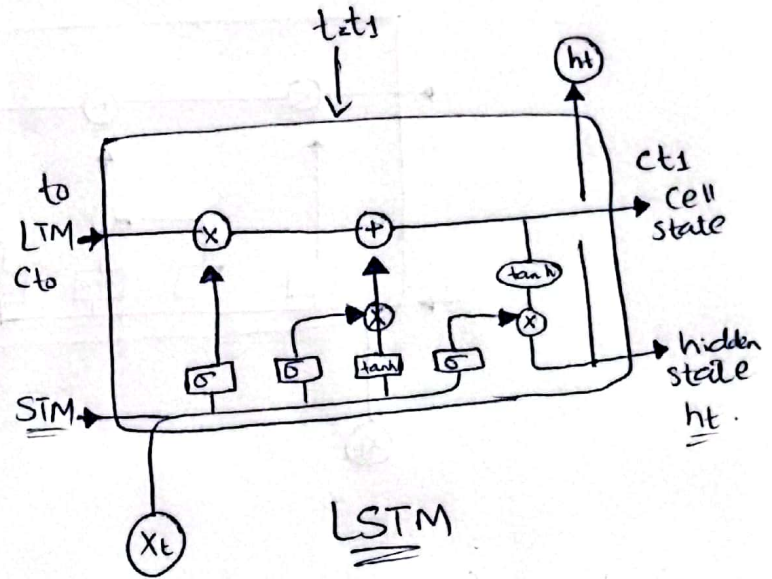
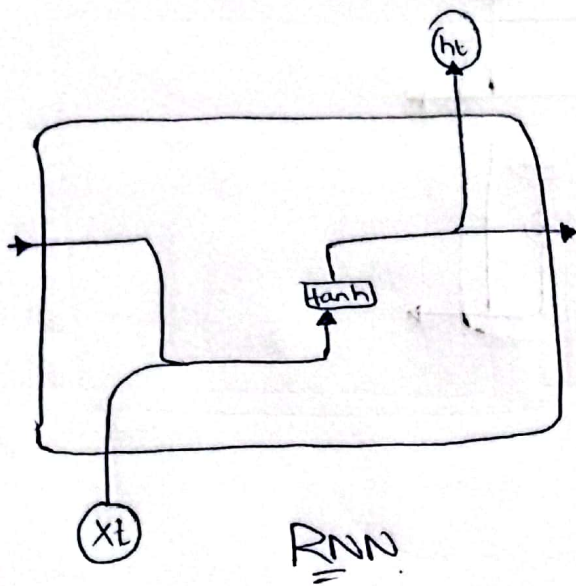
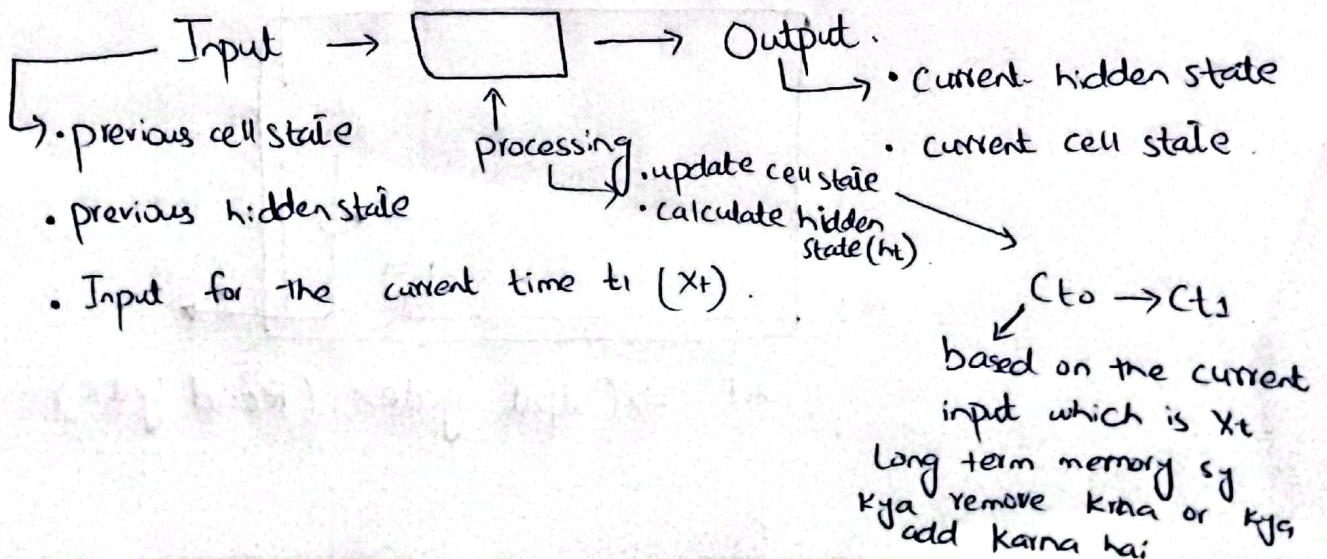
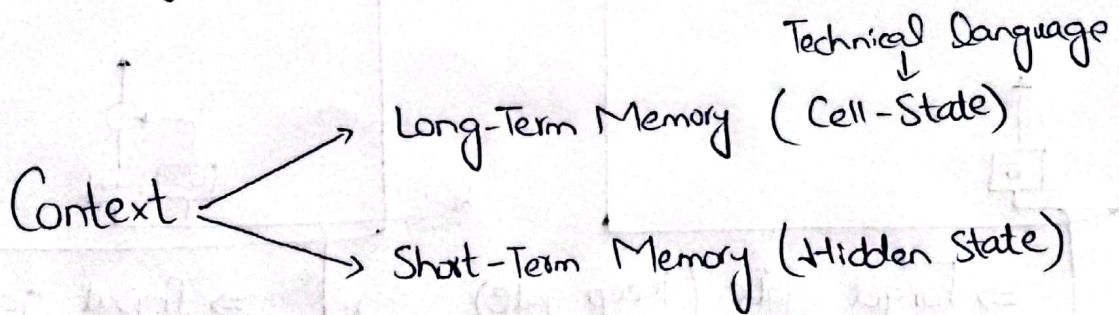


