CS 4063 – Natural Language Processing

Lecture Notes – week 7 Lec 1+2

Muhammad Hannan Farooq

Today’s Lecture

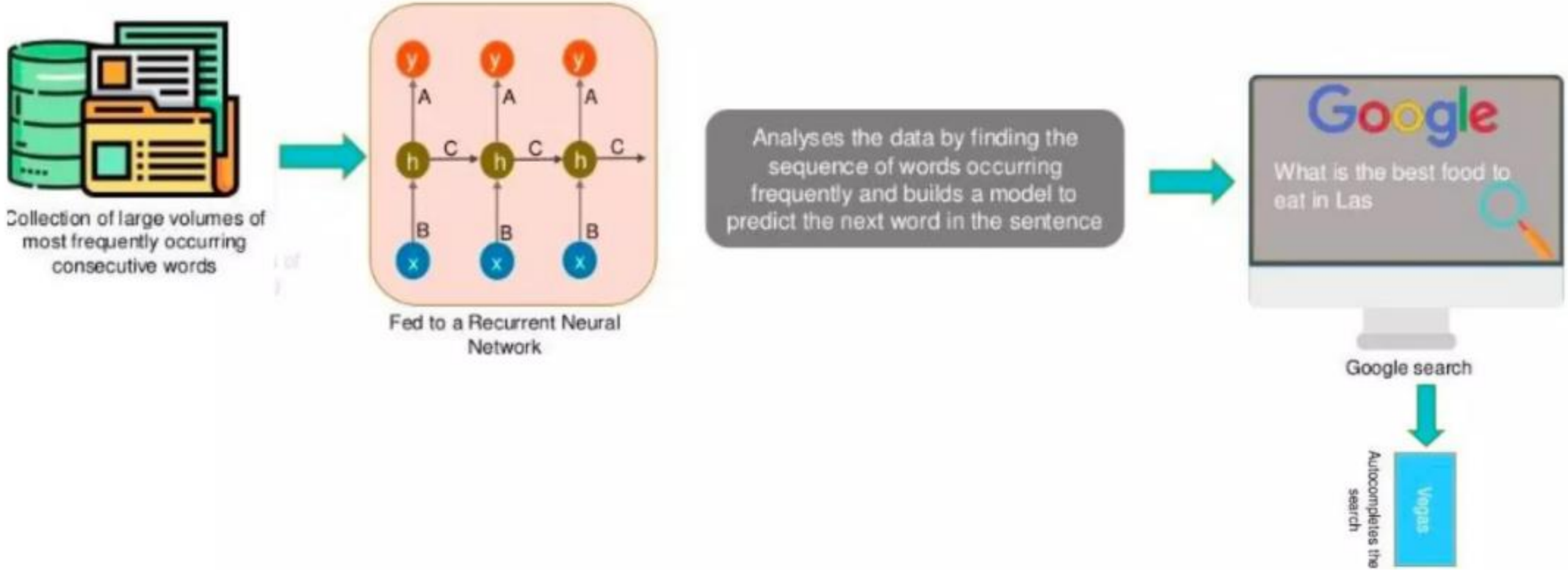
- Revision of RNN
- LSTM

RNN

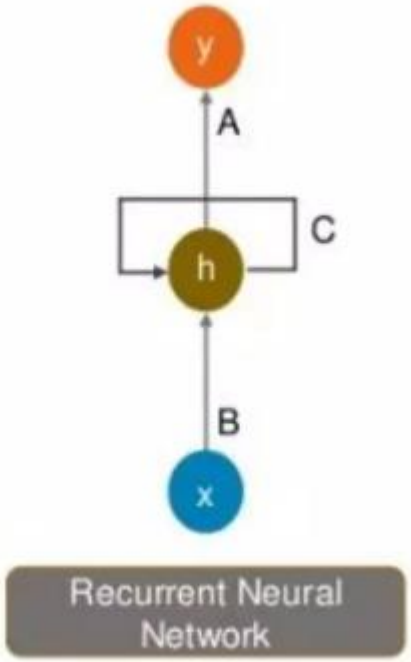
- Handle sequential data
- Each word in a sentence (or token in text) is fed into the network one at a time.
- Maintain a hidden state that captures information from previous steps, allowing them to learn patterns in sequences, such as the relationship between words.

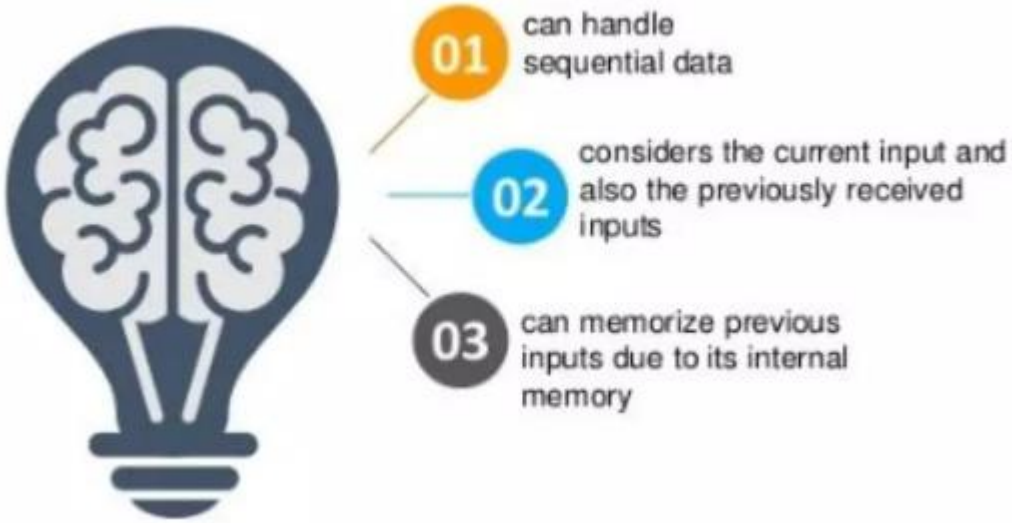
RNN

Do you know how Google's autocomplete feature predicts the rest of the words a user is typing?

