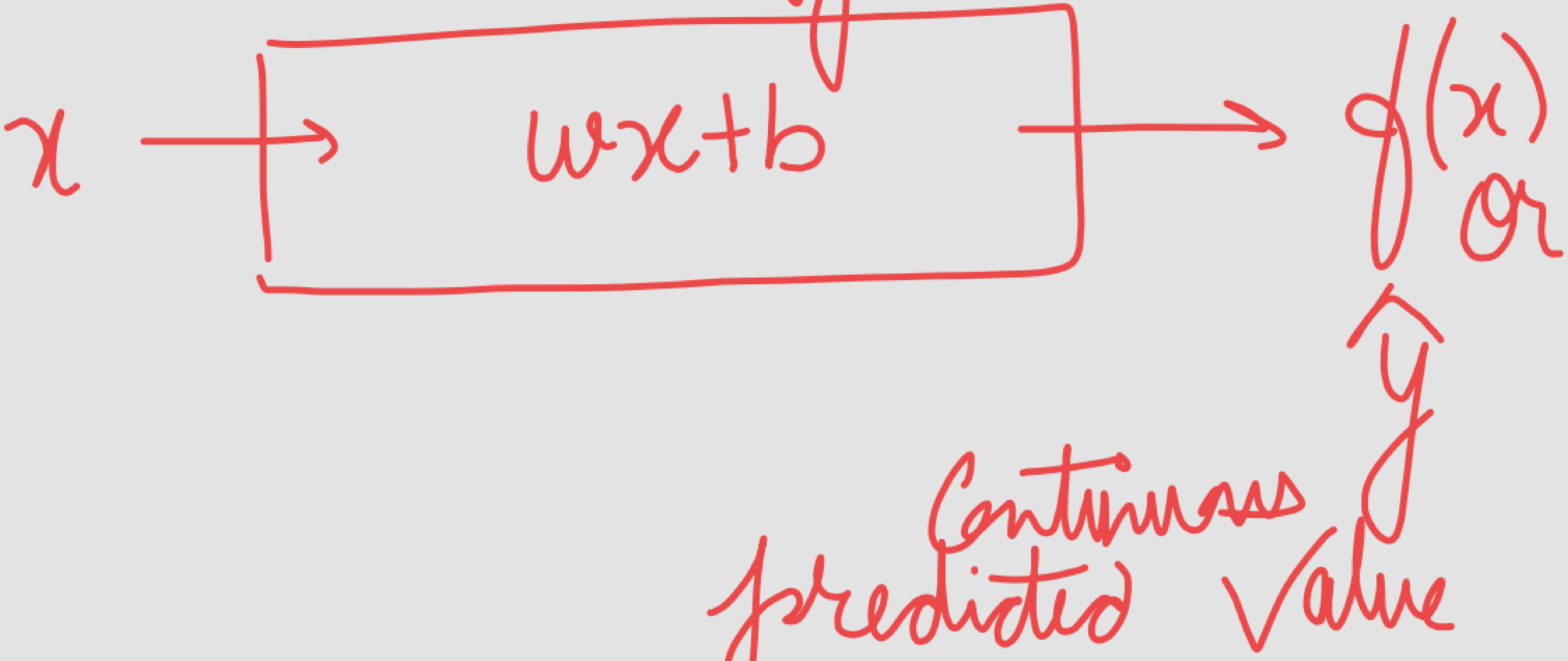
Linear Regression



Logistic Regression

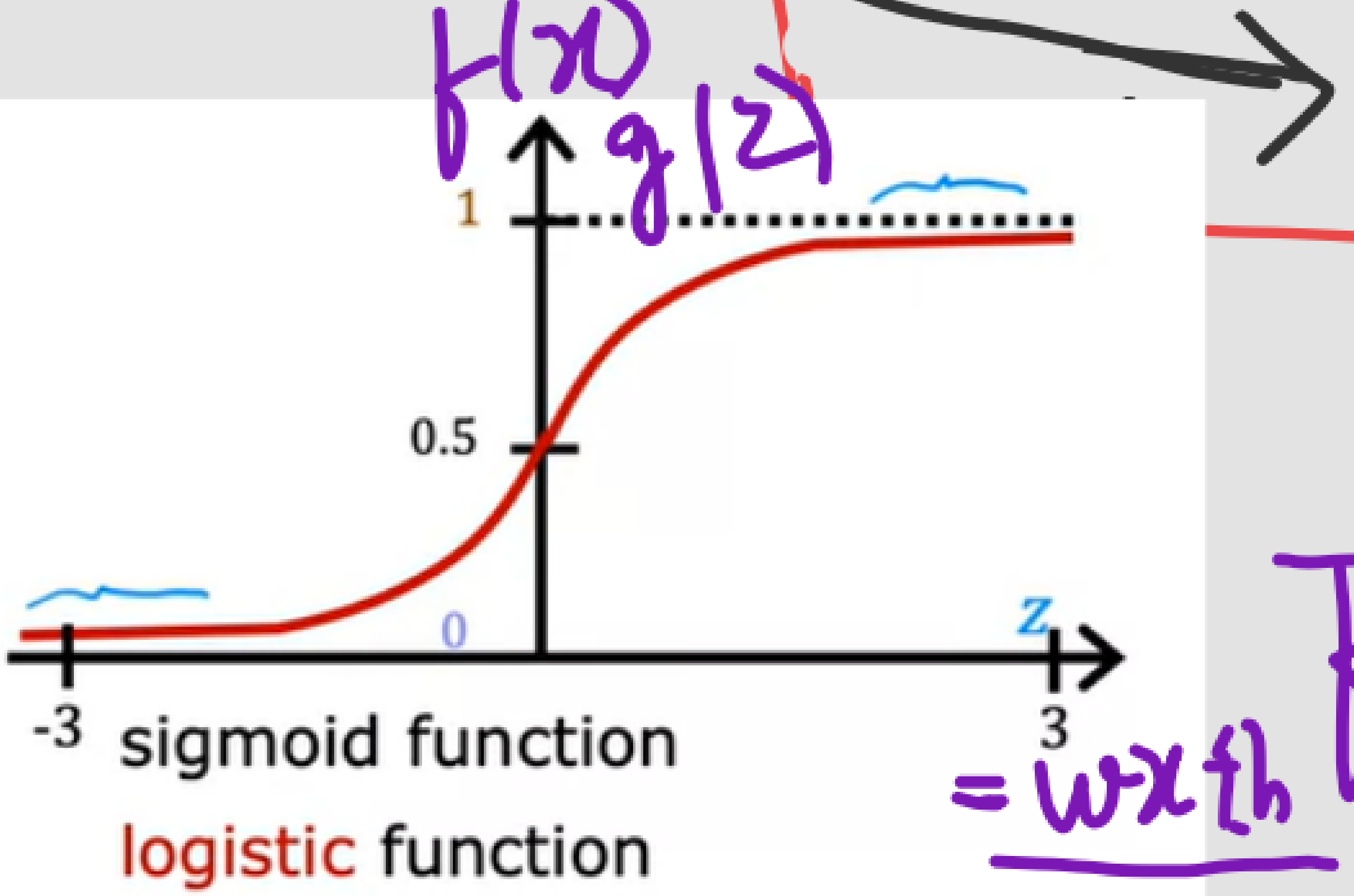
$x \rightarrow$ $\boxed{z = wx + b} \rightarrow \boxed{g(z) = \frac{1}{1+e^{-z}}}$ sigmoid func $\rightarrow f(x)$

$[0, 1]$

Probability that class is 1

OR

$= \frac{1}{1+e^{-(wx+b)}}$



sigmoid function

logistic function

outputs between 0 and 1

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

$f(x) > 0.5$ when $wx+b > 0$

$\Rightarrow$ class 1

$f(x) < 0.5$ when $wx+b < 0$

$\Rightarrow$ class 0

Classification Decision Boundary using Logistic Regression (Optional)

Consider following dataset:

(In 10 years)

Patient Age (x_1)	Tumor Size (x_2)	y
1	1	0
1	1.2	0
1.2	1.1	0
1.3	1.5	0
2	1.5	1
2.5	2	1
2.6	2.5	1
3	3	1

0: Benign, 1: Malignant

O/P $= f(x) = g(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(w_1x_1+w_2x_2+b)}}$

Eq^n of line (from this data)

