20 Oct 2024

LSTM ARCHITECTURE

forget gate + input gate $\longrightarrow$ output gate

to remove some thing from C_t | to add some thing to the C_t | calculate hidden state (h_t)

What are C_t and h_t and x_t:

- these two are vectors. example [0.1, 0.3, 0.9]
- there the dimension of h_t and C_t are equal
- x_t is also a vector

What are f_t, i_t, o_t and $\overline{C_t}$:

- all of them are vectors

$\underbrace{f_t}_{\text{forget gate}}$ $\underbrace{i_t \quad \overline{C_t}}_{\text{input gate}}$ $\underbrace{o_t}_{\text{output gate}}$

- $\overline{C_t}$ is called candidate cell state
- also all of them dimensions are same.

point wise operations:

(i) point wise multiplication operation ⊗:

let $C_{t-1} = [4\ 5\ 6]$ $f_t = [1\ 2\ 3]$

$C_{t-1} \otimes f_t \rightarrow [4\ 10\ 18]$

(ii) pointwise addition operation ⊕:

let $C_{t-1} = [4\ 5\ 6]$ $f_t = [1\ 2\ 3]$

$C_{t-1} \oplus f_t \rightarrow [5\ 7\ 9]$

(iii) pointwise tanh operation:

let $C_{t-1} = [4 \quad 5 \quad 6]$

$\tanh(C_{t-1}) = [0.26 \cdot 0.34 \quad 0.53]$

Neural Network Layers:

$\boxed{\sigma}$: neural network layer with each node has activation function sigmoid

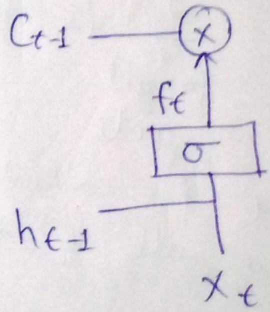
$\boxed{\tanh}$: neural network layer with each node have activation function is tanh()

- the number of ~~b~~ node in each layer is divided by user. like a hyperparameter (based on performance). with all layers have same ~~node~~ number of node

(1.) Forget Gate:

working stage:

- Calculate f_t
- $C_{t-1} \otimes f_t$

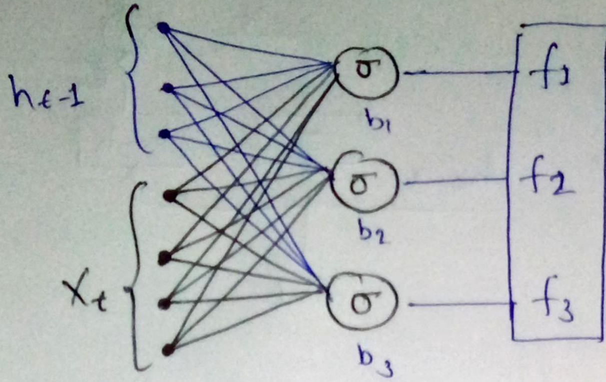


let, x_t is 4 dimension vector.

$[x_{i1}, x_{i2} \quad x_{i3} \quad x_{i4}]$

then number of node in neural network with sigmoid activation function is 3

the vectore dimension of h_{t-1} is 3



$f_t = [f_1, f_2, f_3]$

21 weights.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

and pointwise operation $\boxed{f_t \otimes C_{t-1}}$ ← decide how much information move forward

let $C_{t-1} = [4 \quad 5 \quad 6]$

$f_t = [\tfrac{1}{2} \quad \tfrac{1}{2} \quad \tfrac{1}{2}]$

$C_{t-1} \otimes f_t = [2 \quad 2.5 \quad 3]$ information become half.

if $f_t = [1 \quad 1 \quad 1]$

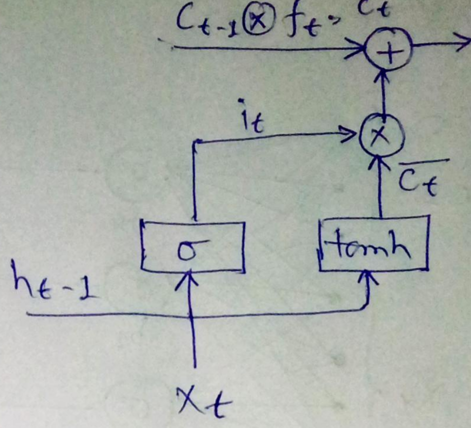
$C_{t-1} \otimes f_t = [4 \quad 5 \quad 6]$ no info loss.

if $f_t = [0 \quad 0 \quad 0]$

$C_{t-1} \otimes f_t = [0 \quad 0 \quad 0]$ all information lost.

