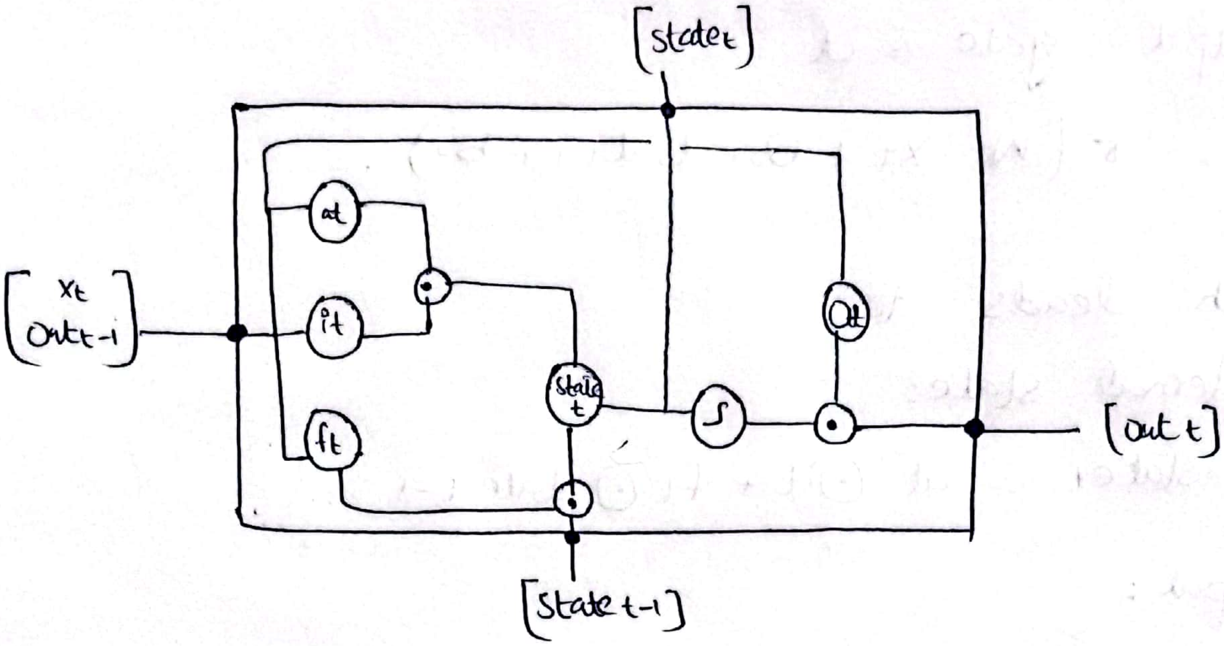
⇒ FORWARD PROPAGATION:-



t = timestamp

t-1 = previous timestamp (timestamp - 1).

Syntatic Notes:

→ Above ⊙ is the element wise product

→ Inner product will be represented as •

→ Outer product will be represented as ⊗

→ σ represents the sigmoid function.

Forward Components:

Input activation = at

$$a_t = \tanh(W_a \cdot x_t + U_a \cdot Out_{t-1} + b_a).$$

Input gate = it

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot Out_{t-1} + b_i)$$

Forget gate = f_t

$$f_t = \sigma(W_f \cdot X_t + U_f \cdot Out_{t-1} + b_f)$$

Output gate = o_t

$$o_t = \sigma(W_o \cdot X_t + U_o \cdot Out_{t-1} + b_o).$$

Which leads to:

Internal state:

$$state_t = a_t \odot i_t + f_t \odot state_{t-1}$$

Output:

$$Out_t = \tanh(state_t) \odot o_t.$$

Note for simplicity we define,

$$gate_t = \begin{bmatrix} a_t \\ i_t \\ f_t \\ o_t \end{bmatrix}, \quad W = \begin{bmatrix} W_a \\ W_i \\ W_f \\ W_o \end{bmatrix}, \quad U = \begin{bmatrix} U_a \\ U_i \\ U_f \\ U_o \end{bmatrix}, \quad b = \begin{bmatrix} b_a \\ b_i \\ b_f \\ b_o \end{bmatrix}.$$

W = It is the weight matrix associated with the hidden state.

U = weight associated with the input.