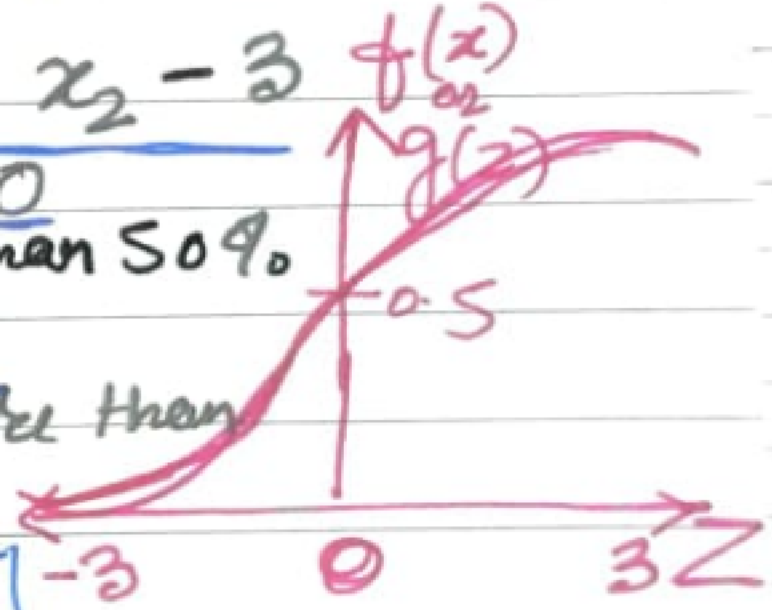
If $x_1+x_2<3$ then $\hat{y}=0$

and

If $x_1+x_2>3$ then $\hat{y}=1$ $\Rightarrow$ $z = x_1+x_2-3$

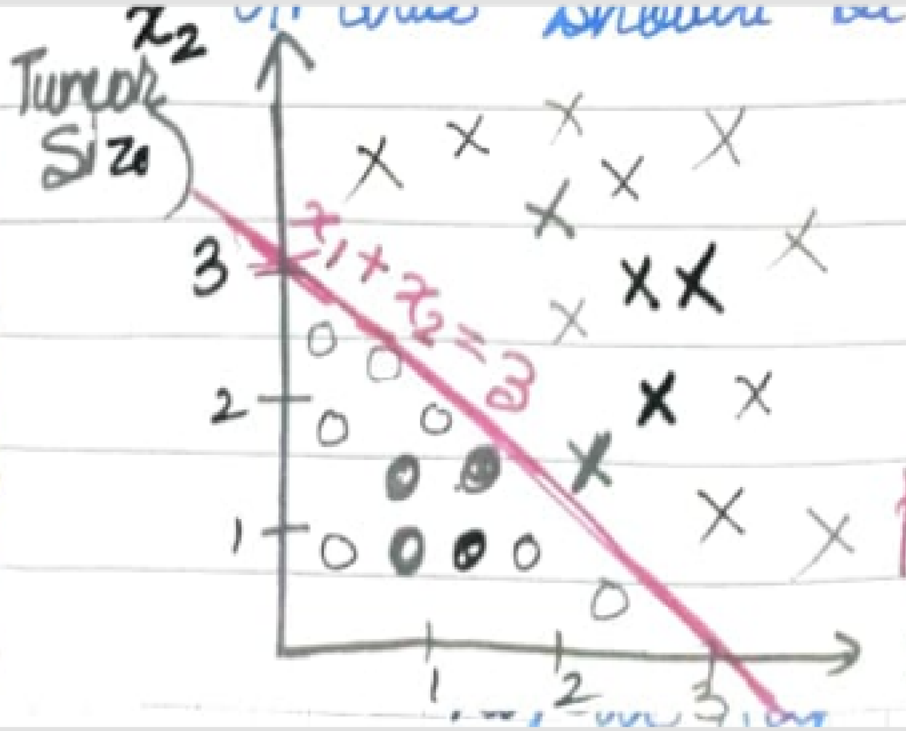
When z is -ve, O/P label is 0, as prob is less than 50%

When z is +ve, O/P label is 1, as prob is more than 50%



$\boxed{\text{Prob. is exactly } 50\% \text{ when } z=0}$

Setting $z=0$ will give decision boundary where we are almost neutral whether O/P class should be 0 or 1