Long - Short Term Memory (LSTM)

Architecture

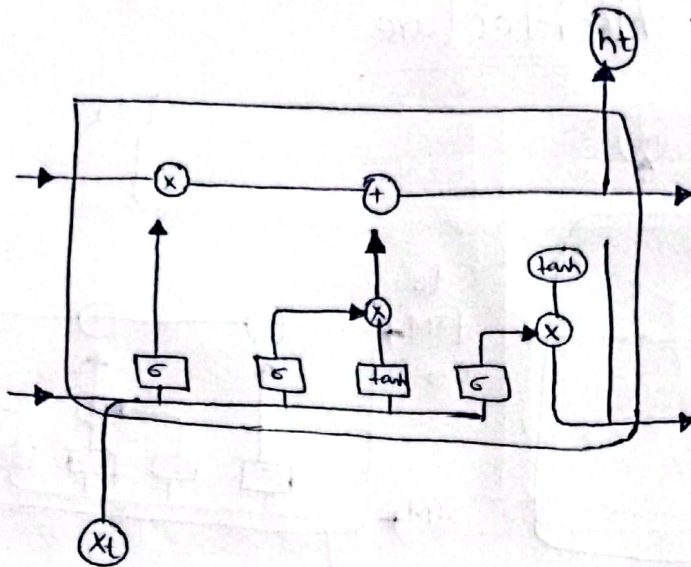


⇒ Two types of context

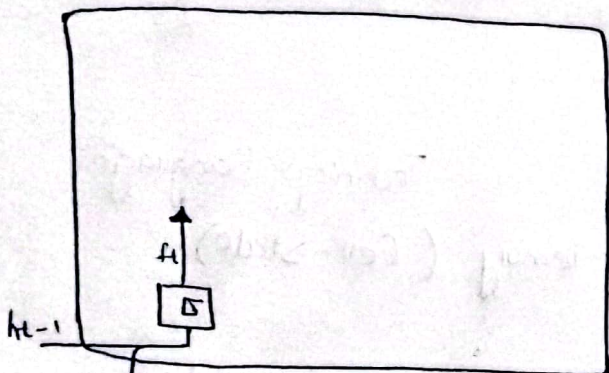


The Gates:-

Based on three gates, to protect and control the cell states.