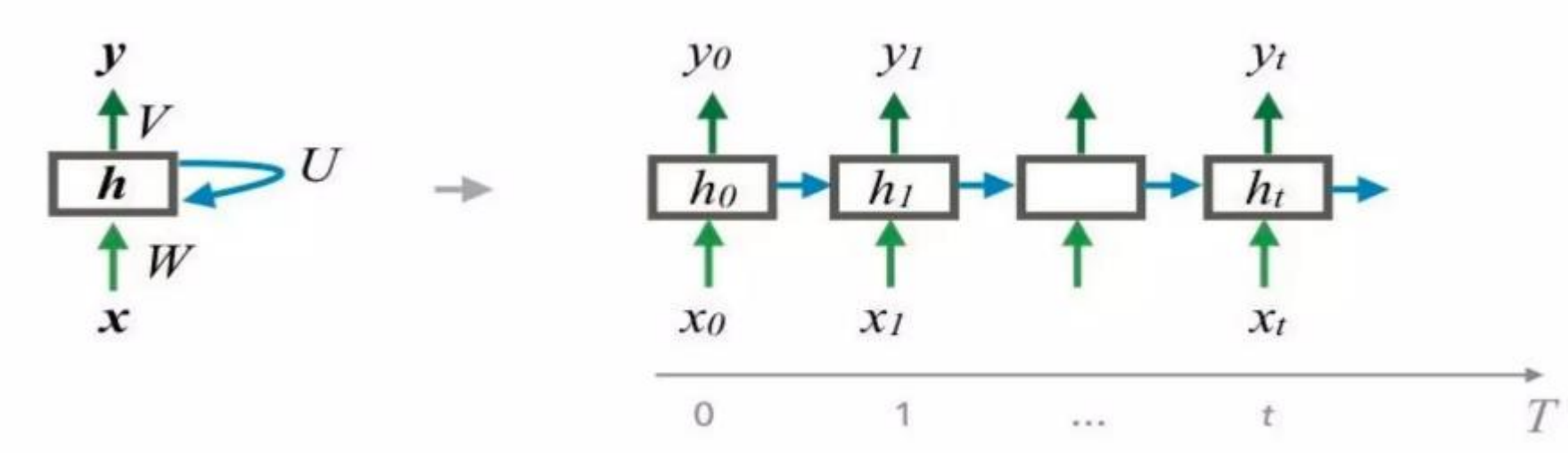


RNN

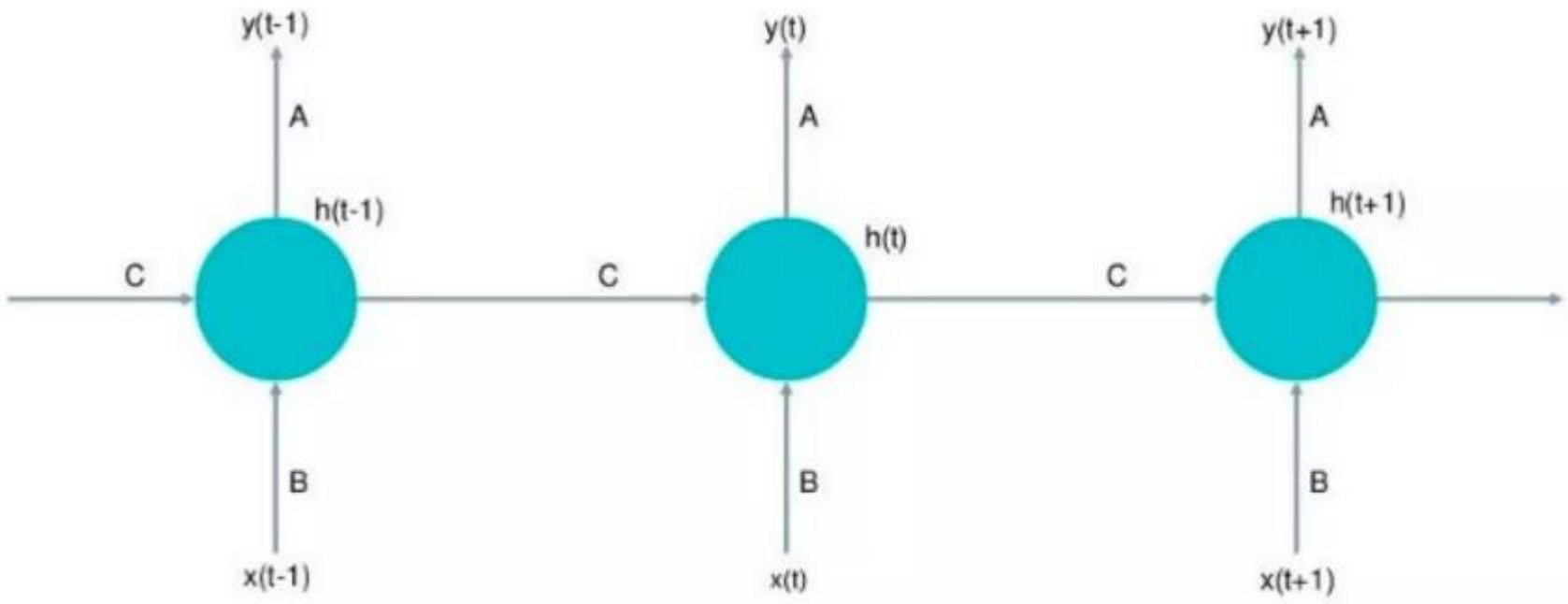

y
A
C
h
B
x
Recurrent Neural Network


01
can handle sequential data
02
considers the current input and also the previously received inputs
03
can memorize previous inputs due to its internal memory

RNN



RNN



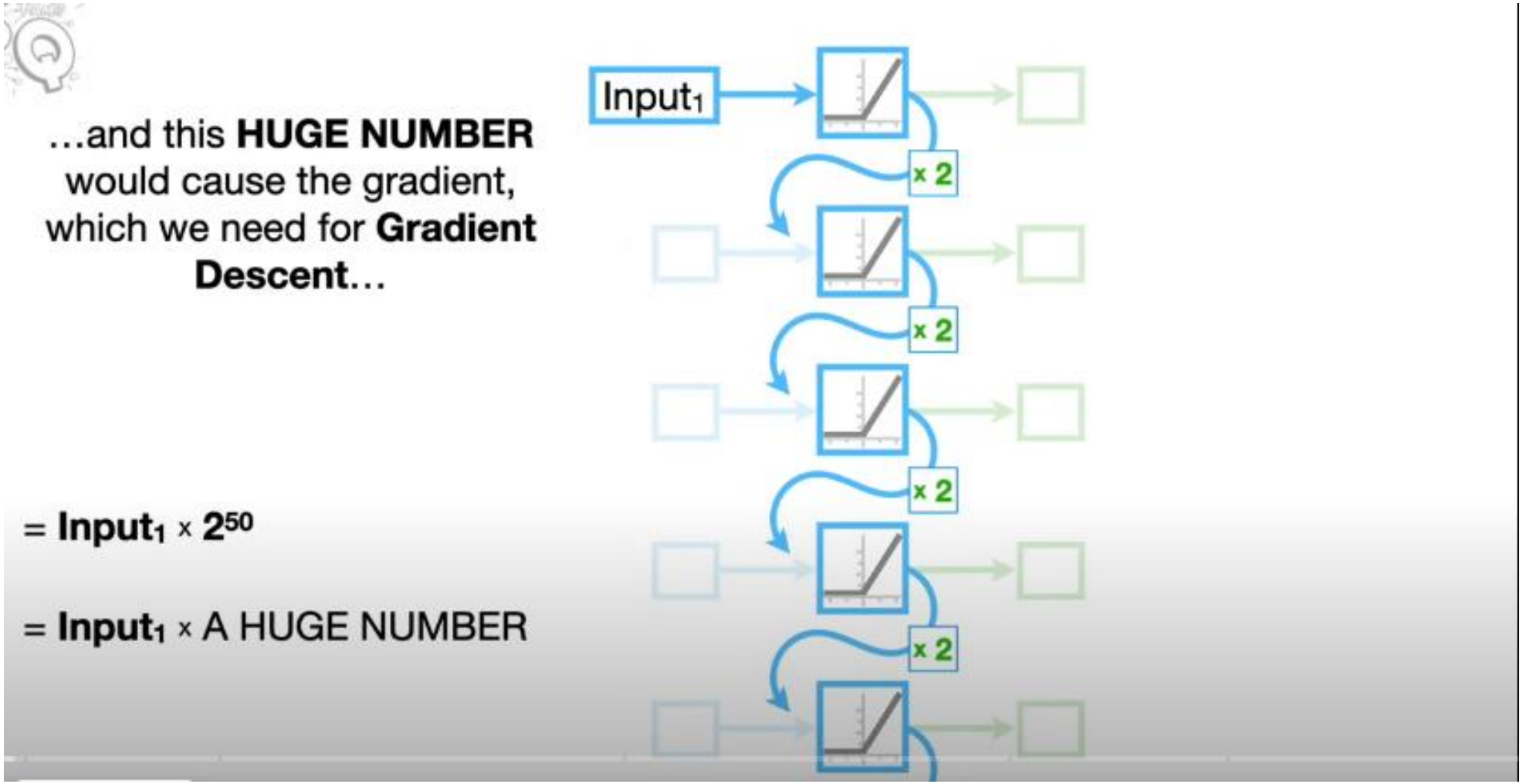
$$h(t) = f_c(h(t-1), x(t))$$

$h(t)$ = new state
f_c = function with parameter c
$h(t-1)$ = old state
$x(t)$ = input vector at time step t

