← Sequence Data →

Speech recognition → Sequence of words

Music generation → output notes.

Machine translation → Voulez-Vous chanter avec moi ?
↓
Will you sing with me?

x_1	x_2	x_3	y

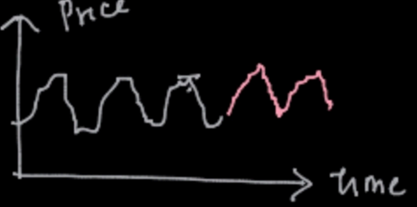
I am speaking.
x_{t-2} x_{t-1} x_t

→ Temporal Data

25×12 → 300

Video activity → □ □ □ □ → Walking.

Stock Market → Stock prices.

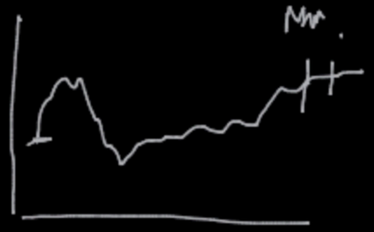


Tom worked in the navy. He had always loved being in the sea.

Recurrent → loop.
→ feedback loop.

Sentence completion → . . . ___

RNN → LSTM → Transformers
→ GRU



Long ago
Past ↑

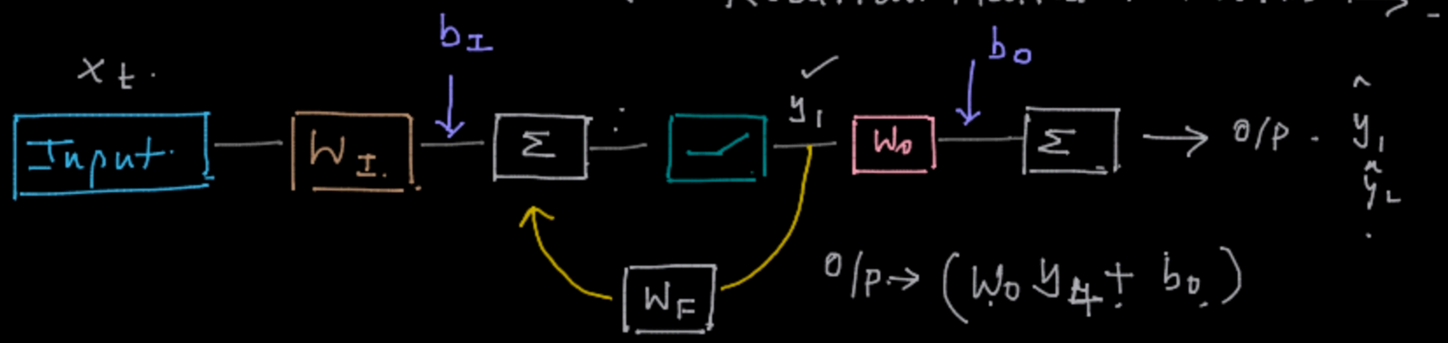
Time	Stock price
Jan 2001 1	–
Feb 2001 2	–
Mar 2001 3	–
→ May 2023	.
→ May 2024	.
→ Mar 2025 ✓	.
→ Apr 2025 ✓	.
→ May 2025 →	–

↑ Past.

60

t → .

← Recurrent Neural Networks →



y_{t-bp} . y . $\hat{y}$.

day 1 — .
y_{t-3} day 2 — —
y_{t-2} day 3 — .
y_{t-1} day 4 — .
y_5 day 5 — .

t = 2 → ✓

$y_1 = Relu(W_I x_t + b_I)$.

$y_2 = Relu(W_I x_t + W_F y_1 + b_I)$

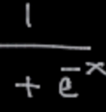
Ignore $y_3 = Relu(W_I x_4 + W_F y_2 + b_I)$. $\hat{y}_5$

