Introduction to logistic Regression

Wednesday, 27 March 2024 10:44 AM

Logistic Regression is a type of regression that is used when the output is binary, that is Yes/No

Probabilities > 0.5 result is likely to 1
" $<= 0.5$ " " " " 0

Logistic regression aims to do the following

① independent variable ② selected observation ③ set of given independent variable

→ will model the probabilities of a particular Event happening or not.

⇒ independent variable can be categorical or non categorical

Logistic Regression Equation

$$z = b + w_1 x_1 + w_2 x_2 + \dots w_n x_n$$

where $x_1, x_2, x_3 \dots$ are independent variables

Logistic Regression is defined by a sigmoid function

output $\boxed{Y = \frac{1}{(1 + e^{-z})}}$

$$1 + e^{-z} = \frac{1}{Y}$$

$$e^{-z} = \frac{1}{Y} - 1$$

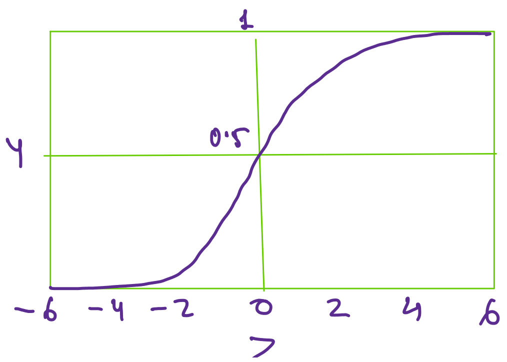
$$e^{-z} = \frac{1 - Y}{Y}$$

$$e^{z} = \frac{Y}{1 - Y}$$

Taking ln both side

$$\ln e^{z} = \ln \frac{Y}{1-Y}$$

$$z = \ln\left(\frac{Y}{1-Y}\right)$$



Gradient descent Example

(1) $Y = mx + c$

(0.5 , 1.4)
(2.3 , 1.9)
(2.9 , 3.2)

Step 1: Sum of Square Residual

$$RSS = (1.4 - (c + m*0.5))^2 + (1.9 - (c + m*2.3))^2 + (3.2 - (c + m*2.9))^2$$

Step 2: Derivatives of RSS with Respect to Intercept (C)

$$d/dc = 2*(c + m*0.5 - 1.4) + 2*(c + m*2.3 - 1.9) + 2*(c + m*2.9 - 3.2)$$

$$d/dm = 2*(c + m*0.5 - 1.4)*0.5 + 2*(c + m*2.3 - 1.9)*2.3 + 2*(c + m*2.9 - 3.2)*2.9$$

assume $m = 1$ $c = 0$

$$\boxed{\frac{d}{dc} = -1.6 \qquad \frac{d}{dm} = -0.8}$$

$$new\_m = m - \alpha \cdot \frac{d}{dm} = 1 + 0.01*0.8 = 1.008$$
$$new\_c = c - \alpha \cdot \frac{d}{dc} = 0 + 0.01*1.6 = 0.016$$

new_m = ?
new_c = ?

1000 iteration

correct m & c

c = 0.94
m = 0.64

V.V.Imp Confusion matrix

		Actual Value	
		Positive	Negative
Predicted value	Positive	TP	FP
	Negative	FN	TN

① Accuracy is a metric that measures how often a machine learning model correctly predicts the outcome

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

② Precision refers to the number of true positives divided by the total number of positive prediction

- It helps us to visualize the reliability of the machine learning model in classifying the model as positive

$$Precision = \frac{TP}{TP + FP}$$

③ Recall: The Recall measures the model's ability to detect positive samples

* higher Recall more positive samples are detected

$$Recall = \frac{TP}{TP + FN}$$

F1 Score F1-score is the evaluation matrix that combines two matrices: Precision and Recall, into a single metric by taking their harmonic mean.

$$F1\ Score = 2 * \frac{recall \times precision}{recall + precision}$$

$$Harmonic\ Mean = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} - - - \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

$$= \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \Rightarrow \frac{2 * (Precision * Recall)}{Precision + Recall}$$