Multiple Linear Regression from scratch

1) Linear Regression
   i) Example Data and Predictions
   ii) Loss function / Cost function
   iii) Gradient Descent

2) Multiple Linear Regression
   i) Multiple x Features
   ii) Updated Cost
   iii) Gradient Descent

~~Rough Work~~

~~$x = [1,3,4]$~~

~~$y = [2,4,8]$~~

~~$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{4-2}{3-1} \approx 1 \qquad \frac{8-4}{4-3} \approx 4 \qquad \frac{4-1}{8-2} \approx 0.5$~~

~~$m_{avg} = \frac{5.5}{3} = 1.7\ldots$~~

~~Just to check assumption~~

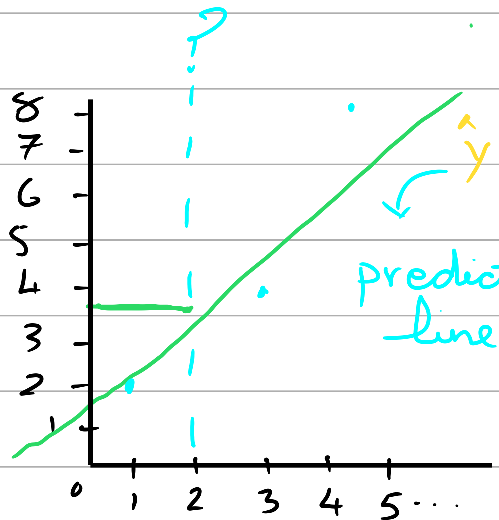
→ Linear Regression :-

* Keeping Small data for easier calculation

	x Size	y Price
i=1	1	2
2	3	4
3	4	8
⋮		

=> x_i => $x_1 = 1$, $x_2 = 3$, $x_3 = 4$ ⋮

=> y_i => $y_1 = 2$, $y_2 = 4$, $y_3 = 8$



$\hat{y} = wx + b \Rightarrow \hat{y} = 1.5x + 1$

prediction line

slope (m) / weight

intercept (c) / bias

Continuing with the assumption that

$\Rightarrow w = 1.5$

$\Rightarrow b = 1$

Finding out Mean Squared Error

$\Rightarrow y = [2, 4, 8]$

& As per assumption :-

$\Rightarrow$ y_hat $= [2.5, 5.5, 7]$

$$\Rightarrow MSE = \frac{1}{2m} \sum_{i=1}^{n} \left(\underbrace{\hat{y}_i}_{\text{y\_hat}} - \overbrace{y_i}^{y} \right)^2$$

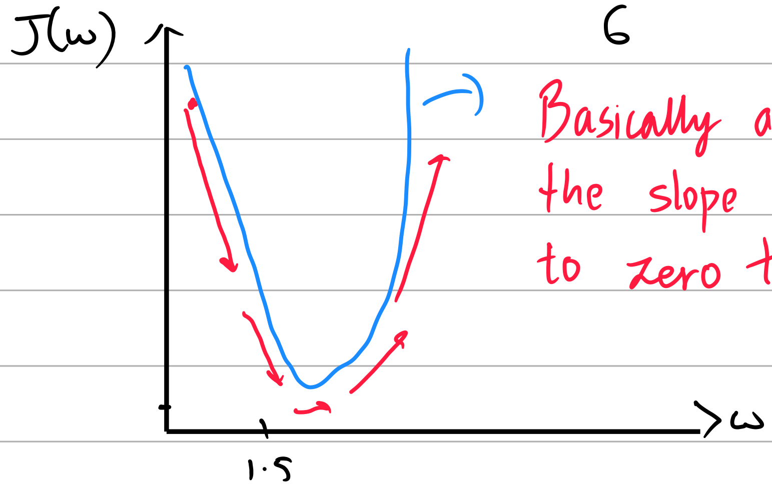
will make sense in later Parts of Derivation

$$= \frac{1}{2 \times 3} \left[(2.5-2)^2 + (5.5-4)^2 + (7-8)^2 \right]$$

$$= \frac{1}{6} \left[(0.5)^2 + (1.5)^2 + (-1)^2 \right]$$

$$= \frac{1}{6} \left[0.25 + 2.25 + 1 \right]$$

$$\text{Cost fxn} = J(w,b) = \frac{1}{6} [3.5] = 0.583\ldots$$



Basically as we go from left to right the slope increases from being negative to zero to positive

So basically, we have to get partial derivative of $J(w,b)$ w.r.t. w and b respectively, to get the global minimum.

