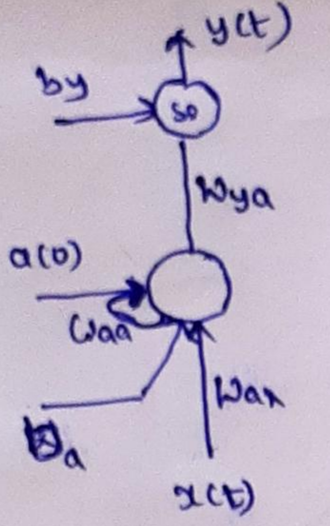
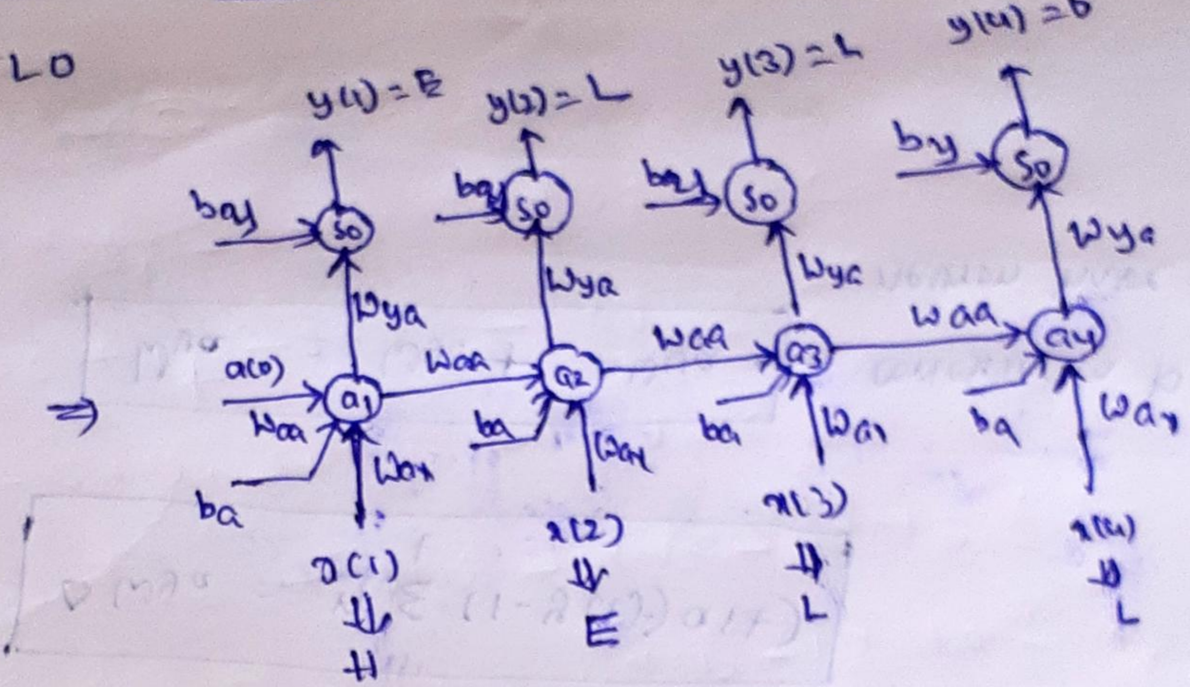
BACK PROPAGATION THROUGH TIME IN RNN

WORD: HELLO



$\Rightarrow$



$$net_y = W_{ya} * a(t) + b_y \quad \text{—— (1)}$$

$$y(t) = \frac{e^{net_y(t)}}{\Sigma e^{net_y(t)}} \quad \text{—— (2)}$$

$$L(t) = -\hat{y}(t) \log(y(t)) \Rightarrow L = -\sum_t \hat{y}(t) \log(y(t))$$

$$= L(1) + L(2) + L(3) + L(4) \quad \text{—— (3)}$$

$$a(t) = \tanh\left[\overbrace{W_{aa}\, a(t-1) + W_{ax}\, a(t) + b_a}^{net_a(t)}\right] \quad \text{—— (4)}$$

Target
↓

$$L = -\hat{y}(1)\log(y(1)) - \hat{y}(2)\log(y(2)) - \hat{y}(3)\log(y(3)) - \hat{y}(4)\log(y(4))$$

↑ predicted.

