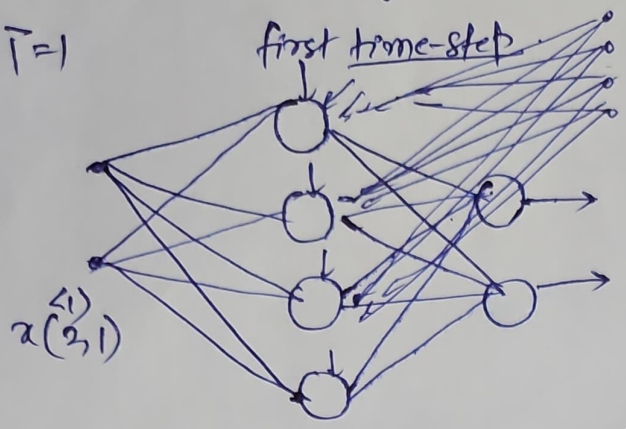
$n_x = 2$, $n_a = 4$, $n_y = 2$, $T = 3$, $m = 1$

$X \Rightarrow$ [illegible]

$T = 1$ first time-step



return $(a^{\langle 1 \rangle}, y^{\langle 1 \rangle}, \text{cache})$

$$a^{\langle 1 \rangle} = \tanh\left(W_{aa} a^{\langle 0 \rangle} + W_{ax} x^{\langle 1 \rangle} + b_a\right)$$

$$y^{\langle 1 \rangle} = g_2\left(W_{ya} a^{\langle 1 \rangle} + b_y\right)$$

↓ Softmax

$y^{\langle 1 \rangle} \Rightarrow (2 \times 1)$

$W_{ax} \Rightarrow (4 \times 2)$

$W_{aa} \Rightarrow (4 \times 4)$

$a_0 \Rightarrow (4 \times 1)$

$b_a \Rightarrow (4 \times 1)$

$W_{ya} \Rightarrow (2 \times 4)$

$b_y \Rightarrow (2 \times 1)$

Cache

W_{ax} $a^{\langle 1 \rangle}$

W_{aa} $a^{\langle 0 \rangle}$

W_{ya} a

b_a $x^{\langle 1 \rangle}$

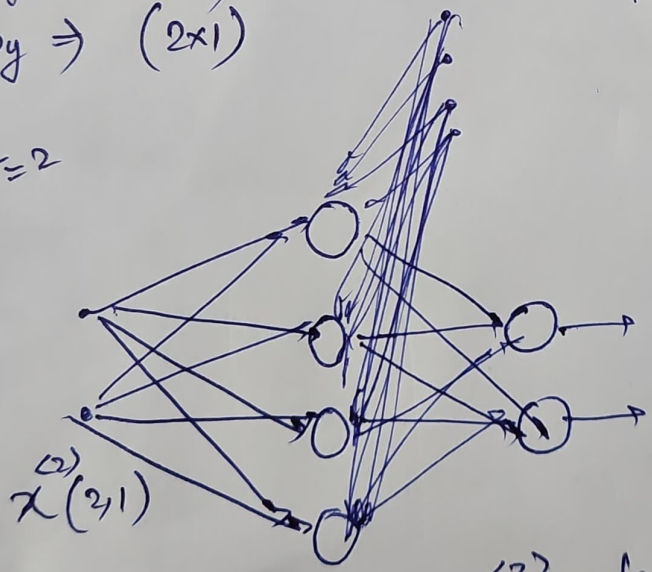
b_y

$a_1^{\langle 1 \rangle}(4 \times 1)$

Softmax $\Rightarrow$ $a_1 = \dfrac{e^{z_1}}{e^{z_1} + e^{z_2}}$

$\Rightarrow$ $a_2 = \dfrac{e^{z_2}}{e^{z_1} + e^{z_2}}$

$T = 2$



$$a^{\langle 2 \rangle} = \tanh\left(W_{aa} a^{\langle 1 \rangle} + W_{ax} x^{\langle 2 \rangle} + b_a\right)$$