Example:

$$W_a = \begin{bmatrix} 0.45 \\ 0.25 \end{bmatrix}, \quad U_a = \begin{bmatrix} 0.15 \end{bmatrix}, \quad b_a = \begin{bmatrix} 0.2 \end{bmatrix}$$

$$W_i = \begin{bmatrix} 0.95 \\ 0.8 \end{bmatrix}, \quad U_i = \begin{bmatrix} 0.8 \end{bmatrix}, \quad b_i = \begin{bmatrix} 0.65 \end{bmatrix}$$

$$W_f = \begin{bmatrix} 0.7 \\ 0.45 \end{bmatrix}, \quad U_f = \begin{bmatrix} 0.1 \end{bmatrix}, \quad b_f = \begin{bmatrix} 0.15 \end{bmatrix}$$

$$W_o = \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}, \quad U_o = \begin{bmatrix} 0.25 \end{bmatrix}, \quad b_o = \begin{bmatrix} 0.1 \end{bmatrix}$$

Now Input data:

$X_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ with label : 0.5

$X_1 = \begin{bmatrix} 0.5 \\ 3 \end{bmatrix}$ with label : 1.25

Sequence length = 2

Forward @ t = 0

→ Input activation:

$$a_t = \tanh(W_a \cdot X_t + U_a \cdot Out_{t-1} + b_a)$$

$$a_t = \tanh\left(\begin{bmatrix} 0.45 & 0.25 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0.15 \end{bmatrix}\begin{bmatrix} 0 \end{bmatrix} + \begin{bmatrix} 0.2 \end{bmatrix}\right)$$

$$a_t = \tanh\begin{bmatrix} 0.45 + 0.5 \end{bmatrix} + 0 + \begin{bmatrix} 0.2 \end{bmatrix}$$

$$a_t = \tanh(1.15)$$

$$\tanh x = \frac{\sinh x}{\cosh x}$$

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\tanh x = \frac{2.718^{1.15} - 2.718^{-1.15}}{2.718^{1.15} + 2.718^{-1.15}}$$

$$\tanh x = 0.8187$$

$$at = 0.8187$$

→ **Input gate:**

$$it = \sigma(Wi \cdot xt + Ui \cdot Out\ t-1 + bi)$$

$$it = \sigma\left(\begin{bmatrix} 0.95 & 0.8 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0.8 \end{bmatrix}\begin{bmatrix} 0 \end{bmatrix} + \begin{bmatrix} 0.65 \end{bmatrix}\right)$$

$$it = \sigma\,[0.95 + 1.6] + 0 + [0.65]$$

$$it = \sigma(3.2)$$

$$\sigma = \frac{1}{1 + e^{-x}}$$

$$\sigma = \frac{1}{1 + 2.718^{-3.2}}$$

$$\sigma = \frac{1}{1 + 0.04} \Rightarrow \frac{1}{1.04}$$

$$\sigma = 0.96$$

$$it = 0.96.$$

→ **Forget gate:**

$f_t = \sigma(W_f \cdot x_t + U_f \cdot Out_{t-1} + b_f)$

$f_t = \sigma\left(\begin{bmatrix}0.7 & 0.45\end{bmatrix} \cdot \begin{bmatrix}1 \\ 2\end{bmatrix} + \begin{bmatrix}0.1\end{bmatrix}\begin{bmatrix}0\end{bmatrix} + \begin{bmatrix}0.15\end{bmatrix}\right)$.

$f_t = \sigma\,[0.7 + 0.9] + 0 + [0.15]$

$f_t = \sigma(1.\overset{x}{7}5)$

$\sigma = \dfrac{1}{1 + e^{-x}}$

$\sigma = \dfrac{1}{1 + 2.718^{-1.75}}$

$\sigma = \dfrac{1}{1 + 0.173}$

$\sigma = \dfrac{1}{1.175}$

$\sigma = 0.85$

$f_t = 0.85$.

→ **Output gate:**

$o_t = \sigma(W_o \cdot x_t + U_o \cdot Out_{t-1} + b_o)$.

$o_t = \sigma\left(\begin{bmatrix}0.6 & 0.4\end{bmatrix}\begin{bmatrix}1 \\ 2\end{bmatrix} + \begin{bmatrix}0.25\end{bmatrix}\begin{bmatrix}0\end{bmatrix} + \begin{bmatrix}0.1\end{bmatrix}\right)$

$o_t = \sigma\,[0.6 + 0.8] + 0 + [0.1]$

$o_t = \sigma(1.\overset{x}{5})$

$$\sigma = \frac{1}{1+e^{-x}}$$

$$\sigma = \frac{1}{1+2.718^{-1.5}}$$

$$\sigma = \frac{1}{1+0.223}$$

$$\sigma = \frac{1}{1.223}$$

$$\sigma = 0.817$$

$$o_t = 0.817$$

→ **Internal state:**

$$state_t = a_t \odot i_t + f_t \odot state_{t-1}$$

$$state_t = 0.8187 \times 0.96 + 0.85 \times 0$$

$$state_t = 0.78$$

→ **Output:**

$$Out_t = \tanh(state_t) \odot o_t.$$

$$Out_t = \tanh(0.78) \times 0.817$$

$$\tanh(0.78)$$

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Scanned with CamScanner

$$\tanh x = \frac{2.718^{0.78} - 2.718^{-0.718}}{2.718^{0.78} + 2.718^{-0.718}}$$

