N

5-10-2021 Tuesday

week#6 lecture#8

Problems in SoftMax:

If we have,

$$\begin{bmatrix} 225 \\ 617 \\ 950 \end{bmatrix} \rightarrow \begin{bmatrix} e^{S_0} \\ e^{S_1} \\ e^{S_2} \end{bmatrix}$$

big Powers to Calculate computationally expensive.

Solution to this:

$$\frac{C e^{f_i}}{C \sum_j e^{f_j}} \rightarrow \frac{e^{\log C} e^{f_i}}{(Ce^{S_0} + Ce^{S_1} + Ce^{S_2})}$$

$$\rightarrow \frac{e^{\log C} e^{f_i}}{e^{\log C} \sum_j e^{f_j}} \rightarrow \frac{e^{(f_i + \log C)}}{\sum_j e^{(f_j + \log C)}}$$

$\log C = -\max(S)$

So in this case

$\log C = -(\max(225, 617, 950)$

$\log C = -950$ so $f_j + \log C$

$225 - 950 = -725$
$617 - 950 = -333$
$950 - 950 = 0$

e.g

X_1 $\qquad\qquad$ X_2 → better

$$\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} \qquad \begin{bmatrix} 2 \\ 2 \\ 13 \end{bmatrix}$$

→ True class $\qquad$ → True class

if it is X_1 ~~&~~ both classifier are different, so 2nd classifier is better bcz it gives bigger score so Soft Max Prefers bigger score.

SVM Treats same

Soft Max treats more score to True Class

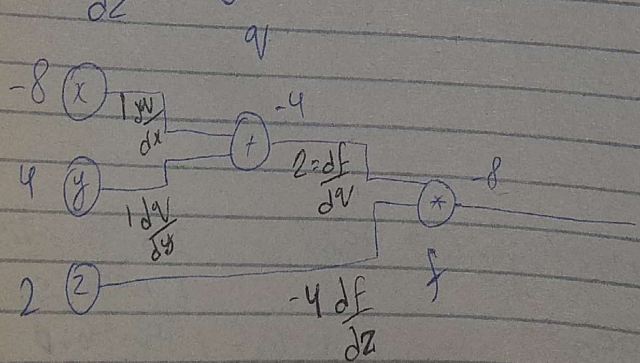
Calculating Gradients:

$$f = (x+y)z$$

$$\frac{df}{dx} = z$$

$$\frac{df}{dy} = z$$

$$\frac{df}{dz} = x+y$$



?

If backward pass, we calculate local gradients.

$f = qz$

$\frac{df}{dq} = z$

$\frac{df}{dz} = q$

$q = x + y$

$\frac{dq}{dx} = 1$

$\frac{dq}{dy} = 1$

Now to calculate $\frac{df}{dx}$ we do ~~for~~ $\frac{df}{dy}$

$$\frac{df}{dx} = \frac{df}{dq} * \frac{dq}{dx}$$

$$= 2 * 1$$

$$\frac{df}{dx} = 2$$

$$\frac{df}{dy} = \frac{df}{dq} * \frac{dq}{dy}$$

$$= 2 * 1$$

$$\frac{df}{dy} = 2$$

Example:

