Logistic Regression

06 April 2022 12:26

When my Target variable is categorical

Ex: Gender, Location, Status, Education, Exam, interview,

1. Binary outcomes/Binary class

Only two outcome

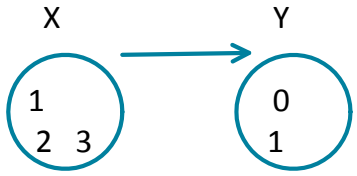
Yes/No, True/ False, 0 /1 , Male/Female, Married/ Unmarried, Positive/Negative

2. Multiclass

Location, Education

If my Target variable is Binary class ---> Logistic Regression

Target variable ----------> 0 or 1



Y = Bo + b1x1 + b2x2 ------> -inf to + inf

Y - Y_pred --> error will be higher

$f(x) = x^2$

$f(x) = 2x$

$f(x) = 2x + 1$ --> mx + c

Bo + b1x1 + b2x2 ---- BX

$$\text{Logit} = \frac{e^{(BX)}}{1 + e^{(BX)}} = 0 \text{ to } 1$$

A	B	C	D	E	F	G	H	I	J	K
age(x)	Y	Y_pred	>0.5							
56	1	0.96	1						1	Pos
23	0	0.45	0						0	Neg
35	0	0.23	0				Predicted values			
45	1	0.15	0			TN	0	1		
36	0	0.89	1			0	4	1	FP	Type I Error
65	1	0.96	1		Actual values	1	1	4		
41	0	0.36	0				FN		TP	
63	1	0.87	1				Type II Error			
23	0	0.16	0							
75	1	0.61	1							

Confusion matrix

Sensitivity : Percentage of positives that are successfully classified as positive.

True positive Rate

TPR : TP / (TP + FN)

Specificity : Percentage of negatives that are successfully classified as negatives.

TNR : TN / (TN+FP)

When Specificity is a High Priority, Testing a Medicine is good or Poisonous

		Actual Classes			
		0 (Negative)	1 (Positive)	1 +ve	Good
				0 -ve	Posinous
Predicted classes	0 (Negative)	True Negative (TN). Actually Medicine is good and predicted them as good . Hence Recommend to use.	False Negative(FN) Actually Medicine is good and predicted them as Poisonous. Hence reject all the bulk.		
	1 (Positive)	False Positive(FP) Actually Medicine is Poisonous and predicted them as good. Recommend to use.	True Positive (TP). Actually Medicine is Poisonous and predicted them as Poisonous . Hence reject all the bulk.		

Precision:

The approach here is to find what percentage of the model's positive (1's) predictions are accurate.

Precision is calculated as **the number of correct positive predictions (TP) divided by the total number of positive predictions (TP + FP).**

(TP) / (TP + FP)

F1 Score:

A good model should have a good precision as well as a high recall. So ideally, I want to have a measure that combines both these aspects in one single metric –the F1 Score.

F1 Score = (2 * Precision * Recall) / (Precision + Recall)