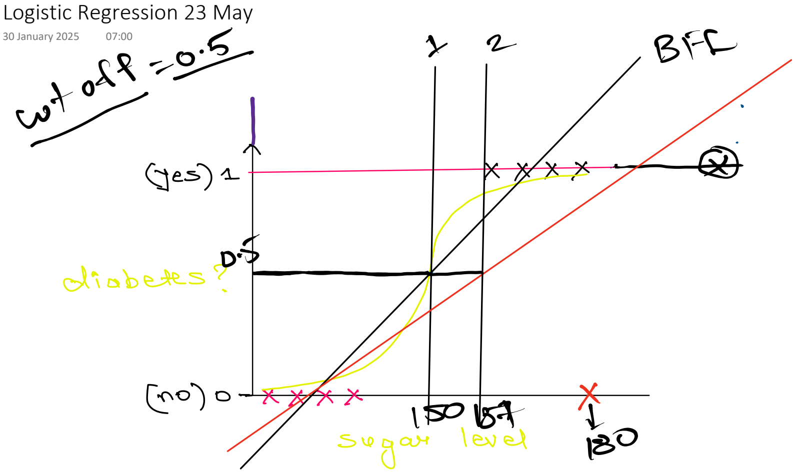
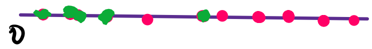


sugar level	Diabetic?
0	yes → 1
1	No → 0
2	yes
3	No
4	No
5	No
6	yes
7	



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

$180 \rightarrow 1.65$

Requirements →

① want output to be 0 or 1.

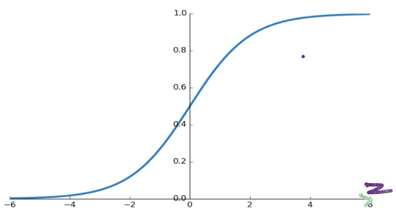
② Not much affected by the outlier.

$\boxed{y = mx + c}$

$-\infty$ to $+\infty$

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

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



50%

Person

medical details

$y = mx + c$

27.3

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

$\boxed{0.78}$

1

78% → Diabetic (1)

22% → Non-Diab (0)

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

↓

Regression Line

↓

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

Person   Yes/1

x_1: CPL → 4

x_2: Chol → 35.8

x_3: BP → 110.5

x_4: Age → 43

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

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

↓

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

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

$$g(z) = \underline{0.9999}$$

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 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,

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

↓

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.