$x_1+x_2-3=0$

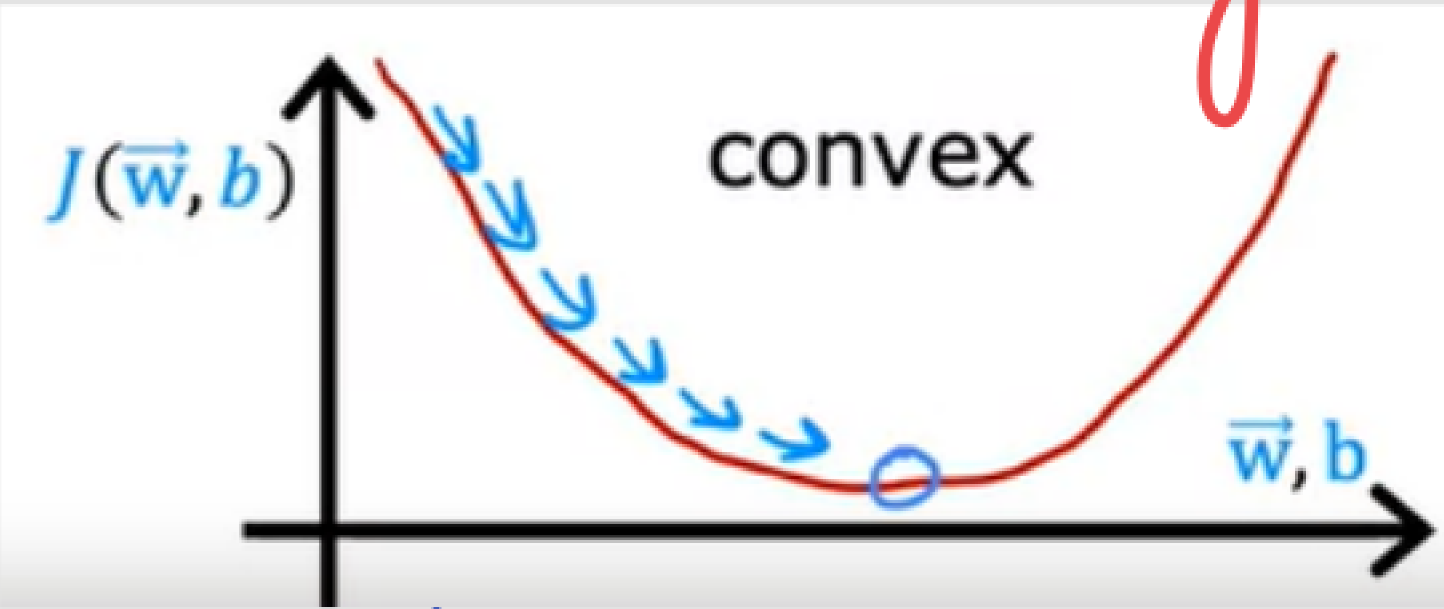
$x_1+x_2=3$

Decision Boundary as points to the left belong to class 0 & to right belong to class

Using MSE (mean squared error) as the cost function

$$J(w,b) = \frac{1}{N}\sum_{i=1}^{N}\left(y^{(i)} - f(x^{(i)})\right)^2$$

Linear Regression || Logistic Regression



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

Idea:

When logistic regression func / sigmoid func is fitted into the cost func, resulting func. should be convex (Single global minimum)