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2$$

POWER RULE

CHAIN RULE

$$\Rightarrow \frac{\partial J}{\partial w} = \frac{\partial}{\partial w}\left[\frac{1}{2m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2\right]$$

$\hat{y}_i \to w \cdot x_i + b$

x_i → given data

b → const w.r.t w

$$= \frac{\partial}{\partial w}\left[\frac{1}{2m} \sum_{i=1}^{m} \left(-y_i + (w \cdot x_i + b)\right)^2\right]$$

When we differentiate a sum, we can just differentiate the term inside the sum.

∴ We need to find

$$\Rightarrow \frac{\partial J}{\partial w} = \frac{1}{2m} \sum_{i=1}^{m} \left[\frac{\partial}{\partial w} \Big(\underbrace{-y_i + (w \cdot x_i + b)}_{u}\Big)^2\right]$$

$\Rightarrow E_i = u^2$

$$\frac{\partial(-\hat{y}_i + (w \cdot x_i + b))}{\partial b} = 1$$

→ Chain Rule:

$$\Rightarrow \frac{\partial E_i}{\partial w} = \frac{\partial E_i}{\partial u} \times \frac{\partial u}{\partial w}$$

$$= \frac{\partial (u^2)}{\partial u} \times \frac{\partial(-y_i + (w \cdot x_i + b))}{\partial w}$$

$= 2u \times (x_i)$

$$\Rightarrow \frac{\partial E_i}{\partial w} = 2\left(\hat{y}_i - (w x_i + b)\right) x_i$$

$$\Rightarrow \frac{\partial J}{\partial w} = \frac{1}{2m} \sum_{i=1}^{m} \left[\frac{\partial}{\partial w} (E_i) \right]$$

$$= \frac{1}{2m} \sum_{i=1}^{m} \left[2(-y_i + w x_i + b) x_i \right]$$

$\Rightarrow$ Similarly,

$$\Rightarrow \frac{\partial J}{\partial b} = \frac{1}{\cancel{2}m} \sum_{i=1}^{m} \left[\cancel{2}(-y_i + w x_i + b) \right]$$

On Repeat (Iterations/Epochs) :-

$$\left.\begin{aligned} w &= w - \alpha \frac{\partial J}{\partial w} \\ b &= b - \alpha \frac{\partial J}{\partial b} \end{aligned}\right\} \alpha = \text{learning Rate (a very small number.)}$$

→ Multiple Linear Regression :-

x_1 size	x_2 age	price
1	2	3
$\boxed{2}$	1	6
3	1	8
1	3	7

x_{12}

x_j : jth feature

x_1 = size

x_2 = age

x_{ji} = jth feature of the ith training example.

$$\hat{y}_i = w x_i + b \Rightarrow \text{Linear Regression}$$

$$\hat{y}_i = w_1 x_1 + w_2 x_2 + b \Rightarrow \text{Multiple}$$

$$\hat{y}_i = \bar{w} \cdot \bar{x}_i + b$$

- Cost Fxn :-

$$\Rightarrow J(w,b) = \frac{1}{2n} \sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2 = \frac{1}{n} \sum_{i=1}^{n} \left(wx + b - y_i\right)^2$$

