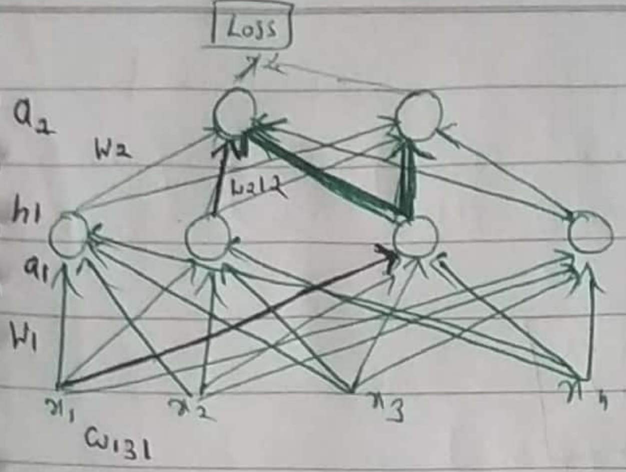
Partial Derivative Computation w.r.t W

Date:



$h_1 = \text{Sigmoid}(a_1)$

W_2 matrix $\rightarrow$ $\cancel{4 \times 2}$ 2×4

W_1 matrix $\rightarrow$ 4×4

$x = [2 \quad 5 \quad 3 \quad 3]$ $\quad y = [1 \quad 0]$

$b = [0 \quad 0]$

$$W_1 = \begin{bmatrix} 0.1 & 0.3 & 0.8 & -0.4 \\ -0.3 & 0.2 & 0.5 & 0.5 \\ -0.3 & 0 & 0.5 & 0.4 \\ 0.2 & 0.5 & -0.9 & 0.7 \end{bmatrix}$$

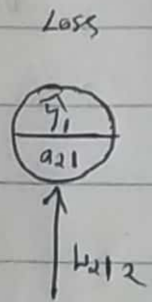
Connected to 2nd neuron in Previous Layer

w_{212} $\rightarrow$ weight of interest, which we want to update.

Layer

Connected to 1st neuron in next Layer.

$$W_2 = \begin{bmatrix} 0.5 & 0.8 & 0.2 & 0.4 \\ 0.5 & 0.2 & 0.3 & -0.5 \end{bmatrix}$$



$$\frac{dL}{dw_{212}} = \frac{dL}{da_{21}} \cdot \frac{da_{21}}{dw_{212}}$$ # w_{212} influences a_{21} $= \cancel{\frac{dy}{d\hat{y}}} \cdot \cancel{\frac{d\hat{y}}{da_{21}}} \cdot \cancel{\frac{da_{21}}{dw_{212}}}$

$$= \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{da_{21}} \cdot \frac{da_{21}}{dw_{212}}$$ # a_{21} influences output and then output influences Loss.

$\hat{y}_1 = \text{Softmax}(a_{21})$

∴ Finally $$\boxed{\frac{dL}{dw_{212}} = \frac{dL}{d\hat{y}_1} \cdot \frac{d\hat{y}_1}{da_{21}} \cdot \frac{da_{21}}{dw_{212}}}$$

Date :

- 1st. $\frac{dL}{d\hat{y}_1}$

$\because L = \sum_{i=1}^{2} (y_i - \hat{y}_i)^2$

$\frac{dL}{d\hat{y}_1} = \frac{d}{d\hat{y}_1}\left[(y_1 - \hat{y}_1)^2 + (y_1 - \hat{y}_2)^2\right]$

$$\boxed{\frac{dL}{d\hat{y}_1} = -2(y_1 - \hat{y}_1)}$$

- 2nd $\frac{d\hat{y}_1}{da_{21}}$

$\because \hat{y}_1 = \text{Softmax}(a_{21})$

$\hat{y}_1 = \frac{e^{a_{21}}}{e^{a_{21}} + e^{a_{22}}} = \frac{1}{\cancel{\quad}^{-a_{22}+a_{21}}} \quad \frac{1}{1+e^{a_{22}-a_{21}}}$

$\because \frac{d\hat{y}_1}{da_{21}} = \frac{d}{da_{21}}\left[\frac{1}{1+e^{a_{22}-a_{21}}}\right]$

