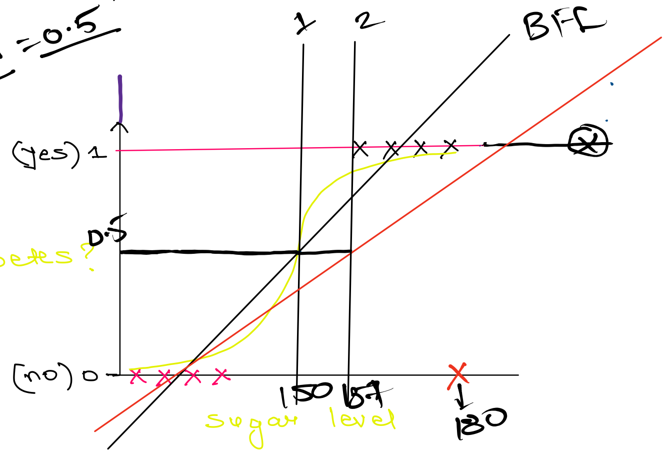
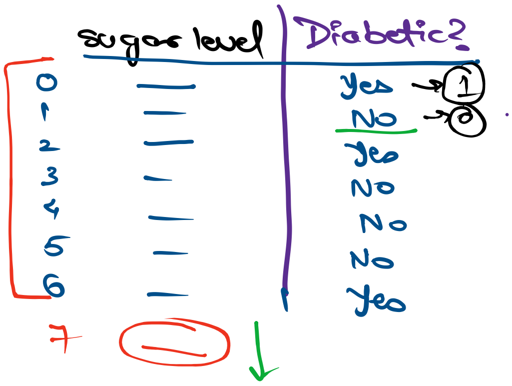
Logistic Regression 23 May

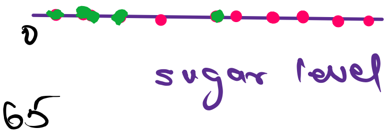
30 January 2025 07:00

cut off = 0.5





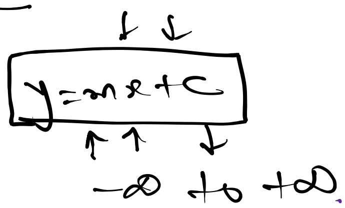
In logistic regression, we fit 'S' like curve to make classification.



180 → 1.65

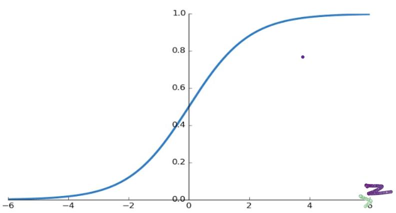
Requirements →

1. want output to be 0 or 1.
2. Not much affected by the outlier.

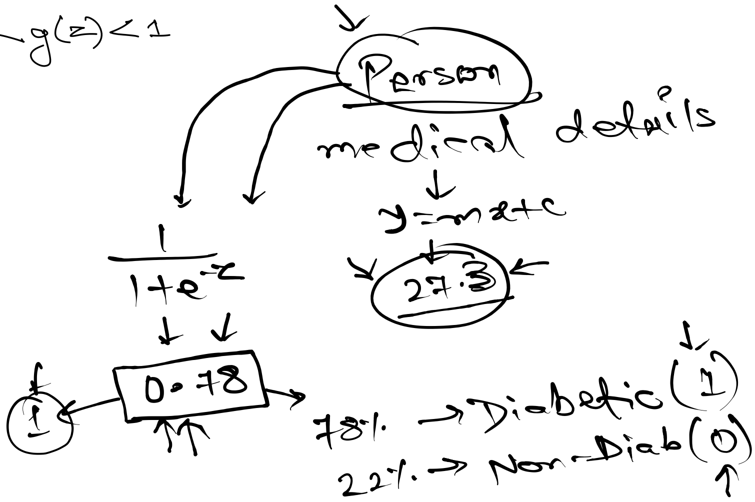


We use 'sigmoid function' to fulfill the above requirements →

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



50%



Linear Regression → Equation of Best Fit Line

↓

$\hat{y} = mx + c$

↓

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

↓

because of this it is not suitable for classification

Logit

Logistic Regression → Sigmoid Function

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

↓

gives an output b/w $\underline{0 \text{ to } 1}$ → Probability

Mathematical Intuition →

Heart Disease Prediction → Training Data

Training Data → Regression Line

$$\hat{y} = 1.1x_1 + 0.7x_2 + 0.15x_3 + 0.29x_4 + 1.8$$

1st Training row

Person	Yes/1
x_1: CPL	→ 4
x_2: Chol	→ 35.8
x_3: BP	→ 110.5
x_4: Age	→ 43

$$\hat{y} = 1.1x_1 + 0.7x_2 + 0.15x_3 + 0.29x_4 + 1.8$$

$z = \hat{y} = 14.2$ (suppose)

↓

$g(z) = 0.5$

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

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

$g(z) = 0.9999$

e = euler's number = 2.71

Yes → HDP

No → Healthy

Always represent the probability that the predicted output is "Yes/1".