$$z = \hat{y} = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + c]$$

↓

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

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

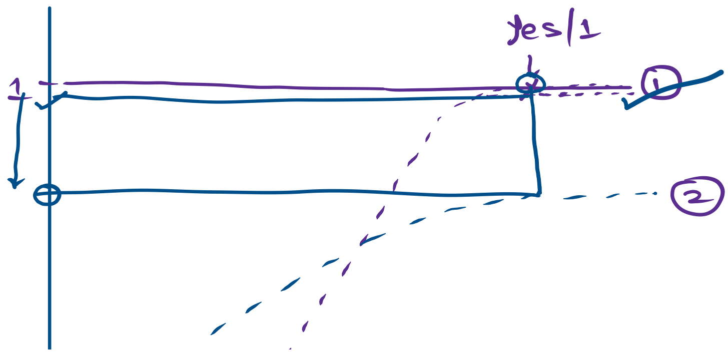
50 training Data

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

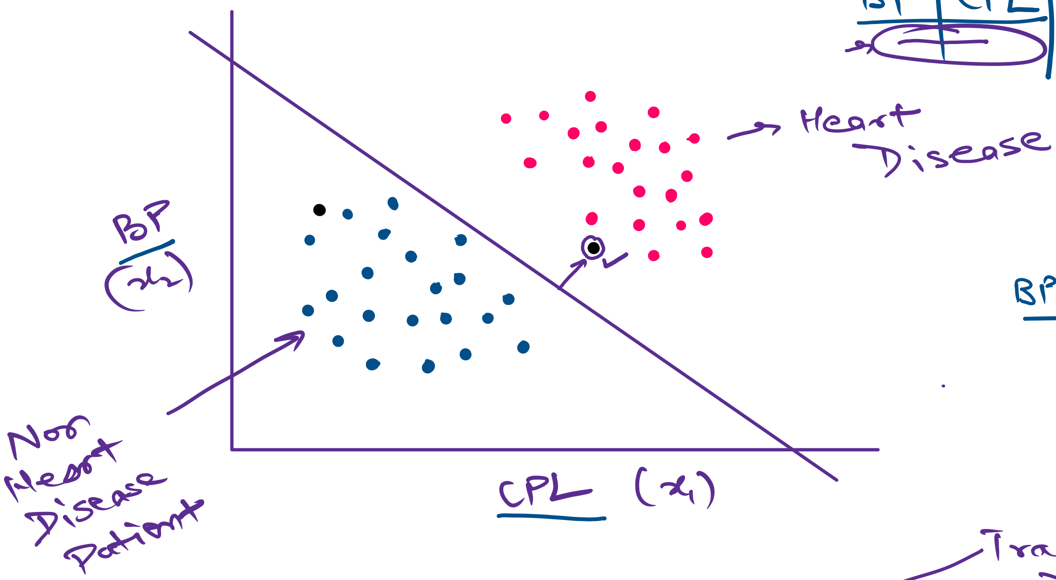
Line 2: $0.91x_1 + 0.42x_2 + 1.13x_3 + 0.89x_4 + 1.35$

50 training data



Decision Boundary :

BP	CPL	HD or Not?
		Yes



BP CPL

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

Training Data

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

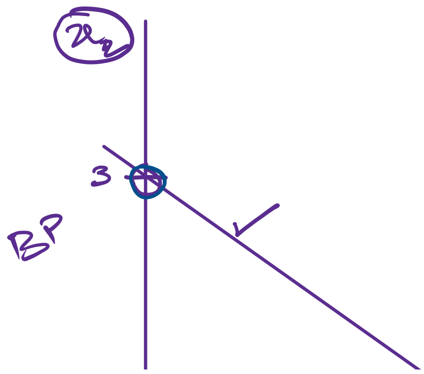
$$m_1x_1 + m_2x_2 + C = 0$$

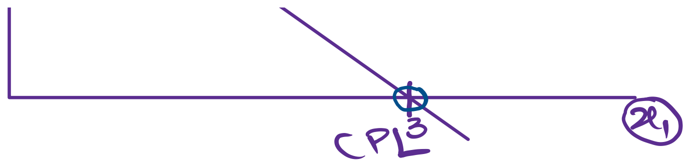
$$x_1 + x_2 - 3 = 0$$

Or

$$x_1 + x_2 = 3$$

$$2x_1 + \frac{5}{7}x_2 = 19.7$$

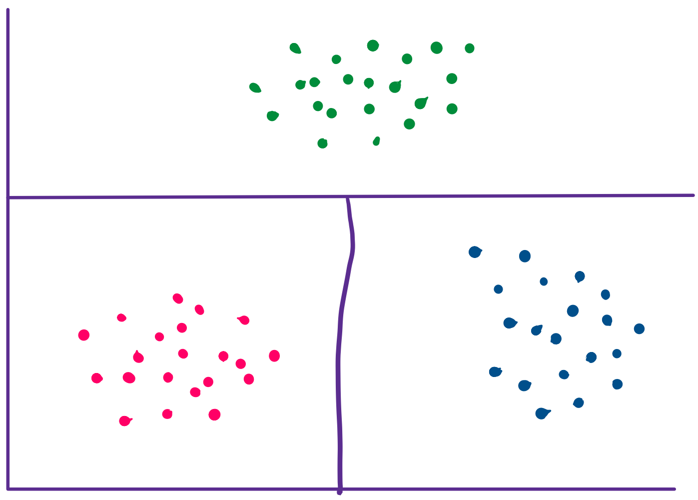




$$g(z) = \left[\frac{1}{1+e^{-(1.1x_1+0.7x_2+0.15x_3+0.29x_4+1.8)}}\right]$$

or

$$g(z) = \frac{1}{1+e^{-(0.41x_1+0.42x_2+1.13x_3+0.89x_4+1.35)}}$$



Decision Tree