- Linear Regression, which can be extended to

$$= \frac{1}{2n} \sum_{i=1}^{n} \left(\bar{w} \cdot \bar{x}_i + b - y_i\right)^2$$

$$\Rightarrow \frac{\partial J}{\partial w} = \frac{1}{2n} \sum_{i=1}^{n} \frac{\partial}{\partial w} \Big[\underbrace{\bar{w} \cdot \bar{x}_i + b - y_i}_{u}\Big]^2$$

$$\Rightarrow E_i = u^2$$

$$\Rightarrow \frac{\partial E_i}{\partial w} = \frac{\partial E_i}{\partial u} \times \frac{\partial u}{\partial w} \rightarrow \text{Chain Rule}$$

$$\Rightarrow \frac{\partial E_i}{\partial w} = \frac{\partial u^2}{\partial u} \times \frac{\partial(\bar{w}\cdot\bar{x} + b - y_i)}{\partial w}$$

$$\Rightarrow \frac{\partial E_i}{\partial w} = 2u \times \bar{X}$$

$$\Rightarrow \frac{\partial J}{\partial w} = \frac{1}{2n} \sum_{i=1}^{n} \frac{\partial E_i}{\partial w}$$

$$\Rightarrow \frac{\partial J}{\partial w} = \frac{1}{2n} \sum_{i=1}^{n} 2u\bar{X}$$

$$\Rightarrow \frac{\partial J}{\partial w_i} = \frac{\cancel{2}}{\cancel{2}n} \sum_{i=1}^{n} (\underbrace{\bar{w}\cdot\bar{x}_i}_{w_1x_1 + w_2x_2 + \dots} + b - y_i)\,\bar{X}_{ji} \Rightarrow \bar{w} = [w_1, w_2, w_3, \dots]$$

w_j : jth weight

$$\therefore \frac{\partial J}{\partial b} = \frac{1}{\cancel{2}n} \sum_{i=1}^{n} \cancel{2}u$$

$$\Rightarrow \frac{\partial J}{\partial b} = \frac{\cancel{2}}{\cancel{2}n} \sum_{i=1}^{n} (\bar{w}\cdot\bar{x}_i + b - y_i)$$

On Repeat (Iterations/Epochs) :-

$$w_j = w - \alpha \frac{\partial J}{\partial w_j} \rightarrow \text{for } j = 1, 2, \dots n$$

→ no. of features

$$b = b - \alpha \frac{\partial J}{\partial b}$$

α = learning Rate (a very small number.)

Multiple Logistic Regression :-

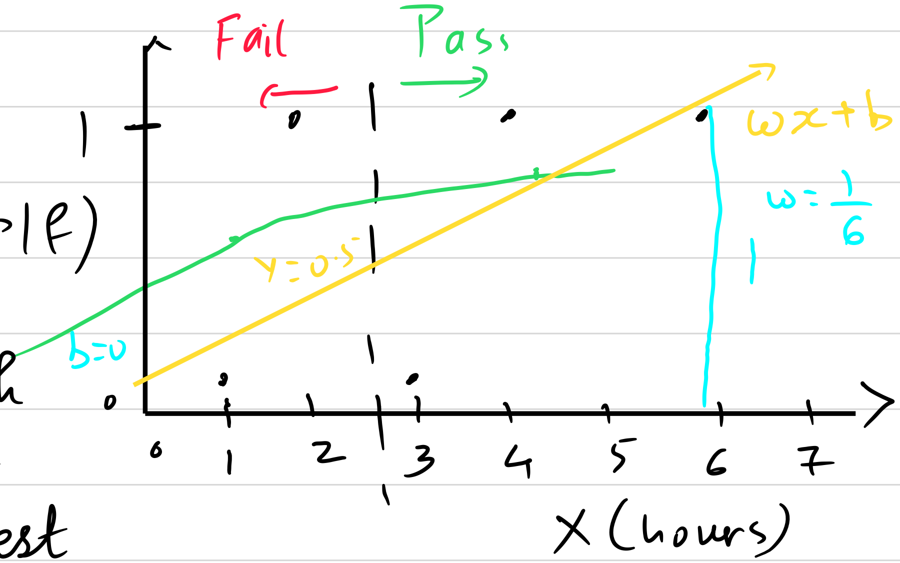
1) Logistic Regression :-
   i) Example Data and Predictions
   ii) Cost f×n
   iii) Gradient Descent

