LSTM

* Long term memory and short term memory introduced to reduce gradient explosion/vanish

(cell state)

LONG TERM MEMORY

(Input Gate)

(output Gate)

PRODUCT

SUM

SIGMOID | B

SIGM | B

[0,1]

TANH | B

SIGM | B

TANH

W_4 W_3

W_6 W_5

W_8 W_7

(Hidden State) W_1 W_2

SHORT TERM MEMORY

(FORGET GATE)

Input

New Short Term Memory
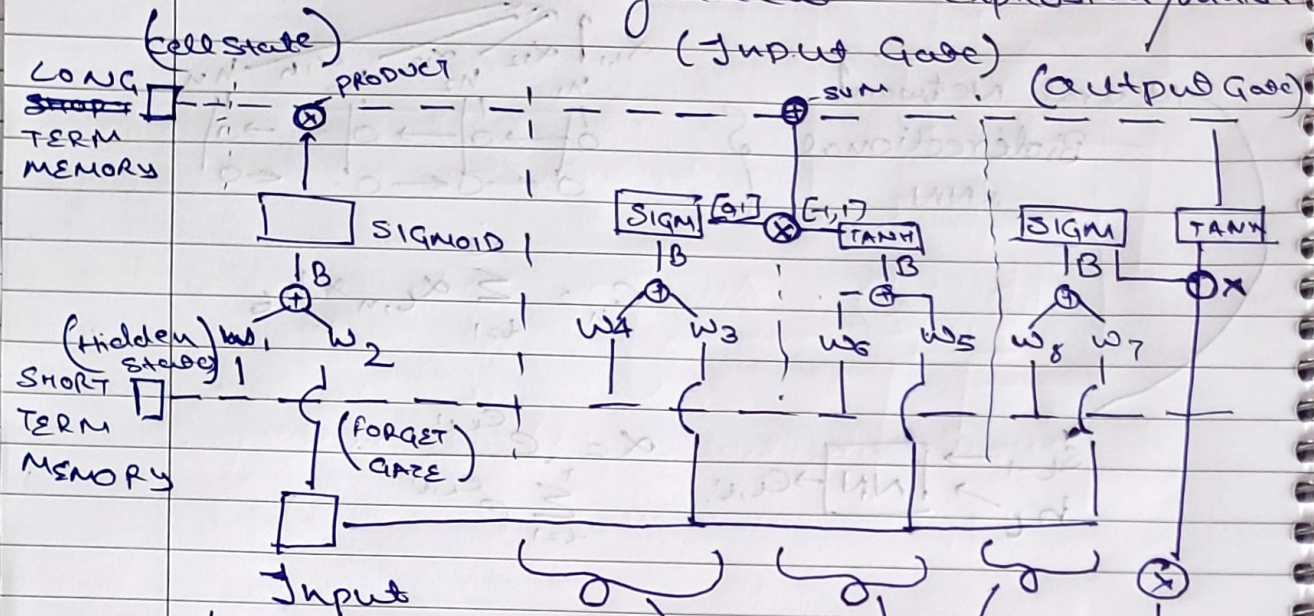
Forget Gate
Determines what percentage of LT memory to be forgotten.

% Potential Memory ST.

% Potential Memory LT

Amount of memory to be updated

Potential Memory

A new longterm memory

* LSTMS can face vanishing / exploding gradients on:

① Inaccurate initialization
② Extremely deep networks.

* Units are like neurons. more units mean more neurons in the same layer.