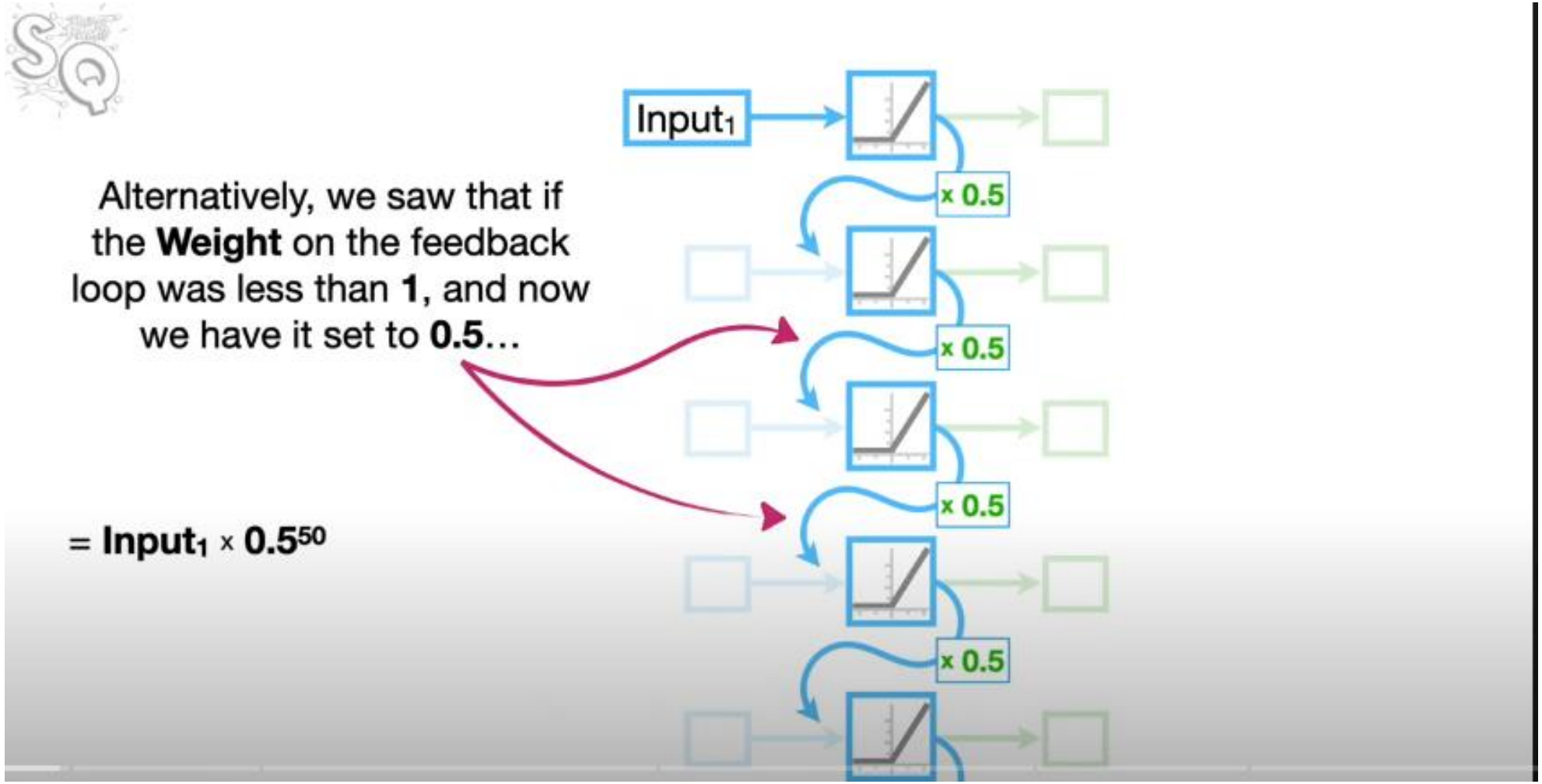
Limitation

- **Vanishing Gradient Problem:** When training RNNs, the gradients used in backpropagation can become too small (vanish), making it difficult to capture long-term dependencies.
- **Long-Term Dependencies:** Basic RNNs struggle to remember information over long sequences, which is why advanced variations like LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Units) are often used.

