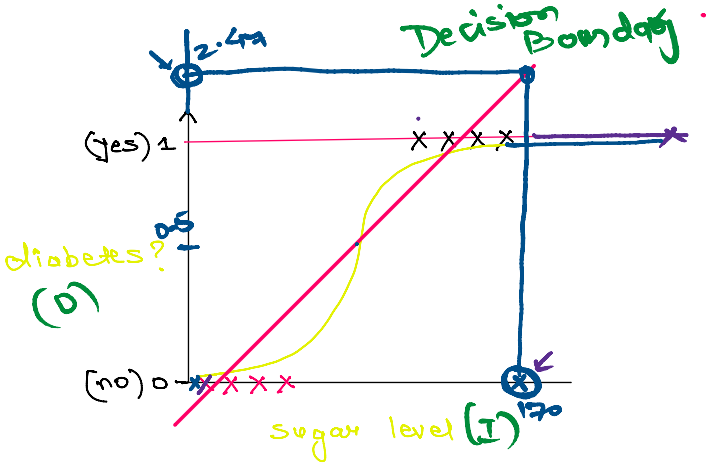
Logistic Regression

24 November 2024 19:58



Inp	output
Sugar level	Diabetic?
0 K5	Yes
1 110	No
2 150	Yes
3 95	No
⋮	⋮

Yes/1 No/0

↓

Encoded

$y = mx + c$

↓

Line

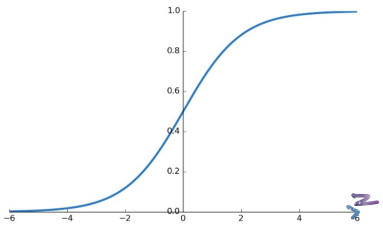
In logistic regression, we fit 'S' like curve to make classification.

Requirements →

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

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

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



Lin Reg → $y = mx + c$ → $-\infty$ to $+\infty$

↓

but we want only 0 or 1 in classification.

Log Reg → $g(z) = \frac{1}{1+e^{-z}}$ → 0 to 1

↓

↓

we can easily interpret it get the final output as 0 or 1.

Mathematical Intuition →

$$\hat{y} = mx + c$$

↓

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

↓

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

↓

Keeping this in mind, we will use 'Sigmoid Function' instead to achieve the result easily.

↓

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

↓

output will be $0 < g(z) < 1$.

Training Data (classification)

↓

Regression Line

↓

$c + m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + \dots$

↓

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

HD Pred

x_1 : CPL → 4

x_2 : Chol → 35.8

x_3 : BP → 110.5

Test Data

↓

give it to

↓

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

$z = \underline{14.2}$ (suppose)

↓

But the above output is not useful for classification so we give this output (14.2) to sigmoid function

↓

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

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

e = euler's constant

e = 2.71

$g(z) = 0.9999$

Always represents the probability that the final output is Yes/1.

99.99% possibility that the person is a Heart Disease Patient or Yes/1.

Compare the output of Sigmoid Function which we got above as 0.9999 with a pre-defined

we got above as 0.9999 with a pre-defined cut-off / threshold value.

0.5

if $g(z) > 0.5 \rightarrow$ Prediction is Yes / 1.

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

Patient 2 →

x_1 : CP = 1

x_2 : chol = 10.2

x_3 : BP = 75

↓

Reg Line

↓

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

$z = -3.49$

↓

$$g(z) = \frac{1}{1+e^{-(-3.49)}}$$

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

$$= 0.0295$$

$$= 2.95\%.$$

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

100 − 2.95% possibility that this person is a Healthy (Non Heart Disease) Person.

↓

Compare this with 0.5

$g(z) = 0.0295$

↓

here, $g(z) \leq 0.5$

↓

Final predicted output is 0/ Healthy Person.

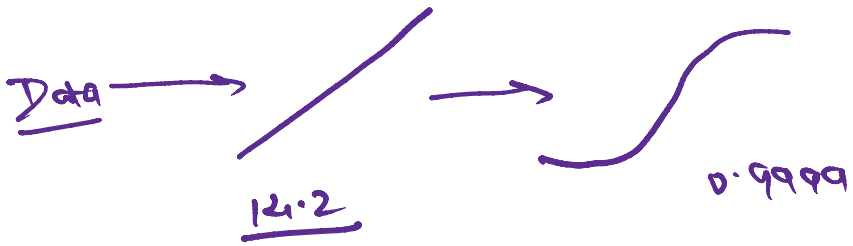
if $z >= 0 \rightarrow \frac{1}{1+e^{-z}} \rightarrow$ will always give the $g(z) \geq 0.5$

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

Final Form of equation →

$$z = \hat{y} = m_1 x_1 + m_2 x_2 + m_3 x_3 + c$$

$$g(z) = \frac{1}{1 + e^{-(m_1x_1 + m_2x_2 + m_3x_3 + \dots + c)}}$$



Variance & SD

100 200 300 600 ←

10 20 30 60 55 ←

age

IQR