2) Multiple Logistic Regression :-
   i) Multiple x features
   ii) Updated Cost & Gradient Descent

→ Logistic Regression :-

	X: hours	Y: pass/fail (1/0)
i=0	6	1
1	1	0
2	4	1
3	3	0
4	2	1
⋮		
n		

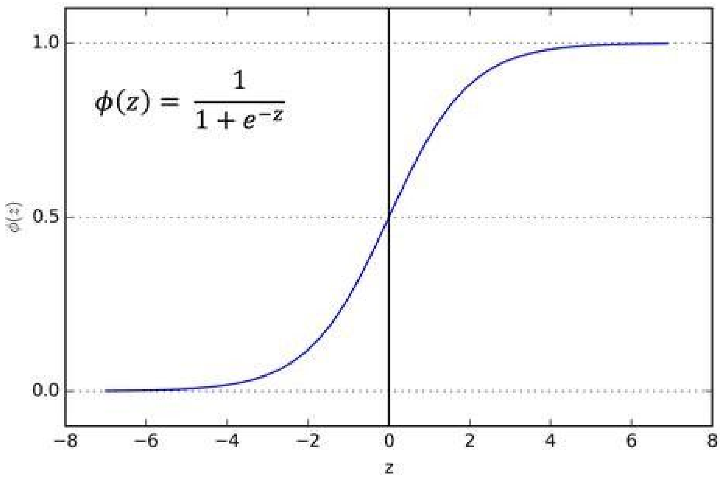
x_i - hours of i^{th} student
y_i : i^{th} student Result
$x_0 = 6$ $y_0 = 1$
$x_3 = 3$ $y_3 = 0$



We could go through Linear Regression & figure how to best fit a line to the data & then just take $y = 0.5$ where anything above it is a pass &

below it is a fail, but it turns out with Logistic Regression, there's another eqⁿ we can use, i.e., Sigmoid function, which will increase the accuracy even more.

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



$$\Rightarrow g(0) = \frac{1}{1+e^{-0}} = \frac{1}{2} = 0.5$$

$$g(1) = \frac{1}{1+e^{-1}} \approx 0.73$$

$$g(-1) = \frac{1}{1+e} \approx 0.27$$

$$\Rightarrow f_{wb}(x) = g(wx+b)$$

let $w = \frac{1}{6}$ & $b=0$

$$\Rightarrow \frac{1}{6}x + 0 \Rightarrow \frac{1}{6}(1) = 0.16\ldots \quad \Rightarrow f_{wb}(1) = g(0.1\bar{6}) = \frac{1}{1+e^{-(0.1\bar{6})}}$$

0.54

$$\Rightarrow \frac{1}{6}(4) = 0.66\ldots \quad \Rightarrow f_{wb}(4) = g(0.\bar{6}) = \frac{1}{1+e^{-(0.\bar{6})}}$$

0.66

$$\Rightarrow \hat{y}_i = \begin{cases} f_{wb}(x_i) \geq 0.5 \Rightarrow 1 \text{ (pass)} \\ f_{wb}(x_i) < 0.5 \Rightarrow 0 \text{ (fail)} \end{cases}$$

$\Rightarrow P(y=1 \mid x) = f_{wb}(x)$

$\Rightarrow$ For $x = 1$, the Probability of y being 1 is 0.54
& for $x = 4$, " " " " " " 0.66