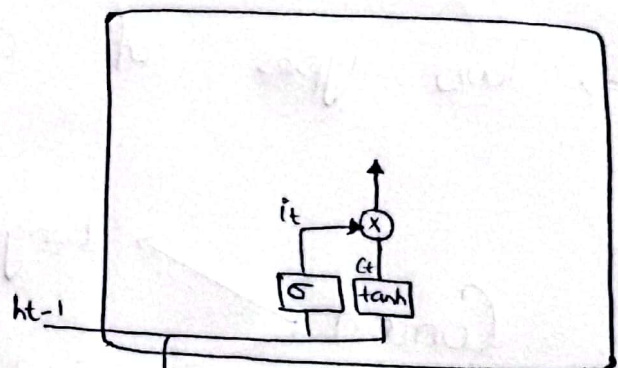


→ Remove something from cell state



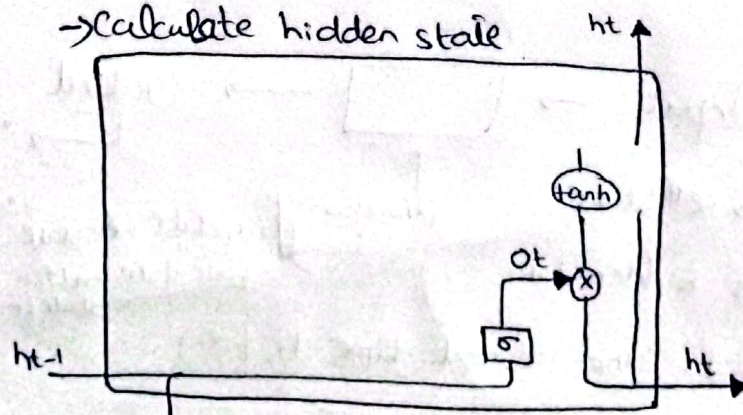
\Rightarrow Forget gate (keep gate)

→ Add something to the cell state.



\Rightarrow Input gate (write gate)

→ calculate hidden state



\Rightarrow Output gate (Read gate)

$C_t \quad h_t \rightarrow$ These are vectors

$$[0.1 \quad 0.3 \quad 0.9]$$

\hookrightarrow 3 dimensional vector

$$h_t = [0.1 \quad 0.2 \quad 0.3]$$

$$C_t = [0.4 \quad 0.5 \quad 0.6]$$

} dimensions are equal
(Always)

$X_t \rightarrow$ is a vector

Sentiment analysis problem

Predict $[0, 1]$

1 = +ive

0 = -ive

$$\begin{array}{ccc|c} \text{cat} & \text{mat} & \text{rat} & 0 \\ \text{cat} & \text{rat} & \text{rat} & 0 \\ \text{mat} & \text{mat} & \text{cat} & 1 \end{array}$$

\Rightarrow Lstm only numbers process

\Rightarrow How to calculate text into numbers?

• Text vectorization

\rightarrow Use one hot encoding

\rightarrow Bag of word

\rightarrow TF-IDF

$$\begin{array}{ccc|ccc} \text{cat} & \text{mat} & \text{rat} & & & \\ [100] & [010] & [001] & 1 & 0 & 0 \\ & & & 0 & 1 & 0 \\ & & & 0 & 0 & 1 \end{array}$$

$$\begin{array}{ccc} t=1 & t=2 & t=3 \\ [100] & [010] & [001] \end{array}$$

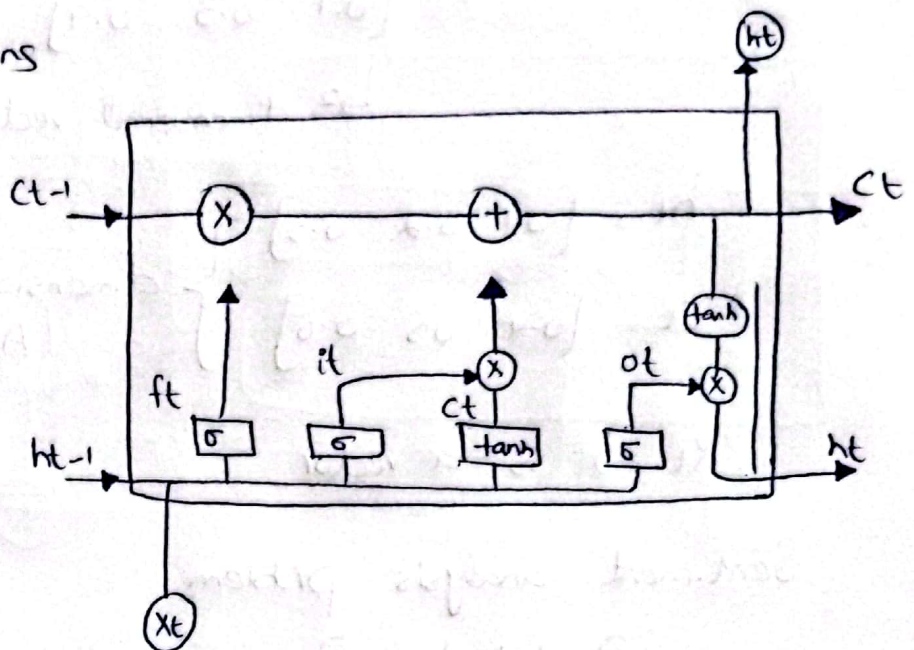
X_t is just one word, converted into a vector

f_t \downarrow forget gate
 i_t \bar{c}_t \downarrow Input gate
 o_t \downarrow Output gate

\Rightarrow These are vectors
 • Dimensions are same.

