Mithilesh singh

Logistic Regression: building classification model

Logistic regression is a machine learning classification model with similar as Linear Regression

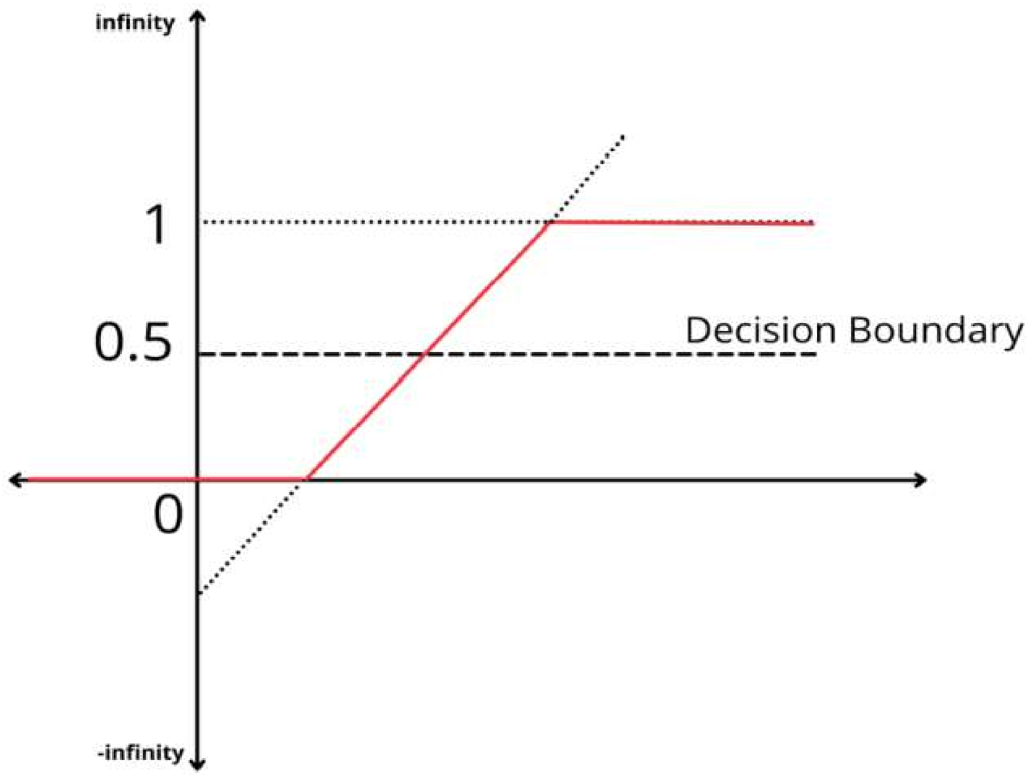
but it’s not used to predict a continuous outcome. -Not for Regression.

Instead, it is a statistical BINARY classification model.

Binary classes defined as success and failure: -

0 (Failure) and 1 (Success), Yes / No ,Selected / Rejected ,True/ False

Decision boundary:



The S-shaped curve is a sigmoid curve. Logistic regression function is also called sigmoid function.

p is a non-linear function, say, an exponential function.

Sigmoid Function: $$\boldsymbol{p = \frac{1}{1+e^{-y}}}$$

For -∞

$$p = \frac{1}{1+e^{-(-\infty)}}$$

$$p = \frac{1}{\infty}$$

Anything divided by ∞ is 0, Hence for -∞ it is **p = 0**

$$p = \frac{1}{1+e^{-y}}$$

For ∞

$$p = \frac{1}{1 + e^{-(+\infty)}}$$

$$p = \frac{1}{1 + 0}$$

Hence for ∞ it **is p = 1**

logistic regression formula

Sigmoid function:-

Mathematical function having a characteristic "S"shaped curve or sigmoid curve.

$$p = \frac{1}{1 + e^{-y}}$$

Transformation of the sigmoid function

$$p(1 + e^{-y}) = 1$$

$$p + pe^{-y} = 1$$

$$pe^{-y} = 1 - p$$

$$e^{-y} = \frac{1 - p}{p}$$

$$log(e^{-y}) = log\left(\frac{1 - p}{p}\right)$$

$$-y = ln\left(\frac{1 - p}{p}\right)$$

$$y = ln\left(\frac{p}{1 - p}\right)$$

Hence we have converted the sigmoid function such that now it is expressed as providing the value of Y that is the target variable

$$y = ln\left(\frac{p}{1 - p}\right)$$

Here (P/ (1-P)) is called "odds ratio".

Hence Y can be defined as the log of the odds ratio.