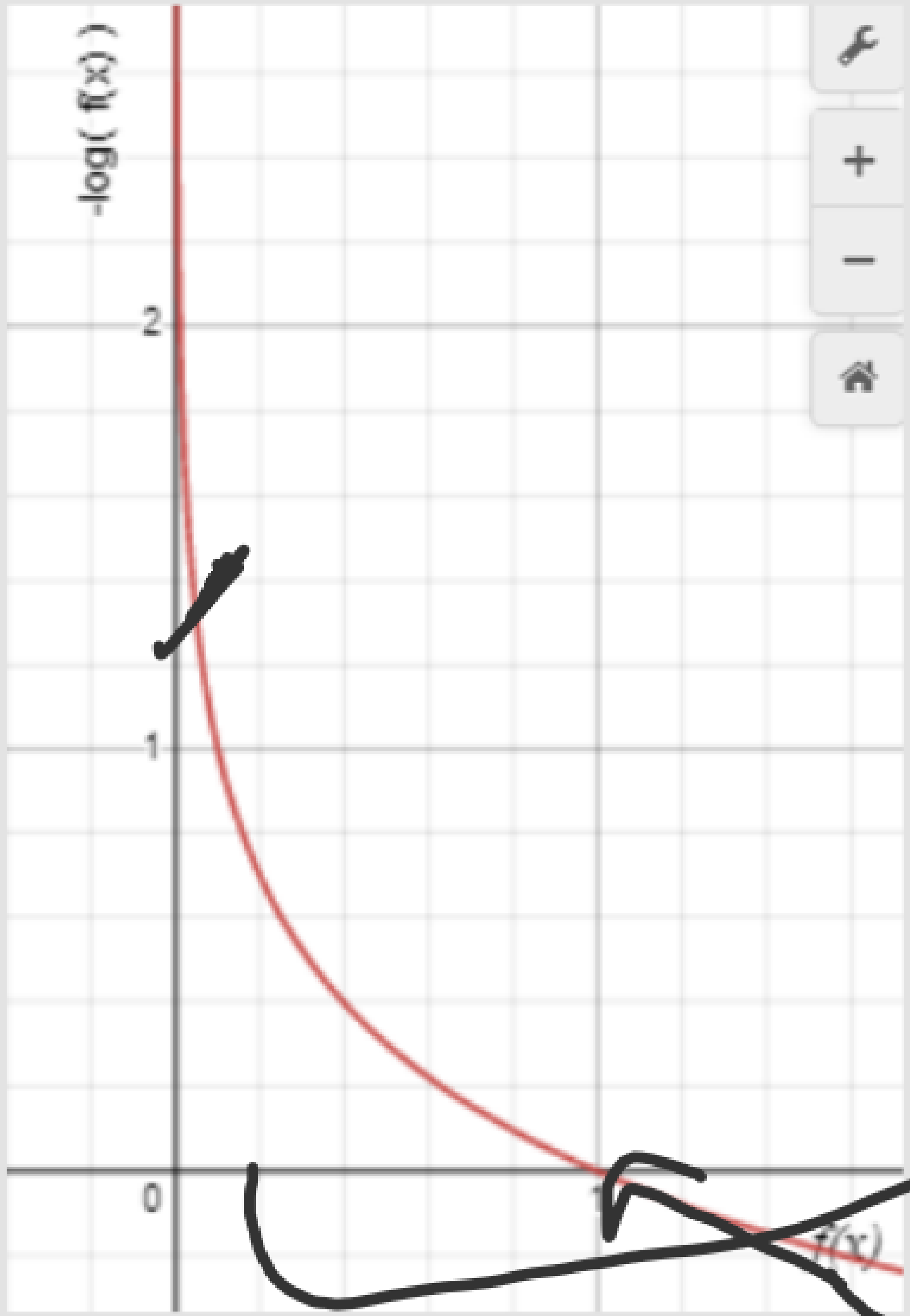
Cost func $= \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(y^{(i)}, f\left(x^{(i)}\right)\right)$

Logistic Loss function

$$L\left(y^{(i)}, f\left(x^{(i)}\right)\right) = \begin{cases} -\log\left(f\left(x^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f\left(x^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$