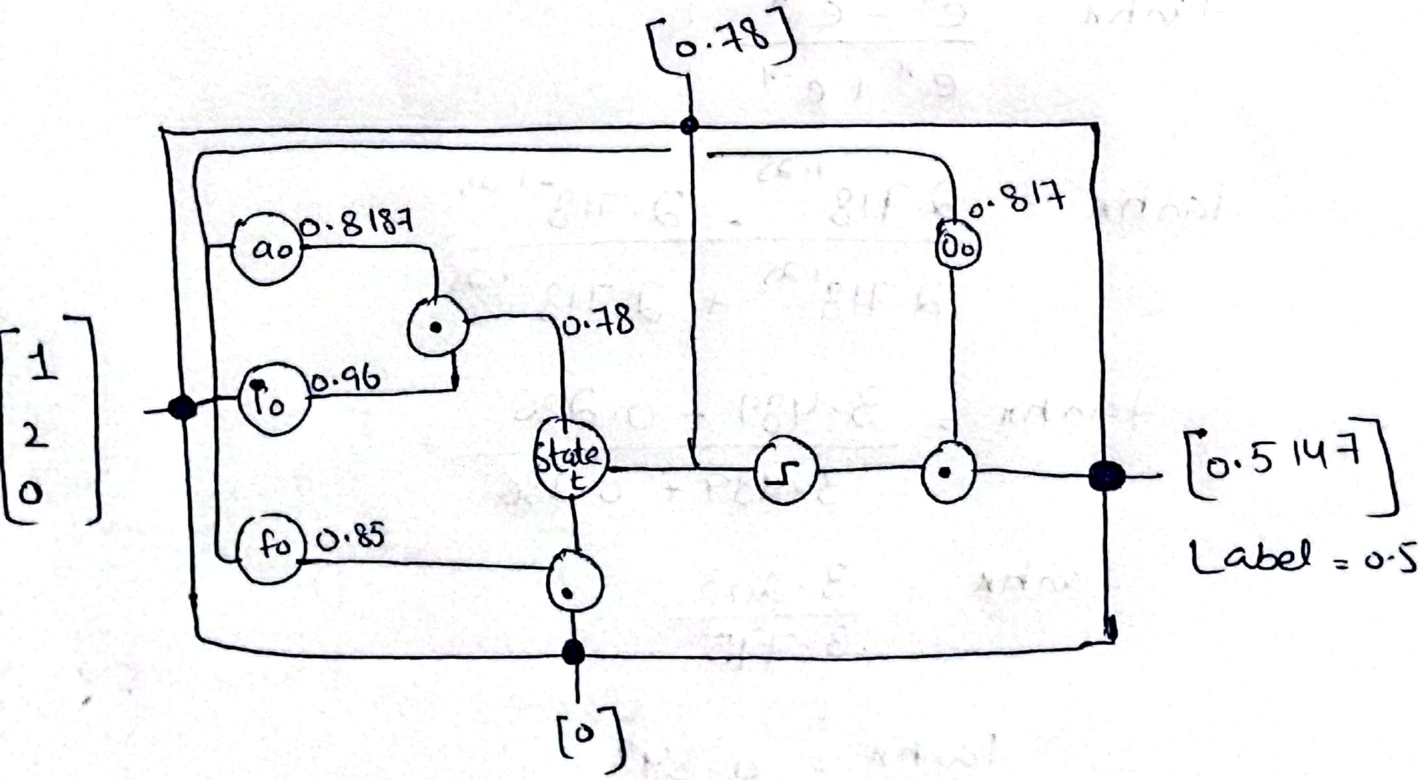
$$\tanh x = \frac{2.181 - 0.487}{2.181 + 0.487}$$

$$\tanh x = \frac{1.694}{2.668}$$

$$\tanh x = 0.63$$

$$\text{Out } t = 0.63 \times 0.817$$

$$\text{Out } t = 0.5147$$



Forward @ $t = 1$

→ Input activation:

$$at = \tanh(Wa \cdot xt + Ua \cdot Out_{t-1} + ba)$$

$$at = \tanh\left(\begin{bmatrix} 0.45 & 0.25 \end{bmatrix}\begin{bmatrix} 0.5 \\ 3 \end{bmatrix} + \begin{bmatrix} 0.15 \end{bmatrix}\begin{bmatrix} 0.5147 \end{bmatrix} + \begin{bmatrix} 0.2 \end{bmatrix}\right)$$

$$at = \tanh\left(\begin{bmatrix} 0.225 + 0.75 \end{bmatrix} + \begin{bmatrix} 0.0772 \end{bmatrix} + \begin{bmatrix} 0.2 \end{bmatrix}\right)$$