$$y^{\langle 2 \rangle} = g_2\left(W_{ya} a^{\langle 2 \rangle} + b_y\right)$$

$W_{ax} \Rightarrow 4 \times 2$

$W_{aa} \Rightarrow 4 \times 4$

$a_0 \Rightarrow 4 \times 1$

$b_a \Rightarrow 4 \times 1$

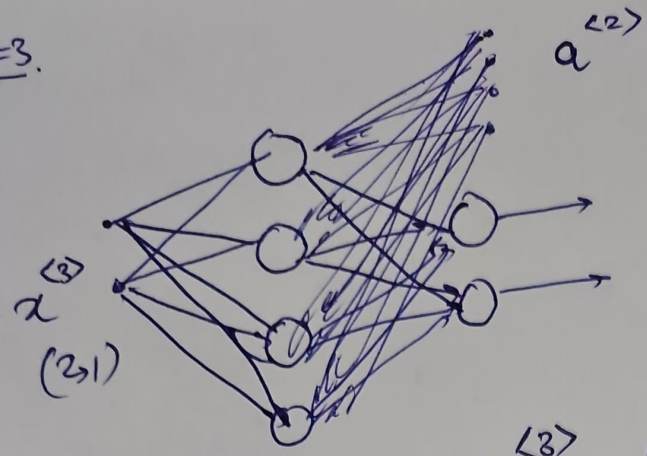
$W_{ya} \Rightarrow (2 \times 4)$

$b_y \Rightarrow 2 \times 1$

$a^{\langle 2 \rangle} \Rightarrow 4 \times 1$

$y^{\langle 2 \rangle} \Rightarrow 2 \times 1$

$T=3$.



$W_{ax} \Rightarrow 4 \times 2$

$W_{aa} \Rightarrow 4 \times 4$

$a_2 \Rightarrow 4 \times 1$

$b_a \Rightarrow 4 \times 1$

$W_{ya} \Rightarrow (2 \times 4)$

$b_y \Rightarrow (2 \times 1)$

$$a^{\langle 3 \rangle} = \tanh\left(W_{aa} a^{\langle 2 \rangle} + W_{ax} x^{\langle 3 \rangle} + b_a\right)$$

$$y^{\langle 3 \rangle} = g_2\left(W_{ya} a^{\langle 3 \rangle} + b_y\right)$$

$a^{\langle 3 \rangle} \Rightarrow 4 \times 1$

$y^{\langle 2 \rangle} \Rightarrow 2 \times 1$

Single ~~step~~ Time step. $x_t, a_{prev}, W_{ax}, W_{aa}, W_{ya}, b_a, b_y$

$$a_{next} = \tanh\left(W_{ax} \cdot x_t + W_{aa} \cdot a_{prev} + b_a\right)$$

$$y_t = \text{softmax}\left(W_{ya} \cdot a_{next} + b_y\right)$$

Cache :- $a_{next}, a_{prev}, x_t, W_{ax}, W_{aa}, W_{ya}, b_a, b_y$

return a_{next}, y_t, Cache

Full time step. $x \Rightarrow (2,1,3)$ $a_0 \Rightarrow (4 \times 1 \times 3)$ zeroes random. $W_{ax}, W_{aa}, W_{ya}, b_a, b_y$

Initi.

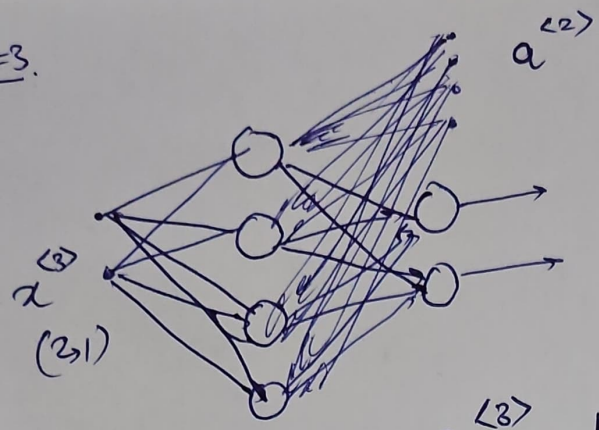
$a \Rightarrow$ zero $(4,1,3)$

$y_{pred} \Rightarrow$ zero $(2,1,3)$

$a_{next} = a_0$

caches[ ]

T=3.



$$a^{\langle 3 \rangle} = \tanh\left(W_{aa} a^{\langle 2 \rangle} + W_{ax} x^{\langle 3 \rangle} + b_a\right)$$

