Logistic Regression (Classification)

↓ most popular learning algo.

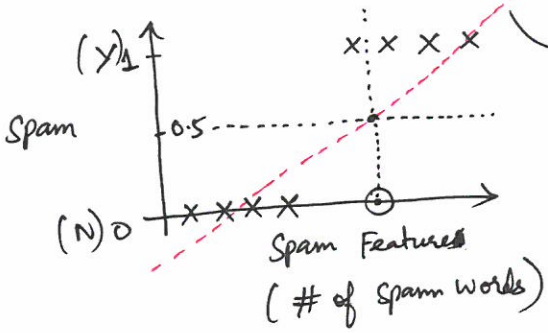
↓ $y \rightarrow$ discrete values.

Eg: Spam / Not Spam ?

Credit Card Transactions Fraud (Yes / No) ?

Tumor Malignant / Benign ?

Output variable $y = \{0, 1\}$ or $\{-1, +1\}$ or $\{0, 1, 2, 3\}$

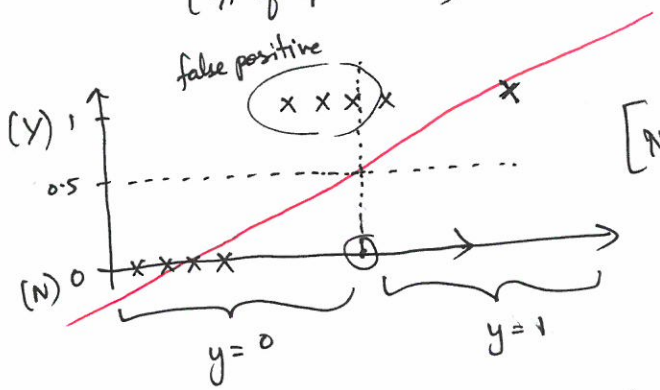


$\hat{y} = B^T X$ (Linear Reg)

↓

$\hat{y}(x)$ at 0.5: if $\hat{y} \geq 0.5$, then predict "y=1"

if $\hat{y} < 0.5$, then predict "y=0".



[Not good prediction] (worst hypothesis)

If Regression is used for classification $\Rightarrow$ $y > 1$ or $y < 0$

Logistic Regression $\Rightarrow 0 \leq \hat{y}(x) \leq 1$

↓

- Not a regression algo
- its a classification algo.

Hypothesis fn. for Classification Problem.

$y_b(x) = y = f_b(x, b)$.

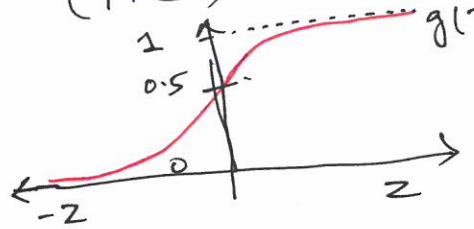
We want $0 \leq \hat{y}_b(x) \leq 1$

In linear reg: $\hat{y}_B(x) = B^T x$.

For logistic reg: $\hat{y}_B(x) = g(B^T \cdot X) \Rightarrow \hat{y}(x) = \left(\frac{1}{1+e^{-B^T \cdot x}}\right)$

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

sigmoid fn. or logistic fn.



$0 \leq g(z) \leq 1$.

We need to find the parameters 'B' in the following Eq: using our training data.

$B = ?$

$$\hat{y}_B(x) = \frac{1}{1+e^{-B^T \cdot x}}$$

$\hat{y}_B(x)$ = estimated probability that $y = 1$ on input x.

Eg: if $x = [x_0 \; x_1]^T = [1 \quad \#spam\_word\_count]$ then,

$\hat{y}_B(x) = 0.7$.

70% of chance that email is a spam $\Rightarrow$ ie., $y = 1$.

$\hat{y}_B(x) = P(y=1 \mid x, B)$ → probability that $y=1$, given x, and parameter B.

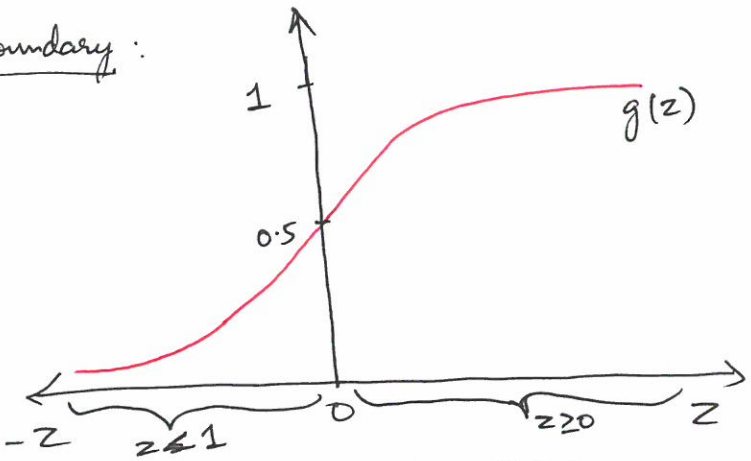
Since $y \rightarrow 0$ or 1,

We have $P(y=0 \mid x, B) = 1 - P(y=1 \mid x, B)$.