Case 1 : If $y^{(i)} = 1$

$$L = -\log\left(f(x^{(i)})\right)$$



$$\boxed{f(x^{(i)}) \gtrsim 0.5}$$

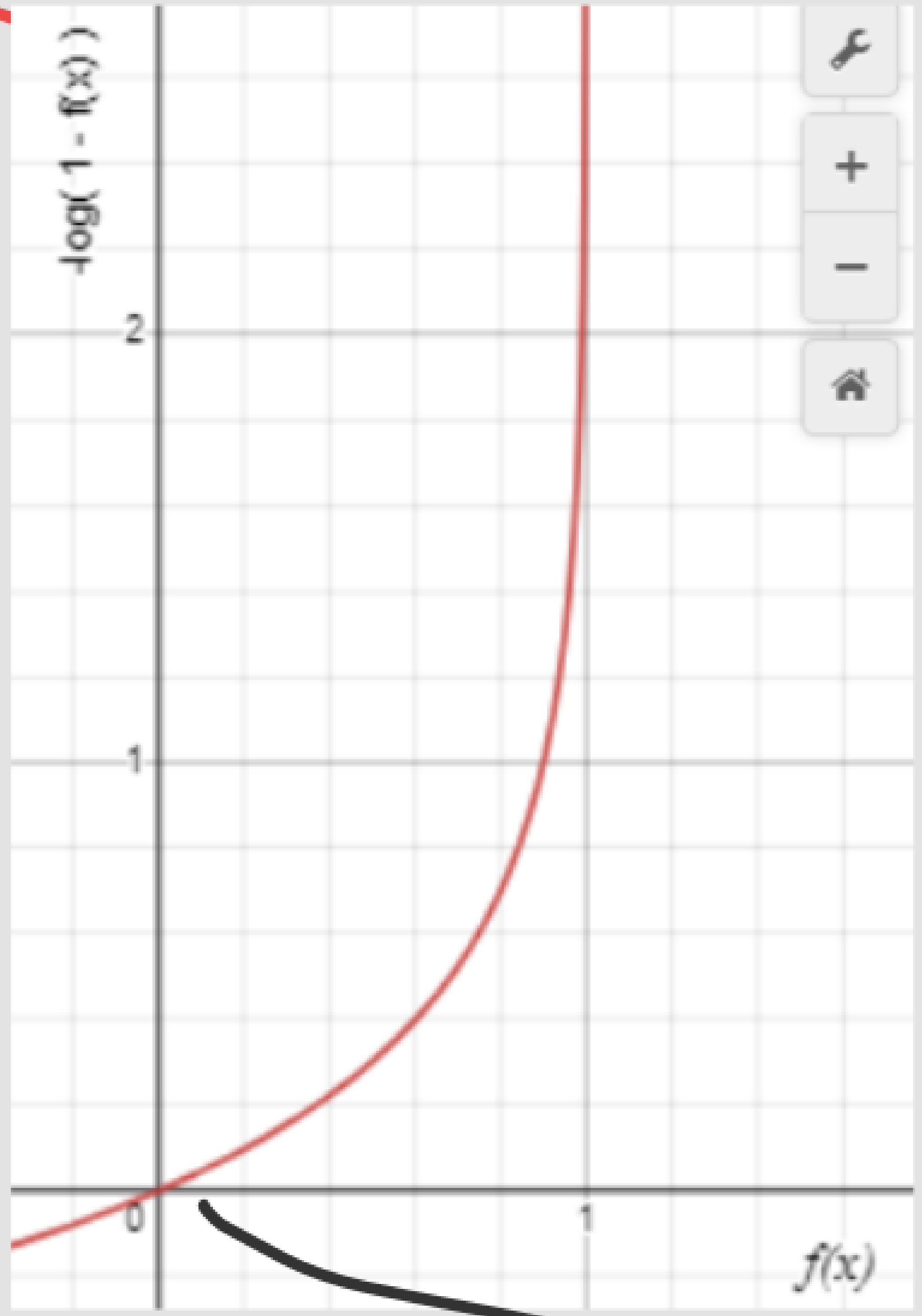
0.1

1

$f(x^{(i)}) \longrightarrow 1$, Loss $\rightarrow 0$

$f(x^{(i)}) \rightarrow 0$, Loss $\rightarrow$ High

Case 2 : If $y^{(1)} = 0$



$$\mathcal{L} = -\log\left(1 - f(x^{(1)})\right)$$

$f(x^{(1)}) \to 1$, Loss $\to$ High

$f(x^{(1)}) \to 0$, Loss $\to 0$

Simplified Cost func

$$L\left(y^{(i)}, f(x^{(i)})\right) = \begin{cases} -\log\left(f(x^{(i)})\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f(x^{(i)})\right) & \text{if } y^{(i)} = 0 \end{cases}$$

$$L\left(y^{(i)}, f(x^{(i)})\right) = \boxed{-y^{(i)} \log\left(f(x^{(i)})\right) - \left(1 - y^{(i)}\right) \log\left(1 - f(x^{(i)})\right)}$$