Limitation-Exploding Gradient Problem

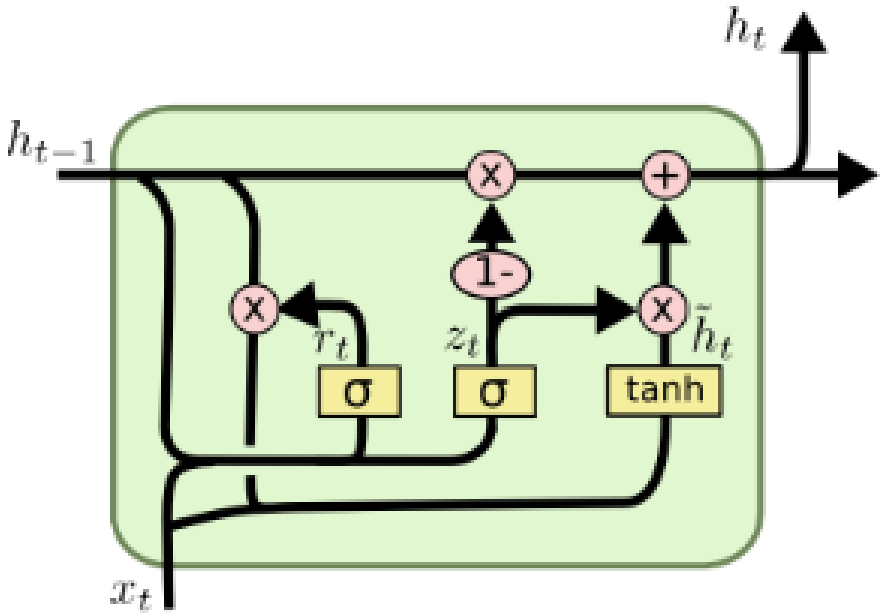


Limitation-Vanishing Gradient Problem

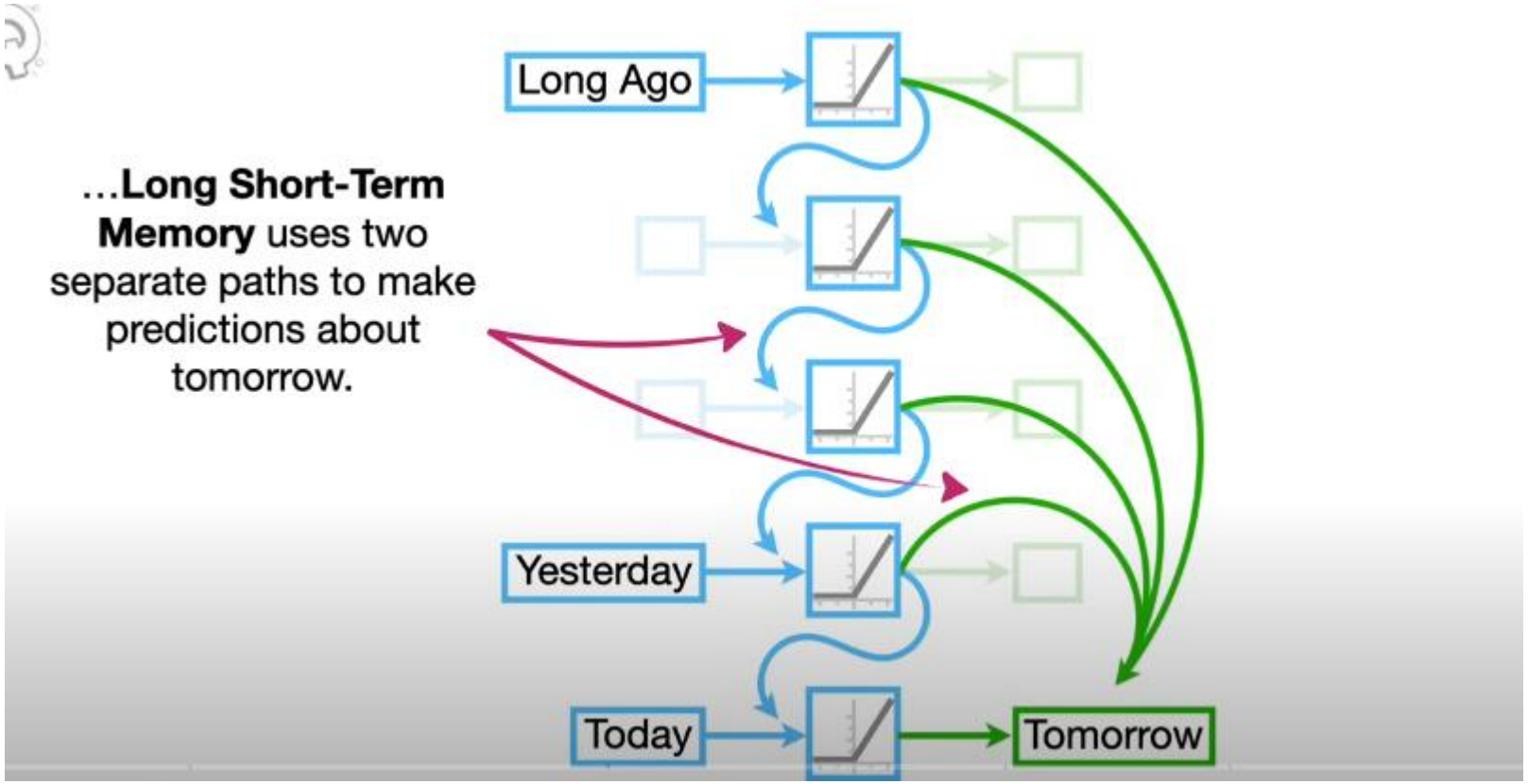


LSTM

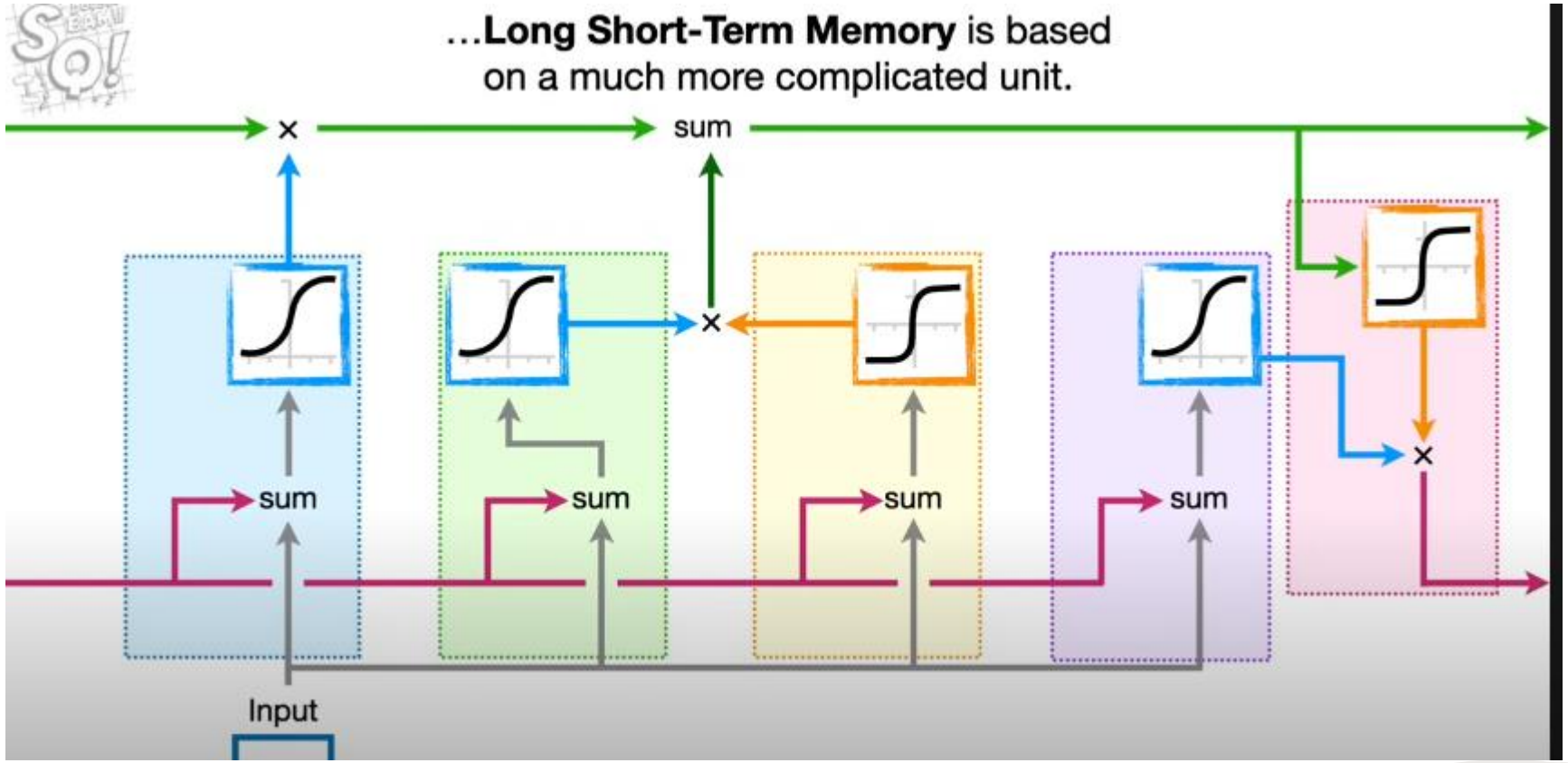
- The LSTM architectures involves the memory cell which is controlled by three gates: the input gate, the forget gate, and the output gate.
- These gates decide what information to add to, remove from, and output from the memory cell.



