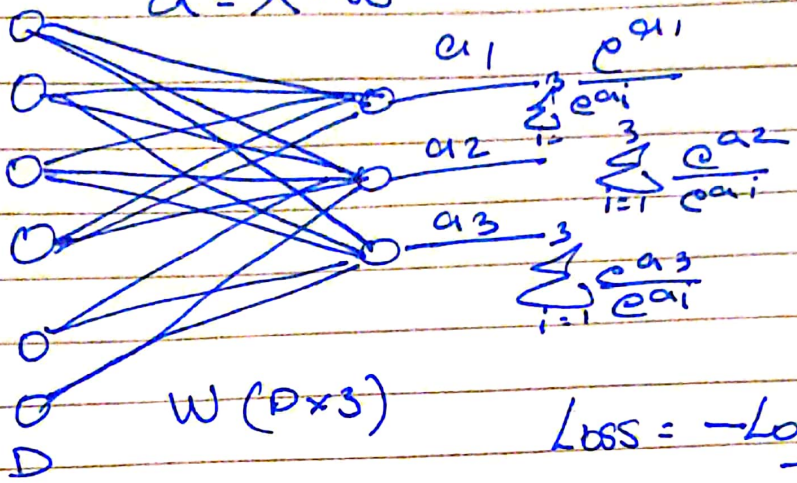
One Image : Assume Class 2

$a = X \cdot W$



$1 \times D$ $\quad W(D \times 3)$

$$Loss = -Log\left(\frac{e^{a_2}}{e^{a_1}+e^{a_2}+e^{a_3}}\right)$$ (Correct class)

Case 1 : When considering a not of the correct class

$$L_{a_3} = -Log\left(\frac{e^{a_2}}{e^{a_1}+e^{a_2}+e^{a_3}}\right)$$

$$\frac{\partial L}{\partial W_{:,1}} = -\frac{1}{\frac{e^{a_2}}{e^{a_1}+e^{a_2}+e^{a_3}}} \times e^{a_2}\left[(-1)\left(e^{a_1}+e^{a_2}+e^{a_3}\right)^{-2}\right] \times e^{a_1} \times X$$

$\frac{e^{a_2}}{e^{a_1}+e^{a_2}+e^{a_3}}$: Constant wrt. (only need to consider a_3)

[Chain Rule]

$$\left[\frac{\partial L}{\partial a_1} \cdot \frac{\partial a_1}{\partial W_{:,1}}\right]$$

$$\frac{\partial L}{\partial W_{:,1}} = \left[\frac{e^{a_1}}{e^{a_1}+e^{a_2}+e^{a_3}}\right] \times X$$

(Incorrect)

Chandra

Case 2: When considering a_2 (correct class)

$$L = -\log\left(\frac{e^{a_2}}{e^{a_1}+e^{a_2}+e^{a_3}}\right)$$

$$\frac{\partial L}{\partial w_{:,2}} = \frac{\partial L}{\partial a_2} \times \frac{\partial a_2}{\partial w_{:,2}}$$

[U/V rule]

$$= -\frac{1}{\frac{e^{a_2}}{e^{a_1}+e^{a_2}+e^{a_3}}} \times \left[\frac{e^{a_2} \cdot X \cdot (e^{a_1}+e^{a_2}+e^{a_3}) - e^{a_2} \cdot (e^{a_2}) \cdot X}{(e^{a_1}+e^{a_2}+e^{a_3})^2}\right]$$

$$\boxed{\frac{\partial L}{\partial w_{correct}} = \left[\frac{e^{a_2}}{e^{a_1}+e^{a_2}+e^{a_3}}\right] X - X}$$