$P(y=0 \mid x, B) + P(y=1 \mid x, B) = 1$.

(3)

Decision Boundary :



$$y_B(x) = g(B^T x)$$

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

$g(z) \geq 0.5$ when $z \geq 0$

$$\boxed{y_B(x) = g(B^T x) \geq 0.5} \text{ when } \boxed{B^T x \geq 0}$$

That is,

predict "$y=1$", if $\hat{y}_B(x) \geq 0.5$

$$B^T x \geq 0$$

predict "$y=0$", if $\hat{y}_B(x) < 0.5$

$$B^T x < 0$$

Example: $\hat{y}_B(x) = g(b_0 + b_1 x_1 + b_2 x_2)$.

$$B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

predict "$y=1$" if $(-3 + x_1 + x_2) \geq 0$

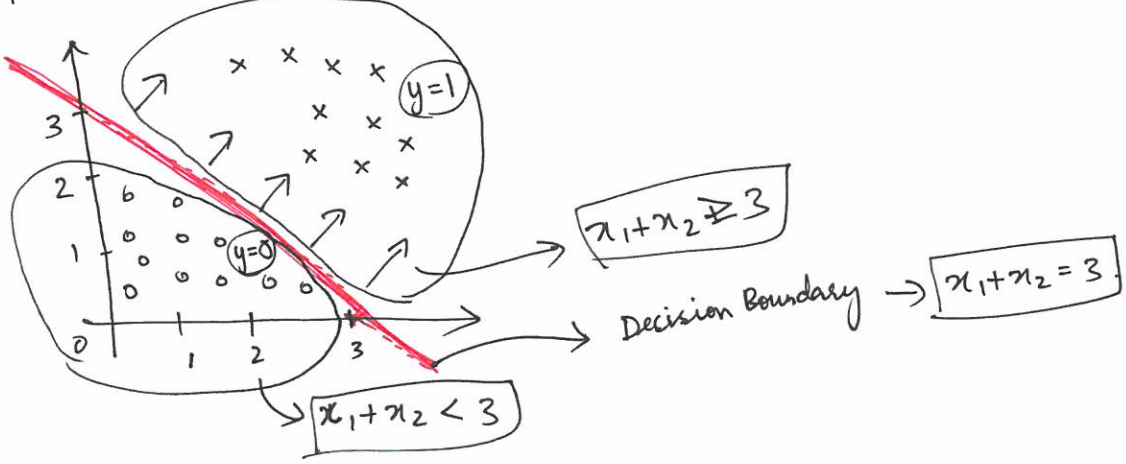
$\Downarrow$

$B^T \cdot X$

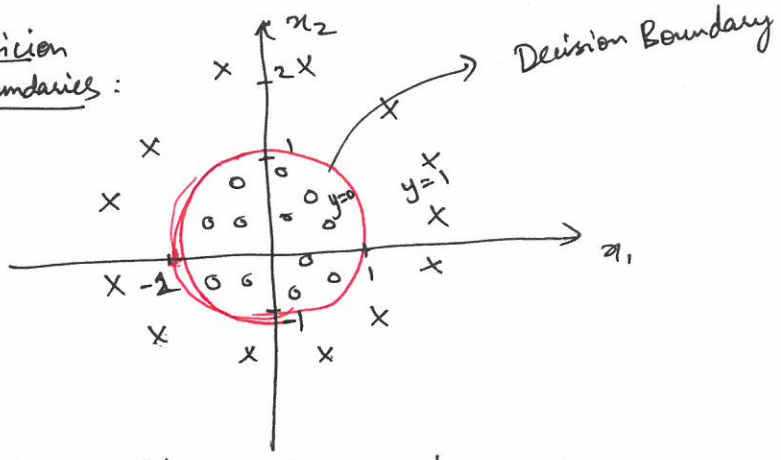
$\Downarrow$

$$\boxed{x_1 + x_2 \geq 3}$$

$x_1 + x_2 = 3$ is a straight line.



Non-linear division Boundaries:



$$\hat{y}_B(x) = g(b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_2 x_2^2)$$

$$B = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Predict "y = 1" if $-1 + x_1^2 + x_2^2 \geq 0$

$$\Downarrow$$

$$x_1^2 + x_2^2 \geq 1$$

Plot the $x_1^2 + x_2^2 = 1$ in the figure, we get a circle.

$$\downarrow$$

Decision Boundary.

Note: - Decision Boundary is property of our $\hat{y}_B(x)$ function and parameter 'B'.

- Nothing to do with the data from training set.

More-complex Boundaries:

$$\hat{y}_B(x) = g(b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_1^2 x_2 + \cdots)$$