Change in weight b/w H/L & o/L $\Rightarrow$ $\Delta w_{ya} = -\eta \frac{\partial L}{\partial w_{ya}}$.

$$\frac{\partial L}{\partial w_{ya}} = \frac{\partial L(1)}{\partial w_{ya}} + \frac{\partial L(2)}{\partial w_{ya}} + \frac{\partial L(3)}{\partial w_{ya}} + \frac{\partial L(4)}{\partial w_{ya}} = \sum_t \frac{\partial (L(t))}{\partial w_{ya}}$$

$$\Rightarrow \frac{\partial L(t)}{\partial w_{ya}} = \cancel{\frac{\partial L t}{\partial w_a}} \frac{\partial L(t)}{\partial y(t)} \frac{\partial y(t)}{\partial net_y(t)} \frac{\partial net_y(t)}{\partial w_{ya}}$$

$$= \left(\frac{-\hat{y}(t)}{y(t)}\right) y(t)(1-y(t))\, a(t)$$

$$\frac{\partial L(t)}{\partial w_{ya}} = -\hat{y}(t)(1-y(t))\, a(t) \Rightarrow -(1-y(t))\, a(t)$$

$$\frac{\partial L}{\partial w_{ya}} = \sum_t \frac{\partial L(t)}{\partial w_{ya}} = -\sum_t (1-y(t))\, a(t)$$

$$\boxed{\Delta w_{ya} = \eta \sum_t (1-y(t))\, a(t)}$$

$$\boxed{w_{ya} = w_{ya} + \Delta w_{ya}}$$ $\rightarrow$ corresponding to the neuron have target "1"

change in bias (Δb_y) at output layer $\Rightarrow$ $\Delta b_y = -\eta \frac{\partial L}{\partial b_y}$

$$\frac{\partial L}{\partial b_y} = \frac{\partial L(1)}{\partial b_y} + \frac{\partial L(2)}{\partial b_y} + \frac{\partial L(3)}{\partial b_y} + \frac{\partial L(4)}{\partial b_y} = \sum_t \frac{\partial L(t)}{\partial b_y}$$

$$\frac{\partial L(t)}{\partial b_y} = \frac{\partial L(t)}{\partial y(t)} \frac{\partial y(t)}{\partial net_y(t)} \frac{\partial net_y(t)}{\partial b_y}$$

$$= -\frac{\hat{y}(t)}{y(t)} y(t)(1-y(t)) * (1)$$

$$\frac{\partial L(t)}{\partial b_y} = -\hat{y}(t)(1-y(t))$$

$$\frac{\partial L}{\partial b_y} = \sum_t \frac{\partial L(t)}{\partial b_y} = -\sum_t \hat{y}(t)(1-y(t))$$

$$\Delta b_y = \eta \sum_t \hat{y}(t)(1-y(t))$$

$$\boxed{\Delta b_y = \eta \sum_t (1-y(t))}$$

$$\boxed{b_y = b_y + \Delta b_y}$$

<u>Change in weights for activation</u> $\Rightarrow$ $\Delta W_{aa} = -\eta \dfrac{\partial L}{\partial W_{aa}}$

$$\frac{\partial L}{\partial W_{aa}} = \frac{\partial L(1)}{\partial W_{aa}} + \frac{\partial L(2)}{\partial W_{aa}} + \frac{\partial L(3)}{\partial W_{aa}} + \frac{\partial L(4)}{\partial W_{aa}} \Rightarrow \sum_t \frac{\partial L(t)}{\partial W_{aa}}$$

$$\left.\frac{\partial L(t)}{\partial W_{aa}}\right|_{t=1} \Rightarrow \frac{\partial L(1)}{\partial W_{aa}} = \frac{\partial L(1)}{\partial y(1)} \frac{\partial y(1)}{\partial nety(1)} \left[\frac{\partial nety(1)}{\partial a(1)}\right] \boxed{\frac{\partial nety(1)}{\partial a(1)} \cdot \frac{\partial a(1)}{\partial W_{aa}}} * \frac{\partial a(1)}{\partial W_{aa}}$$