↑ activation

$o \Rightarrow fn(W_1 x_1 \oplus W_2 x_2 \oplus W_3 x_3)$

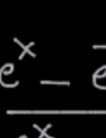
ReLu →

Sigmoid → $\frac{1}{1+e^{-x}}$

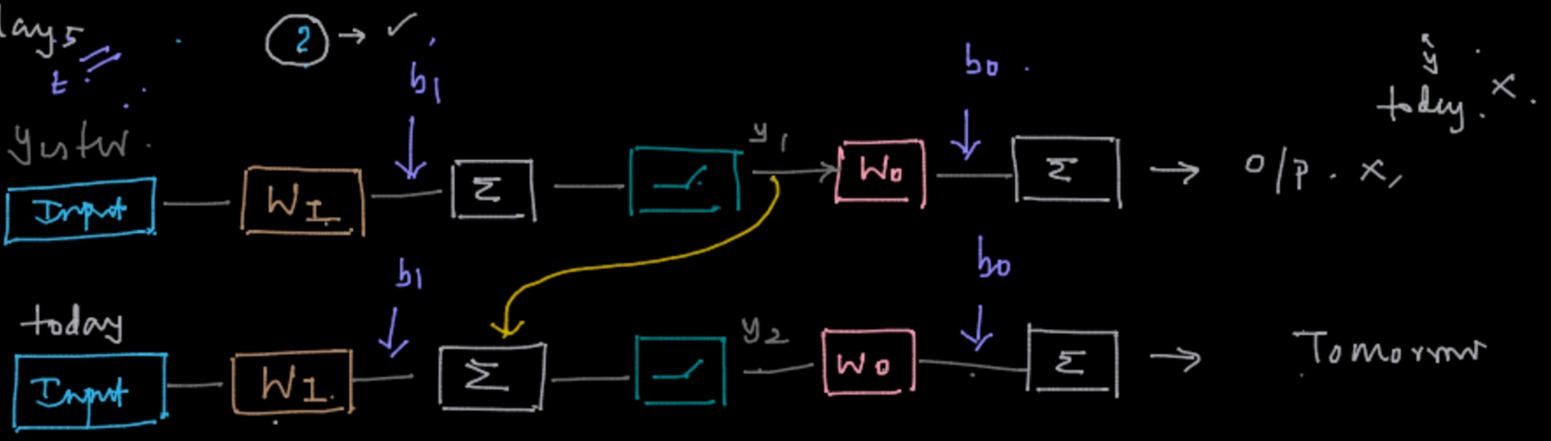


0 → 1 .

tanh → $\frac{e^x - e^{-x}}{e^x + e^{-x}}$



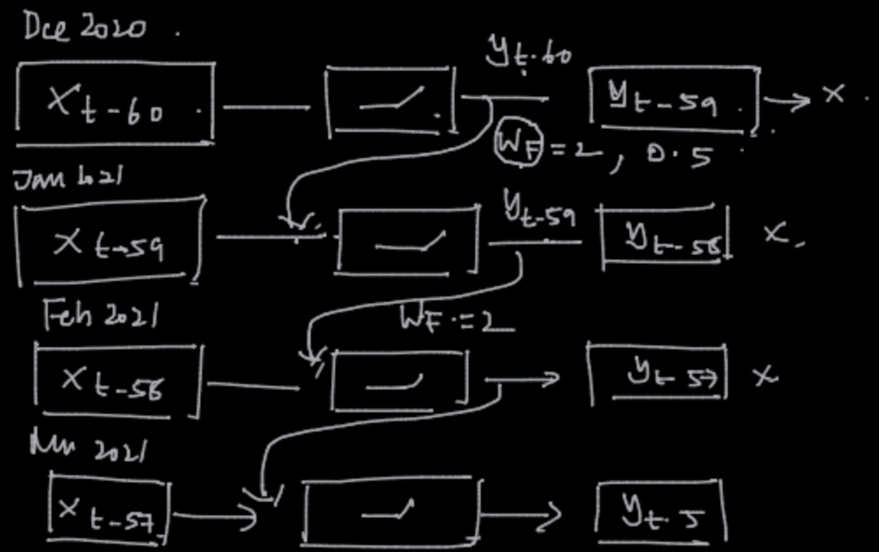
→ −1 to +1 .



I/p .
Yesterday → Tomorrow .
today .

Weights and biases are shared

long upto 1/2



$(2 \times y_1)\ (2\ y_1)$

$2 \times 2 \times 2 \rightarrow 2^3$

$2^{60} \times \ldots$

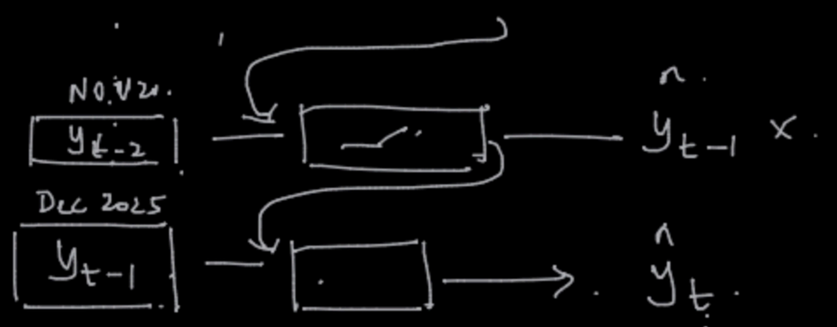
$(1/2)^{60} \rightarrow 1/2^{60}$

$\rightarrow 0$

$W > 1$

< 1

Shortterm.