LSTM

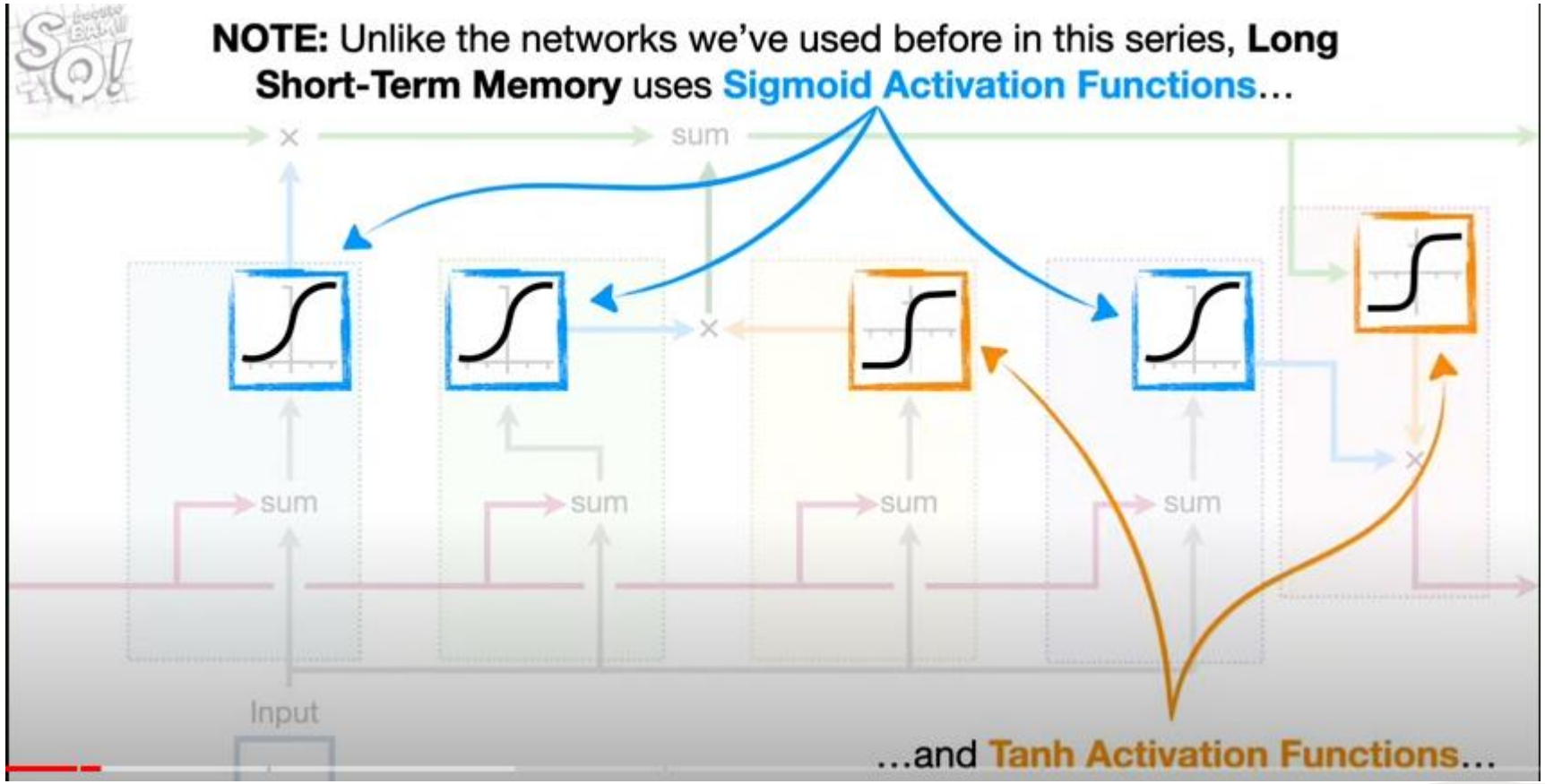

...Long Short-Term Memory uses two separate paths to make predictions about tomorrow.
Long Ago
Yesterday
Today
Tomorrow

LSTM


...Long Short-Term Memory is based
on a much more complicated unit.
×
sum
×
×
sum
sum
sum
sum
Input