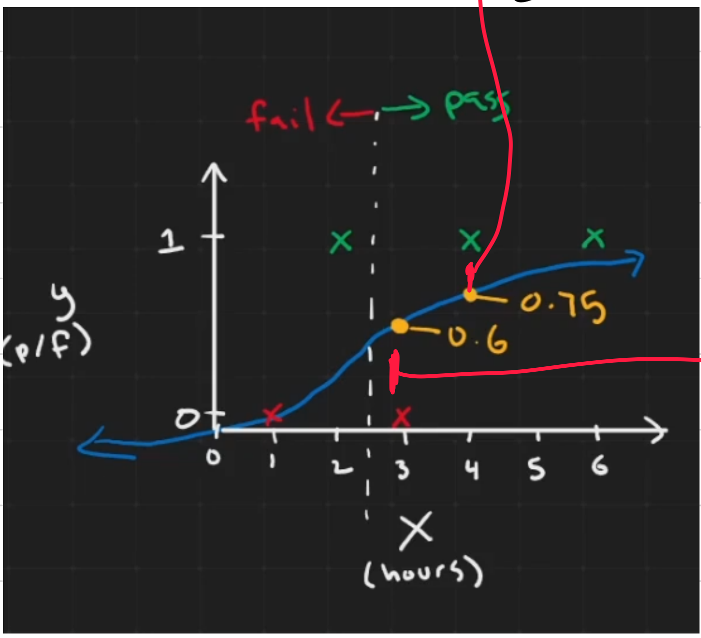
For linear Regression, we used MSE,
for logistic Regression, since MSE can cause issues, we'll use Log loss function

$$L(f_{wb}(x_i) - y_i) = \begin{cases} -\log(f_{wb}(x_i)), & \text{if } y_i = 1 \\ -\log(1 - f_{wb}(x_i)), & \text{if } y_i = 0 \end{cases}$$

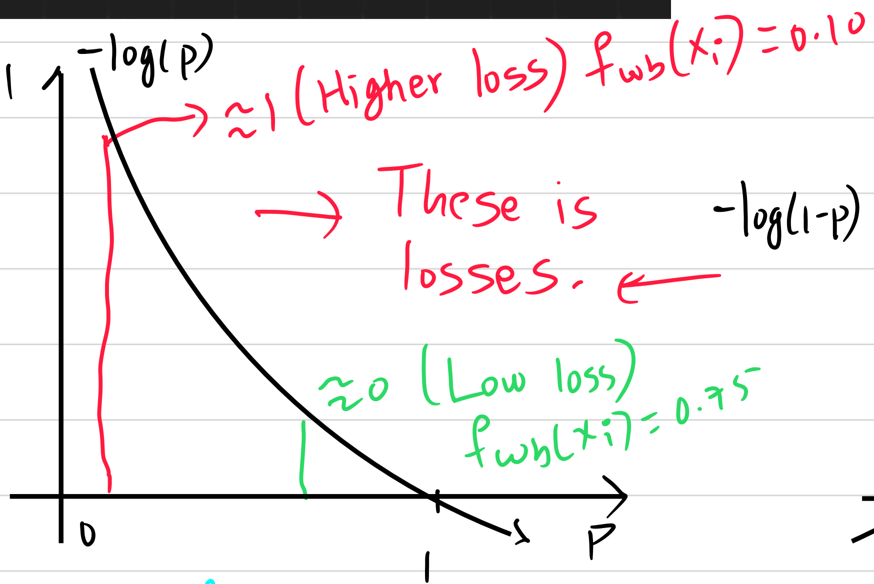
$\because y=1$

$\rightarrow P$

$\because y=0$

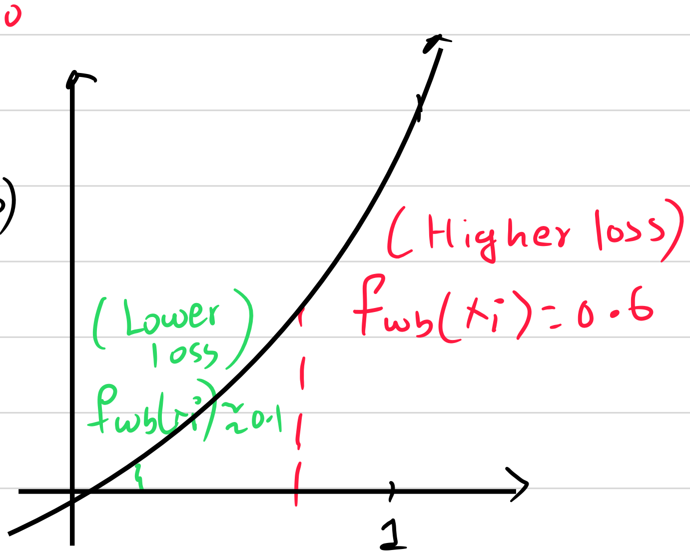


error



if $y_i = 1$

-log(1-p)



