Derivatives

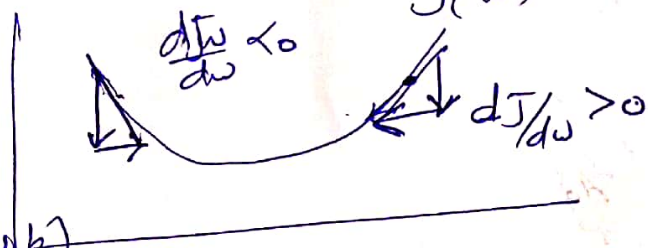
Gradient Descent

$= \sigma(\theta^T x + \theta_0)$ $\quad \sigma(z) = \frac{1}{1+e^{-\theta^T x + \theta_0}}$

$h_\theta(x) = \hat{y} = \sigma(w^T x + b)$, $\quad \sigma(z) = \frac{1}{1+e^{-z}}$ ✓

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

- To optimized $J(w,b)$ we ~~need~~ use gradient descent ✓
- GD requires Derivative to update weights



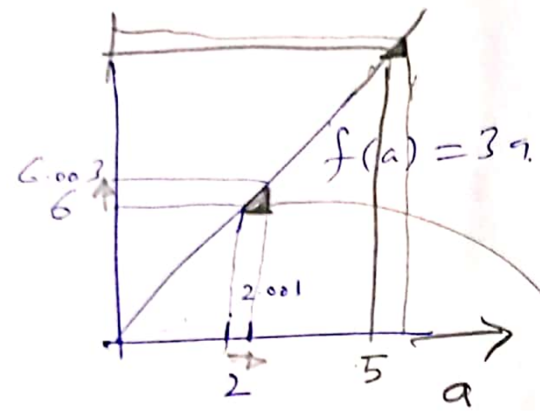
$w = w - \alpha \frac{dJ(w,b)}{dw}$ $\qquad w := w - \alpha \frac{dJ(w)}{dw}$

$\boxed{b = b - \alpha \frac{dJ(w,b)}{db}}$ $\qquad \boxed{w := w - \alpha\, dw}$

$\boxed{\text{Slope of function / Derivative of function}}$

~~Intuitio~~

Intuitive Understanding of derivatives



(A)

Intuition

6.003

$a = 2 \quad f(a) = 6$ value

① If we nudge/pump up a variable a by smaller) $\Delta x = 0.001$ $a = 2.001 \quad f(2.001) = 6.003$

② Then f(2.001), then f(a) is go up 3 times

$\therefore \Delta x = 0.001 \Rightarrow \Delta h = 0.003$

$\Rightarrow$ Slope at 2 $= \frac{\Delta H}{\Delta x} = \frac{0.003}{0.001} = 3$

$\therefore$ Slope of the function at $a = 2$ is 3

Slope = Derivative 2

It can be visually seen on small triangle

$\Rightarrow$ Slope of this small Triangle is $\frac{\Delta H}{\Delta x}$

(B)

$\Rightarrow$ calculus formula

$f(a) = 3a$

$\frac{d\,f(a)}{da} = 3$

$\Rightarrow$ Slope of f(a) is remained same at everywhere

→ let see this function at Slope at different points

$a = 5, \quad f(5) = 15$

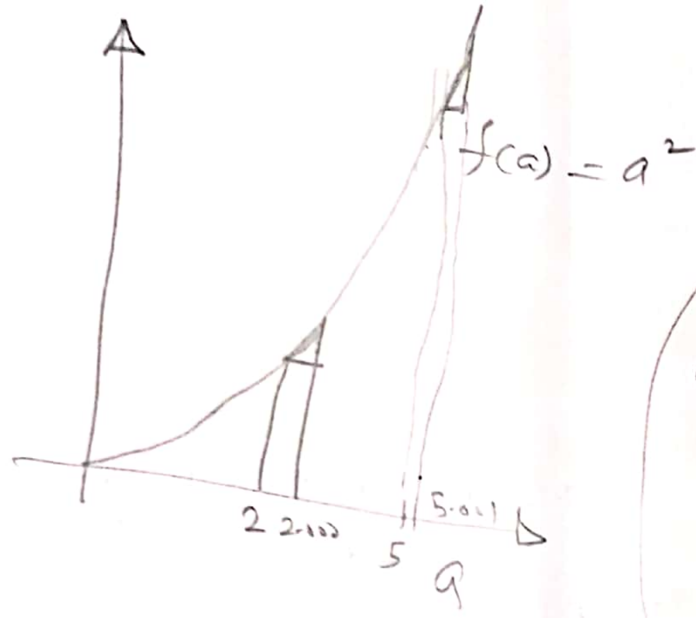
$a = 5.001 \Rightarrow f(5.001) = 3 \times 5.001 = 15.003$