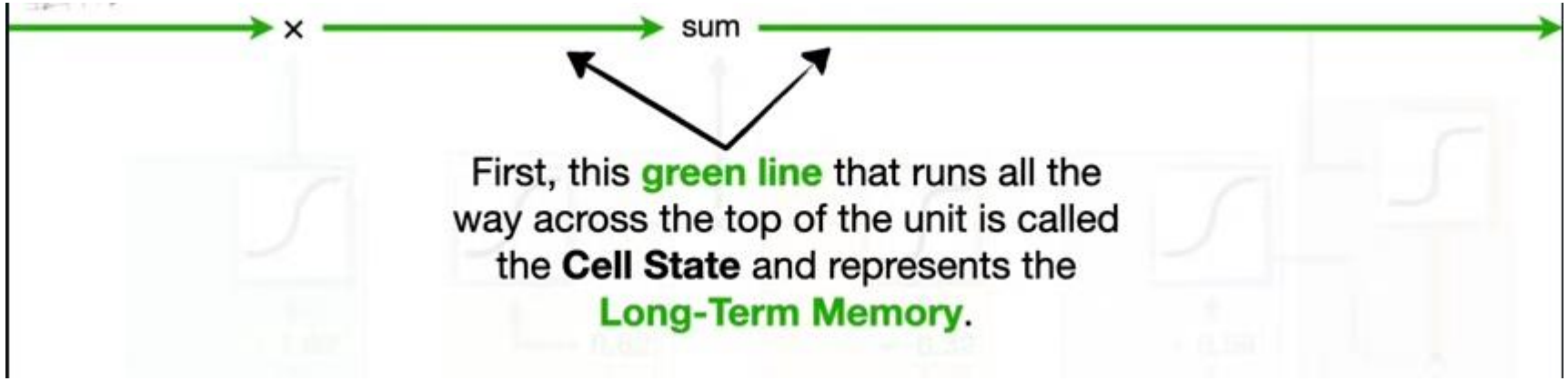
LSTM

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

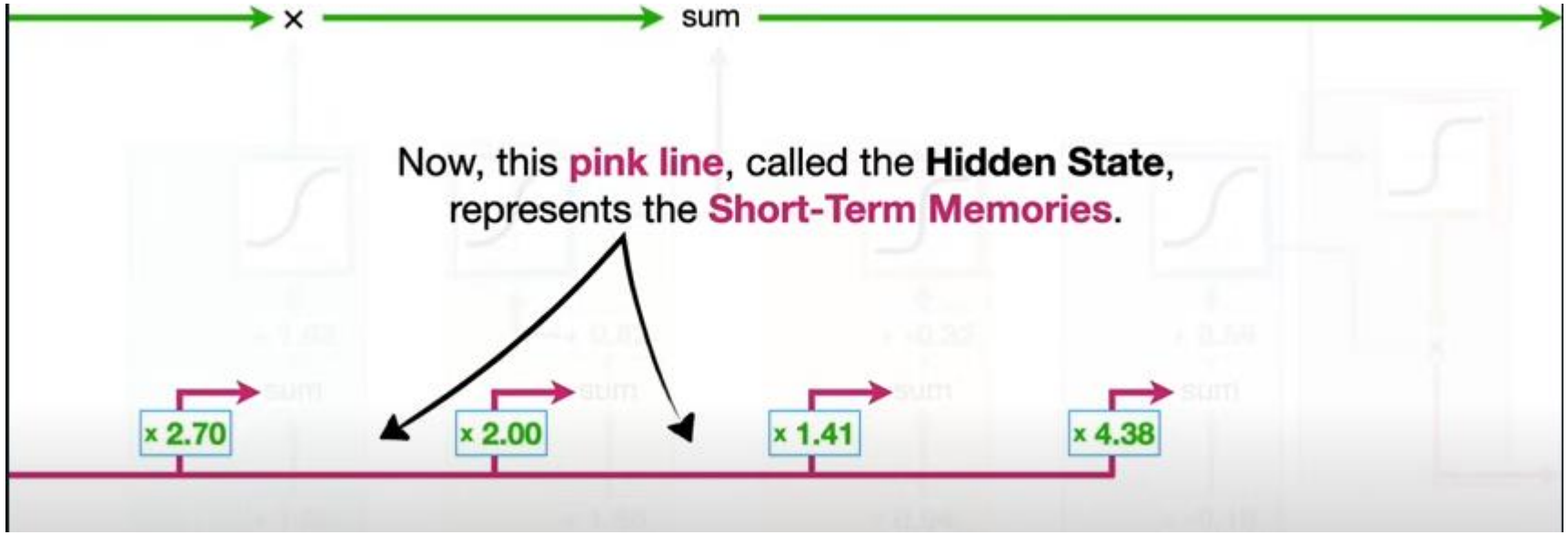


$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

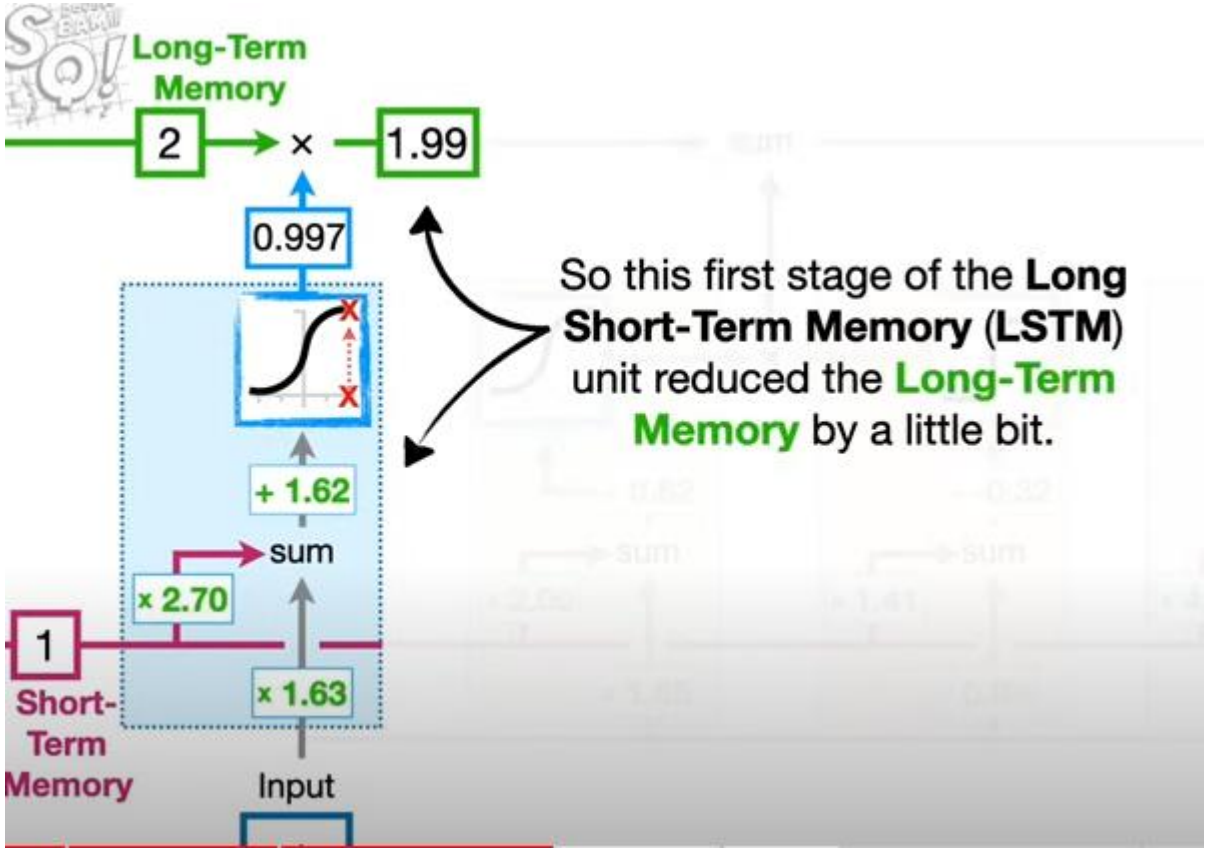
LSTM


×
sum
First, this green line that runs all the way across the top of the unit is called the Cell State and represents the Long-Term Memory.

LSTM


×
sum
Now, this pink line, called the Hidden State,
represents the Short-Term Memories.
× 2.70
× 2.00
× 1.41
× 4.38

LSTM- Forget gate

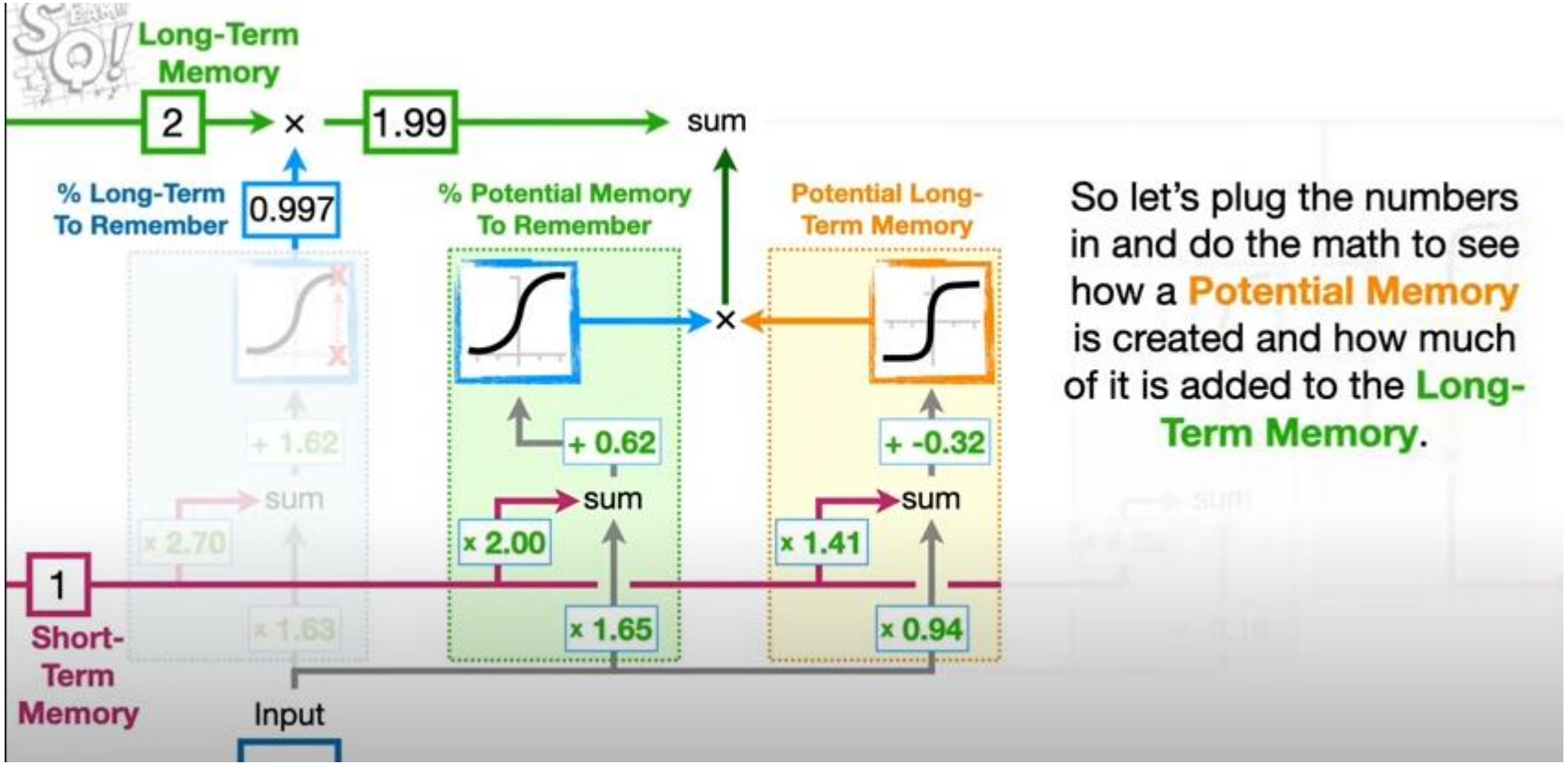


TERMINOLOGY ALERT!!!

