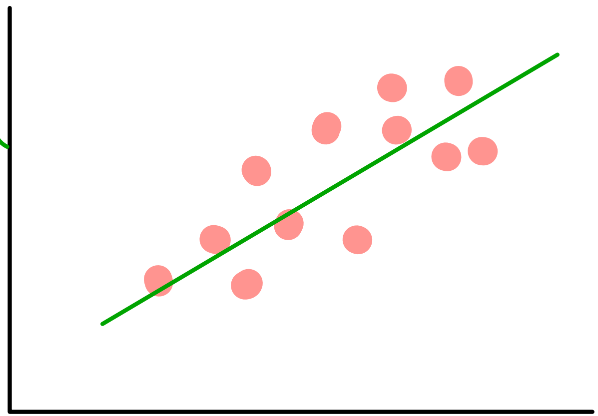
LOGISTIC REGRESSION |

Logistic Regression

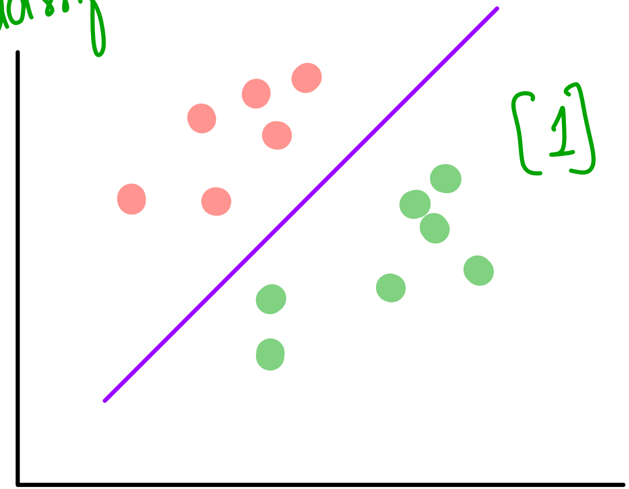
Regression



Line of best fit

Classification

[0]



Line of best separation

{0, 1}

Recap

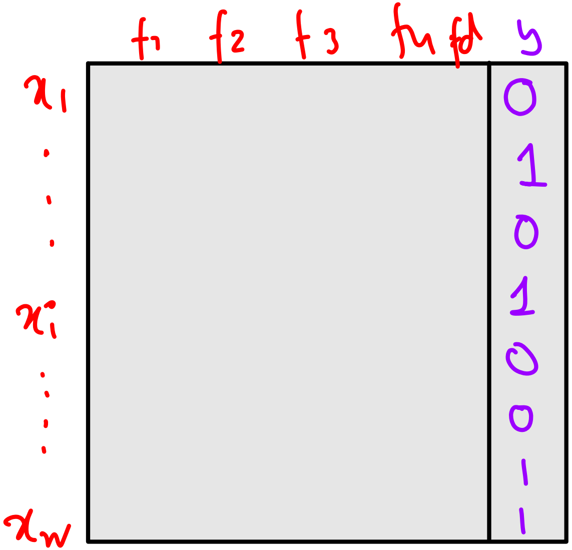
Linear Reg → Linear Model

1. hyperplane $w^T x + w_0 = 0$
2. $x_q \rightarrow \hat{y}_q \mid \hat{y}_q = w^T x_q + w_0$
3. $LOSS = MSE = \frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2$

Log Regression

↳ Classification

↳ Supervised ✓ / ~~Unsupervised~~



$$\mathcal{D} : \{ (x_i, y_i) \}_{i=1}^{n}, \quad x_i \in \mathbb{R}^d, \quad \underbrace{y_i \in \{0, 1\}}_{\text{classification}} \}$$

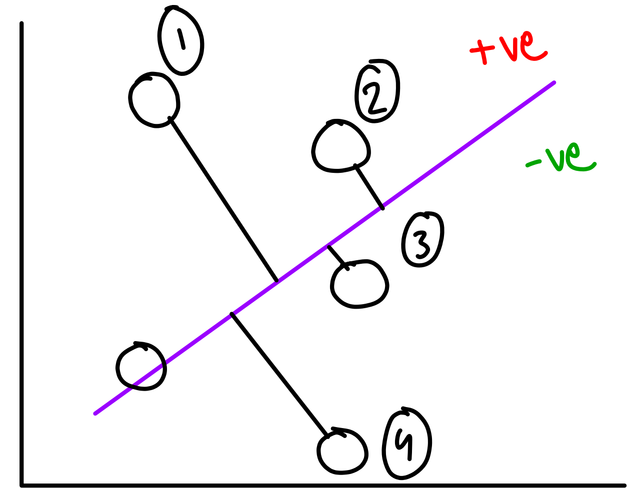
Reg. $\Rightarrow y_i \in \mathbb{R}$

Logistic Reg.

