Let's start with logistic regression mathematically

$\hat{y}$ (y hat) is the o/p or probabilities.

$$\hat{y} = P(y=1 \mid x)$$

Read: Probability of $y=1$, given x

$x \in R^{n_x}$ [x is a set of Real numbers, 'm' training examples or i/p vector, $n \rightarrow$ dimension of i/p 'x'. number of variables.]

Parameters

$w \in R^{n_x}$

$b \in \mathbb{R}$

$$X = \begin{bmatrix} | & | & | & & \\ x^1 & x^2 & x^3 & \cdots & m \\ | & | & | & & \end{bmatrix} \quad n$$

m

$$Y = [y^1 \; y^2 \; \cdots \; y^m]$$

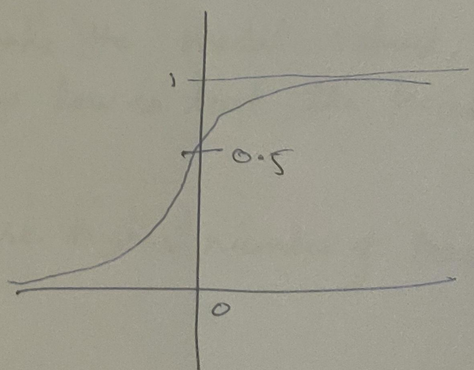
$Y \in R^{1 \times m}$

Linear reg

$$\hat{y} = w^T x + b \qquad w = [w_1 \; w_2 \; w_3 \; w_4 \cdots]$$

Logistic regression

$$\hat{y} = \sigma(\underbrace{w^T x + b}_{z})$$



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

If z is large

If z is small

Loss function

A function that captures the loss between predicted and actual.

Anybody knows the loss function of linear reg?

$$\mathcal{L}(\hat{y}, y) = -(y \log \hat{y} + (1-y) \log(1-\hat{y}))$$

we want to minimize the loss function. but how?
we want to find an optimal value for w and b which would minimize the cost function

Cost function

$$J(w,b) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(\hat{y}; y).$$

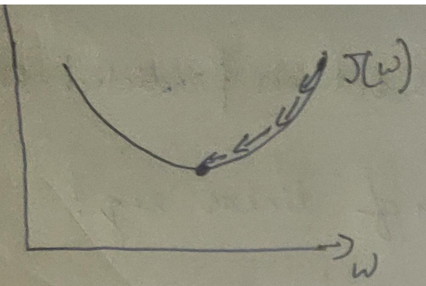
$$= \frac{1}{m} \sum_{i=1}^{m} \left(-(y \log \hat{y} + (1-y) \log(1-\hat{y})) \right)$$

If $y=0$ $-\log(1-\hat{y})$ to have it minimized we want $\hat{y}$ to be small

If $y=1$ $-\log \hat{y}$ to make this minimum, we wan $\hat{y}$ to be large

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

we want w, b that minimizes $J(w,b)$



Repeat

$$w = w - \alpha \frac{dJ(w)}{\partial w}$$

or $\boxed{w = w - \alpha dw}$

for J(w,b)

$$w = w - \alpha \frac{\partial J(w,b)}{dw}$$

$$b = b - \alpha \frac{dJ(w,b)}{db}$$

Let's look at computational graph

Inorder to feedback the new computations, that come from the Loss we need to know 2 things

1. Forward propagation
2. Backward propagation.

Lets understand this with a simple example, then we will go back to logistic reg

Suppose

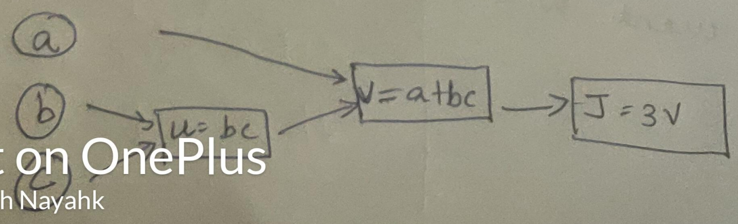
$J(a,b,c) = 3(a+bc)$

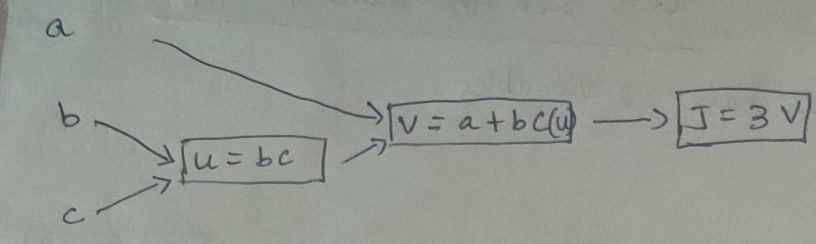
lets store them in variables

$u = bc$

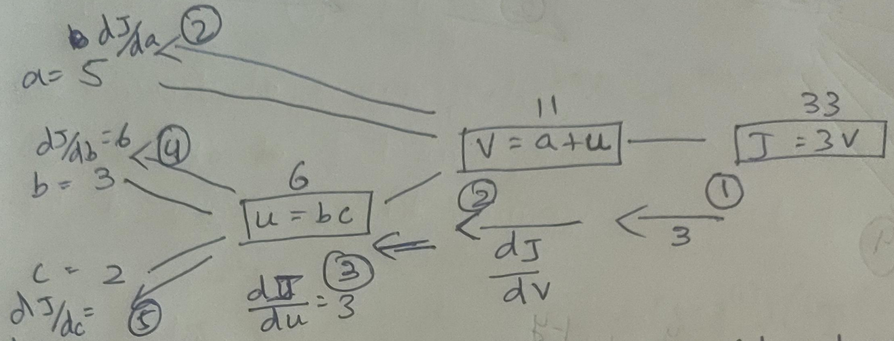
$V = a + bc \; [a+u]$

$J = 3(a+bc) \; [3V]$.





