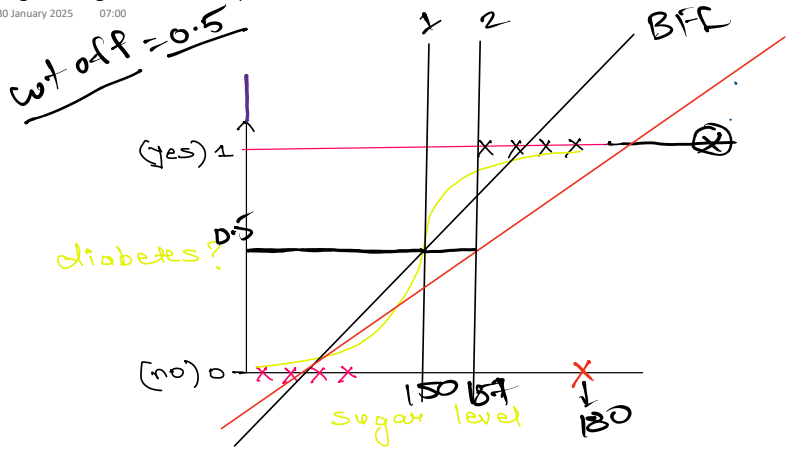
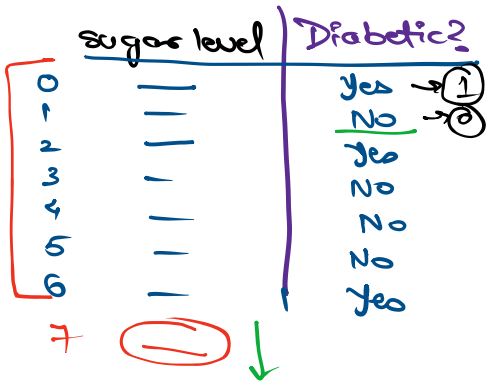
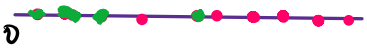
Logistic Regression 23 May

30 January 2025 07:00





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



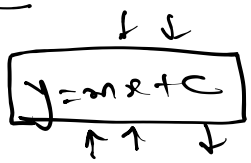
Requirements →

(1) Want output to be 0 or 1.

(2) Not much affected by the outlier.

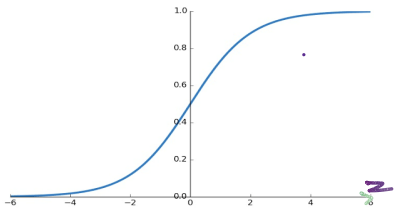
We use 'sigmoid function' to fulfill the above requirements →

180 → 1.65

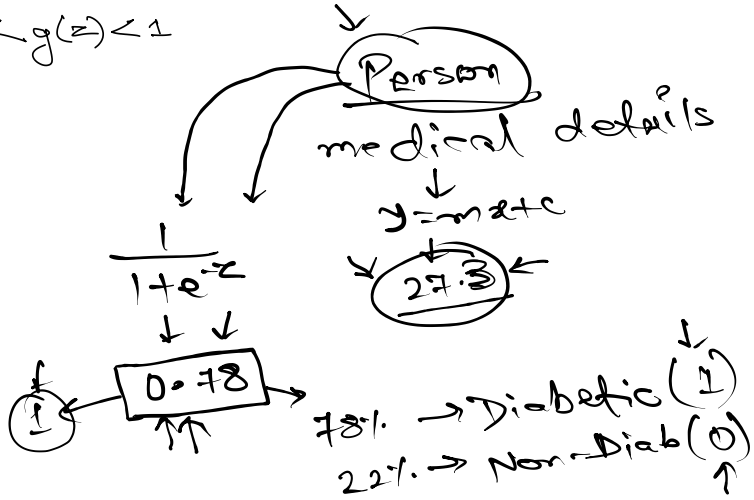


−∞ to +∞

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



50%



Linear Regression → Equation of Best Fit Line

↓

$\hat{y} = mx+c$

↓

gives us the output b/w −∞ to +∞

↓

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 0 to 1 → Probability

Mathematical Intuition →

Heart Disease Prediction

→ 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)

$g(z) = 0.5$

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

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

e = euler's number = 2.71

Yes → HDP

No → Healthy

$g(z) = 0.9999$

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 $g(z) > 0.5$ ⟶ output is yes/1

if $g(z) \leq 0.5$ ⟶ output is No/0.