In Lin Reg

$$\hat{y} = W^T x_q + w_0 \longrightarrow (-\infty, \infty)$$

x_q



$$\sum \frac{W^T x_1 + w_0}{\|W\|} + \sum \frac{W^T x_3 + w_0}{\|W\|} = \text{Sum of all distances}$$

distance

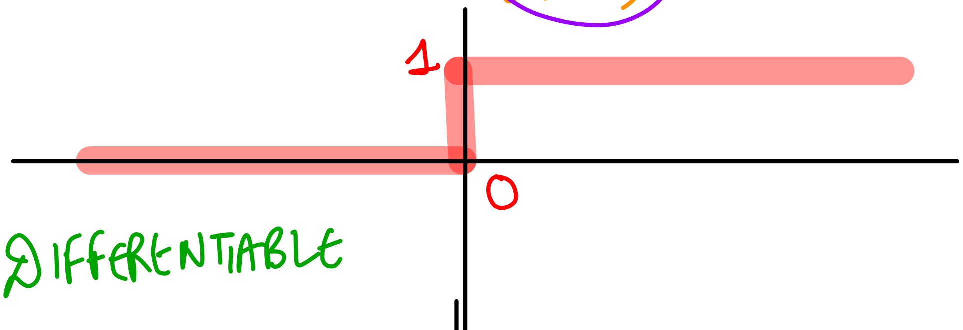
Reality: $(-\infty, \infty)$ Required: $\{0, 1\}$

$$z(i) \Rightarrow \frac{w^T x + w_0}{\cancel{\|w\|}} \Rightarrow (-\infty, \infty) \Rightarrow \{0, 1\}$$

ignore

① $(-\infty, \infty) \longrightarrow g(\;) \longrightarrow \{0, 1\}$

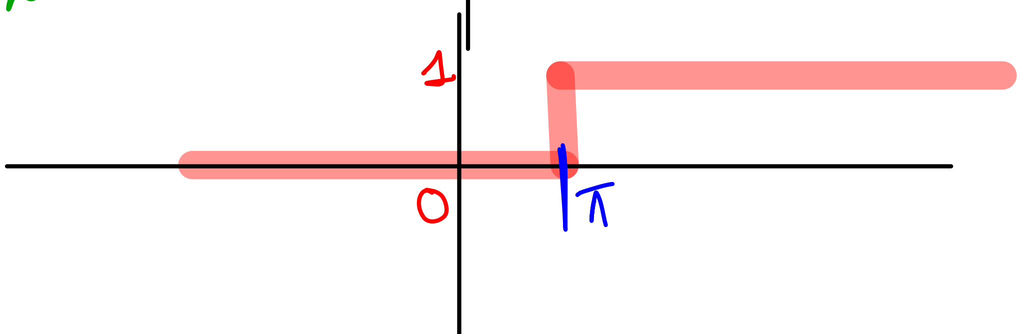
$g \rightarrow f$



if $z(i) \geq 0$
$= 1$

if $z(i) < 0$
$\hookrightarrow = 0$

NON DIFFERENTIABLE

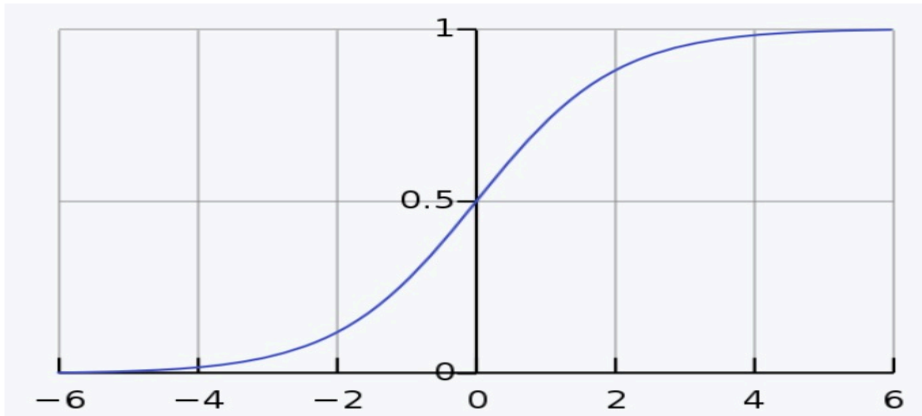


if $z_i \geq \bar{T}$
$\Rightarrow 1$

$z_i < \bar{T}$
$= 0$

② $(-\infty, \infty) \Rightarrow g(\ ) \Rightarrow (0, 1)$ — $> 0.5 \Rightarrow$ +ve 1; $\leq 0.5 \Rightarrow$ −ve 0 $\Rightarrow \{0, 1\}$