$$\boxed{\frac{\partial L(1)}{\partial W_{aa}} = \frac{\partial L(1)}{\partial y(1)} \frac{\partial y(1)}{\partial nety(1)} \frac{\partial nety(1)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}}}$$

$$\left.\frac{\partial L(t)}{\partial W_{aa}}\right|_{t=2} \Rightarrow \frac{\partial L(2)}{\partial W_{aa}} = \frac{\partial L(2)}{\partial y(2)} \frac{\partial y(2)}{\partial nety(2)} \frac{\partial nety(2)}{\partial a(2)} \left[\frac{\partial a(2)}{\partial W_{aa}} + \frac{\partial a(2)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}}\right]$$

$$\therefore \boxed{\frac{\partial L(2)}{\partial W_{aa}} = \frac{\partial L(2)}{\partial y(2)} \frac{\partial y(2)}{\partial nety(2)} \frac{\partial nety(2)}{\partial a(2)} \left[\frac{\partial a(2)}{\partial W_{aa}} + \frac{\partial a(2)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}}\right]}$$

$$= \frac{\partial L(2)}{\partial y(2)} \frac{\partial y(2)}{\partial nety(2)} \frac{\partial nety(2)}{\partial a(2)} \frac{\partial a(2)}{\partial W_{aa}} + \frac{\partial L(2)}{\partial y(2)} \frac{\partial y(2)}{\partial nety(2)} \frac{\partial nety(2)}{\partial a(2)} \frac{\partial a(2)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}}$$

$$\left.\frac{\partial L(t)}{\partial W_{aa}}\right|_{t=3} \Rightarrow \frac{\partial L(3)}{\partial W_{aa}} = \frac{\partial L(3)}{\partial y(3)} \frac{\partial y(3)}{\partial nety(3)} \frac{\partial nety(3)}{\partial a(3)} \left[\frac{\partial a(3)}{\partial W_{aa}} + \frac{\partial a(3)}{\partial a(2)} \cdot \left\{\frac{\partial a(2)}{\partial W_{aa}}\right\}\right]$$

$$\frac{\partial L(3)}{\partial W_{aa}} = \frac{\partial L(3)}{\partial y(3)} \frac{\partial y(3)}{\partial nety(3)} \frac{\partial nety(3)}{\partial a(3)} \frac{\partial a(3)}{\partial W_{aa}} + \frac{\partial L(3)}{\partial y(3)} \frac{\partial y(3)}{\partial nety(3)} \frac{\partial nety(3)}{\partial a(3)} \frac{\partial a(3)}{\partial a(2)} \left[\frac{\partial a(2)}{\partial W_{aa}} + \frac{\partial a(2)}{\partial a(1)} \cdot \frac{\partial a(1)}{\partial W_{aa}}\right]$$

$$\frac{\partial L(3)}{\partial W_{aa}} = \frac{\partial L(3)}{\partial y(3)} \frac{\partial y(3)}{\partial nety(3)} \frac{\partial nety(3)}{\partial a(3)} \frac{\partial a(3)}{\partial W_{aa}} +$$

$$\frac{\partial L(3)}{\partial y(3)} \frac{\partial y(3)}{\partial nety(3)} \frac{\partial nety(3)}{\partial a(3)} \frac{\partial a(3)}{\partial a(2)} \frac{\partial a(2)}{\partial W_{aa}} +$$

$$\frac{\partial L(3)}{\partial y(3)} \frac{\partial y(3)}{\partial nety(3)} \frac{\partial nety(3)}{\partial a(3)} \frac{\partial a(3)}{\partial a(2)} \frac{\partial a(2)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}}$$

$$\left.\frac{\partial L(t)}{\partial W_{aa}}\right|_{t=4} \Rightarrow \frac{\partial L(4)}{\partial W_{aa}} = \frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \left[\frac{\partial a(4)}{\partial W_{aa}} + \frac{\partial a(4)}{\partial a(3)} \frac{\partial a(3)}{\partial W_{aa}} \right]$$

$$\frac{\partial L(4)}{\partial W_{aa}} = \frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial W_{aa}} +$$

$$\frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial a_3} \left[\frac{\partial a(3)}{\partial W_{aa}} + \frac{\partial a(3)}{\partial a(2)} \frac{\partial a(2)}{\partial W_{aa}} \right]$$