$\Delta X = 0.001, \quad \Delta Y = 0.003$

Slope of the function at 5 is show with small triangle. and

Slope of f(a) at 5 $= \frac{\Delta Y}{\Delta X} = \frac{0.003}{0.001} = 3$

$\Rightarrow \frac{d f(a)}{da} = 3 = \frac{d}{da} f(a)$



$a = 2 \Rightarrow f(a) = 4$

$a = 2.001 \Rightarrow f(2.001) = 4.004$

~~Slop~~ $\Delta x = 0.001$

$\Delta Y = 0.004$

$slop = 4 = \frac{\Delta Y}{\Delta X} = \frac{0.004}{0.001}$

$a = 5 \quad f(a) = 25$

$a = 5.001 \quad f(a) = 25.010$

$slop \quad \frac{\Delta Y}{\Delta X} = \frac{0.010}{0.001} = 10$

slope of this function is different at different points

⇒ In calculus table

$f(a) = a^2$

$\boxed{\frac{d}{da} f(a) = 2a}$ i.e. is formula of derivative at a specific point

Now if $a = 2$

$\boxed{\text{derivative} = \text{slope} = 4}$

which is same as we have.

See by smaller pumping up in horizontal direction & see its effect on vertical direction.

$\boxed{\frac{d}{da}(f(5)) = 2 \times 5 = 10}$

- $f(a) = a^3$

In calculus

$$\frac{d}{da} f(a) = 3a^2$$

II $\Rightarrow$ at $a = 2$

$$\frac{d}{da} f(a) = 3 \times 4 = 12$$

$\Rightarrow$ Intuition

$a = 2$ $\quad f(a) = 8$

$a = 2.001$ $\quad f(a) = 8.012$

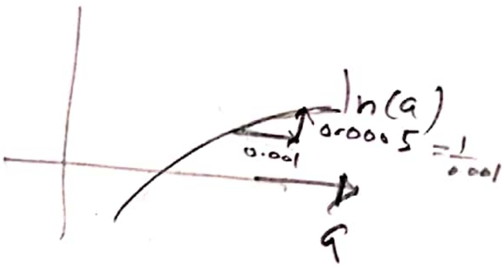
$$\frac{\Delta y}{\Delta x} = \frac{0.012}{0.001} = 12$$

Same

- $f(a) = \log_e(a) = \ln(a) = \log(a)$

In Calculus (formulation) vs Intuition

$$\frac{d}{da} f(a) = \frac{1}{a}$$



at $a = 2$

$$f(2) = \frac{1}{2}$$

$a = 2$ $\quad f(a) = \ln(2) = 0.693$
$= 0.693147$

- prompt up x by small value.

$a = 2.001$ $\quad f(a) = \ln(2.001) = 0.693$
$= 0.693647$

$\Delta x = 0.001$ $\quad \Delta y = 0.0005$

$$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{0.0005}{0.001} = 0.5 = \frac{1}{2}$$

TakeWAY MESSAGES

2 Message =
1. Derivative of function is slope of line (fact)
2. /
3. $a^2, a^3,$ slope different at different points
4. Formulas of derivative can be found from
Rn Bo

*

Computation of Neural Network is organised in forward propagation steps to compute the output of the network. — followed by backward propagation step which is used to compute gradients / derivative

* Computation graph tell us why NN is organised this way

× let consider a simple function instead of a complex loss function of NN.

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

① $v = bc$

② $u = a + v$

③ $J = 3 * u$

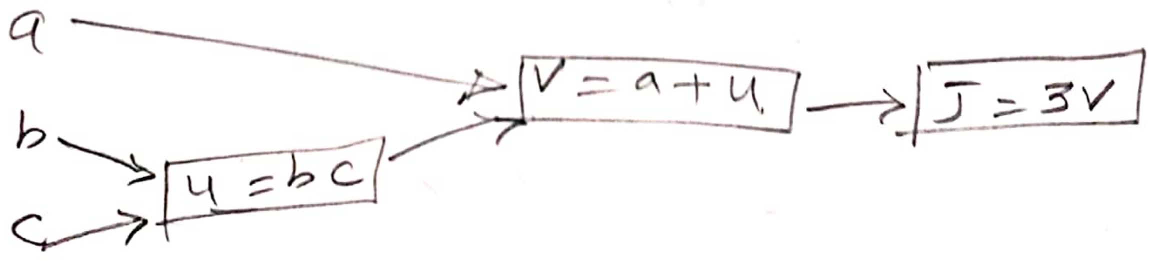
→ we can compute this function in 3 steps