differentiable



(σ) $f(x) = \dfrac{1}{1+e^{-x}}$

Sigmoid function

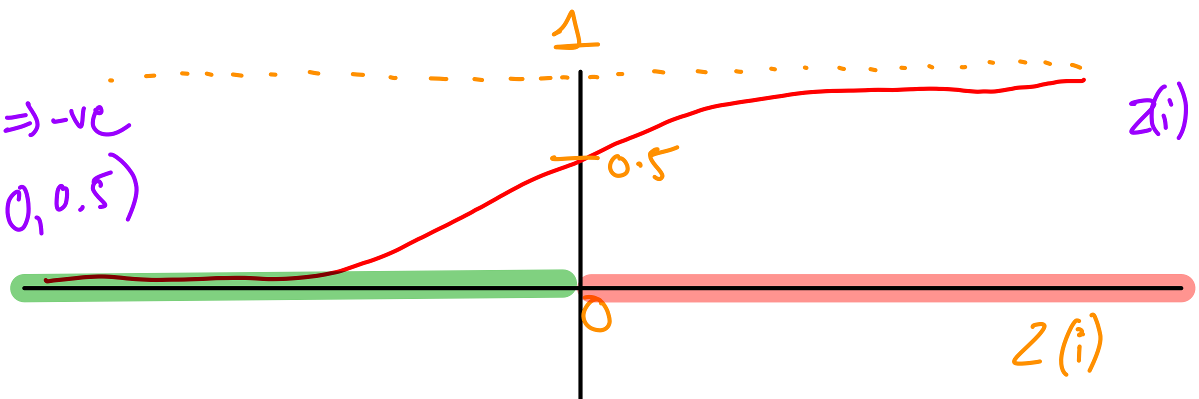
Logistic function

$z(i) \Rightarrow$ −ve

$(0, 0.5)$

$z(i) \Rightarrow$ +ve

$(0.5, 1)$



$z(i) = W^T x + w_0$

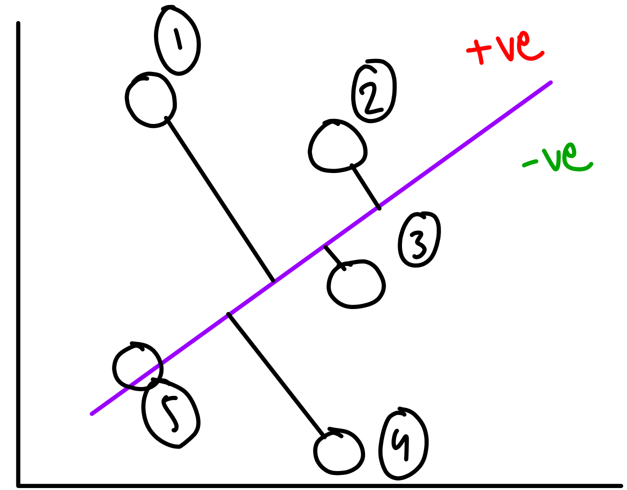
$z(i) = \infty \Rightarrow \sigma(z_i) \Rightarrow 1$

$z(i) = -\infty \Rightarrow \sigma(z_i) \Rightarrow 0$

$z(i) = 0 \Rightarrow \sigma(z_i) \Rightarrow 0.5$

$\sigma(z_i) \geq 0.5 \Rightarrow$ Red +ve

$\sigma(z_i) < 0.5 \Rightarrow$ Green −ve



	$z(i)$	$\sigma(z(i))$	
x_1	+5	0.9	Red
x_2	+2	0.7	Red
x_3	-1	0.3	Green
x_4	-4	0.1	Green
x_5	0	0.5	Red

$P[x_1 = R] = 0.9$ $\quad P[x_1 = G] = 0.1$

$P[x_2 = R] = 0.7$ $\quad P[x_2 = G] = 0.3$

$P[x_3 = R] = 0.3$ $\quad P[x_3 = G] = 0.7$

$P[x_4 = R] = 0.1$ $\quad P[x_4 = G] = 0.9$

$P[x_5 = R] = 0.5$ $\quad P[x_5 = G] = 0.5$

$\underbrace{\hat{y}_i}_{\text{predicted}} = P[\underbrace{y_i}_{\text{actual}} = 1 \mid x_i]$

$x_i \Rightarrow \underset{z_i}{\underset{(-\infty,\infty)}{w^T x + w_0}} \overset{\sigma(z_i)}{\Longrightarrow} \underset{(0,1)}{\hat{y}_i} \rightarrow P[y_i = 1 \mid x_i]$

$\hat{y}_i \geq 0.5 = 1$ +ve

$\hat{y}_i < 0.5 = 0$ -ve

Sigmoid $\rightarrow$

1. Squashing func $(-\infty, \infty) \rightarrow (0,1)$
2. Smooth & differentiable.
3. $P[y_i = 1 \mid x]$
4. Monotonic

Quiz time!

