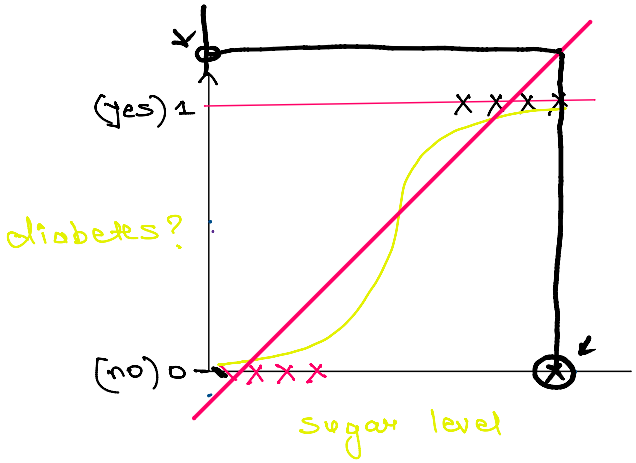
Logistic Regression

21 August 2024 06:48

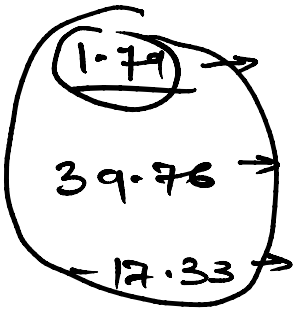


encoding

$-\infty$ to $+\infty$

yes → 1

No → 0



In logistic regression, we fit 'S' like curve to make classification.

Requirements →

1. Want output to be 0 or 1.
2. Not much affected by the outliers.

$y = mx + c$

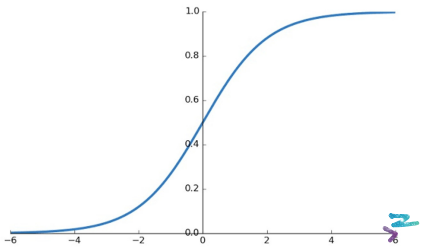
$\hat{y} = \beta_1 x + \beta_0$

$-\infty$ to $+\infty$

0 to 1

We use 'sigmoid function' to fulfill the above requirements →

$$g(z) = \frac{1}{1+e^{-z}}, \quad 0 < g(z) < 1$$



Mathematical Intuition →

y-pred → $\hat{y} = \beta_0 + \beta_1 x$

↓

gives output b/w $-\infty$ to $+\infty$

↓

But in classification we need the

But in classification we need the output to be Yes(1) or No(0).

↓

Then we use Sigmoid Functions to achieve this.

↓

$$g(z) = \frac{1}{1+e^{-z}}$$

↓

will give us the output b/w 0 and 1.

Training Data

↓

Linear Regression

↓

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots$$

BFL eqn. → $1.8 + 1.1x_1 + 0.7x_2 + (-0.15)x_3 + \dots$

Test Data / Future Data.

↓

give to

↓

$$z = \hat{y} = 1.8 + 1.1x_1 + 0.7x_2 + (-0.15)x_3 + \dots$$

$z = 14.2$ (suppose)

Heart Dis Pat.

CPL → 4

Chol → 35.8

BP → 110.5

$z = 14.2$ (suppose)

↓

But the above output is not very useful for classification so we give this value to 'sigmoid function' further.

e = euler's constant = 2.71

$$g(z) = \frac{1}{1+e^{-z}}$$ → only changeable value

$$= \frac{1}{1+e^{-14.2}}$$

$$= 0.99 = g(z)$$

Always represents the probability of final output being Yes or 1

99% possibility that this person is Heart Disease Patient.

↓

Compare the above probability with a pre-defined cut-off / threshold value.

0.5

if, $g(z) > 0.5$ → Prediction is Yes / 1.

if, $g(z) \leq 0.5$ → Prediction is No / 0.

Patient 2 →

CP = 1

Chol = 19.2

BP = 56.5

↓

BFL

↓

$z = \hat{y} = -3.49$ (suppose)

↓

$$g(z) = \frac{1}{1+e^{-z}}$$

$$= \frac{1}{1+e^{3.49}}$$

$$g(z) = 0.0296$$

↓

2.95% possibility that this person is a Heart Disease Patient.

100 − 2.95% possibity that this person is Non Heart Disease Patient.

↓

Compare with 0.5

$$g(z) = 0.0296$$

here, $g(z) \leq 0.5$

↓

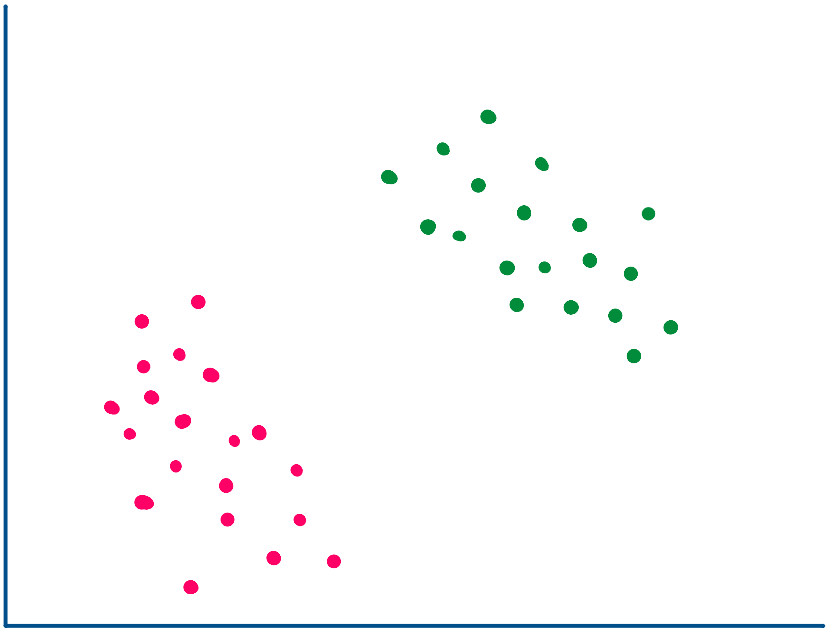
↓

Final output is 0/Non Heart Dis Patient.

if $z >= 0 \rightarrow \frac{1}{1+e^{-z}} \rightarrow$ will always give the $g(z)$ as probability ≥ 0.5.

if $z < 0 \rightarrow \frac{1}{1+e^{-z}} \rightarrow$ will always give the $g(z)$ as probability < 0.5.

$0 < g(z) < 1$



$z = \hat{y} = \beta_0 + \beta_1 x$

↓

$\frac{1}{-z}$

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

↓

$$\frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

$\beta_0 = 1.2 \quad \beta_1 = 2.8$

↓

Adjust these betas