y_{t-60} Dec 2020

y_{t-59} Jan 2021

y_{t-58} Feb 2021 →

1 period → 1 month -

→ (Jan 25) Feb 25 →

y_{t-1} Dec 2025 — — 2 → Jan 2026

Week 1 — y_{t-3}

Week 2 — y_{t-2}

Week 3 — y_{t-1}

Week 4 → y_t

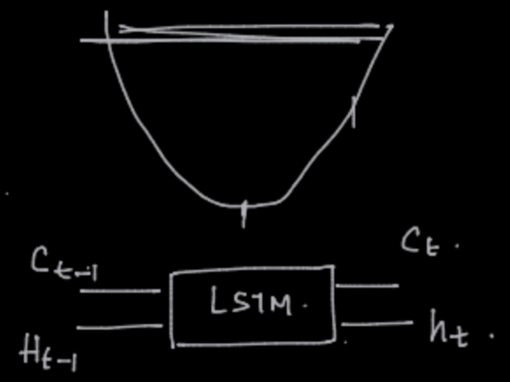
$\hat{y}_{t-1}$ 1 period ⇒ 1 week -

$W_0 \rightarrow$

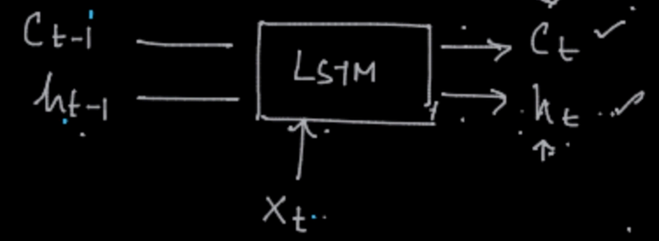
Exploding / Vanishing Gradients -

Cellstate → No Wts

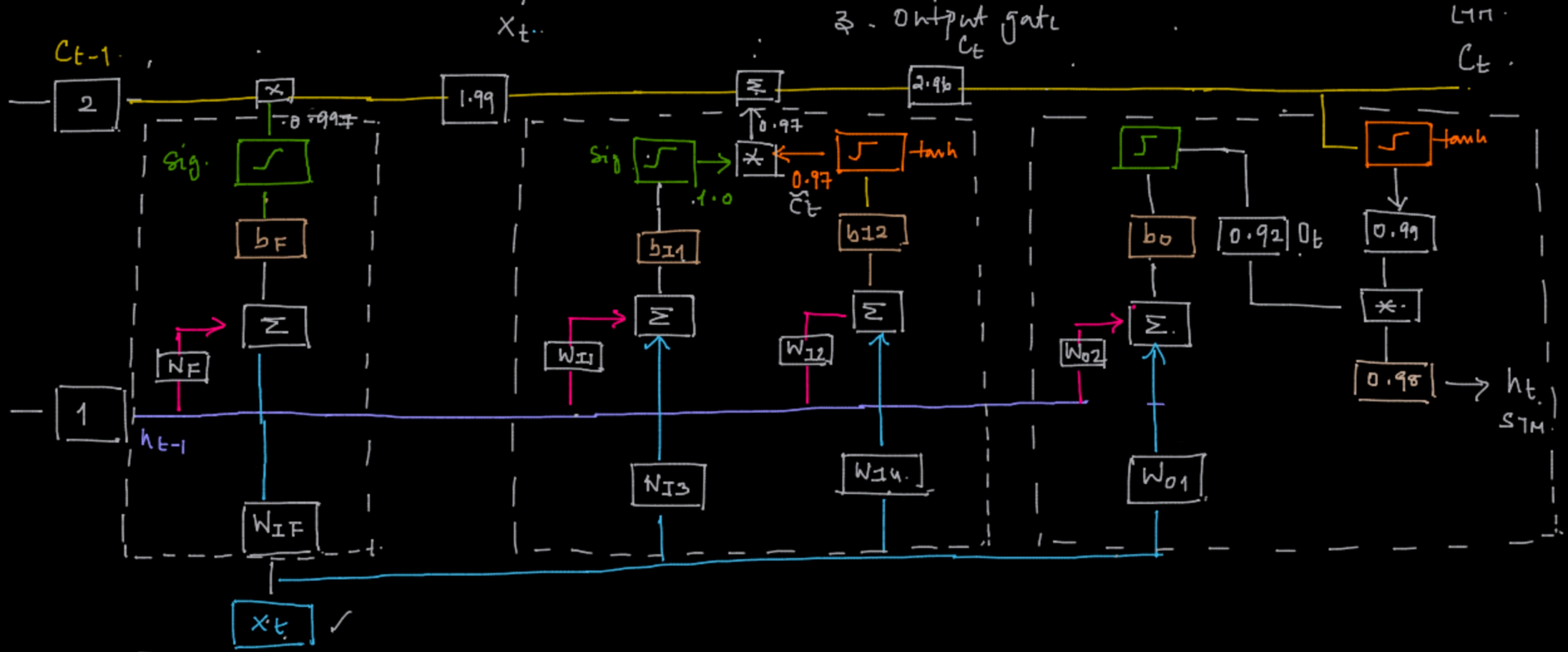
Hiddenstate wt



← LSTM → long Short term Memory →



1. Forget gate → How much of the past is remembered
   0 → 100%.
2. Input gate.
3. Output gate



Forget gate

Sigmoid → LTM.
tanh → STM ✓

1. Forget Gate: How much of the past information is remembered.

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

$\Rightarrow 0 \rightarrow 1$. % to be remembered

2. Input Gate: Tries to learn the new info that is given as input

$$\tilde{C}_t = \tanh\left(W_I\left[h_{t-1}, x_t\right] + b_{I2}\right) \rightarrow \text{candidate / potential Cell state}$$

$$i_t = \sigma\left(W_I\left[h_{t-1}, x_t\right] + b_{I1}\right)$$

3. output: Updated past and current info is passed to the next period

$$C_t = C_{t-1} + i_t \times \tilde{C}_t$$