\Rightarrow Pointwise Operations

- $\rightarrow \otimes$
- $\rightarrow +$
- $\rightarrow \tanh$



$C_{t-1} \otimes f_t$

$$C_{t-1} = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}$$

$$f_t = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

$$C_{t-1} \otimes f_t = \begin{bmatrix} 4 & 10 & 18 \end{bmatrix}$$

$C_{t-1} + C_t$

$$C_{t-1} = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}$$

$$C_t = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

$$C_{t-1} + C_t = \begin{bmatrix} 5 & 7 & 9 \end{bmatrix}$$

$C_{t-1} \tanh$

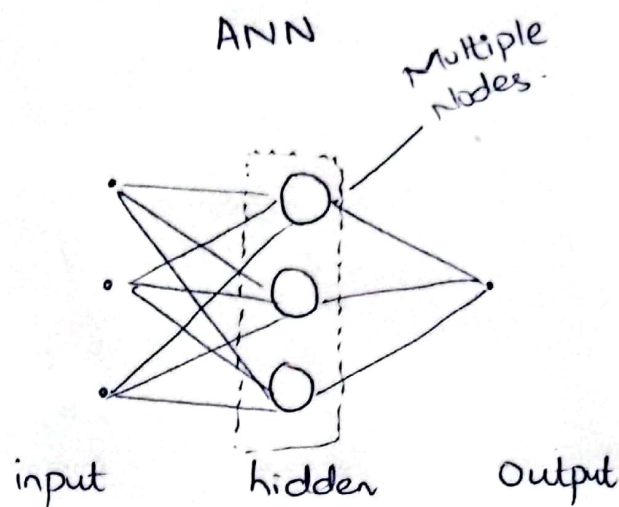
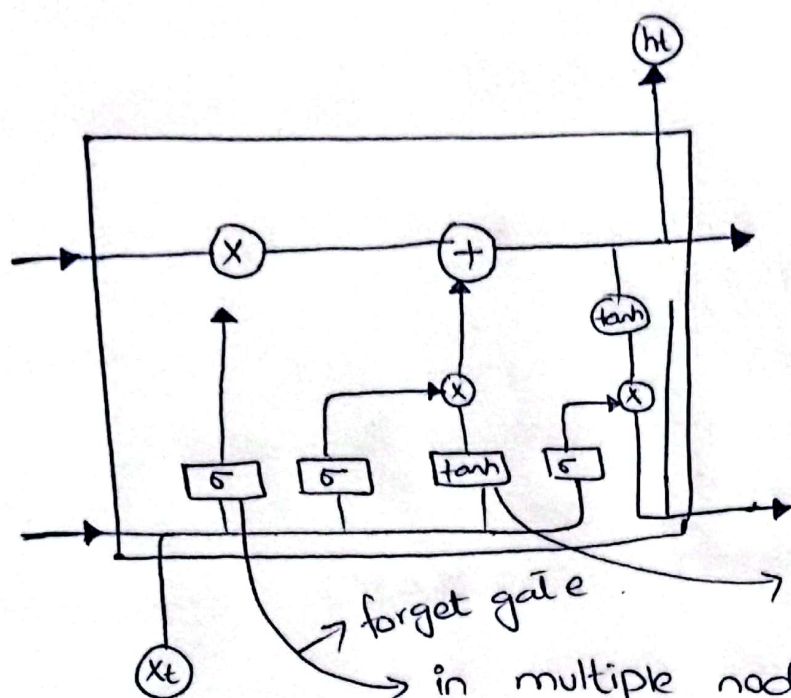
$$C_{t-1} = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}$$

$$\tanh(4) = 0.069 \quad \tanh(5) = 0.087 \quad \tanh(6) = 0.105$$

$$= \begin{bmatrix} 0.06 & 0.08 & 0.1 \end{bmatrix}$$

$$\tanh(6) = 0.105$$

⇒ Neural Network Layers



⑤ ⑤ ⑤

Activation function is tanh.

Activation function which is sigmoid.