$$\frac{\partial L(4)}{\partial W_{aa}} = \frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial W_{aa}} +$$

$$\frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial a(3)} \frac{\partial a(3)}{\partial W_{aa}} +$$

$$\frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial a(3)} \frac{\partial a(3)}{\partial a(2)} \left[\frac{\partial a(2)}{\partial W_{aa}} + \frac{\partial a(2)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}} \right]$$

$$\frac{\partial L(4)}{\partial W_{aa}} = \frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial W_{aa}} +$$

$$\frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial a(3)} \frac{\partial a(3)}{\partial W_{aa}} +$$

$$\frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial a(3)} \frac{\partial a(3)}{\partial a(2)} \frac{\partial a(2)}{\partial W_{aa}} +$$

$$\frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial a(3)} \frac{\partial a(3)}{\partial a(2)} \frac{\partial a(2)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}}$$

$$\frac{\partial L}{\partial W_{aa}} = \frac{\partial L(1)}{\partial y(1)} \frac{\partial y(1)}{\partial net_y(1)} \frac{\partial net_y(1)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}} + \quad \} \rightarrow t=1$$

$$\left.\begin{aligned} &\frac{\partial L(2)}{\partial y(2)} \frac{\partial y(2)}{\partial net_y(2)} \frac{\partial net_y(2)}{\partial a(2)} \frac{\partial a(2)}{\partial W_{aa}} + \\ &\frac{\partial L(2)}{\partial y(2)} \frac{\partial y(2)}{\partial net_y(2)} \frac{\partial net_y(2)}{\partial a(2)} \frac{\partial a(2)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}} + \end{aligned}\right\} \rightarrow t=3$$

$$\left.\begin{aligned} &\frac{\partial L(3)}{\partial y(3)} \frac{\partial y(3)}{\partial net_y(3)} \frac{\partial net_y(3)}{\partial a(3)} \frac{\partial a(3)}{\partial W_{aa}} + \\ &\frac{\partial L(3)}{\partial y(3)} \frac{\partial y(3)}{\partial net_y(3)} \frac{\partial net_y(3)}{\partial a(3)} \frac{\partial a(3)}{\partial a(2)} \frac{\partial a(2)}{\partial W_{aa}} + \\ &\frac{\partial L(3)}{\partial y(3)} \frac{\partial y(3)}{\partial net_y(3)} \frac{\partial net_y(3)}{\partial a(3)} \frac{\partial a(3)}{\partial a(2)} \frac{\partial a(2)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}} + \end{aligned}\right\} \rightarrow t=3$$

$$\left.\begin{aligned} &\frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial W_{aa}} + \\ &\frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial a(3)} \frac{\partial a(3)}{\partial W_{aa}} + \\ &\frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial a(3)} \frac{\partial a(3)}{\partial a(2)} \frac{\partial a(2)}{\partial W_{aa}} + \\ &\frac{\partial L(4)}{\partial y(4)} \frac{\partial y(4)}{\partial net_y(4)} \frac{\partial net_y(4)}{\partial a(4)} \frac{\partial a(4)}{\partial a(3)} \frac{\partial a(3)}{\partial a(2)} \frac{\partial a(2)}{\partial a(1)} \frac{\partial a(1)}{\partial W_{aa}} \end{aligned}\right\} t=4$$

$$\boxed{\frac{\partial L}{\partial W_{aa}} = \sum_{i=1}^{T} \sum_{k=1}^{i} \frac{\partial L(i)}{\partial y(i)} \frac{\partial y(i)}{\partial net_y(i)} \frac{\partial net_y(i)}{\partial a(i)} \prod_{m=k+1}^{i} \frac{\partial a(m)}{\partial a(m-1)} \frac{\partial a(k)}{\partial W_{aa}}}$$

(*)