Quiz Ended!

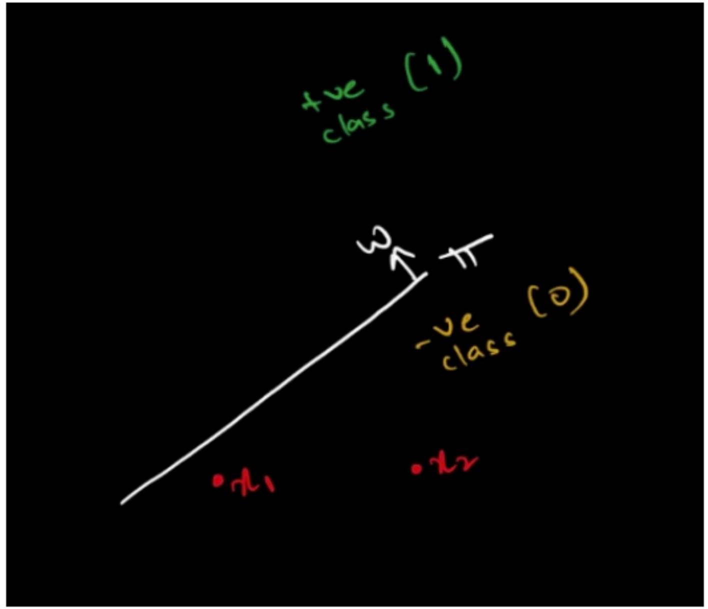
What happens when the input to the sigmoid function is a very large negative value?

29 users have participated

A	The output becomes negative	3%
B	The output approaches 0	86%
C	The output approaches 1	7%
D	The output becomes undefined.	3%

Quiz time!

Quiz Ended!



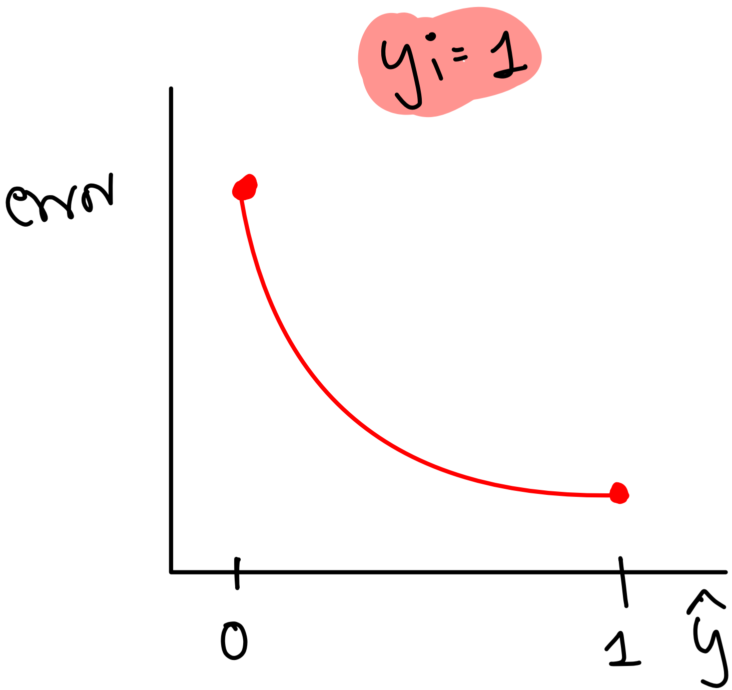
Which point will have a higher probability of belonging to class 1?