$$y^{\langle 3 \rangle} = g_2\left(W_{ya} a^{\langle 3 \rangle} + b_y\right)$$

$W_{ax} \Rightarrow 4 \times 2$

$W_{aa} \Rightarrow 4 \times 4$

$a_2 \Rightarrow 4 \times 1$

$b_a \Rightarrow 4 \times 1$

$W_{ya} \Rightarrow (2 \times 4)$

$b_y \Rightarrow (2 \times 1)$

$a^{\langle 3 \rangle} \Rightarrow 4 \times 1$

$y^{\langle 2 \rangle} \Rightarrow 2 \times 1$

Single ~~step~~ Time step. $x_t, a_{prev}, W_{ax}, W_{aa}, W_{ya}, b_a, b_y$

$$a_{next} = \tanh\left(W_{ax} x_t + W_{aa} a_{prev} + b_a\right)$$

$$y_t = \text{softmax}\left(W_{ya} \cdot a_{next} + b_y\right)$$

Cache :- $a_{next}, a_{prev}, x_t, W_{ax}, W_{aa}, W_{ya}, b_a, b_y$

return a_{next}, y_t, Cache

Full time step. $x \Rightarrow (2,1,3)$ $a_0 \Rightarrow (4 \times 1 \times 3)$ zeroes or random. $W_{ax}, W_{aa}, W_{ya}, b_a, b_y$

Initi.

$a \Rightarrow$ zero $(4,1,3)$

$y_{pred} \Rightarrow$ zero $(2,1,3)$

$a_{next} = a_0$

Caches [ ]

for every t.

$xt = x[:,:,t]$

a_next, yt, cache → through single time step

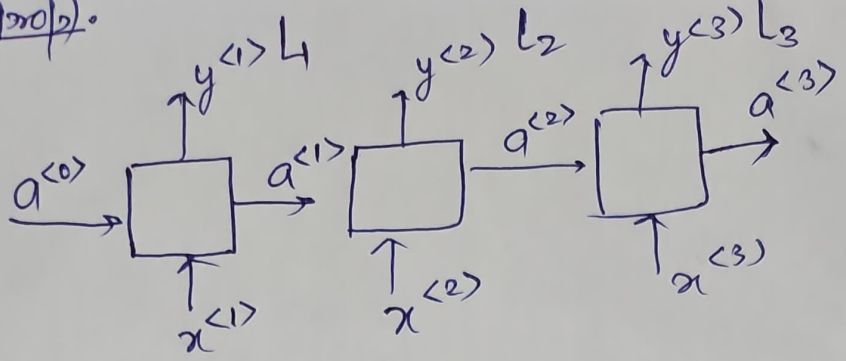
$a[:,:,t] = a\_next$ → To get full value of a

$y\_pred[:,:,t] = yt$ → y_pred of all time st.

caches = $[(a^{\langle 1\rangle}, a^{\langle 0\rangle}, x^{\langle 1\rangle}), (a^{\langle 2\rangle}, a^{\langle 1\rangle}, x^{\langle 2\rangle}), (a^{\langle 3\rangle}, a^{\langle 2\rangle}, x^{\langle 3\rangle})]$
par. par. par.

caches = $([(), (), ()], x)$

(Backprop).



$$a^{\langle 1\rangle} = \tanh\left(W_{aa}a^{\langle 0\rangle} + W_{ax}x^{\langle 1\rangle} + b_a\right)$$
$$a^{\langle 2\rangle} = \tanh\left(W_{aa}a^{\langle 1\rangle} + W_{ax}x^{\langle 2\rangle} + b_a\right)$$
$$a^{\langle 3\rangle} = \tanh\left(W_{aa}a^{\langle 2\rangle} + W_{ax}x^{\langle 3\rangle} + b_a\right)$$

Single backprop.

(da_1, cache $W_{ax}, a^{\langle 0\rangle}, x^{\langle 1\rangle}, W_{ax}, W_{aa}, W_{ya}, b_a, b_y$)

$$\frac{d}{dx}\tanh(x) = 1 - \tanh^2(x)$$

$$\frac{dL_1}{\partial W_{ax}} \Rightarrow \frac{dL_1}{\partial W_{ax}} = \left(1 - \tanh^2(\quad)\right) \cdot x^{\langle 1\rangle T} + \qquad \boxed{da_1 \Rightarrow \frac{dL}{da_1}}$$

$$dW_{ax} = \left(1 - (a^{\langle 1\rangle})^2\right) \cdot x^{\langle 1\rangle T} \quad -(1)$$

$$\partial W_{aa} = \left(1 - (a^{\langle 1\rangle})^2\right) \cdot a^{\langle 0\rangle T} \quad -(2)$$

$$\partial b_a = \text{sum of}\left(da_1 * \left(1 - \tanh^2(\quad)\right)\right) \quad -(3)$$

dictionary

$$dx_t = W_{ax}{}^T \cdot \left(1 - \tanh^2(\quad)\right)$$

$$da_0 = W_{aa}{}^T \cdot \left(da_1 * \left(1 - \tanh^2(\quad)\right)\right)$$