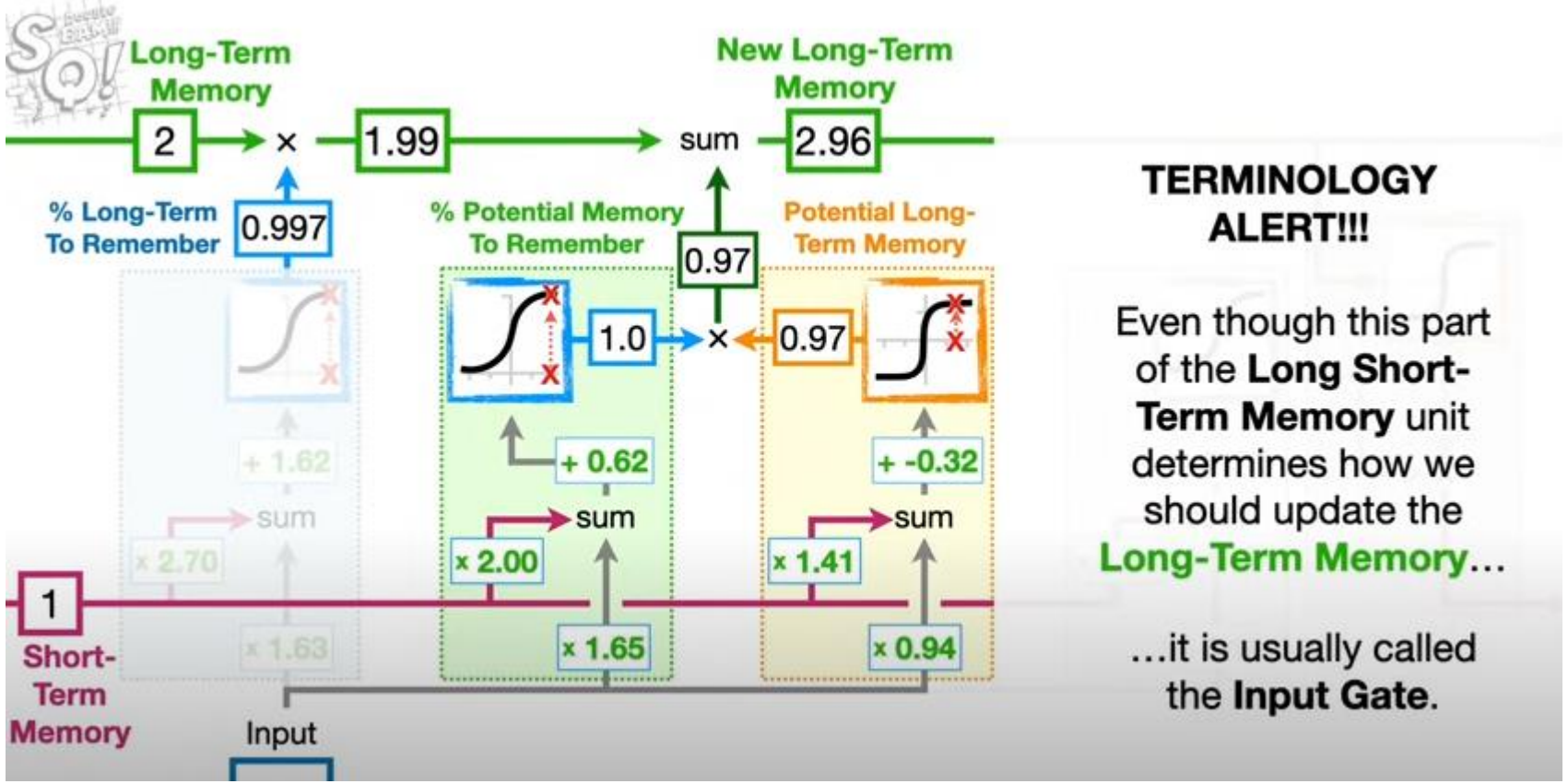
Even though this part of the **Long Short-Term Memory** unit determines what percentage of the **Long-Term Memory** will be remembered…

…it is usually called the **Forget Gate**.

