Date..........

Softmax Regression

Softmax Regression

cgpa	iq	Placement
7.1	71	Yes
8.5	85	No
9.5	95	May Not

Yes ⇒ 1
No ⇒ 0
Opt out ⇒ 2
(May Not)

Q Why Softmax Regression when we have the Loss Function Logistic Regression?

Ans Softmax Regression ⇒ Softmax Regression is required even we have the Loss Function Logistic Regression because the Loss Function Logistic Regression works only for the Binary data prediction.

- But, let's say we have to Predict more than 2 values i.e; Yes, No, Opt-out etc. So in that Case Loss Function Logistic Regression don't works. Here, we use the Softmax Regression.

Softmax Function

where:- K = No. of Classes..

$$\Rightarrow \sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

For Example :- Yes ⇒ $\sigma(z)_1 = \frac{e^{z_1}}{e^{z_1}+e^{z_0}+e^{z_2}}$ Probability of Class of 'Yes'..

No ⇒ $\sigma(z)_0 = \frac{e^{z_0}}{e^{z_0}+e^{z_1}+e^{z_2}}$ Probability of Class of 'No'...

Opt-out ⇒ $\sigma(z)_2 \Rightarrow \frac{e^{z_2}}{e^{z_0}+e^{z_1}+e^{z_2}}$ Probability of Class of 'Opt Out'...

"Note:- $P(\sigma(z)_1) + P(\sigma(z)_2) + P(\sigma(z)_3) = 1$"

⊙ Training Intuition

cgpa	iq	placement
—	—	0
—	—	1
data	—	2

transform → OHE [One Hot Encoding]

Complete "dataSet"

cgpa	iq	=0	=1	=2
—	—	1	0	0
—	—	0	1	0
—	—	0	0	1

where 1 → True
0 → False

"We have divide into the form that the data is converted into Binary Distribution."

0 → Yes
1 → No
2 → Opt Out

dataSet → d1, d2, d3

d1

cgpa	iq	=0
—	—	1
—	—	0
—	—	0

→ Model 1

$w_1^{(0)}, w_2^{(0)}, w_0^{(0)}$

d2

cgpa	iq	=1
—	—	0
—	—	1
—	—	0

→ Model 2

$w_1^{(1)}, w_2^{(1)}, w_3^{(1)}$

d3

cgpa	iq	=2
—	—	0
—	—	0
—	—	1

→ Model 3

$w_1^{(2)}, w_2^{(2)}, w_3^{(2)}$

Prediction

$\text{Student}_x \Rightarrow \{7, 70\} \Rightarrow$ Yes/No/Opt Out

Model 1

$Z_1 = \sum W_i * X_i$

$\boxed{Z_1 \Rightarrow 7 \times W_1^{(0)} + 70 \times W_2^{(0)} + W_0^{(0)}}$

Model 2

$\boxed{Z_2 \Rightarrow 7 \times W_1^{(1)} + 70 \times W_2^{(1)} + W_0^{(1)}}$

Model 3

$\boxed{Z_3 = 7 \times W_1^{(2)} + 70 \times W_2^{(2)} + W_0^{(2)}}$

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

Model 1

Applying Softmax Function:-

$\Rightarrow \boxed{\sigma(Z)_0 = \frac{e^{Z_0}}{e^{Z_1}+e^{Z_2}+e^{Z_0}}}$

Assuming:-

$\Rightarrow \boxed{0.40}$

Model 2

Applying Softmax Function:-

$\Rightarrow \boxed{\sigma(Z)_1 = \frac{e^{Z_1}}{e^{Z_1}+e^{Z_2}+e^{Z_0}}}$

$\Rightarrow 0.35$

Model 3

Applying Softmax:-

$\Rightarrow \boxed{\sigma(Z)_2 \Rightarrow \frac{e^{Z_2}}{e^{Z_0}+e^{Z_1}+e^{Z_2}}}$

$\Rightarrow 0.25$

Highest Softmax Value i.e., Model 1 is having the Highest Value means that Student Placement is Yes...

Loss Function for Softmax Regression

Q Why we are using the Loss Function when we have the Training Intuition?

Sol The Training Intuition includes the Classification of column into different columns and make different Models. But, it includes too much of time because, many models have their different Coefficients and Prediction will be too much time Consuming.

So, Solving this problem as $\Rightarrow$ We make the Changes in the Loss Function so that using 1 Function, It Classifies the Classes and make Prediction.