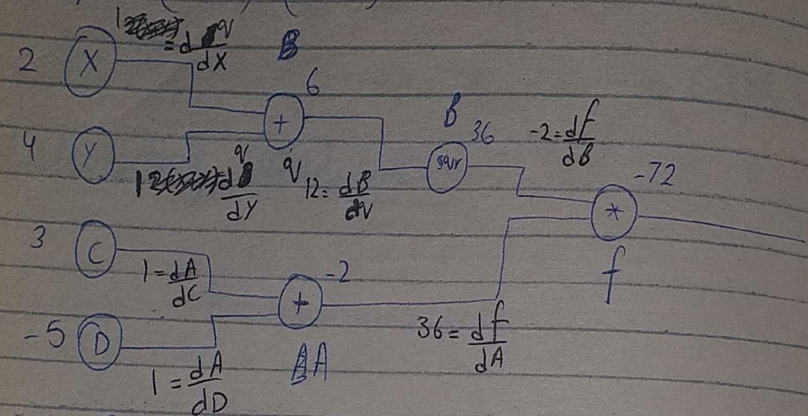
$f = AB$

$A = C + D$

$B = (X+Y)^2$ $\qquad$ $B = q^2$

Sol: $\qquad$ $q = X + Y$

$f = (C+D) \cdot (X+Y)^2$



$C = 3$

$D = -5$

$X = 2$

$Y = 4$

$$\frac{df}{dq} = \frac{df}{dB} \cdot \frac{dB}{dq}$$

$= -2 * 12$

$= -24$

$$\frac{df}{dc} = \frac{df}{dA} \cdot \frac{dA}{dc}$$

$$= 36 * 1$$

$$\boxed{= 36}$$

$$\frac{df}{dD} = \frac{df}{d\cancel{B}A} * \frac{d\cancel{f}A}{dD}$$

$$= 36 * 1$$

$$\boxed{= 36}$$

$$\frac{df}{dx} = \frac{df}{dV} * \frac{dfV}{dx}$$

$$= -24 * 1$$

$$\boxed{= -24}$$

$$\frac{df}{dy} = \frac{df}{dV} * \frac{dfV}{dy}$$

$$= -24 * 1$$

$$\boxed{= -24}$$

Example

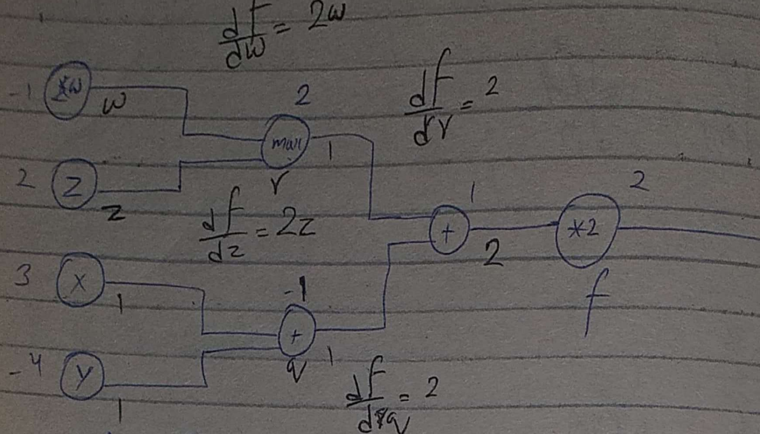
$$f = [x + y + \max(z, w)] * 2$$

$w = -1$

$x = 3$

$y = -4$

$z = 2$



$f = [q + r] * 2$

$t = q + r$

$f = 2t$

$\frac{df}{dx} = 2$

$\frac{df}{dy} = 2$

Zero bcz $2 > -1$

$\frac{df}{dw} = 2w\,[\max(z, w)] = 2 \times 0 = 0$

$\frac{df}{dz} = 2z\,[\max(z, w)] = 2 \times 1 = 2$

C

max (2,1)

1+0.00005

then function output would remain same

so Gradient wouldn't matter.

if both values are same in max, then return 1.

Example.

$$f = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$

Solution.

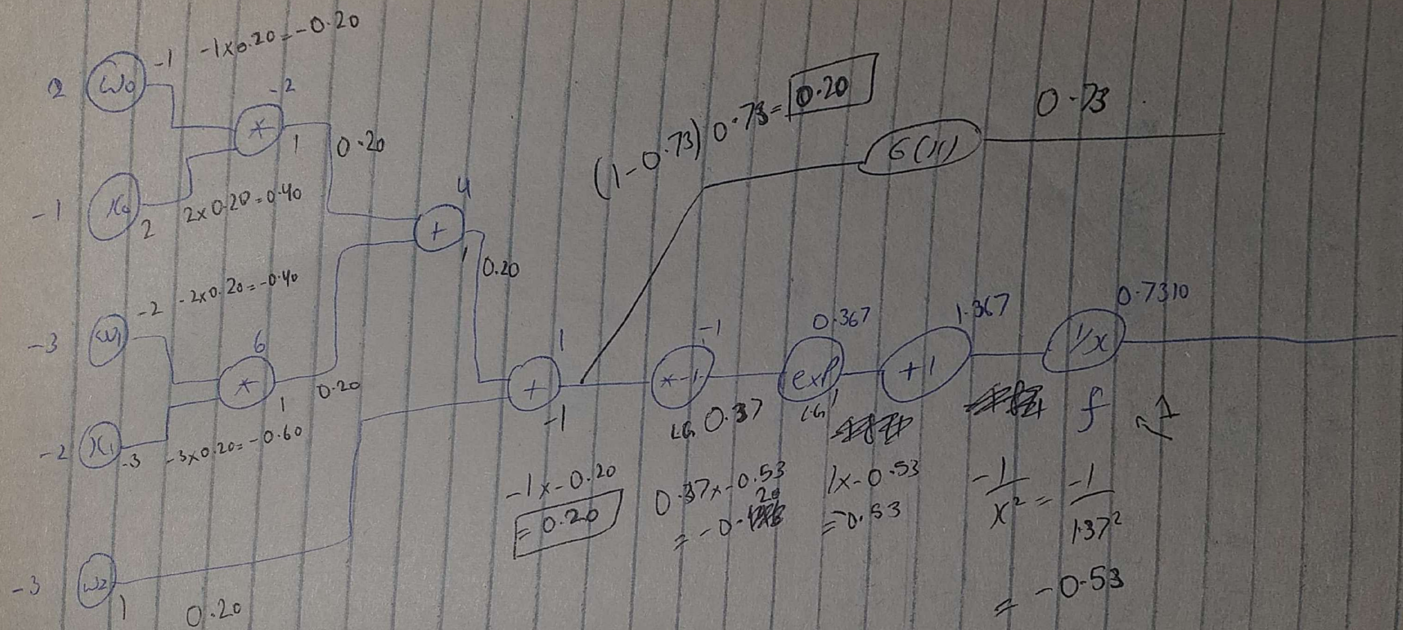
$w_0 = 2$

$x_0 = -1$

$w_1 = -3$

$x_1 = -2$

$w_2 = -3$



$f = \frac{1}{x} \Rightarrow \frac{df}{dx} = \frac{-1}{x^2}$

Directly Take Gradients of B/w the functions.

$f = \dfrac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$ is basically

$\therefore f = \dfrac{1}{1+e^{-x}}$ sigmoid function

$$\frac{d\sigma(x)}{dx} = \frac{-1}{(1+e^{-x})^2} \times e^{-x} \times -1$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

Adding & substracting 1 in numerator

$$= \frac{1+e^{-x}-1}{(1+e^{-x})^2}$$

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

$$= \left(\frac{1+e^{-x}}{(1+e^{-x})} - \frac{1}{(1+e^{-x})}\right) * \frac{1}{(1+e^{-x})}$$

$$\frac{d\sigma(x)}{dx} = (1-\sigma(x))(\sigma(x))$$