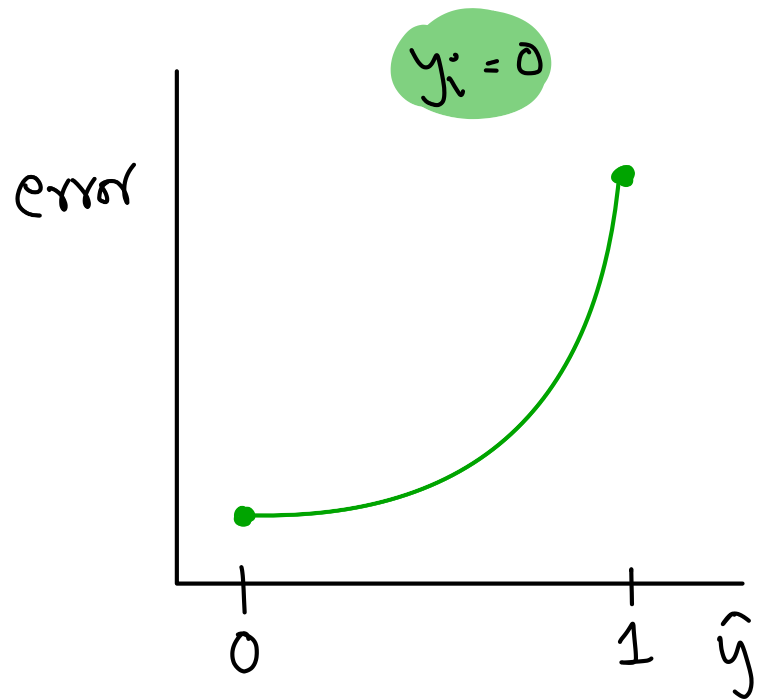
34 users have participated

A	x1	91%
B	x2	9%

$$z(i) = \sigma(z_i) = P[y_i=1 \mid x_i] = p = \sigma(z_i)$$

$$= P[y_i=0 \mid x_i] = 1-p = 1-\sigma(z_i)$$





$$\text{log loss} = \begin{cases} -\log(\hat{y}_i) & y_i = 1 \\ -\log(1-\hat{y}_i) & y_i = 0 \end{cases}$$

$$\text{log loss} = y_i\left(-\log(\hat{y}_i)\right) + (1-y_i)\left(-\log(1-\hat{y}_i)\right)$$

when $y_i = 1$

$$\text{log loss} = \boxed{y_i\left(-\log(\hat{y}_i)\right)} + \cancelto{0}{\left[\overset{1-1\,=\,0}{(1-y_i)}\left(-\log(1-\hat{y}_i)\right)\right]}$$

when $y_i = 0$

$$\text{log loss} = \cancel{y_i\left(-\log(\hat{y}_i)\right)} + \boxed{(1-y_i)\left(-\log(1-\hat{y}_i)\right)}$$

$$\underline{\text{LOSS}} = \frac{1}{n} \sum_{i=1}^{n} -y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i)$$

$$\underset{\underline{w_j}}{\text{Argmin}} \; [\text{LOSS}]$$

$$\underset{w_j}{\text{Argmin}} \left[-\frac{1}{n} \sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i) \right]$$

$w_1 \quad w_2 \quad w_3$

$$\hat{y}_i = \sigma(z_i) = \sigma\left[w^T x + w_0\right) = \sigma\left(w_1 x_1 + w_2 x_2 \cdots w_j x_j + w_0\right)$$

$$\sigma(z_i) = \frac{1}{1+e^{-z_i}} = \frac{1}{1+e^{-(w_1 x_1 + w_2 x_2 \cdots w_j x_j + w_0)}}$$

$$f(x) = \frac{1}{1+e^{-x}}$$

$$f'(x) = f(x)\left(1 - f(x)\right)$$

$$f(x) = \frac{1}{1+e^{-x}} \quad \begin{matrix} \} u \\ \} v \end{matrix}$$

$$f'(x) = \frac{0\left[1+e^{-x}\right] - 1\left[-e^{-x}\right]}{\left(1+e^{-x}\right)^2} = \frac{e^{-x}}{\left(1+e^{-x}\right)^2}$$

$$\frac{e^{-x}}{\left(1+e^{-x}\right)} \times \underbrace{\frac{1}{\left(1+e^{-x}\right)}}_{f(x)}$$

$$f(x)\left[\frac{1+e^{-x}}{\left(1+e^{-x}\right)} - \frac{1}{\left(1+e^{-x}\right)}\right] \Rightarrow f(x)\left[1 - f(x)\right]$$

$$f'(x) = f(x)\left(1-f(x)\right)$$

$y = 0$

$\hat{y} = 0.01$

log loss

$\uparrow$ $\downarrow$

Quiz time!

Time Left: 0s

In logistic regression, the output of the sigmoid function is interpreted as:

30 users have participated

A	Class probabilities	70%
B	Raw scores	7%
C	Error rates	17%
D	Regression coefficients	7%

End Quiz Now