In our case,

$0.9999 > 0.5$

$g(z)$

↓

yes/1

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

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

Final Form of Logistic Regression equation:

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

↓

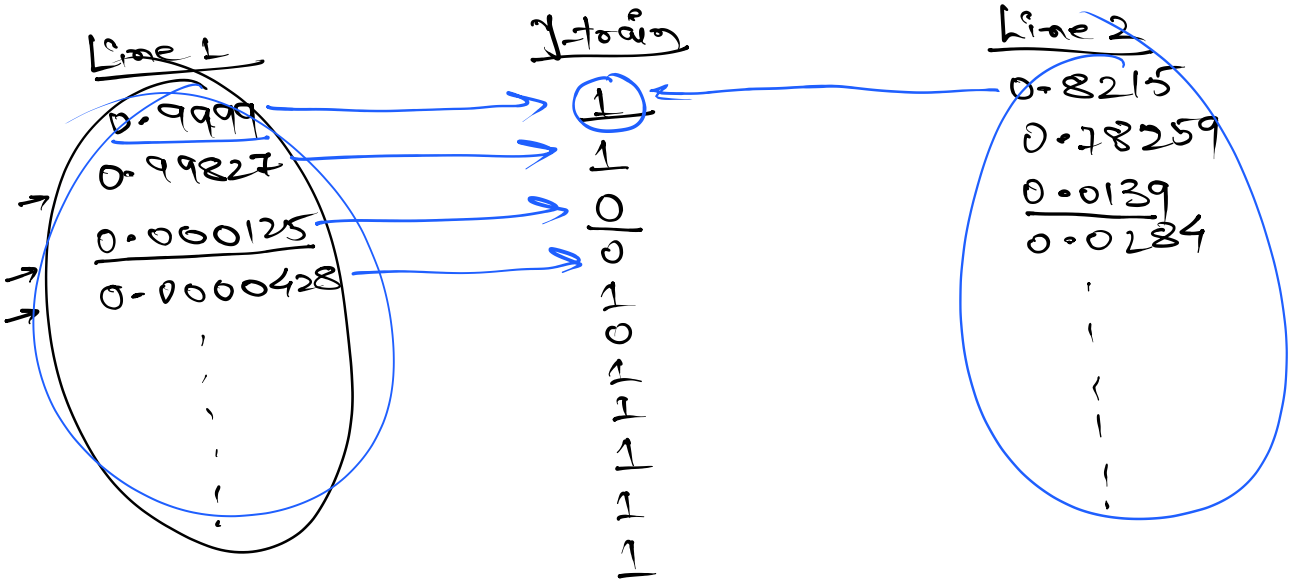
$$g(z) = \frac{1}{1+e^{-z}}$$

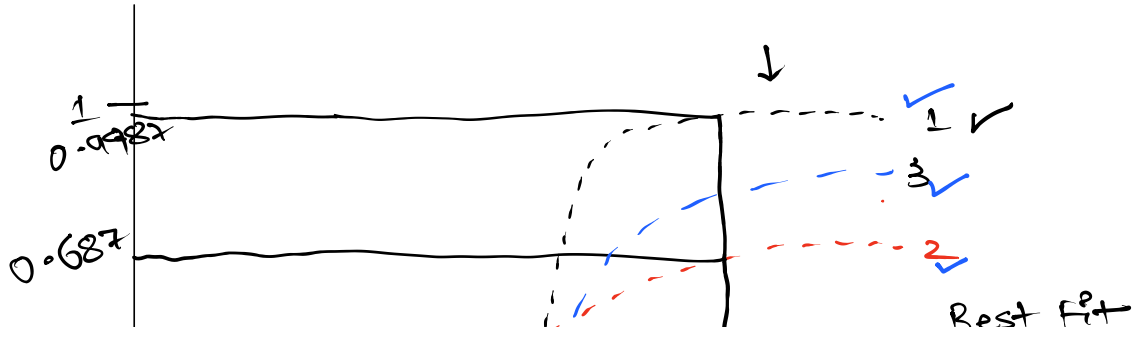
$$g(z) = \frac{1}{1 + e^{-(m_1 x_1 + m_2 x_2 + m_3 x_3 + m_4 x_4 + C)}}$$