From equation (*)

$$\frac{\partial L}{\partial W_{aa}} = \sum_{i=1}^{T} \sum_{k=1}^{i} \frac{\partial L(i)}{\partial y(i)} \frac{\partial y(i)}{\partial net_y(i)} \frac{\partial net_y(i)}{\partial a(i)} \prod_{m=k+1}^{i} \frac{\partial a(m)}{\partial a(m-1)} \frac{\partial a(k)}{\partial W_{aa}}$$

$$\frac{\partial L}{\partial y(i)} = -\frac{\hat{y}(t)}{y(i)}, \quad \frac{\partial y(i)}{\partial net_y(i)} = y(i)(1-y(i)), \quad \frac{\partial net_y(i)}{\partial a(i)} = W_{ya}$$

$$\frac{\partial a(m)}{\partial a(m-1)} = W_{aa}[1-[a(m)]^2], \quad \frac{\partial a(k)}{\partial W_{aa}} = a(k-1)(1-[a(k)]^2)$$

$$\therefore \boxed{\frac{\partial L}{\partial W_{aa}} = -\sum_{i=1}^{T} (1-y(i))\, W_{ya} \sum_{k=1}^{i} \left[\prod_{m=k+1}^{i} W_{aa}[1-[a(m)]^2]\right] a(k-1)(1-[a(k)]^2)}$$

$$\Delta W_{aa} = -\eta \frac{\partial L}{\partial W_{aa}}$$

$$W_{aa} = W_{aa} + \Delta W_{aa}$$

7

Change in bias at activation state (b_a) $\Rightarrow$ $\Delta b_a = -\eta \frac{\partial L}{\partial b_a}$

$$\frac{\partial L}{\partial b_a} = \frac{\partial L(1)}{\partial b_a} + \frac{\partial L(2)}{\partial b_a} + \frac{\partial L(3)}{\partial b_a} + \frac{\partial L(4)}{\partial b_a} \Rightarrow \sum_t \frac{\partial L(t)}{\partial b_a}$$

$$\frac{\partial L(t)}{\partial b_a} = \frac{\partial L(t)}{\partial y(t)} \cdot \frac{\partial y(t)}{\partial nety(t)} \frac{\partial nety(t)}{\partial a(t)} \frac{\partial a(t)}{\partial b_a}$$

$$= \frac{-\hat{y}(t)}{y(t)} \, y(t)(1-y(t)) \, w_{ya} \left(1 - [\tanh(neta(t))]^2\right)$$

$$= -\hat{y}(t)(1-y(t)) \, w_{ya} \left(1 - [\tanh(neta(t))]^2\right)$$

$$\frac{\partial L}{\partial b_a} = \sum_t \frac{\partial L(t)}{\partial b_a} = -\sum_t (1-y(t))\left(1-\tanh^2(neta(t))\right) w_{ya}$$

$$\therefore \boxed{\Delta b_a = \eta \sum_t (1-y(t))\left(1-(a(t))^2\right) w_{ya}}$$

$$\boxed{b_a = b_a + \Delta b_a}$$

Change in weight (w_{ax}): $\Delta w_{ax} = -\eta \frac{\partial L}{\partial w_{ax}}$

$$\frac{\partial L}{\partial w_{ax}} = \frac{\partial L(1)}{\partial w_{ax}} + \frac{\partial L(2)}{\partial w_{ax}} + \frac{\partial L(3)}{\partial w_{ax}} + \frac{\partial L(4)}{\partial w_{ax}} = \sum_t \frac{\partial L(t)}{\partial w_{ax}}$$

$$\frac{\partial L(t)}{\partial w_{ax}} = \frac{\partial L(t)}{\partial y(t)} \frac{\partial y(t)}{\partial nety(t)} \frac{\partial nety(t)}{\partial a(t)} \frac{\partial a(t)}{\partial w_{ax}}$$

$$= \frac{-\hat{y}(t)}{y(t)} \, y(t)(1-y(t)) \, w_{ya} \left(1-(a(t))^2\right) x(t)$$

$$= -\hat{y}(t)(1-y(t))\left(1-(a(t))^2\right) w_{ya} \, x(t)$$

$$\frac{\partial L}{\partial w_{ax}} = \sum_t \frac{\partial L(t)}{\partial w_{ax}}$$

$$= -\sum_t (1-y(t))\left(1-(a(t))^2\right) w_{ya} \, x(t)$$

$$\boxed{\Delta w_{ax} = -\eta \frac{\partial L}{\partial w_{ax}} = \eta \sum_t (1-y(t))\left(1-(a(t))^2\right) w_{ya} \, x(t)}$$

$$\boxed{w_{ax} = w_{ax} + \Delta w_{ax}}$$