$\because$ & It's a sigmoid function

$$\boxed{\frac{d\hat{y}_1}{da_{21}} = \hat{y}_1(1 - \hat{y}_1)}$$

Scanned with CamScanner

Date :

3rd. $\dfrac{da_{21}}{dw_{212}}$

$$a_{21} = w_{211} \cdot h_{11} + w_{212} h_{12} + w_{213} h_{13} + w_{214} h_{14},$$

$$\boxed{\dfrac{da_{21}}{dw_{212}} = \cancel{w_{21}} \cdot h_{12}}$$

Now $\rightarrow$ Output $\Rightarrow$

$$a_1 = W_1 x_n + b_1 = [2.9 \quad 1.4 \quad 2.1 \quad 2.3]$$

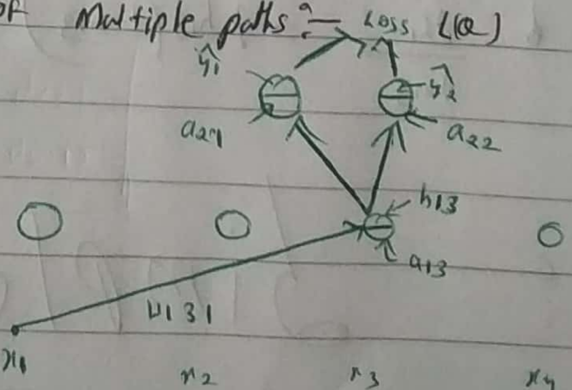
$$h_1 = \text{Sigmoid}(a_1) = [0.95 \quad 0.80 \quad 0.89 \quad 0.91]$$

$$a_2 = W_2 h_1 + b_2 = [1.66 \quad 0.45]$$

$$\hat{y} = \text{Softmax}(a_2) = [0.77 \quad 0.23]$$

$\therefore \dfrac{\partial L}{dw_{212}}$ = Substitute Values calculated in forward pass and get the update values.

Computing Partial derivative in case of Multiple paths :-



$$\frac{dL}{dw_{131}} = \frac{dL}{da_{13}} \cdot \frac{da_{13}}{dw_{131}}$$

$$= \frac{dL}{dh_{13}} \cdot \frac{dh_{13}}{da_{13}} \cdot \frac{da_{13}}{dw_{131}}$$

$$= \left(\frac{dL}{da_{21}} \cdot \frac{da_{21}}{dh_{13}} + \frac{dL}{da_{22}} \cdot \frac{da_{22}}{dh_{13}}\right) \frac{dh_{13}}{da_{13}} \cdot \frac{da_{13}}{dw_{131}}$$

Date :

$$\frac{dL}{dw_{131}} = \left(\frac{dL}{d\hat{y_1}} \cdot \frac{d\hat{y_1}}{da_{21}} \cdot \frac{da_{21}}{dh_{13}} + \frac{dL}{d\hat{y_2}} \cdot \frac{d\hat{y_2}}{da_{22}} \cdot \frac{da_{22}}{dh_{13}}\right) \cdot \left(\frac{dh_{13}}{da_{13}}\right) \cdot \left(\frac{da_{13}}{dw_{131}}\right)$$

$$\frac{dL}{d\hat{y_1}} = -2(y - \hat{y_1}) \qquad \frac{dL}{dy_2} = -2(y - y_2)$$

$$\frac{d\hat{y_1}}{da_{21}} = \hat{y_1}(1 - \hat{y_1}) \qquad \frac{d\hat{y_1}}{da_{22}} = \hat{y_2}(1 - \hat{y_2})$$

$$\frac{da_{21}}{dh_{13}} = ?$$

$$a_{21} = \cancel{w_{2}h_{11}}\ w_{211} h_{11} + w_{212} h_{12} + w_{213} h_{13} + w_{214} h_{14}$$

$$\frac{da_{21}}{dh_{13}} = w_{213} \qquad \frac{da_{22}}{dh_{13}} = w_{223}$$

$$\frac{dh_{13}}{da_{13}} = h_{13}(1 - h_{13}) \quad h_{13} = \text{sigmoid}(a_{13})$$

Simple ! ! !