$u = bc$

$v = a + u$

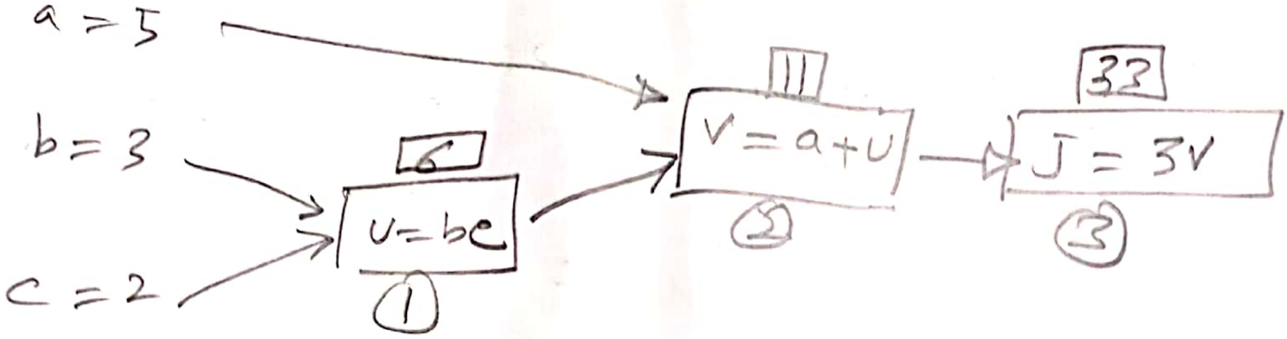
$J = 3v$

- we [illegible] these 3 steps in a computation graph & compute result of J.



⇒ Now Consider following inputs to graph.

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



① $U = bc$

② $V = a + U$

③ $J = 3V$.

$$J = 3(\underbrace{a+\overbrace{bc}^{U}}_{V})$$
J

we can verify it:

$$J(a,b,c) = 3(a+bc) = 3(5+3\times2) = 33$$

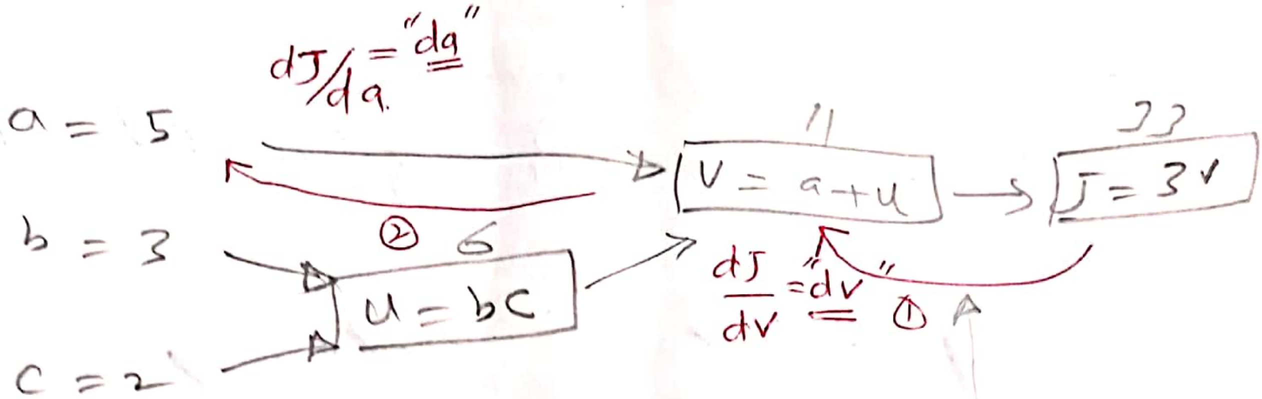
⇒ Forward pass results

* Computational Graph is handy, when we want to optimized the output variable, as J in our case, — in

— In case of logistic regression J is cost function that we want to optimized.

— we see in left to right pass, we compute the value of J (also called forward pass / propagation).