if $y_i = 0$

$x = 4$
$y = 1$
$f_{wb}(x_i) = 0.75 = y\_prob$

$x = 3$
$y = 0$
$f_{wb}(x_i) = 0.6 = y\_prob$

$$\Rightarrow \boxed{L(f_{wb}(x_i) - y_i) = \underbrace{-y_i \log(f_{wb}(x_i))} - \underbrace{(1-y_i)\log(1-f_{wb}(x_i))}}$$

if $y_i = 1$	1	0
if $y_i = 0$	0	1

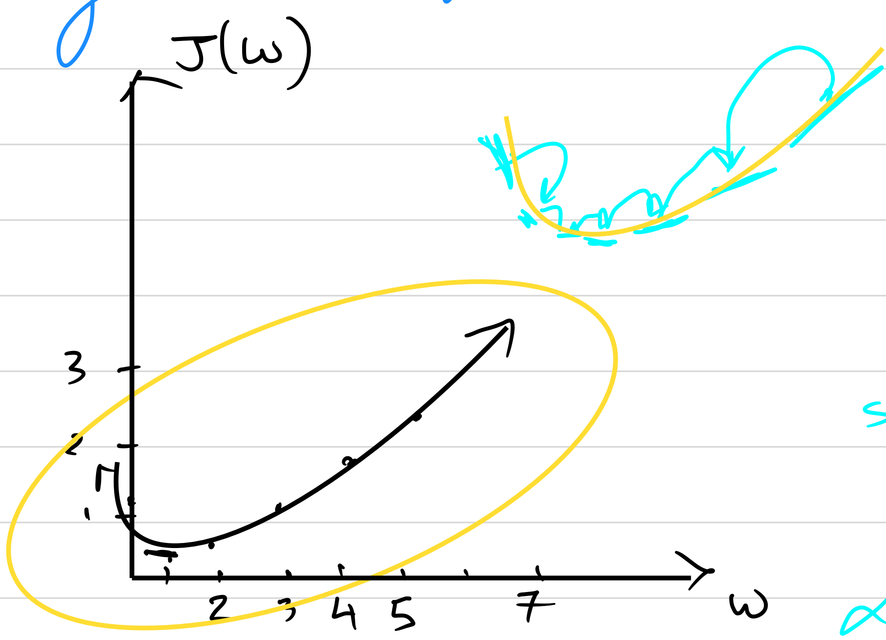
loss fxn, now ready to be used to get

Cost fxn :-

$$\Rightarrow J(w,b) = \frac{1}{m}\sum_{i=0}^{m} L(f_{wb}(x_i), y_i)$$

$$\Rightarrow J(w,b) = \frac{1}{m}\sum_{i=0}^{m}\left[-y_i \log(f_{wb}(x_i)) - (1-y_i)\log(1-f_{wb}(x_i))\right]$$

Gradient Descent :-



+ve slope → w↓
−ve slope → w↑

$$\text{slope} = \frac{\partial J(w)}{\partial w} \quad \begin{array}{ll} > 0 & w\downarrow \\ < 0 & w\uparrow \end{array}$$

α = learning Rate

Repeat :-

$$w = w - \alpha \frac{\partial J}{\partial w}$$

Similarly, $b = b - \alpha \frac{\partial J}{\partial b}$

$$J(w,b) = \frac{1}{m} \sum_{i=1}^{m} \Big[\underbrace{\left(-y_i \times \log(f_{wb}(x_i)\right) - (1-y_i)\left(\log(1-f_{wb}(x_i))\right)}_{= a} \Big]$$

$$\frac{\partial J(w,b)}{\partial (w)} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial a}{\partial w}$$

$$a = \left(-\overset{c}{y_i} \times \log\left(\underline{f_{wb}(x_i)}\right)\right) - \left(1-\overset{c}{y_i}\right)\left(\log\left(1-\underline{f_{wb}(x_i)}\right)\right)$$

