Softmax Regression

for multi class in logistic regression we use softmax Regression / multinomial logistic Regression.

softmax function

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

Here k = number of class

Dataset example

cgpa	iq	placed
7.3	130	1 (Yes)
4.0	120	2 (No)
8.0	110	3 (optout)

Here

$$\sigma(z)_1 = \frac{e^{z_1}}{e^{z_1}+e^{z_2}+e^{z_3}} \quad \sigma(z)_2 = \frac{e^{z_2}}{e^{z_1}+e^{z_2}+e^{z_3}} \quad ; \sigma(z_3) = \frac{e^{z_3}}{e^{z_1}+e^{z_2}+e^{z_3}}$$

Here $0 < \sigma < 1$

and $\sigma(z)_1 + \sigma(z)_2 + \sigma(z)_3 = 1$

example we have a dataset with three columns

1st approach for training

cgpa	iq	placed
		0
		1
		2

0 → Yes
1 – No
2 – optout

↓ data transformation using OHE (one hot encoding)

let query point $S_x = \{7, 70\}$

cgpa	iq	= 0 class	= 1 class	= 2 class
		1	0	0
		0	1	0
		0	0	1

Here the data is divided into three parts.

D1

cgpa	iq	= 0 class

↓

logistic Regression model-1

↓

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

↓

$Z_1 = 7 \times w_1^{(1)} + 70 \times w_2^{(1)} + w_0^{(1)}$

↓

$$\sigma(Y) = \frac{e^{z_1}}{e^{z_1} + e^{z_2} + e^{z_3}} = 0.40$$

D2

cgpa	iq	= 1 class

↓

logistic Regression model-2

↓

$w_1^{(1)}, w_2^{(1)}, w_0^{(1)}$

↓

$Z_2 = 7 \times w_1^{(2)} + 70 \cdot w_2^{(2)} + w_0^{(2)}$

↓

$$\sigma(N) = \frac{e^{z_2}}{e^{z_1} + e^{z_2} + e^{z_3}} = 0.35$$

D3

cgpa	iq	= 2 class

↓

logistic regression m-3

↓

$w_1^{(2)}, w_2^{(2)}, w_0^{(2)}$

↓

$Z_3 = 7 \times w_1^{(3)} + 70 \times w_2^{(3)} + w_0^{(3)}$

↓

$$\sigma(O) = \frac{e^{z_3}}{e^{z_1} + e^{z_2} + e^{z_3}} = 0.25$$

Hence here this student $S_x = \{7, 70\}$ has placed because $\sigma(Y) > \sigma(N)$ and $\sigma(Y) > \sigma(O)$

* in that way approach in the softmax prediction the prediction is done.
* The above training approach is very complex

2nd approach for training the model

using loss function:

Here the another approach is that the we modify the loss function in such a way it is useful in softmax Regression

previous loss function

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

in softmax Regression

$$L = \frac{-1}{m} \underbrace{\sum_{i=1}^{m}}_{\text{for number of raw.}} \underbrace{\sum_{k=1}^{K}}_{\text{for class in each raw}} Y_k^{(i)} \log(\hat{Y}_K^{(i)})$$

example:

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

$$L = Y_1^{(1)}\log(\hat{Y}_1^{(1)}) + Y_2^{(1)}\log(\hat{Y}_2^{(1)}) + Y_3^{(1)}\log(\hat{Y}_3^{(1)}) +$$
$$Y_1^{(2)}\log(\hat{Y}_1^{(2)}) + Y_2^{(2)}\log(\hat{Y}_2^{(2)}) + Y_3^{(2)}\log(\hat{Y}_3^{(2)}) +$$
$$Y_1^{(3)}\log(\hat{Y}_1^{(3)}) + Y_2^{(3)}\log(\hat{Y}_2^{(3)}) + Y_3^{(3)}\log(\hat{Y}_3^{(3)})$$

$$L = Y_1^{(1)}\log(\hat{Y}_1^{(1)}) + Y_2^{(2)}\log(\hat{Y}_2^{(2)}) + Y_3^{(3)}\log(\hat{Y}_3^{(3)})$$

Here

$$\hat{Y}_1^{(1)} = \sigma\left(W_1^{(1)} X_{11} + W_2^{(1)} X_{12} + W_0^{(1)}\right)$$
$$\hat{Y}_2^{(2)} = \sigma\left(W_1^{(2)} X_{21} + W_2^{(2)} X_{22} * W_0^{(2)}\right)$$
$$\hat{Y}_3^{(3)} = \sigma\left(W_1^{(3)} X_{31} + W_2^{(3)} X_{32} + W_0^{(3)}\right)$$

softmax function

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

Now calculate

$$\frac{\partial L}{\partial W_1^{(1)}}, \frac{\partial L}{\partial W_2^{(1)}}, \frac{\partial L}{\partial W_0^{(1)}} \text{ ------- } 9 \text{ times}$$

Now then

loop → 1000 epochs (approx)

$$W_1^{(1)} = W_1^{(1)} - \eta \frac{\partial L}{\partial W_1^{(1)}}$$

$$W_2^{(1)} = W_2^{(1)} - \eta \frac{\partial L}{\partial W_2^{(1)}}$$

$$\vdots \qquad \vdots \qquad \vdots$$