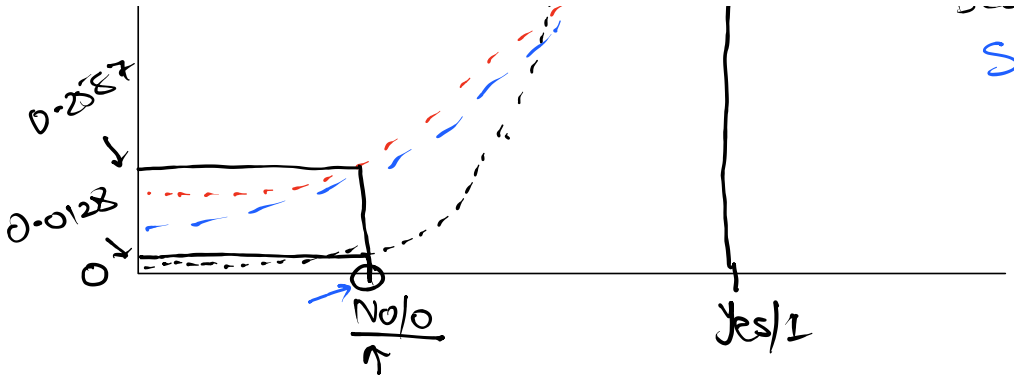
How do we finalize the best values for the above equation →

Line 1 : $1.1x_1 + 0.7x_2 + 0.15x_3 + 0.29x_4 + 1.8$

Line 2 : $0.91x_1 + 0.42x_2 + 1.13x_3 + 0.89x_4 + 1.35$







S-curve

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

↓

Testing Data

Decision Boundary →

$$\rightarrow m_1 x_1 + m_2 x_2 + c \leftarrow$$

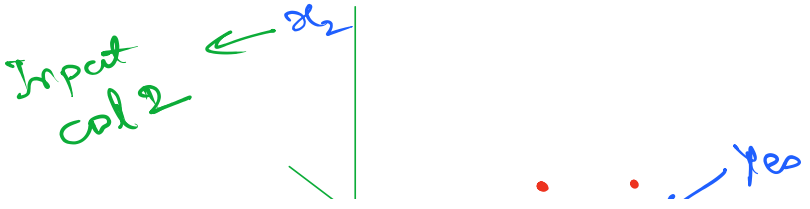
$$m_1 = 1, \; m_2 = 1, \; c = -3$$

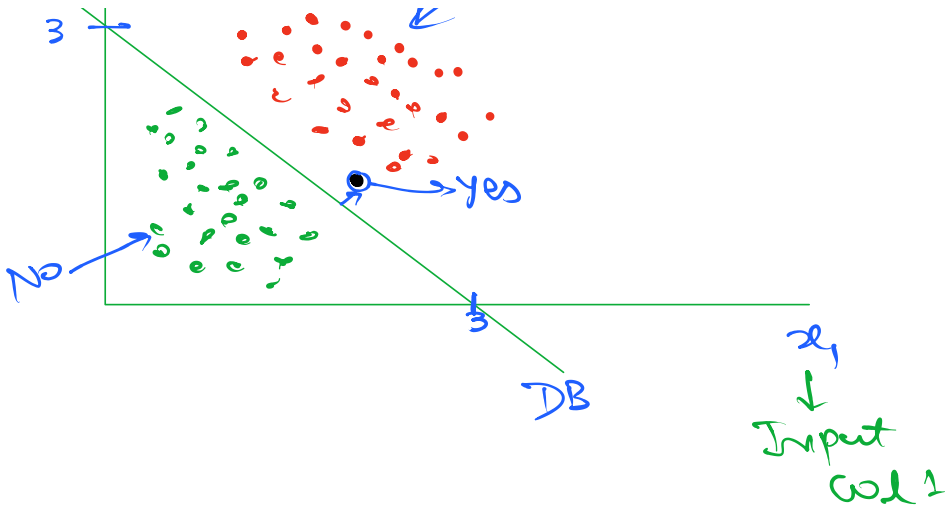
$$m_1 x_1 + m_2 x_2 + c = 0$$

$$1 \cdot x_1 + 1 \cdot x_2 - 3 = 0$$

$$x_1 + x_2 - 3 = 0$$

$$x_1 + x_2 = 3$$





Age	chol	HD or Not?
31	—	No
—	—	Yes
—	—	No