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

$$L \Rightarrow \frac{-1}{m}\boxed{\sum_{i=1}^{m}}\boxed{\sum_{K=1}^{K}} Y_K^{(i)} \log(\hat{Y}_K^{(i)}) \Longrightarrow \boxed{L = \frac{-1}{m}\sum_{i=1}^{m}\sum_{K=1}^{K} Y_K^{(i)} \log(\hat{Y}_K^{(i)})}$$

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

$K = \{1, 2, 3\}$ $\quad [W_1, W_2, W_0]$

X_1	X_2	Y	$Y_{K=1}$	$Y_{K=2}$	$Y_{K=3}$
X_{11}	X_{12}	1	1	0	0
X_{21}	X_{22}	2	0	1	0
X_{31}	X_{32}	3	0	0	1

$$L = \frac{-1}{m} \sum_{i=1}^{m} \sum_{K=1}^{K} Y_K^{(i)} \log \left(\hat{Y}_K^{(i)}\right)$$

Expanding the Complete Formula:-

$$\underbrace{Y_1^{(1)}}_{=1} \log\left(\hat{Y}_1^{(1)}\right) + \boxed{\underbrace{Y_2^{(1)}}_{=0} \log\left(\hat{Y}_2^{(1)}\right)} + \boxed{\underbrace{Y_3^{(1)}}_{=0} \log\left(\hat{Y}_3^{(1)}\right)} +$$

$$\boxed{Y_1^{(2)} \log\left(\hat{Y}_1^{(2)}\right)} + Y_2^{(2)} \log\left(\hat{Y}_2^{(2)}\right) + \boxed{Y_3^{(2)} \log\left(\hat{Y}_3^{(2)}\right)} +$$

$$\boxed{Y_1^{(3)} \log\left(\hat{Y}_1^{(3)}\right)} + \boxed{Y_2^{(3)} \log\left(\hat{Y}_2^{(3)}\right)} + Y_3^{(3)} \log\left(\hat{Y}_3^{(3)}\right)$$

$\to 0 \quad 0 \quad 0 \quad \to 0$

$$L = \underbrace{Y_1^{(1)}}_{\text{Given}} \log\left(\hat{Y}_1^{(1)}\right) + \underbrace{Y_2^{(2)}}_{\text{Given}} \log\left(\hat{Y}_2^{(2)}\right) + \underbrace{Y_3^{(3)}}_{\text{Given}} \log\left(\hat{Y}_3^{(3)}\right)$$

Calculate....

$$\hat{Y}_1^{(1)} \Rightarrow \sigma\left(W_1^{(1)} * X_{11} + W_2^{(1)} * X_{12} + W_0^{(1)}\right)$$

Softmax Function

$$\boxed{\hat{Y}_1^{(1)} \Rightarrow \sigma\left(W_1^{(1)} * X_{11} + W_2^{(1)} * X_{12} + W_0^{(1)}\right)}$$

$$\boxed{\hat{Y}_2^{(2)} \Rightarrow \sigma\left(W_1^{(2)} * X_{21} + W_2^{(2)} * X_{22} + W_0^{(2)}\right)}$$

$$\boxed{\hat{Y}_3^{(3)} \Rightarrow \sigma\left(W_1^{(3)} * X_{31} + W_2^{(3)} * X_{32} + W_0^{(3)}\right)}$$

$$\begin{bmatrix} W_1^{(1)} & W_2^{(1)} & W_0^{(1)} \\ W_1^{(2)} & W_2^{(2)} & W_0^{(2)} \\ W_1^{(3)} & W_2^{(3)} & W_0^{(3)} \end{bmatrix}$$

$$\frac{\Delta L}{\Delta W_1^{(1)}}, \frac{\Delta L}{\Delta W_2^{(1)}}, \frac{\Delta L}{\Delta W_0^{(1)}}, \frac{\Delta L}{\Delta W_1^{(2)}}, \frac{\Delta L}{\Delta W_2^{(2)}}, \frac{\Delta L}{\Delta W_0^{(2)}}, \frac{\Delta L}{\Delta W_1^{(3)}},$$

$$\frac{\Delta L}{\Delta W_2^{(3)}}, \frac{\Delta L}{\Delta W_0^{(3)}}$$

"Derivates to be find, while Calculating the Gradient Descent."

Spiral