Final Cost function for logistic regression

$$\text{Cost} = J(w,b) = \frac{-1}{N}\sum_{i=1}^{N}\left(y^{(i)}\log\left(f(x^{(i)})\right) + (1-y^{(i)})\log\left(1-f(x^{(i)})\right)\right)$$

where $f(x^{(i)}) = \dfrac{1}{1+e^{-(wx+b)}}$

gradient descent

$$\left\{ \begin{aligned} w &= w - \alpha \frac{\partial J}{\partial w} \\ b &= b - \alpha \frac{\partial J}{\partial b} \end{aligned} \right\}$$

Gradient Descent Algorithm for Logistic Regression

→ Algo to find values of parameters w & b that minimizes the cost J.

Say $z = w_1x_1 + w_2x_2 + \dots w_dx_d + b$

repeat
{
$$w_j = w_j - \alpha \boxed{\frac{\partial J(w,b)}{\partial w_j}} \qquad \frac{\partial J(w,b)}{\partial w_j} = \frac{1}{N}\sum_i \left(f(x^{(i)}) - y^{(i)}\right) * x_j^{(i)}$$

$$b = b - \alpha \boxed{\frac{\partial J(w,b)}{\partial b}} \qquad \frac{\partial J(w,b)}{\partial b} = \frac{1}{N}\sum_i \left(f(x^{(i)}) - y^{(i)}\right)$$
}

where $f(x^{(i)}) = \dfrac{1}{1 + e^{-(wx+b)}}$

$$J(w,b) = -\frac{1}{N}\sum_{i}\left(y^{(i)}\log\left(f(x^{(i)})\right) + (1-y^{(i)})\log\left(1-f(x^{(i)})\right)\right)$$

$$J(w,b) = -\frac{1}{N}\sum_{i=1}^{N}\left(y^{(i)}\log\left(\frac{1}{1+e^{-z}}\right) + (1-y^{(i)})\log\left(1-\frac{1}{1+e^{-z}}\right)\right)$$

where $z = w_1x_1 + w_2x_2 + \cdots + w_dx_d + b$

Derivation of gradients in gradient descent (optional)

$$J(w,b) = \frac{-1}{N} \sum \left(y^{(i)} \lg\left(f(x^{(i)})\right) + (1-y^{(i)}) \lg\left(1 - f(x^{(i)})\right) \right)$$

where $f(x^{(i)}) = \frac{1}{1+e^{-z}}$ and $z = w_1x_1 + w_2x_2 + \cdots w_d x_d + b$

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

$$\frac{\partial J(w,b)}{\partial w_1} = \frac{\partial J}{\partial f} * \frac{\partial f}{\partial z} * \frac{\partial z}{\partial w_1}$$

$$\frac{-1}{N} \sum \left(\left(\frac{y^{(i)}}{f(x^{(i)})} * \frac{\cancel{-1} * e^{-z} * \cancel{-1}}{(1+e^{-z})^2} * x_1 \right) + \left(\frac{1-y^{(i)}}{1-f(x^{(i)})} * \frac{-1 * \cancel{-1} * e^{-z} * \cancel{-1}}{(1+e^{-z})^2} * x_1 \right) \right)$$

$$= \frac{-x_1}{N} \sum \left(\left(y^{(i)} * \frac{\cancel{1+e^{-z}}}{1} * \frac{e^{-z}}{(1+e^{-z})^{\cancel{2}}} \right) + \left(-(1-y^{(i)}) * \frac{\cancel{1+e^{-z}}}{\cancel{e^{-z}}} * \frac{\cancel{e^{-z}}}{(1+e^{-z})^{\cancel{2}}} \right) \right)$$

$$= \frac{-x_1}{N} \sum \left(y^{(i)} * \left(1 - f(x^{(i)})\right) - (1-y^{(i)})\left(f(x^{(i)})\right) \right)$$

$$= \frac{x_1}{N} \sum \left(f(x^{(i)}) - \cancel{y^{(i)} f(x^{(i)})} - y^{(i)} + \cancel{y^{(i)} f(x^{(i)})} \right)$$

$$= \frac{x_1}{N} \sum \left(f(x^{(i)}) - y^{(i)} \right)$$