$$f_{wb}(x_i) = \frac{1}{1-e^{-(wx_i+b)}}$$

$$a = -y_i \log\left(\frac{1}{1+e^{(wx_i+b)}}\right) - (1-y_i)\left(\log\left(1-\frac{1}{1+e^{-(wx_i+b)}}\right)\right)$$

$$\frac{\partial a}{\partial w} = ? = \frac{\partial a}{\partial f_{wb}(x_i)} \times \frac{\partial f_{wb}(x_i)}{\partial w}$$

$$\frac{\partial a}{\partial f_{wb}(x_i)} = -y_i \frac{\partial \log(f_{wb}(x_i))}{\partial f_{wb}(x_i)} - (1-y_i)\frac{\partial \log(1-f_{wb}(x_i))}{\partial f_{wb}(x_i)}$$

$$= \frac{-y_i}{f_{wb}(x_i)} - (1-y_i)\left[\frac{1}{1-f_{wb}(x_i)}(-1)\right]$$

$$= \frac{-y_i}{f_{wb}(x_i)} + \frac{1}{1-f_{wb}(x_i)} - \frac{y_i}{1-f_{wb}(x_i)}$$

$$= \frac{-y_i(1-f_{wb}(x_i)) + f_{wb}(x_i) - y_i f_{wb}(x_i)}{f_{wb}(x_i)(1-f_{wb}(x_i))}$$

$$= \frac{-y_i + \cancel{y_i f_{wb}(x_i)} + f_{wb}(x_i) - \cancel{y_i f_{wb}(x_i)}}{f_{wb}(x_i)(1-f_{wb}(x_i))}$$

$$\frac{\partial a}{\partial f_{wb}(x_i)} = \frac{f_{wb}(x_i) - y_i}{f_{wb}(x_i)(1-f_{wb}(x_i))}$$

$z = wx_i + b$

$$\Rightarrow \frac{\partial (f_{wb}(x_i))}{\partial w} = \frac{\partial \sigma(z)}{\partial w} \quad \left(\sigma(wx_i+b) = \frac{1}{1+e^{-(wx_i+b)}}\right)$$

$$= \frac{\partial \sigma(z)}{\partial z} \times \frac{\partial z}{\partial w}$$

$$= \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) \times \frac{\partial (wx_i+b)}{\partial w}$$

$$= \frac{e^{-z}}{(1+e^{-z})^2} \times x_i$$

$$= \frac{1}{1+e^{-z}} \times \frac{1+e^{-z}-1}{1+e^{-z}} \times x_i$$

$$= \frac{1}{1+e^{-z}}\left[\frac{\cancel{1+e^{-z}}}{\cancel{1+e^{-z}}} - \frac{1}{1+e^{-z}}\right] x_i$$

$$= \sigma(z)(1-\sigma(z))\, x_i$$

$$= f_{wb}(x_i)(1-f_{wb}(x_i))\, x_i$$

$$\frac{\partial a}{\partial w} = ? = \frac{\partial a}{\partial f_{wb}(x_i)} \times \frac{\partial f_{wb}(x_i)}{\partial w}$$

$$= \frac{f_{wb}(x_i) - y_i}{\cancel{f_{wb}(x_i)(1-f_{wb}(x_i))}} \times \cancel{f_{wb}(x_i)(1-f_{wb}(x_i))}\, x_i$$

$$= x_i f_{wb}(x_i) - x_i y_i = \text{error} * x_i$$

$$\frac{\partial a}{\partial b} = \frac{\partial a}{\partial f_{wb}(x_i)} \times \frac{\partial f_{wb}(x_i)}{\partial b}$$

$$= f_{wb}(x_i) - y_i \quad = \text{error}$$

$$\frac{\partial J(w,b)}{\partial w} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial a}{\partial w}$$

$$\frac{\partial J(w,b)}{\partial b} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial a}{\partial b}$$

Repeat

$$w = w - \alpha \frac{\partial J}{\partial w}$$

$$b = b - \alpha \frac{\partial J}{\partial b}$$