Y is also called as Log Odd Function or Logistic Function or Log it Function

Applying Linear Equation Formula

$$exp\left(log\left(\frac{p}{1-p}\right)\right) = exp\left(a + b_1 x_1\right)$$

$$\Rightarrow \frac{p}{1-p} = exp\left(a + b_1 x_1\right) \quad [Since\ exp(logx) = x]$$

Solving the above equation for p, we get

$$\mathrm{p} = exp\left(a + b_1 x_1\right) - \mathrm{p}\ exp\left(a + b_1 x_1\right)$$

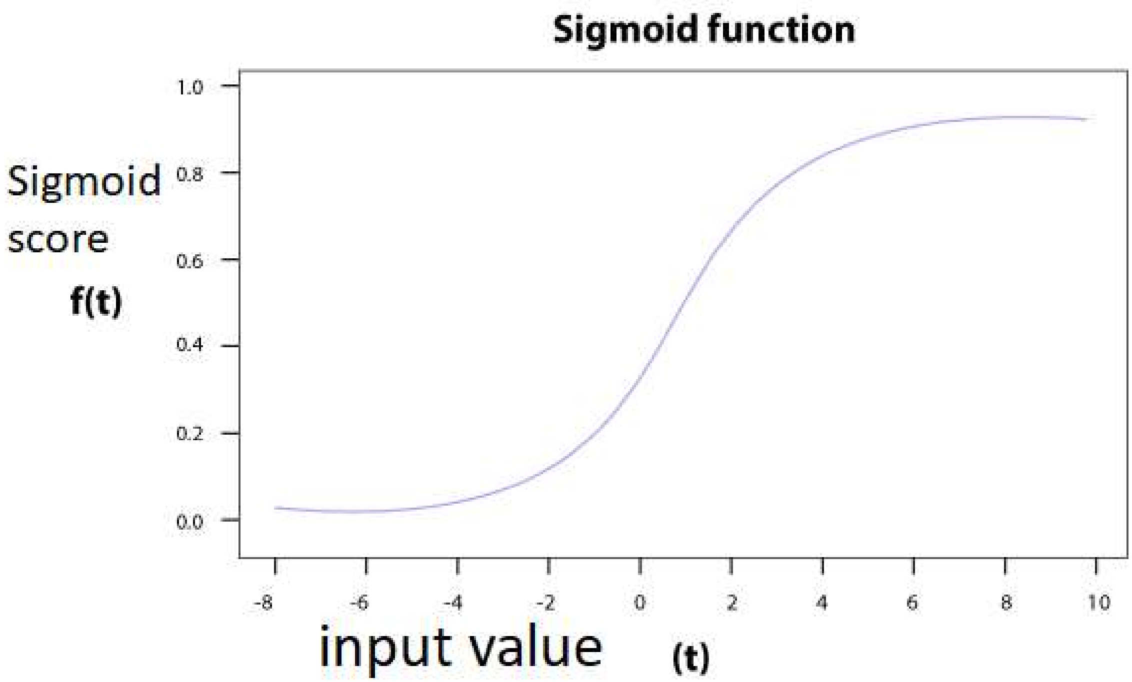
$$p = \frac{exp(a+b_1x_1)}{1+exp(a+b_1x_1)}$$

Finally, we will get the value of probability p in the range of 0 to 1.

Once we calculate the probabilities, we will plot them to get an s-curve. Now,
We just keep rotation the log(odds) line and projecting the data points onto it and then transforming it to probabilities and calculating the log-likelihood. We will repeat this process until we maximize the log-likelihood.

Keeping Threshold =0.5
Above 0.5 output belongs to class 1

Below 0.5 output belongs to class 0




```
from sklearn.linear_model import LogisticRegression
lr=LogisticRegression()
lr.fit(X_train1,Y_train)

Y_pred_lr_train=lr.predict(X_train1)
Y_pred_lr_test=lr.predict(X_test1)

print("LOGISTIC REGRESSION TRAINING ACCURACY ",accuracy_score(Y_train,Y_pred_lr_train))

print("######"*20)

print("LOGISTIC REGRESSION TESTING ACCURACY ",accuracy_score(Y_test,Y_pred_lr_test))
```

MATHEMATICS of Logistic Regression: -

Maximum Likelihood Estimation

logistic regression formula
Sigmoid function:-
Mathematical function having a characteristic "S"shaped curve or sigmoid curve.

$$p = \frac{1}{1 + e^{-y}}$$

The logistic regression function converts the values of **logits** also called **log-odds** that range from **−∞ to +∞** to a range between **0 and 1**.

