Applying G.D on Loss function

No. of Rows = m , columns = n

	1	2	3	...	n	y
1	X_{11}	X_{12}	X_{13}	----	X_{1n}	Y_1
2	X_{21}	X_{22}	X_{23}	----	X_{2n}	Y_2
3	X_{31}	X_{32}	X_{33}	----	X_{3n}	Y_3
⋮	⋮	⋮	⋮			⋮
m	X_{m1}	X_{m2}	X_{m3}	----	X_{mn}	Y_m

In logistic Regression
no. of input columns = n+1 coefficient

$w_1, w_2, w_3 \ldots w_n$
w_0

we want to calculate $\hat{y}$

① $\hat{y}$ for 1st row

$$\hat{y}_1 = \sigma(w_1 x_{11} + w_2 x_{12} + w_3 x_{13} \ldots + w_n x_{1n} + w_0)$$

② $\hat{y}$ for 2nd row.

$$\hat{y}_2 = \sigma(w_1 x_{21} + w_2 x_{22} + w_3 x_{23} \ldots + w_n x_{2n} + w_0)$$

③ $\hat{y}$ for 3rd row

$$\hat{y}_3 = \sigma(w_1 x_{31} + w_2 x_{32} + w_3 x_{33} + \ldots + w_n x_{3n} + w_0)$$

$$\hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_m \end{bmatrix} = \begin{bmatrix} \sigma(w_0 + w_1 x_{11} + w_2 x_{12} + \dots w_n x_{1n}) \\ \sigma(w_0 + w_1 x_{21} + w_2 x_{22} + \dots w_n x_{2n}) \\ \sigma(w_0 + w_1 x_{m1} + w_2 x_{m2} + \dots + w_n x_{mn}) \end{bmatrix}$$

$$\hat{Y} = \sigma\left(\begin{bmatrix} w_0 + w_1 x_{11} + w_2 x_{12} \dots + w_n x_{1n} \\ w_0 + w_1 x_{21} + w_2 x_{22} \dots + w_n x_{2n} \\ \vdots \\ w_0 + w_1 x_{m1} + w_2 x_{m2} \dots + w_n x_{mn} \end{bmatrix}\right)$$

to the dot product

$$\hat{Y} = \sigma\left(\underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ \vdots \\ 1 & x_{21} & x_{22} & \dots & x_{2n} \\ \vdots \\ 1 & x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}}_{B} \right)$$

Added 1 for w_0

X (input data)

W (coefficients)

$$\boxed{\hat{Y} = \sigma(XW)} \quad \text{--- } (1)$$

$$L = -\frac{1}{m} \sum_{i=1}^{m} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)$$

$$L = -\frac{1}{m}\left[\underbrace{\sum_{i=1}^{m} y_i \log(\hat{y}_i)}_{\text{converting this to matrix form}} + \sum_{i=1}^{m}(1-y_i)\log(1-\hat{y}_i)\right]$$

$$\sum_{i=1}^{m} y_i \log(\hat{y}_i) = y_1 \log \hat{y}_1 + y_2 \log \hat{y}_2 \cdots + y_m \log \hat{y}_m$$

1st row output column value

1st row prediction

then we can write like this

$$\begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_m \end{bmatrix} \begin{bmatrix} \log \hat{y}_1 \\ \log \hat{y}_2 \\ \log \hat{y}_3 \\ \vdots \\ \log \hat{y}_m \end{bmatrix}$$

Dot product

Rewrite like this

$$[y_1 \; y_2 \; y_3 \ldots y_m] \log \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_m \end{bmatrix}$$

$$Y \qquad \hat{Y}$$

$$Y \log \hat{Y} \qquad \sigma(XW)$$

$$\boxed{Y \log(\sigma(XW))} \text{ from } (1)$$

then this part will become

$$\boxed{(1-Y)\log\sigma(XW)}$$

Then our Loss function is

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

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

Matrix form of Loss function

* Loss Function in Matrix Form

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

coefficient

we have to calculate / find all the (w) coefficient value that makes the whole equation (minimum)

To solve this problem we will use Gradient Descent.

① Initialize W with Random value

② Loop for i in epochs

$$W = W - \alpha \left(\frac{\Delta L}{\Delta W}\right)$$

Learning Rate

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

Derivative of Loss function w.r.t all coefficient

$\left(\frac{\Delta L}{\Delta W}\right)$ How to calculate this?

Let find $\left\{\frac{\Delta L}{\Delta W}\right\}$

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

$\frac{dL}{dw}$ Derivative of this

$$\frac{d}{dw} y \log \hat{y} \Rightarrow y \frac{d}{dw} \log \hat{y}$$

y was constant that wing took out

$$\Downarrow$$

$$\boxed{\log = \frac{1}{y}} \rightarrow \frac{y}{\hat{y}} \frac{d}{dL}(\hat{y})$$

$$\Downarrow$$

$$\boxed{\hat{y} = \sigma(wx)} \rightarrow \frac{y}{\hat{y}} \frac{d}{dL} \sigma(wx)$$

$$\Downarrow$$

$$\frac{y}{\hat{y}} \underbrace{\sigma(wx)}_{\hat{y}} \overbrace{[1 - \sigma(wx)]}^{\hat{y}} \frac{d}{dw}(wx)$$

$$\Downarrow$$

$$\frac{y}{\cancel{\hat{y}}} \cancel{\hat{y}} (1-\hat{y})\, x$$

$$\Downarrow$$

Derivation

$$\boxed{y(1-\hat{y})x} \longrightarrow (2)$$

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

$$\Downarrow$$

$$\frac{(1-y)}{(1-\hat{y})}\frac{d}{dw}[1-\hat{y}]$$

$$\Downarrow$$

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

$$\Downarrow$$

$$-\frac{(1-y)}{[1-\hat{y}]}[\sigma(wx)[1-\sigma(wx)]]\frac{d}{dw}(wx)$$

$$\Downarrow$$

$$-\frac{(1-y)}{\cancel{(1-\hat{y})}}\hat{y}\cancel{(1-\hat{y})}\,x$$

$$\Downarrow$$

$$\boxed{=\hat{y}(1-y)x} \text{ ——— } (3)$$

derivation

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

From (2) and (3)

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

Multiply

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

$$\frac{\Delta L}{\Delta w} = \boxed{\frac{-1}{m}\left[y - \hat{y}\right]x} \longrightarrow (4)$$

Now Gradient Descent update is

$$\boxed{w = w + \alpha \frac{1}{m}(y - \hat{y})x} \leftarrow \text{from } (4)$$

Mechanic code this

$$w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix} \quad x = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & & \ddots & & \\ \vdots & & & \ddots & \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$(n+1, 1)$ Shape $(m, n+1)$ $(m, 1)$

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

$(m, 1)$