> 0.5

99.99% possibility that the person is a Heart Disease Patient.

↓

Compare the output "g(z)" given by the sigmoid function with a pre-defined cut-off value of 0.5.

if $g(z) > 0.5 \longrightarrow$ output is Yes/1

if $g(z) \leq 0.5 \longrightarrow$ output is No/0.

In our case,

$$\underset{g(z)}{0.9999} > 0.5$$

↓

Yes/1

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

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

Final Form of Logistic Regression equation.

1D

$$\left[z = \hat{y} = m_1 x_1 + m_2 x_2 + m_3 x_3 + m_4 x_4 + c \right]$$

↓

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

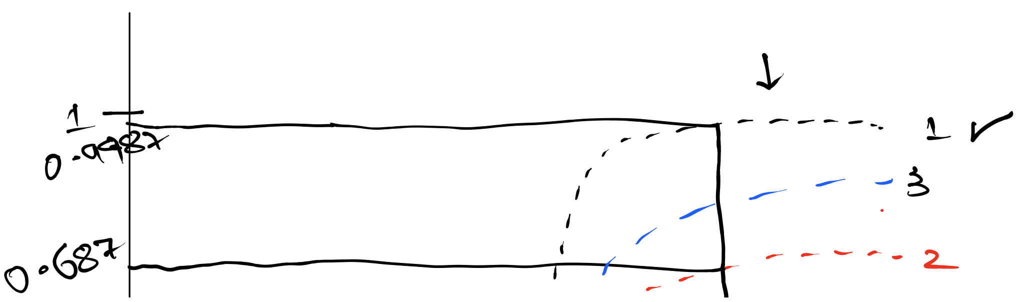
$$\boxed{g(z) = \frac{1}{1+e^{-(m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + c)}}}$$

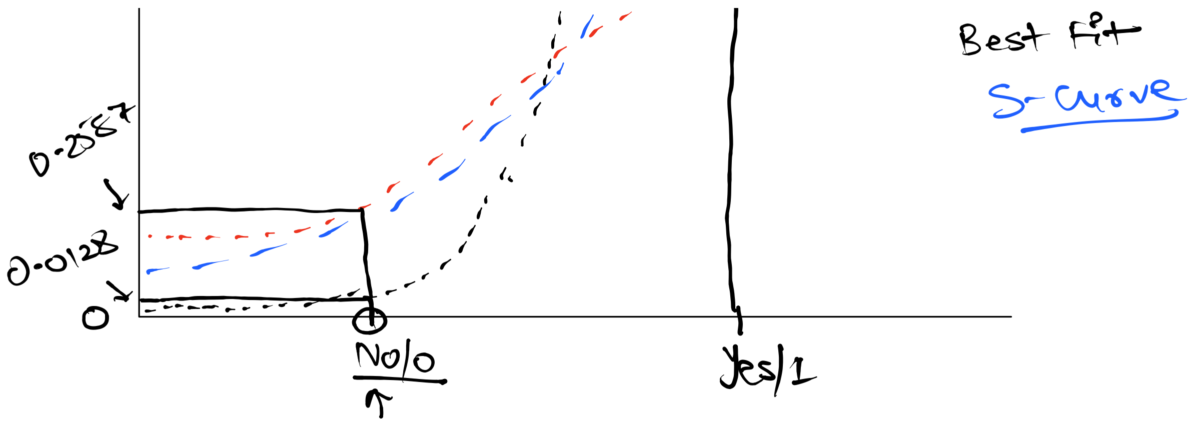
How do we finalize the best values for the above equation →

<u>Line 1</u>: $1.1x_1 + 0.7x_2 + 0.15x_3 + 0.29x_4 + 1.8$

<u>Line 2</u>: $0.91x_1 + 0.42x_2 + 1.13x_3 + 0.89x_4 + 1.35$

Line 1	y-train	Line 2
0.9999	1	0.8215
0.99827	1	0.78259
0.000125	0	0.0139
0.0000428	0	0.0284
⋮	1	⋮
	0	
	1	
	1	
	1	
	1	
	1	





$$\rightarrow \quad = \frac{1}{1+e^{-(1.1x_1+0.7x_2+0.15x_3+0.29x_4+1.8)}}$$

$\downarrow$

Testing Data