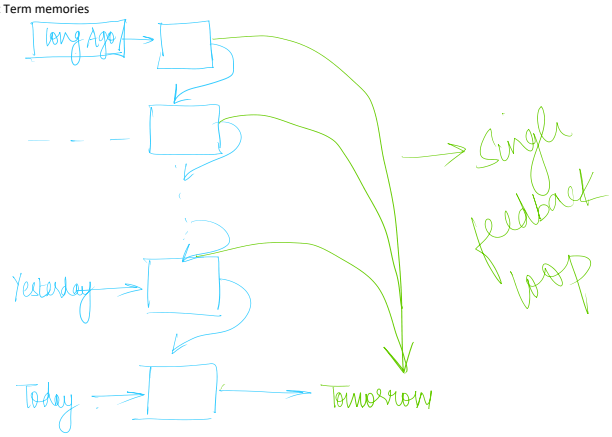
LSTM - Long Short-Term Memory

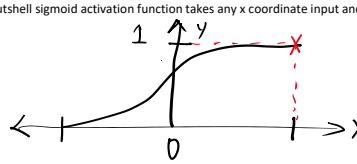
21 May 2023 21:13

- The main idea behind the Long Short-Term Memory Network works is that it uses the same feedback loop connection for events that happened long ago and events that happened just yesterday to **make a prediction about TOMORROW**
- LSTM uses 2 separate paths for making predictions about tomorrow.
  - One path is for Long Term and
  - One for Short Term memories



- There is one problem; compared to the basic NN, RNN, which unrolls from a very simple basic unit, LSTM is based on a much more complicated unit.
- NOTE: unlike the networks we used previously, LSTM uses SIGMOID ACTIVATION FUNCTION & TANH AF.

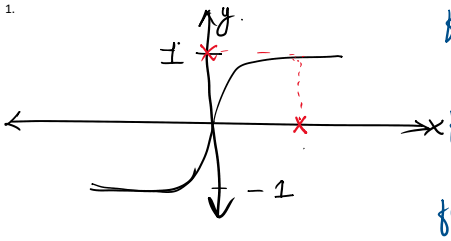
SIGMOID ACTIVATION FUNCTION:
In a nutshell sigmoid activation function takes any x coordinate input and turns it into a y axis coordinate between 0 and 1.



$f(x) = \frac{e^x}{e^x+1}$

$f(10) = \frac{e^{10}}{e^{10}+1} = 0.9995$, $f(-5) = \frac{e^{-5}}{e^{-5}+1} = 0.01$

TANH(HYPERBOLIC TANGENT) ACTIVATION FUNCTION:
In a nutshell sigmoid activation function takes any x coordinate input and turns it into a y axis coordinate between -1 and 1.



$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

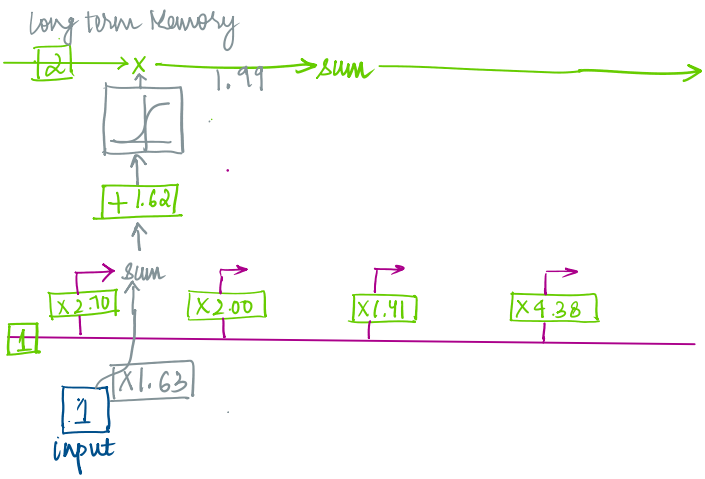
$f(2) = \frac{e^2 - e^{-2}}{e^2 + e^{-2}} = 0.96$

$f(-5) = -1$

HOW THE LSTM WORKS?
This green line that runs all along the top of the unit is called as **Cell State** and represents **LONG TERM MEMORY.**

- **The lack of weights in the Cell State or Long memory, leads to allow flowing of the unrolling units w/o causing Vanishing/Exploding gradients.**

FIRST STAGE OF LSTM UNIT

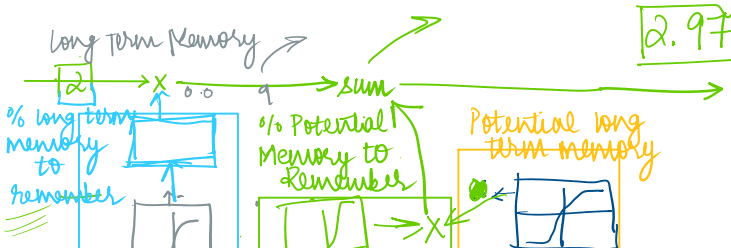


We start w LSTM, 1

$(1 \times 2.70) + (1 \times 1.63) + 1.62 = 5.95$

$f(x) = \frac{e^x}{e^x+1} = 0.997$

To summarize the first stage of LSTM determines that what amount of long term memory will be used or basically remembered.
-> **FORGET GATE:** Even though this part of LSTM determines the % of Long term memory to be remembered.
SECOND STAGE OF LSTM UNIT



