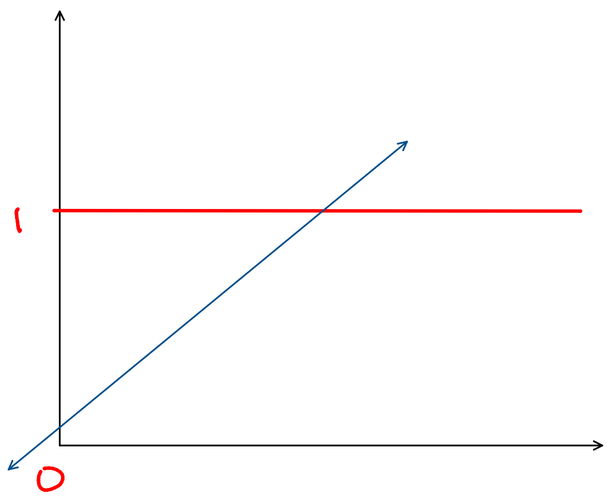
Logistic Regression

27 January 2024 09:16 AM

- Logistic regression is a supervised learning algorithm.
- It is used for binary classification
- It is based on the regression approach



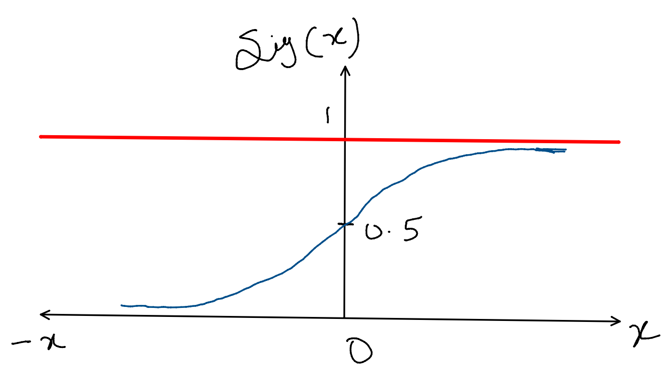
- A Regression approach can be used to predict values 0 to 1.
- But the predictions may go above 1 or below 0.
- In logistic regression the model gives a probability which is converted to labels based on threshold value.
- Probability above 1 or below 0 does not make sense.
- We use sigmoid function to limit the output between 0 to 1.

Input $\longrightarrow$ Regression $\longrightarrow$ Sigmoid $\longrightarrow$ Output

* SIGMOID FUNCTION :-

$$\boxed{Sig(x) = \frac{1}{1 + e^{-x}}}$$

$e = 2.71..$



CASE 1 : x is very large +ve value

$\Rightarrow e^{-x} \approx 0$

$\Rightarrow 1 + e^{-x} \approx 1$

$\Rightarrow Sig(x) \approx 1$ (still loss)

CASE 2: x is very large -ve value

$\Rightarrow e^{-x} \approx +\infty$ (very large value)

$\Rightarrow 1+e^{-x} \approx +\infty$

$\Rightarrow Sig(x) \approx 0$

CASE 3: x is 0

$\Rightarrow e^{-x} = e^{0} = 1$

$\Rightarrow 1+e^{-x} = 1+1 = 2$

$\Rightarrow Sig(x) = 0.5$

* LOGISTIC REGRESSION:

$$\boxed{\begin{aligned} Z &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots \beta_n x_n \\ \hat{y} &= Sig(z) \end{aligned}}$$

$$\hat{y} = Sig(\beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots \beta_n x_n)$$

* The Cost function used for logistic regression is log loss.

$$\boxed{LogLoss = -\frac{1}{n} \Sigma \, Log(P_c(Y_i))}$$

ACTUAL LABEL	PRED. PROB.	CORRECTED PROB.
1	0.9	0.9
1	0.8	0.8
0	0.1	0.9
0	0.3	0.7
1	0.2	0.2
0	0.7	0.3

$$P_c = \begin{cases} P, & y = 1 \\ 1 - P, & y = 0 \end{cases}$$

$$\text{Log Loss} = -\frac{1}{6}\left[\log(0.9) + \log(0.8) + \log(0.9) + \log(0.7) + \log(0.2) + \log(0.3)\right]$$

$$= 0.87$$

$$\boxed{\text{Log Loss} = -\frac{1}{n}\sum\left[y_i \cdot \log(p) + (1 - y_i)\log(1 - p)\right]}$$