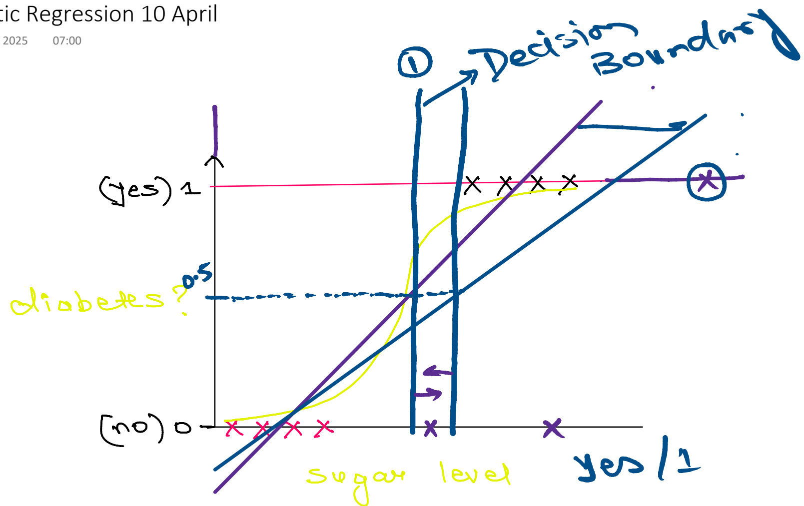
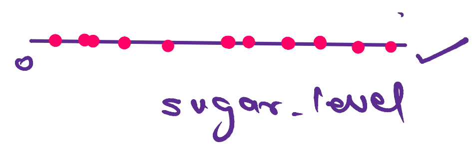
Logistic Regression 10 April

30 January 2025 07:00



	sugar level	Diabetic?
0	—	yes
1	—	No
2	—	yes
3	—	No
4	—	No
5	—	No
6	—	yes
7	—	?

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



Requirements →

① want output to be 0 or 1.

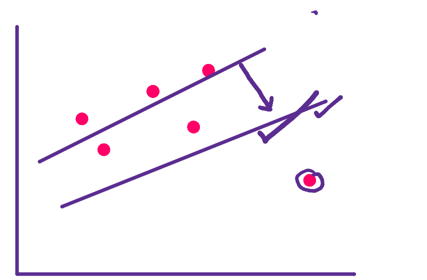
② Not much affected by the outlier.

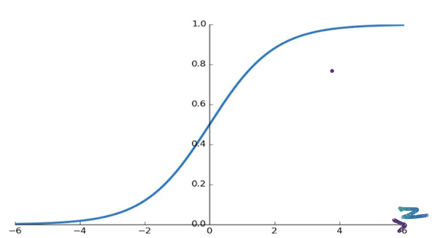
1.33 No ⓪

0 or 1 yes ①

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

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





SF
↓
0 to 1
↓
0.75 ✓

$> 0.5 \rightarrow 1$

$\leq 0.5 \rightarrow 0$

$\rightarrow \hat{y} = mx + c$ ✓
↓
$-\infty$ to $+\infty$
↓
0 or 1

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

↓

because of this it is not suitable for classification

Logit → $-\infty$ to $+\infty$

↓

Logistic Regression → Sigmoid Function SF

↓ SL : 170 — 0 to 1

$$g(z) = \frac{1}{1+e^{-z}} \rightarrow 0.83 \qquad 0.17$$

↓

gives an output b/w 0 to 1 — 0.5

↓

Probability — Yes

Mathematical Intuition →

Heart Disease Prediction

Training Data

↓

Regression Line

↓

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

↓

Yes/1

Person

x_1 : CPL → 4

x_2 : Chol → 35.8

x_3 : BP → 110.5

x_4 : Age → 43

Testing Data

↓

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

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

1

↓

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

e = euler's number
= $\underline{2.71}$

Yes HD

No Healthy

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

$$g(z) = 0.9999$$

0.023

0.023 $\underline{2.3\%}$

Always represent the probability that the predi--cted output is "Yes/1".

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 $\underline{0.5}$.

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

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

In our case,

$\underline{0.9999} > 0.5$

$\downarrow$

$\underline{\text{Yes}/1}$

will always be ≥ 0.5.

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

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

Final Form of Logistic Regression equation.

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

$$\downarrow$$

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

$$\downarrow$$

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

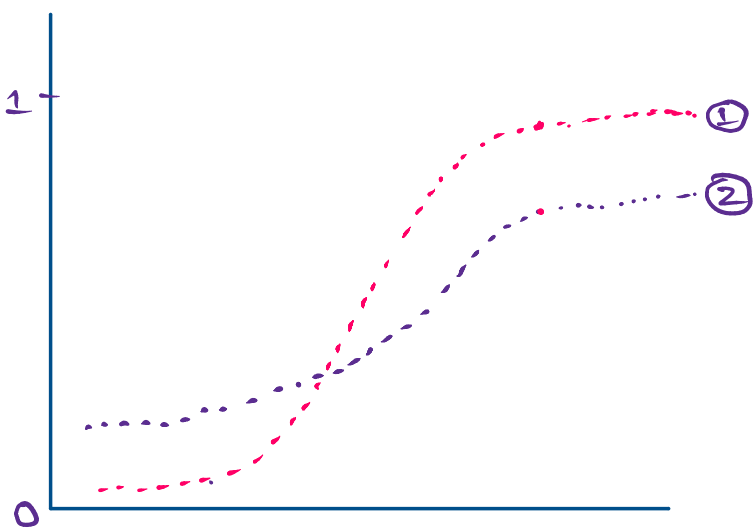
50 training Data

Line 1	y-train	Line 2
0.9999	1	0.8215
0.999827	1	0.78259
0.000125	0	0.0139
—	0	
—	1	
	1	
	1	
	0	
	0	

