

## LSTM Architecture:

→ 3 gates + cell state.

→ \*Forget gate, \*Input gate, \*output gate

\*Intermediate cell state \*cell state

Each gate has got their weights.

### Forget Gate:

→ rep as  $f_t$

→ computed through sigmoid

(weight of forget gate \* old state + weight \* forget gate + input)

$$\rightarrow f_t = \sigma(w_f * s_{t-1} + w_{f,x} * x_t)$$

### Why Forget Gate?

→ The 1st sigmoid activation fun in the n/w is forget gate

→ As the name says, this gate shall decide, which info has to be retained or dropped.

→ The info from pre hidden state & current inp gets through the sigmoid fun and the o/p arrives. It's b/w 0 to 1.

→ So, closer the value to 0, it is to be <sup>(dropped)</sup> forgotten, closer it to one, it is to be remembered.

### Input Gate:

→ This is and sigmoid fun & 1st tanh activation function.

→ This decides which information should be saved to the cell state & which should be dropped.



→ rep as  $i_t$

$$\rightarrow i_t = \sigma(w_i * s_{t-1} + w_i * x_t)$$

Output gate:

It highlights which info should be goint to the next hidden state

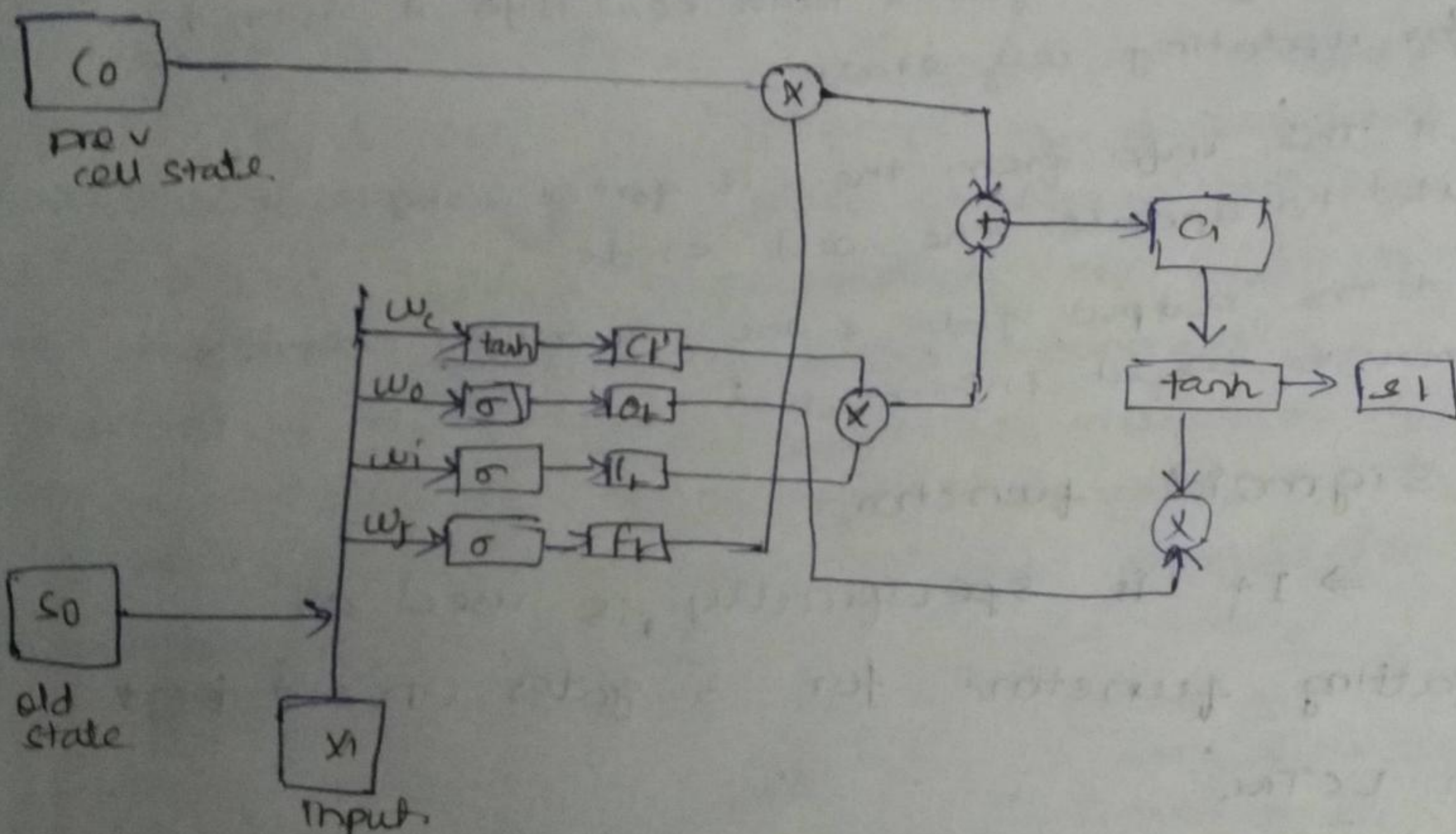
$$\rightarrow O_t = \sigma(w_o * s_{t-1} + w_o * x_t)$$

Intermediate cell state:-

$$C_t' = \tanh(w_c * s_{t-1} + w_c * x_t)$$

$$\text{cell state} \Rightarrow C_t = (i_t * C_t') + (F_t * C_{t-1})$$

$$\text{New state} = O_t * \tanh(C_t)$$



→ LSTM are unsupervised learning methods but technically they are trained using a supervised learning method called self-monitoring.

Why do LSTM use Tanh?

Tanh fun adjusts the output of neural n/w so tat the value stays b/w -1 and 1. You can see how the same value from above remains b/w the boundaries allowed by the tanh fun.



How are LSTM trained?

To train an LSTM neural net to generate text, you need to do following:

i) Processing the text data, so that it can be used on the network.

In this case, the neural net takes the vector as input, so we need a way to convert the text to a vector.

LSTM Cycle:

→ Divided into 4 steps:

\* Using forget gate, info to be forgotten is identified from a prior time step.

\* Using input gate & tanh, new info is sought for updating cell state.

\* The info from the 2 gates above is used to update the cell state.

\* The output gate & the squashing operation provide useful information.