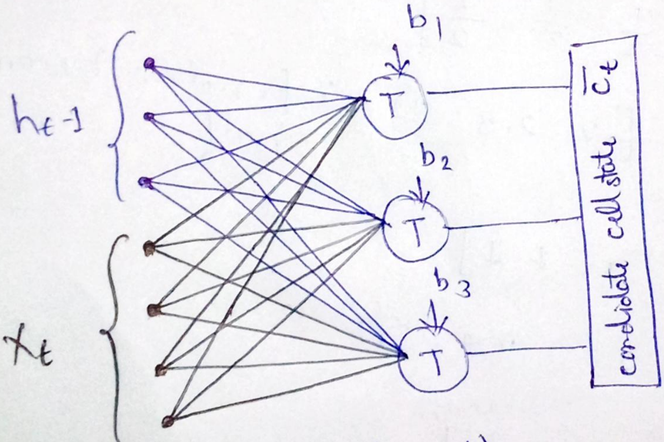
2. Input Gate:

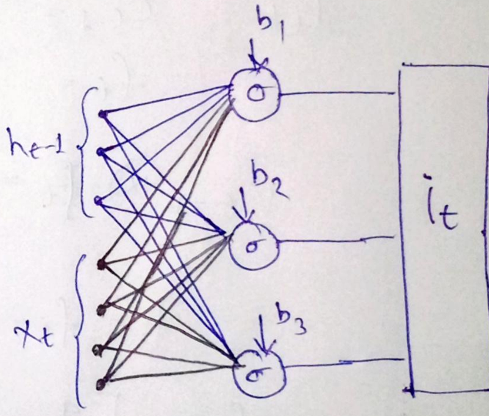
The reson for use of input gate is to add some new important information to the cell state



working stage:

- calculate candidate cell state $\bar{C}_t$
- calculate i_t (it decide what values of $\bar{C}_t$ is added in cell state (filtering)).
- Calculate $C_t = f_t \otimes C_{t-1}$ = current cell state





21 W_c (weights)

$$\bar{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

potential important information.

T = Tanh activation function

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

σ = sigmoid activation function.
filter the $\bar{C}_t$ values.

Now point wise operation between the i_t and $\bar{C}_t$ (filtered candidate cell state)

let $\bar{C}_t$ = [9 5 6]

i_t = [0.5 0.5 0.5]

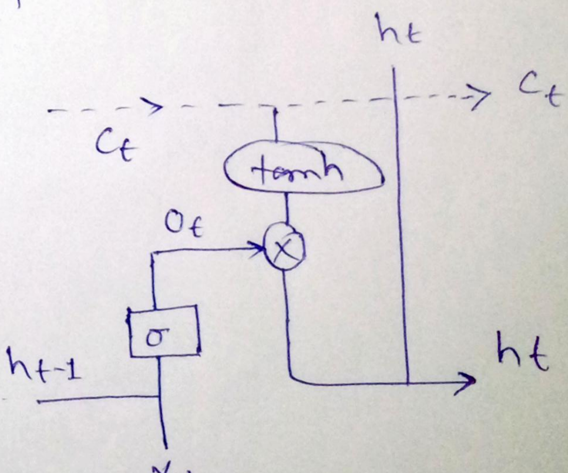
$$i_t \otimes \bar{C}_t = [2, 2.5, 3]$$

Now point wise addition operation between the $f_t \otimes C_{t-1}$ and $i_t \otimes \bar{C}_t$

$$C_t = (f_t \otimes C_{t-1}) \oplus (i_t \otimes \bar{C}_t)$$

(3) Output Gate:

used for current time step के लिए output decide करना (find the value of hidden state (h_t))



working stage:

- apply tanh on C_t
- filteration on tanh(C_t) done by O_t
- point wise multiplication of $O_t \otimes \tanh(C_t)$

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

$$h_t = O_t \otimes \tanh(C_t)$$ = output for current time step