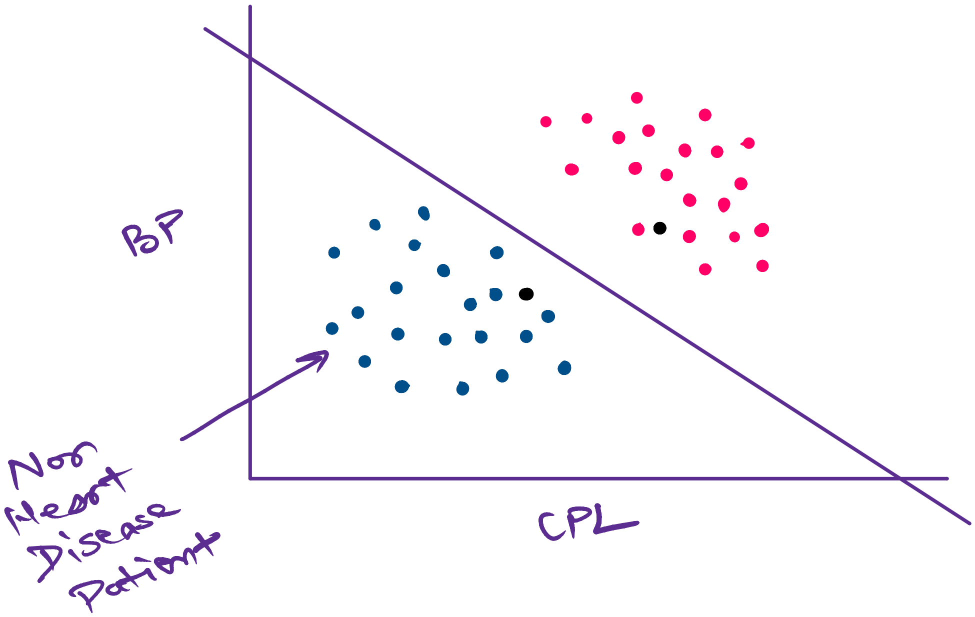
Line 1: $1.1x_1 + 0.7x_2 + 0.15x_3 + 0.29x_4 + 1.8$

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

50 training data



Decision Boundary :



$$g(z) = \frac{1}{1+e^{-(x_1+x_2-3)}}$$

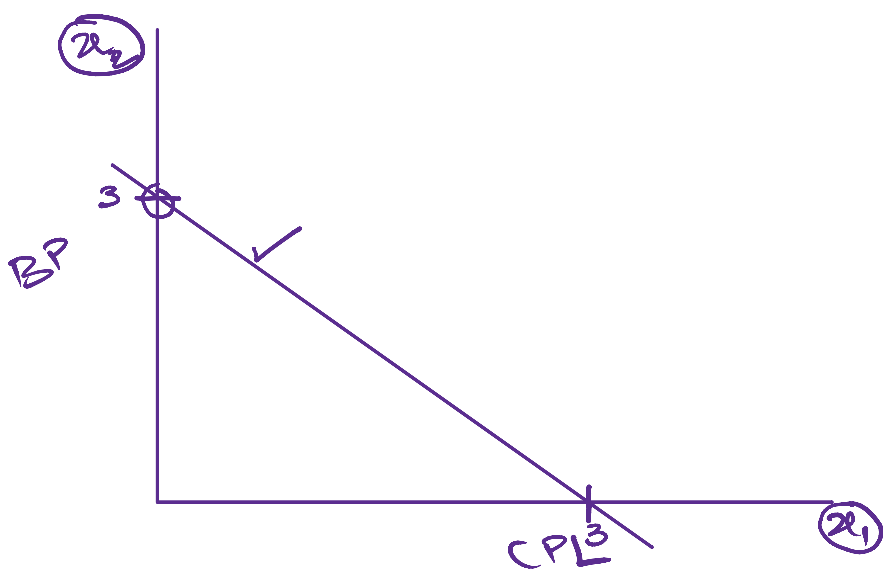
$$m_1 = 1, \quad m_2 = 1, \quad c = -3$$

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

$$x_1 + x_2 - 3 = 0$$

or

$$x_1 + x_2 = 3$$



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

or

$$g(z) = \frac{1}{1 + e^{-(0.91x_1 + 0.42x_2 + 1.13x_3 + 0.89x_4 + 1.35)}}$$