Backward Pass ⇒ to compute derivatives



① $\boxed{\dfrac{dJ}{dv} = ?}$

As $\underline{J = 3v}$

If $v = 11 \Rightarrow J = 33$ ↕ Δy

↕ Δx

If pump up $v = 11.001 \Rightarrow J = 33.003$

Intuit.

⇒ slope $= \dfrac{0.003}{0.001} = 3$ ✓

$\therefore \dfrac{dJ}{dv} = 3$ ✓

Formula

⇒ $J = 3v$

$\dfrac{dJ}{dv} = 3$

just like that $f(a) = 3a$

* as previously calculated.

* This is one step backward for this graph.

② $\boxed{\dfrac{dJ}{da} = ?}$ ⇒ it means if we pump up the value of a, ~~it~~ how much it effect J an $a = 5$ & it compute $v = 11$ & $J = 33$

Pumpup it to $a = 5.001 \Rightarrow v = 11.001 \Rightarrow J = 33.003$

$\therefore \Delta X = 0.001$, $\Delta Y = 0.003$

slope $= 3$

It mean J is change by 3 time the a.

or

$\Rightarrow$ If we change a, it change v & ultimately it change J

Change a effect v

$$\boxed{\frac{dJ}{da} = \frac{dJ}{dv}\frac{dv}{da}}$$

is called chain rule = is product of $\frac{dJ}{dv} \times \frac{dv}{da}$

$a \rightarrow v \rightarrow J$

As

$$\frac{dv}{da} = 1$$

$\Rightarrow V = a + u$

$\Delta X = 0.001$

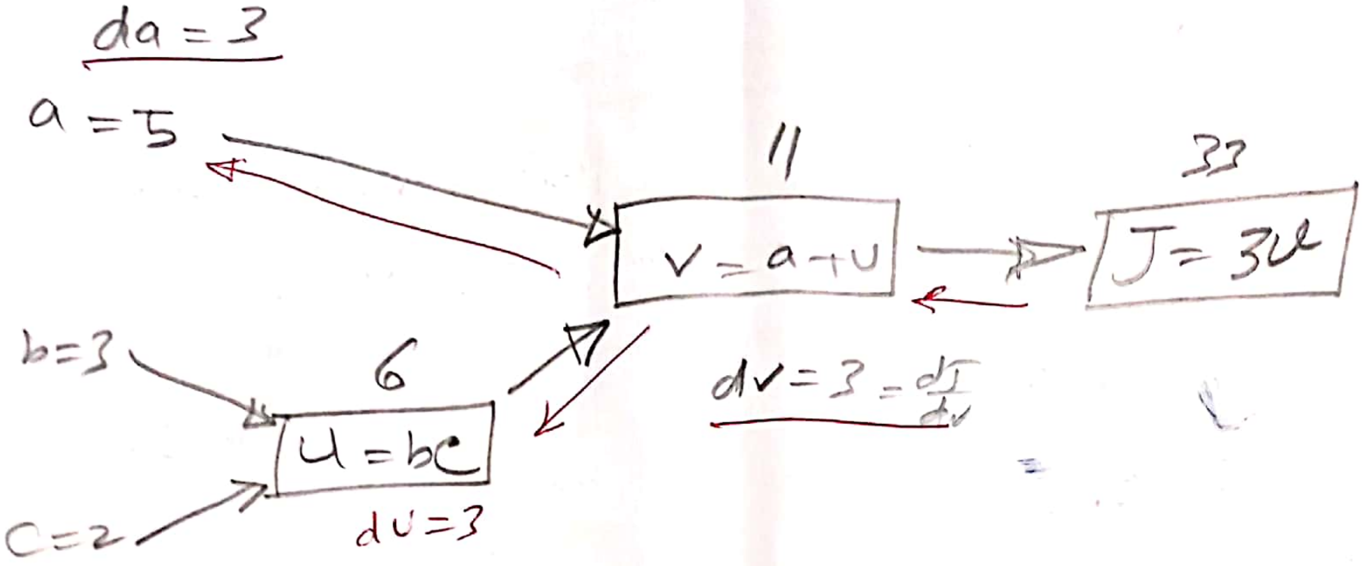
$\Delta Y = 0.001$

$\frac{\Delta Y}{\Delta X} = 1$

$$\therefore \frac{dJ}{da} = \overset{\frac{dJ}{dv} \times \frac{dv}{da}}{3 \times 1} = \underline{\underline{3}}$$

In code we use $\frac{d\text{Final output var}}{d\text{var}} = \underline{\underline{d\text{var}}}$

As we always compute derivative w.r.t J, so we call dvar



$\frac{dJ}{dv} = ?$ Intuition

(vs)

3(a + bc)

u = bc ✓

v = a + u ✓

J = 3v ✓

3(a + u)

as $u = 6 \rightarrow v = 11 \rightarrow J = 33$

$u = 6.001 \rightarrow v = 11.001 \rightarrow J = 33.003$

$\Delta x = 0.001$ $\Rightarrow$ slope = 3

$\Delta Y = 0.003$

Chain rule.