Transformation of the sigmoid function

$$p(1 + e^{-y}) = 1$$

$$p + pe^{-y} = 1$$

$$pe^{-y} = 1 - p$$

$$e^{-y} = \frac{1 - p}{p}$$

$$log(e^{-y}) = log\left(\frac{1 - p}{p}\right)$$

$$-y = ln\left(\frac{1 - p}{p}\right)$$

$$y = ln\left(\frac{p}{1 - p}\right)$$

Odds is defined as the ratio of the probability of occurrence of a particular event to the probability of the event not occurring.

$$Odds = \frac{P}{1-P}$$

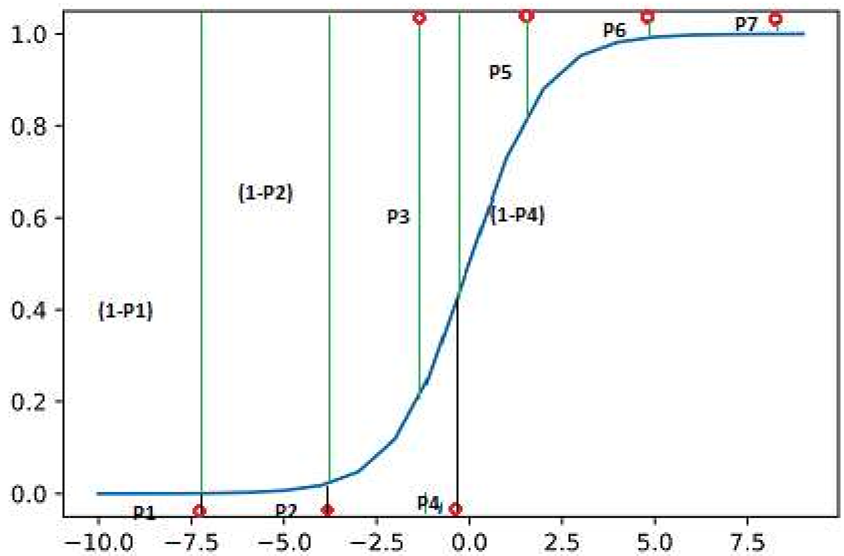
$$\ln(Odds) = \ln\left(\frac{P}{1-P}\right) = y$$

$$\ln\left(\frac{P}{1-P}\right) = \beta 0 \; + \; \beta 1x$$

Logistic Regression predicts output variable as 0 or 1, we need to find the best fit sigmoid curve.

Now a cost function tells how close the values are from actual. So here we need a cost function which maximizes the likelihood of getting desired output values. Such a cost function is called as **Maximum Likelihood Estimation (MLE)** function.

The cost function should be such that it maximizes the probability of predicted values being close to the actual ones.



Sigmoid curve and probabilities

From the above figure, There are 7 points and seven associated probabilities P1 to P7, classified as class 0 and 1.

For points to be class 0, we need the probabilities of P1, P2 and P4 to be as minimum as possible and for points to be class 1, we need the probabilities of P3, P5, P6 and P7 to be as high as possible, for correct classification.

We can also say that (1-P1), (1-P2), P3, (1-P4), P5, P6 and P7 should be as high as possible.
The joint probability is nothing but the product of probabilities. So the product:
[**(1-P1)*(1-P2)* P3*(1-P4)*P5*P6*P7**]
should be maximum.

This joint probability function is our cost function or **the Maximum Likelihood Estimation (MLE)** function.

This joint probability function or the Cost function which should be maximized in order to get a best fit sigmoid curve. Or we can say predicted values to be close to the actual values.

Now coming to our cost function, let **J(z)** be a function of **z** such that

J(z) = P(Y ; z) = P(Y1, Y2 … Yn ; z)

The assumption here is that all Y are independent.

$$J(z) = {}_{i=1}^{n}\Pi P(Yi\ ; z)$$

Cost function is product of all probabilities P(Yi)

Taking natural log on both sides:

$$\ln(J(z)) = \ln({}_{i=1}^{n}\Pi P(Yi\ ; z))$$

Taking log of both sides

Since log of product becomes summation:

$$\ln(J(z)) = \sum_{i=1}^{n} \ln\left(P(Yi\,; z)\right)$$

log of product is summation

J(z) can also be written as L(z|Yi)

(L for Likelihood).

For a given value of **z** and observed sample Yi , this function gives the probability of observing the sample values.

So if Yi=1 the expression becomes **z** and

if Yi is 0 the expression becomes **1-z**:

$$\ln(J(z)) = L(z \mid Yi) = \sum_{i=1}^{n} \ln\left(z^{Yi} * (1-z)^{1-Yi}\right)$$

Solving this equation further we get:

$$\ln(J(z)) = L(z \mid Yi) = \sum_{i=1}^{n} Yi * ln(z) + (1 - Yi) * ln(1 - z)$$

Differentiating this equation with respect to **z** and setting the derivative to zero, we calculate the maxima using closed form solution:

$$\frac{d}{dP}\left(\sum_{i=1}^{n} Yi * ln(z) + (1 - Yi) * ln(1 - z)\right) = 0$$

Solving further, this equation becomes:

$$\frac{z}{1-z} = \sum_{i=1}^{n} \frac{Yi}{1-Yi}$$

The right side term represents the ratio of number of 1 s to number of 0 s. Thus the function achieves a maximum at: -

$$z = \sum_{i=1}^{n} \frac{Yi}{n}$$