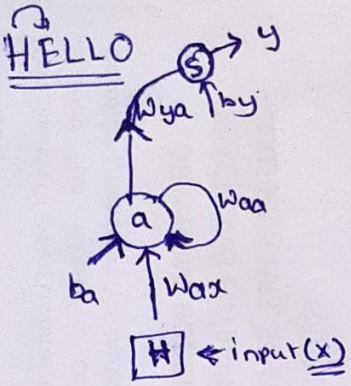
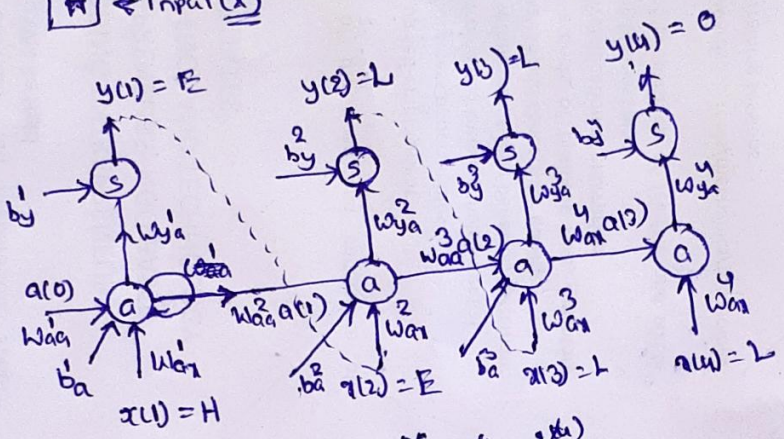
(1)



$$a(t) = \tanh(W_{ax}X + W_{aa}a(t-1) + b_a)$$

$$y(t) = \text{Softmax}(W_{ya}a(t) + b_y)$$

Unroll RNN



Step: (1) One Hot Encoding

	H	E	L	O
H	1	0	0	0
E	0	1	0	0
L	0	0	1	0
O	0	0	0	1

$x = [x(1)\ x(2)\ x(3)\ x(4)]$

$= [H\ E\ L\ L]$

$y = [y(1)\ y(2)\ y(3)\ y(4)]$

$[E\ L\ L\ O]$

(Assume number of Hidden neurons = 3)

Initialize weight matrix

$$W_{ax} = \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ -0.1 & -0.2 & 0.3 & 0.1 \\ 0.1 & 0.2 & -0.3 & -0.4 \end{bmatrix}$$

$$W_{ya} = \begin{bmatrix} 0.1 & 0.2 & 0.3 \\ -0.1 & -0.2 & 0.3 \\ 0.1 & 0.2 & -0.3 \\ -0.1 & -0.2 & 0.3 \end{bmatrix}$$

$$W_{aa} = \begin{bmatrix} 0.1 \\ -0.1 \\ 0.1 \end{bmatrix} \quad b_a = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix} \quad b_y = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix} \quad a(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(2)

(8) →

Step (2): compute activation for $t=1$

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

$$= \tanh\left(\begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ -0.1 & -0.2 & 0.3 & 0.1 \\ 0.1 & 0.2 & -0.3 & -0.4 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + [0] + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix}\right)$$

$$= \tanh\left(\begin{bmatrix} 0.1 \\ -0.1 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix}\right) = \tanh\begin{bmatrix} 0.2 \\ 0.1 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.1 \\ 0.4 \end{bmatrix}$$ → round

Compute $y(1)$

$$y(1) = \text{softmax}\left(W_{ya}\, a(1) + b_y\right)$$

$$= \text{softmax}\left(\begin{bmatrix} 0.1 & 0.2 & 0.3 \\ -0.1 & -0.2 & 0.3 \\ 0.1 & 0.2 & -0.3 \\ -0.1 & -0.2 & 0.3 \end{bmatrix}\begin{bmatrix} 0.2 \\ 0.1 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix}\right)$$

$$= \text{softmax}\begin{bmatrix} 0.26 \\ 0.28 \\ 0.22 \\ 0.48 \end{bmatrix} = \begin{bmatrix} 0.21 \\ 0.23 \\ 0.18 \\ 0.39 \end{bmatrix}$$ ← predicted