$\Rightarrow$ I can also compute using

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

$u = 6 \rightarrow v = 11$

But $\frac{dv}{du} = 1$ as $v = 6.001 \rightarrow v = 11.001$

$\Delta x = 0.001$, $\Delta Y = 0.001$

slope = 1 = $\frac{dv}{du}$

$$\therefore \frac{dJ}{dv} = \frac{dJ}{dv} \times \frac{dv}{du}$$

$$= 3 \times 1 = 3$$

What J

$$\frac{dJ}{db} = ?$$

If we allow to change b, hurt b a little bit in order to minimized / maximized value of J, what is derivative/slope of the function J, when change value of b a little bit.

— it turns out using the chain rule of calculus

we can write

2) Chain Formula — is a way to efficiently compute gradients

$$\frac{dJ}{db} = \frac{dJ}{du} * \frac{du}{db}$$

1) Intuition

slope $= \frac{\Delta Y}{\Delta X} = 6$

$\Delta X = 0.001$ $\Delta Y = 0.006$

$b = 3 \rightarrow u = 6 \rightarrow v = 11 \rightarrow J = 33$

$b = 3.001 \rightarrow u = 6.002 \rightarrow v = 11.002 \rightarrow J = 33.006$

Intuition

Reason is 1) $b = 3 \rightarrow 3.001$

$\Rightarrow$ first effect on U.

2) $U = bc \Rightarrow 6 \rightarrow 6.002$

$\therefore$ slope $= \frac{dU}{db} = 2$

already figured

$$\frac{dJ}{db} = 3 * 2 = 6$$

$$db = 6$$

(2nd reasoning) $\Rightarrow$ we want to know when U is up 0.002, how that effect on J is goes up by 3 time $\rightarrow$ 0.006

3)

$v = 11 \rightarrow 11.002$

$J = 33 \rightarrow 33.006$

$\Delta Y = 0.006$

$\frac{\Delta Y}{\Delta X} = 6$

3) Formulas

3) $J = 3(a + bc)$

$U = bc$
$V = a + U$
$J = 3V$

$J = 3(a + U)$
$dJ/dU = 3$

$J = 3V \Rightarrow \frac{dJ}{dv} = 3$

$J = 3(a + bc)$

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

$dJ/db = 3c = 6$

$dJ/dc = 3b = 9$

Chain Rule: tell us we can efficiently compute gradients / derivatives

$$\frac{dJ}{dc} = \frac{dJ}{dv} * \frac{du}{dc}$$

$dv * 3$

$$\frac{dJ}{dc} = 3 \times 3 = 9$$

allready Known

$c = 2 \rightarrow 2.001$

$\Delta x = 0.001$

$u = 6 \rightarrow 3 \times 2.001 = 6.003$

$\Delta Y = 0.003$

$slop = \frac{du}{dc} = \frac{0.003}{0.001} = 3$

⇒ It means if we change c a little bit it'll change J by 9 times change.

Intuition

$c = 2 \rightarrow 2.001$

$\Delta x = 0.001$

$u = 6 \rightarrow 6.003$

$\Delta Y = 0.003$

$slop = \frac{du}{dc} = \frac{0.003}{0.001} = 3$

$v = 11 \rightarrow 5 + 6.003 = 11.003$

$\Delta Y = 0.003$

$slop\ \frac{dv}{dc} = \frac{0.003}{0.001} = 3$

$j = 33 \rightarrow 3 \times (11.003) = 33.009$

$\Delta Y = 0.009$

$slop\ \frac{dJ}{dc} = \frac{0.009}{0.001} = 9$

key / Take away:

Efficient way to calculate derivative is the left to right direction of ε

$$\frac{dJ}{dc} = \frac{dJ}{dv} \cdot \frac{du}{dc}$$

① already computed, ② compute it now

so it is efficient way to compute derivative in left right direction or (Backward Pass)

Scanned with CamScanner