Suppose $a = 5,\ b = 3,\ c = 2$



$\frac{dJ}{dv}$ = ? [what will the change in J be if value of v is changed] that what a derivative means.

(1) $\boxed{\frac{dJ}{dV} = 3}$

J = 3V
V = 11
if V = 11.001
J = 33.003.

(2) $\frac{dJ}{da}$ = ? [what does that mean?].

a = 5 → 5.001
V = 11 → 11.001
J = 33 → 33.003

$\frac{dJ}{da} = 3$

So, if you change a, it would change v, which will change a

$$\Rightarrow \frac{dJ}{da} = \frac{dJ}{dV}^{\to 3} \cdot \frac{dV}{da}^{\to 1}$$

= (chain rule).

(3)

$\frac{dJ}{du} = 3$

u = 6 → 6.001
V = 11 → 11.001
J = 33 → 33.003

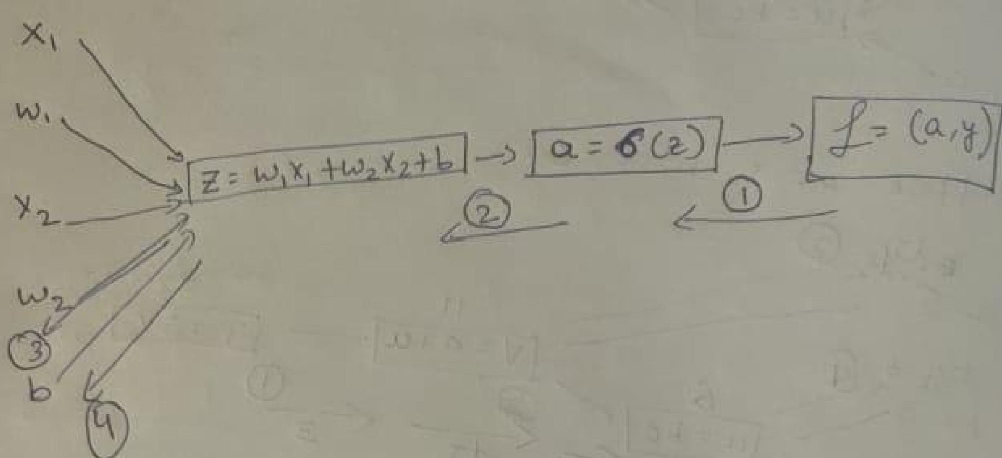
$$\underbrace{\frac{dJ}{dV}}_{3} \cdot \underbrace{\frac{dV}{du}}_{1}$$

(5) $\frac{dJ}{dc} = 9$

(4) $\frac{dJ}{db} = \frac{dJ}{du} \cdot \frac{du}{db} = 6$

Lets see the same for logit

suppose we have 2 variables.



1 $\dfrac{dL}{da} = -\dfrac{y}{a} + \dfrac{1-y}{1-a}$

lets call it `da' $= -\dfrac{y}{a} + \dfrac{1-y}{1-a}$

2 $\dfrac{dL}{dz} = a - y$ $\boxed{\left[\dfrac{dL}{da} \cdot \dfrac{da}{dz}\right] \rightarrow a(1-y)}$

lets call it `dz'

3 $\dfrac{dz}{db} = dz$ lets call `db'

4 $\dfrac{dL}{dw_2} = x_2 \cdot dz$ lets call dw_2

$\dfrac{dL}{dw_1} = x_1 \cdot dz$ lets call dw_1

$$\boxed{\begin{aligned} w_1 &= w_1 - \alpha\, dw_1 \\ w_2 &= w_2 - \alpha\, dw_2 \\ b &= b - \alpha\, db \end{aligned}}$$

Now all that was being done on one row/training examp
we have to do it for all. for that we will minimize the
cost function

$$J = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(a^{(i)}, y^{(i)}) \qquad \left[a^{(i)} = \hat{y}^{(i)} = \sigma(z^{(i)}) = \sigma(w^T x + b) \right.$$

$i \rightarrow$ Training examples.

In computer [Pseudo algo]

$J = 0, \quad dw = 0, \quad dw_2 = 0, \quad db = 0$] Random initializati

for $i = 1$ to m

$$z^{(i)} = w^T x^{(i)} + b^{(i)}$$

$$a^{(i)} = \sigma(z^{(i)})$$

$$J \mathrel{+}= -\left[y^{(i)} \log(a^{(i)}) + (1 - y^{(i)})(\log(1 - a^{(i)})) \right]$$

$$dz^{(i)} = a^{(i)} - y^{(i)}$$

$$\boxed{\begin{aligned} dw_1 &\mathrel{+}= x_1^{(i)} dz^{(i)} \\ dw_2 &\mathrel{+}= x_2^{(i)} dz^{(i)} \\ db &\mathrel{+}= dz^{(i)} \end{aligned}} \quad n = 2$$

$J \mathrel{/}= m \qquad db \mathrel{/}= m$

$dw_1 \mathrel{/}= m \qquad dw_2 \mathrel{/}= m$

update

$$w_1 = w_1 - \alpha dw_1$$

$$w_2 = w_2 - \alpha dw_2$$

$$b = b - \alpha db$$