2.97 New Long Term Memory,

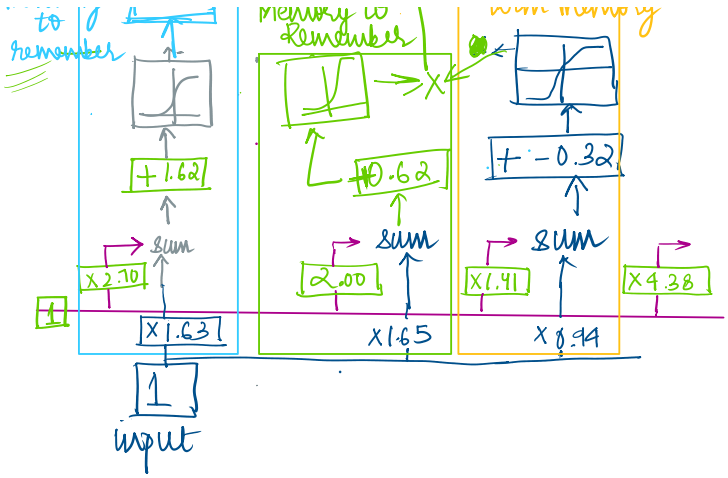
( ) x2) + (1x1.65) + 0.62 (1x0.94 + 1

= 4.27

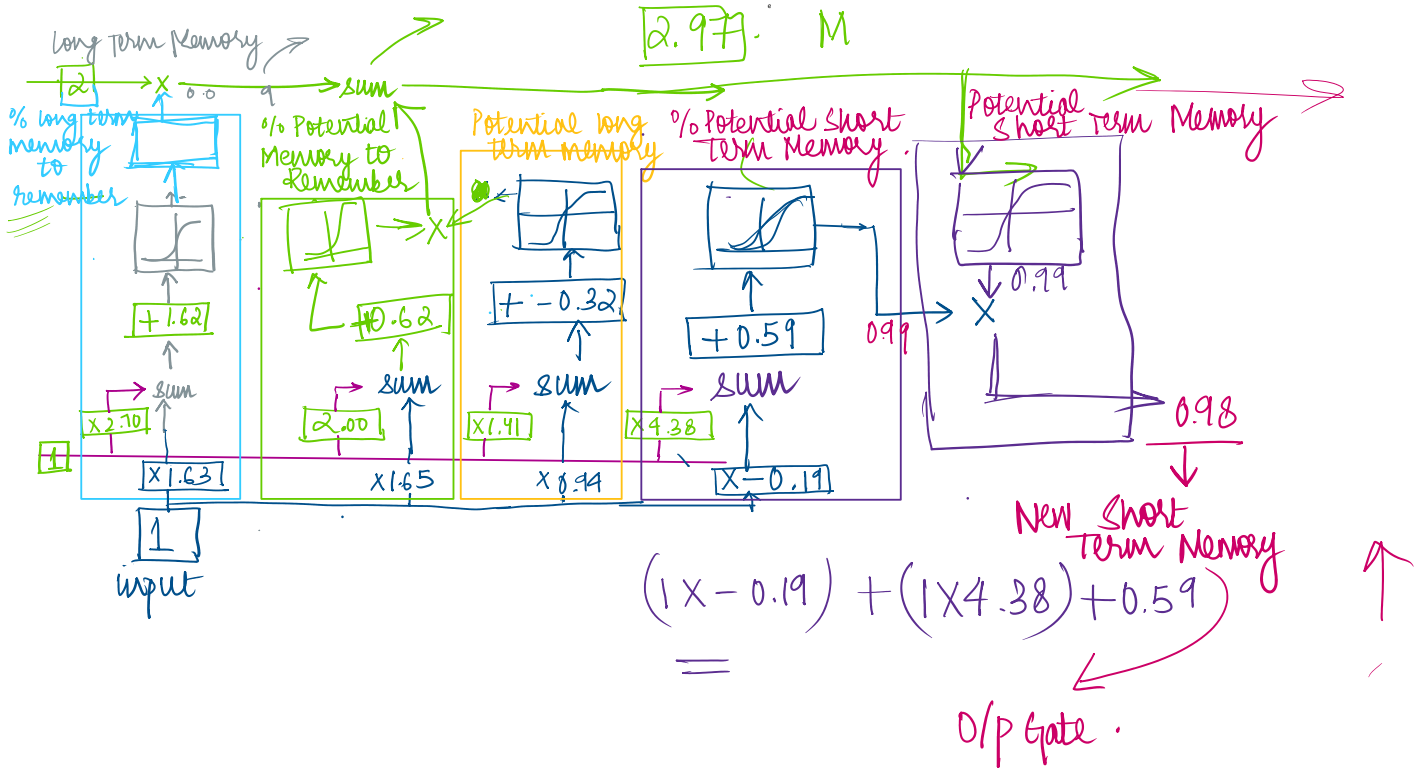
$) + 1 \times 0.91$

(-0.32)

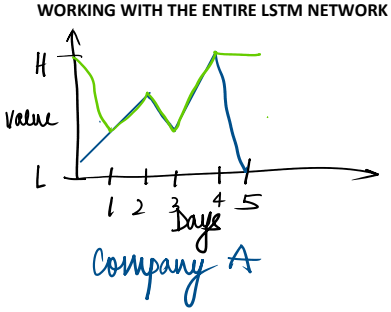
$= 2.03$



THIRD STAGE OF LSTM UNIT



WORKING WITH THE ENTIRE LSTM NETWORK



$= 2.03$

0.97

dinate

⇒ if i/p
was 10