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$ ← actual

$$\text{Error} = -\log(0.23) = 1.4697$$

Step (3): Compute to time stamp $t=2$

Element wise multiplication

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

$$= \tanh\left(\begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ -0.1 & -0.2 & 0.3 & 0.1 \\ 0.1 & 0.2 & -0.3 & -0.4 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.1 \\ -0.1 \\ 0.1 \end{bmatrix}\begin{bmatrix} 0.2 \\ 0.1 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix}\right)$$

$$= \tanh\begin{bmatrix} 0.32 \\ -0.01 \\ 0.54 \end{bmatrix} = \begin{bmatrix} 0.31 \\ -0.01 \\ 0.49 \end{bmatrix}$$

(2)

$$y(2) = \text{Softmax}\left(W_{ya}\, a(2) + b_y\right)$$

$$= \text{Softmax}\left(\begin{bmatrix} 0.1 & 0.2 & 0.3 \\ -0.1 & -0.2 & 0.3 \\ 0.1 & 0.2 & -0.3 \\ -0.1 & -0.2 & 0.3 \end{bmatrix}\begin{bmatrix} 0.31 \\ -0.01 \\ 0.49 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix}\right)$$

$$= \text{Softmax}\begin{bmatrix} 0.28 \\ 0.32 \\ 0.18 \\ 0.52 \end{bmatrix} = \begin{bmatrix} 0.21 \\ 0.25 \\ (0.14) \\ 0.40 \end{bmatrix} \leftarrow \text{predicted} \qquad \begin{bmatrix} 0 \\ 0 \\ (1) \\ 0 \end{bmatrix} \leftarrow \text{Actual}$$

$$E = -\log(0.23) - \log(0.44) = 3.4358$$

Step (4) Compute for time stamp (3) $(t=3)$ $a(3) = 2$, $y(3) = 2$

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

$$= \tanh\left(\begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ -0.1 & -0.2 & 0.3 & 0.1 \\ 0.1 & 0.2 & -0.3 & -0.4 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.1 \\ -0.1 \\ 0.1 \end{bmatrix}\begin{bmatrix} 0.31 \\ -0.01 \\ 0.49 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix}\right)$$

$$= \tanh\begin{bmatrix} 0.43 \\ 0.50 \\ 0.05 \end{bmatrix} = \begin{bmatrix} 0.41 \\ 0.46 \\ 0.05 \end{bmatrix}$$

$$y(3) = \text{Softmax}\left(\begin{bmatrix} 0.1 & 0.2 & 0.3 \\ -0.1 & -0.2 & 0.3 \\ 0.1 & 0.2 & -0.3 \\ -0.1 & -0.2 & 0.3 \end{bmatrix}\begin{bmatrix} 0.41 \\ 0.46 \\ 0.05 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix}\right)$$

$$= \text{Softmax}\begin{bmatrix} 0.25 \\ 0.08 \\ 0.42 \\ 0.28 \end{bmatrix} = \begin{bmatrix} 0.24 \\ 0.08 \\ (0.41) \\ 0.27 \end{bmatrix} \leftarrow \text{predicted} \qquad \begin{bmatrix} 0 \\ 0 \\ (1) \\ 0 \end{bmatrix} \leftarrow \text{Actual}$$

$$E_2 = -\log(0.23) - \log(0.14) - \log(0.41) = 4.3274$$

Step (5) Compute for time stamp (4) $(t=4)$ $a(4) = 2$, $y(4) = 0$

$$a(4) = \tanh\left(W_{ax}\, x(4) + W_{aa}\, a(3) + b_a\right)$$

$$= \tanh\left(\begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ -0.1 & -0.2 & 0.3 & 0.1 \\ 0.1 & 0.2 & -0.3 & -0.4 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.1 \\ -0.1 \\ 0.1 \end{bmatrix}\begin{bmatrix} 0.41 \\ 0.46 \\ 0.05 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix}\right) = \begin{bmatrix} 0.41 \\ 0.42 \\ 0.005 \end{bmatrix}$$

$$y(4) = \text{Softmax}\left(W_{ya}\, a(4) + b_y\right)$$

$$= \text{Softmax}\left(\begin{bmatrix} 0.1 & 0.2 & 0.3 \\ -0.1 & -0.2 & 0.3 \\ 0.1 & 0.2 & -0.3 \\ -0.1 & -0.2 & 0.3 \end{bmatrix}\begin{bmatrix} 0.41 \\ 0.42 \\ 0.00 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix}\right) = \begin{bmatrix} 0.22 \\ 0.07 \\ 0.42 \\ 0.27 \end{bmatrix} \leftarrow \text{predicted} \qquad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \leftarrow \text{actual}$$

$$E = -\log(0.23) - \log(0.14) - \log(0.41) - \log(0.27) = 5.6367$$

Forward pass completed, Now require Backward pass Through Time (BPTT)