C_{t-1} → LTM

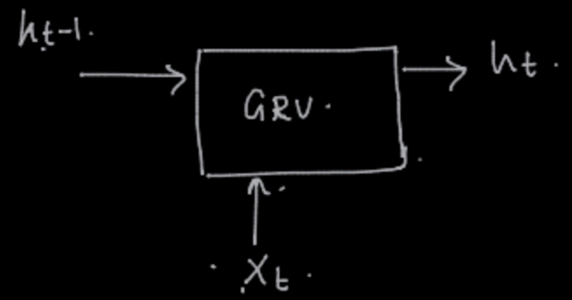
$i_t \times \tilde{C}_t$ → Fraction of potential memory

$$o_t \rightarrow \sigma\left(W_O\left[h_{t-1}, x_t\right] + b_O\right)$$

$$h_t \rightarrow o_t * \tanh\left(C_{t-1}\right)$$

← Gated Recurrent Unit (GRU) →

- Simpler Version of LSTM.
- easier to train.
- Cannot capture complex long term relationships.



→ Reset gate.
update gate (combines the forget gate)

Reset Gate

$$r_t = \sigma(W_R(h_{t-1}, x_t) + b_r)$$

update Gate

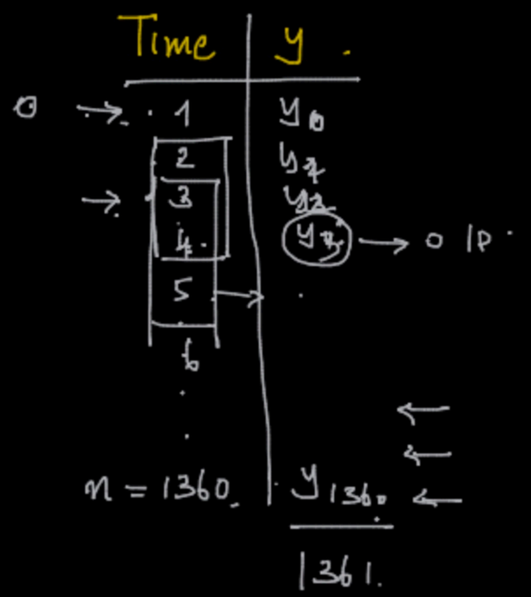
$$u_t = \sigma(W_{u1}(h_{t-1}, x_t) + b_u)$$

Candidate hidden state

$$\tilde{h}_t = \tanh(W_H[x_t, r_t * h_{t-1}] + b_h)$$

→ reset gate controls how much of the past hidden state is used to control the candidate hidden state.

$$h_t = W_{u2} \times h_{t-1} + (1 - W_{u2})\,\tilde{h}_t$$



3 periods → 4th period.

← 3 past periods →

	x_1	x_2	x_3	y
→	y_2	y_1	y_0	y_3
	y_3	y_2	y_1	y_4
	y_4	y_3	y_2	y_5
	y_{1358}	y_{1357}	y_{1356}	→ 1359
	y_{1359}	y_{1358}	y_{1357}	→ 1360
	y_{1360}	y_{1359}	y_{1358}	→ 1361

look_back = 3

→ X →	0 : 0+3	1 : 1+3	2 : 2+3
→ y →	3+0	3+1	2+3

10 − 3

length − lookbc. → records