$$at = \tanh(1.25)^{x}$$

$$\tanh x = \frac{\sinh x}{\cosh x}$$

$$\tanh x = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

$$\tanh x = \frac{2.718^{1.25} - 2.718^{-1.25}}{2.718^{1.25} + 2.718^{-1.25}}$$

$$\tanh x = \frac{3.489 - 0.286}{3.489 + 0.286}$$

$$\tanh x = \frac{3.203}{3.775}$$

$$\tanh x = 0.848$$

$$at = 0.848$$

→ Input gate:-

$$it = \sigma(Wi \cdot xt + Ui \cdot Out\ t-1 + bi)$$

$$it = \sigma\left(\begin{bmatrix}0.95 & 0.8\end{bmatrix}\begin{bmatrix}0.5 \\ 3\end{bmatrix} + \begin{bmatrix}0.8\end{bmatrix}\begin{bmatrix}0.5147\end{bmatrix} + \begin{bmatrix}0.65\end{bmatrix}\right)$$

$$it = \sigma\left(\begin{bmatrix}0.475 + 2.4\end{bmatrix} + \begin{bmatrix}0.41176\end{bmatrix} + \begin{bmatrix}0.65\end{bmatrix}\right)$$

$$it = \sigma(\overset{x}{3.936})$$

$$\sigma = \frac{1}{1+e^{-x}}$$

$$\sigma = \frac{1}{1 + 2.718^{-3.936}}$$

$$\sigma = \frac{1}{1+0.019}$$

$$\sigma = \frac{1}{1.019}$$

$$\sigma = 0.98$$

$$it = 0.98$$

→ Forget gate:

$$ft = \sigma(Wf \cdot xt + Uf \cdot Out\ t-1 + bf)$$

$$ft = \sigma\left(\begin{bmatrix}0.7 & 0.45\end{bmatrix}\begin{bmatrix}0.5 \\ 3\end{bmatrix} + \begin{bmatrix}0.1\end{bmatrix}\begin{bmatrix}0.5147\end{bmatrix} + \begin{bmatrix}0.15\end{bmatrix}\right)$$

$$ft = \sigma\left(\begin{bmatrix}0.35 + 1.35\end{bmatrix} + \begin{bmatrix}0.05147\end{bmatrix} + \begin{bmatrix}0.15\end{bmatrix}\right)$$

$$ft = \sigma(\overset{x}{1.9})$$

$$\sigma = \frac{1}{1+e^{-x}}$$

$$\sigma = \frac{1}{1+2.718^{-1.9}}$$

$$\sigma = \frac{1}{1+0.149}$$

$$\sigma = \frac{1}{1.149}$$

$$\sigma = 0.87$$

$$ft = 0.87$$

→ Output gate:

$$Ot = \sigma(Wo \cdot xt + Uo \cdot Out\ t-1 + bo)$$

$$Ot = \sigma\left(\begin{bmatrix}0.6 & 0.4\end{bmatrix}\begin{bmatrix}0.5 \\ 3\end{bmatrix} + \begin{bmatrix}0.25\end{bmatrix}\begin{bmatrix}0.5147\end{bmatrix} + \begin{bmatrix}0.1\end{bmatrix}\right).$$

$$ot = \sigma\left(\begin{bmatrix}0.3+1.2\end{bmatrix} + \begin{bmatrix}0.128\end{bmatrix} + \begin{bmatrix}0.1\end{bmatrix}\right)$$

$$ot = \sigma(1.\overset{x}{7}28)$$

$$\sigma = \frac{1}{1+e^{-x}}$$

$$\sigma = \frac{1}{1+2.718^{-1.728}}$$

Scanned with CamScanner

$$\sigma = \frac{1}{1 + 0.177}$$

$$\sigma = \frac{1}{1.177}$$

$$\sigma = 0.849$$

$$o_t = 0.849$$

→ **Internal State:**

$$\text{State}_t = a_t \odot i_t + f_t \odot \text{state}_{t-1}$$

$$\text{state}_t = 0.848 \times 0.98 + 0.87 \times 0.78$$ ← previous internal state

$$\text{state}_t = 1.509$$

→ **Output:**

$$\text{Out}_t = \tanh(\text{State}_t) \odot o_t$$

$$\text{Out}_t = \tanh(1.509) \times 0.849$$

$$\tanh(1.509)$$

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\tanh x = \frac{2.718^{1.509} - 2.718^{-1.509}}{2.718^{1.509} + 2.718^{-1.509}}$$

$$\tanh x = \frac{4.52 - 0.22}{4.52 + 0.22} \Rightarrow \frac{4.3}{4.74}$$

$$\tanh u = 0.907$$

$$Out\ t = 0.907 \times 0.849$$

$$Out\ t = 0.77$$