta\frac{\Delta L}{\Delta W}$$

$$\frac{\Delta L}{\Delta W}\left[\frac{DL}{DW_0}, \frac{DL}{DW_1}, \frac{DL}{DW_2}, \ldots \frac{DL}{DL_n}\right]$$

This is the combination of derivates of Loss Function with respect to all Coefficients......

$\left[\frac{\Delta L}{\Delta W}\right]$ Calculation....

$$L = \frac{-1}{m}\left[\boxed{Y_i\log\bar{Y}_i} + \boxed{(1-Y_i)\log(1-\bar{Y}_i)}\right]$$

$\frac{dL}{dW}$ = Firstly Calulating the Derivation of $Y\log\bar{Y}$

$$\Rightarrow \frac{d}{dW}Y\log\bar{Y} \Rightarrow Y\frac{d}{dW}\log\bar{Y} \Rightarrow \frac{Y}{\bar{Y}}\frac{d(\bar{Y})}{dL}$$

$$\Rightarrow \frac{Y}{\hat{y}}\frac{d}{dL}\sigma(WX) \Rightarrow \frac{y}{\hat{y}}\sigma(WX)[1-\sigma(WX)]\frac{d}{dW}(WX)$$

$$\Rightarrow \frac{y}{\cancel{\hat{Y}}}\cancel{\hat{Y}}(1-\hat{Y})X \Rightarrow \boxed{Y(1-\hat{Y})X}\ 1$$

$$\begin{array}{cc} A[X] & B[W] \end{array}$$

Date..........

$$\hat{Y} \Rightarrow \sigma\left(\begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1n} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2n} \\ 1 & X_{31} & X_{32} & X_{33} & \cdots & X_{3n} \\ \vdots & & & & & \vdots \\ 1 & X_{m1} & X_{m2} & X_{m3} & \cdots & X_{mn} \end{bmatrix} \cdot \begin{bmatrix} W_0 \\ W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} \right)$$

$$\boxed{\hat{Y} = \sigma(X * W)}$$

$$L = \boxed{\frac{-1}{m} \sum_{i=1}^{n} Y_i \log(\hat{Y_i}) + (1-Y_i)\log(1-\hat{Y_i})}$$

$$L \Rightarrow \frac{-1}{m}\left[\underbrace{\sum_{i=1}^{n} Y_i \log(\hat{Y_i})}_{\text{Converting into Matrix form}} + \sum_{i=1}^{n} (1-Y_i)\log(1-\hat{Y_i})\right]$$

$$\sum_{i=1}^{n} Y_i \log(\hat{Y_i}) \Rightarrow Y_1 \log \hat{Y_1} + Y_2 \log \hat{Y_2} + Y_3 \log \hat{Y_3} \cdots\cdots Y_m \log \hat{Y_m}$$

$$\Rightarrow \begin{bmatrix} Y_1 & Y_2 & Y_3 & Y_4 & \cdots & Y_m \end{bmatrix} \cdot \begin{bmatrix} \log \hat{Y_1} \\ \log \hat{Y_2} \\ \log \hat{Y_3} \\ \log \hat{Y_m} \end{bmatrix}$$

$$\Downarrow$$

$$\underbrace{\begin{bmatrix} Y_1 & Y_2 & Y_3 & Y_4 & \cdots & Y_m \end{bmatrix}}_{Y} \cdot \log \underbrace{\begin{bmatrix} \hat{Y_1} \\ \hat{Y_2} \\ \hat{Y_3} \\ \vdots \\ \hat{Y_m} \end{bmatrix}}_{\hat{Y} \text{ Matrix}}$$

$$\Rightarrow Y \log \hat{Y}$$

$$\Rightarrow \boxed{Y \log(\sigma(XW))}$$

So, Now our Complete Loss Function will be re-written as:-