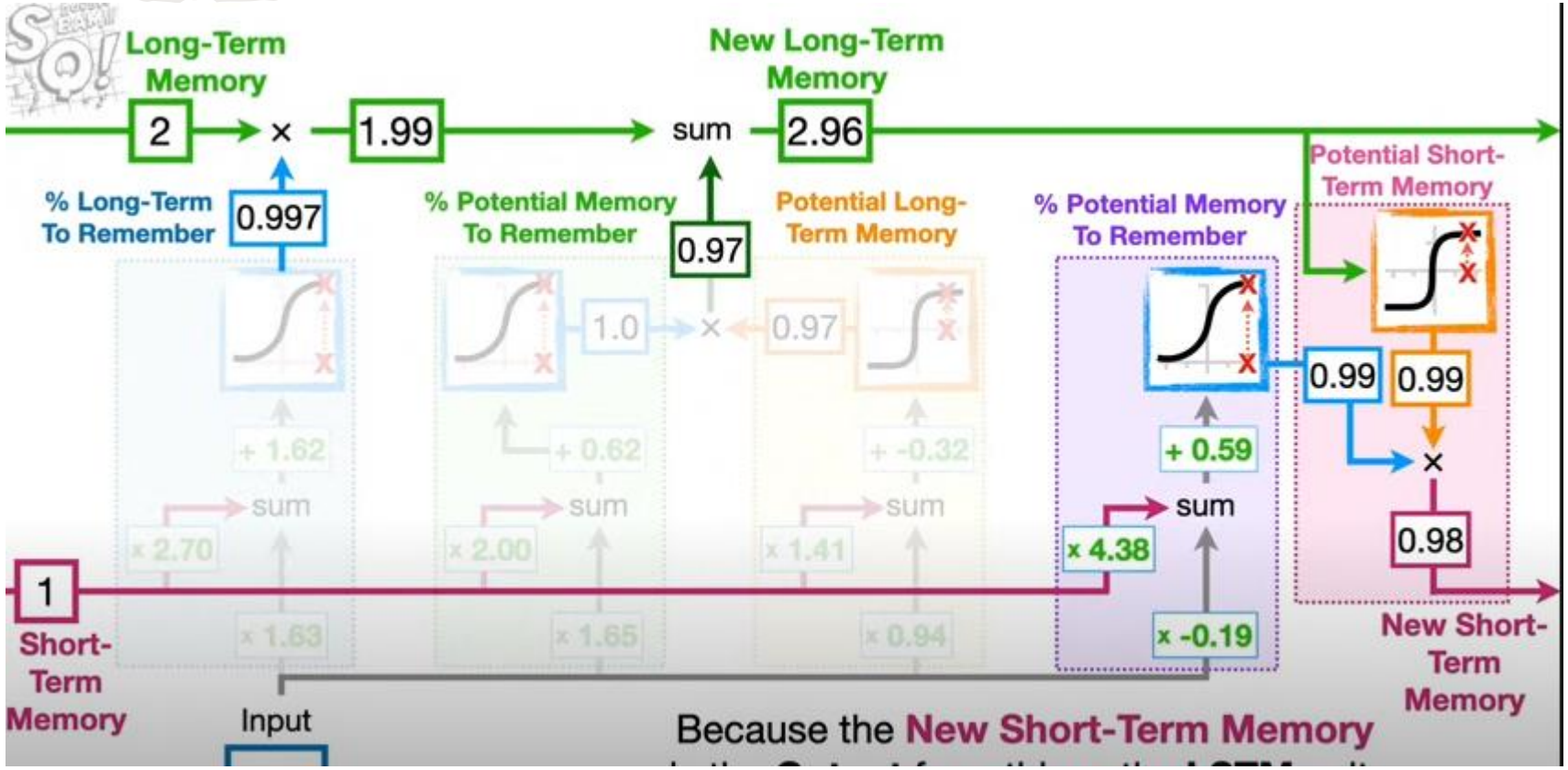
LSTM- Input gate



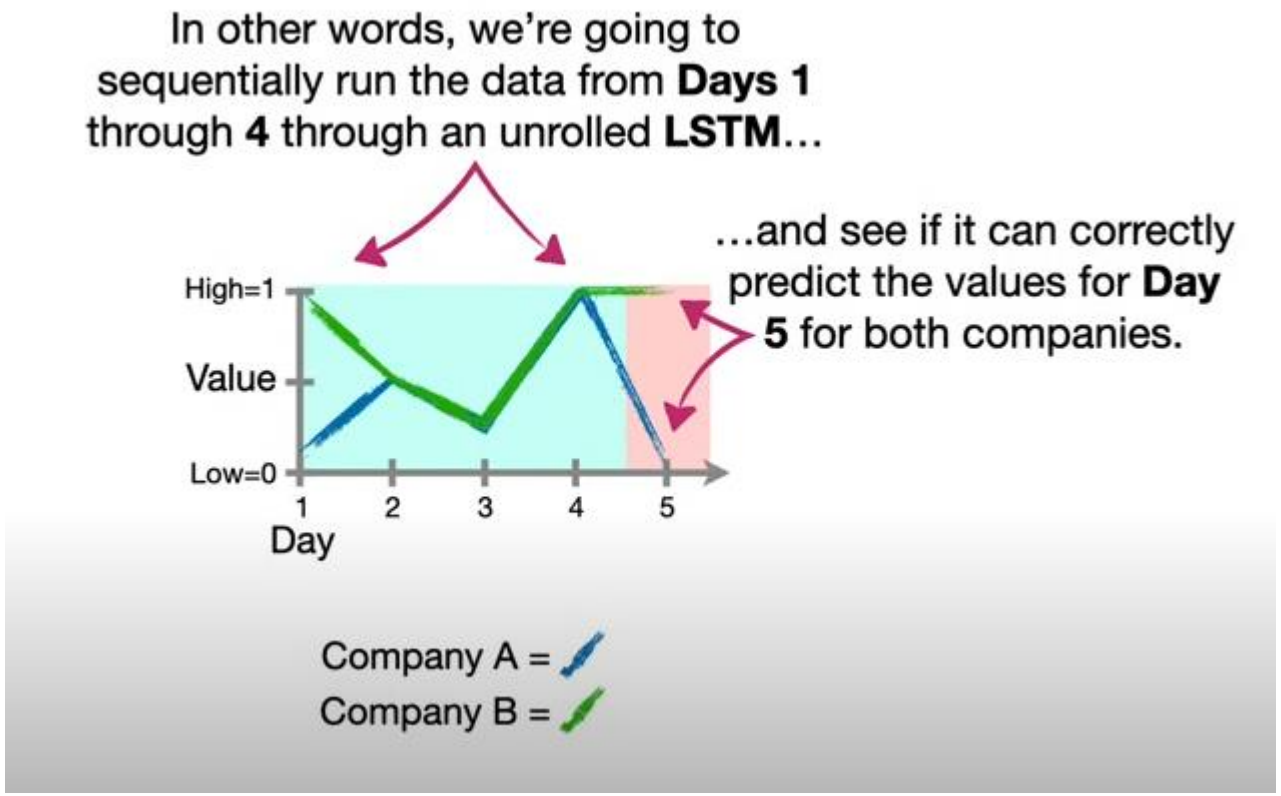
LSTM- Input gate



LSTM- Output gate

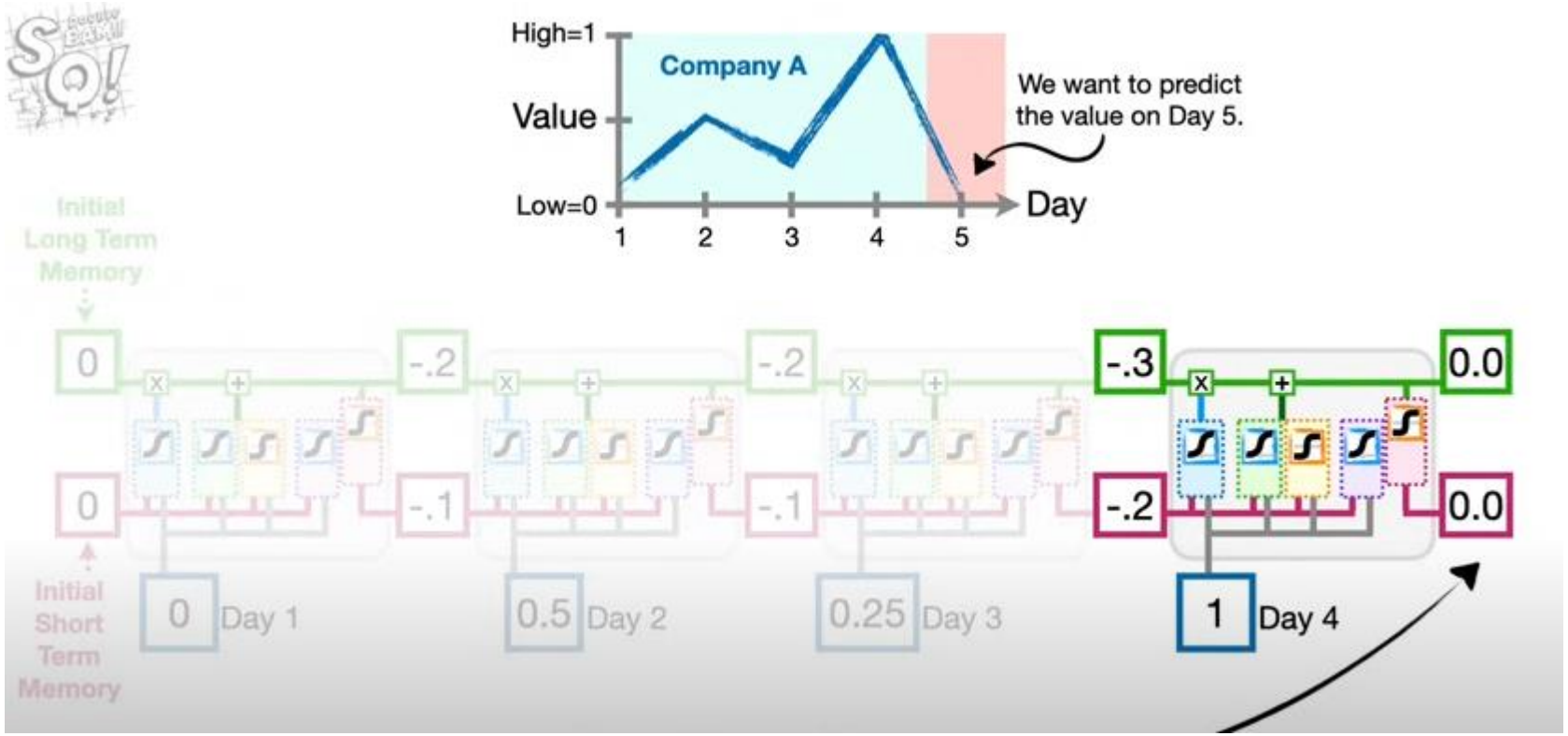


LSTM- Example

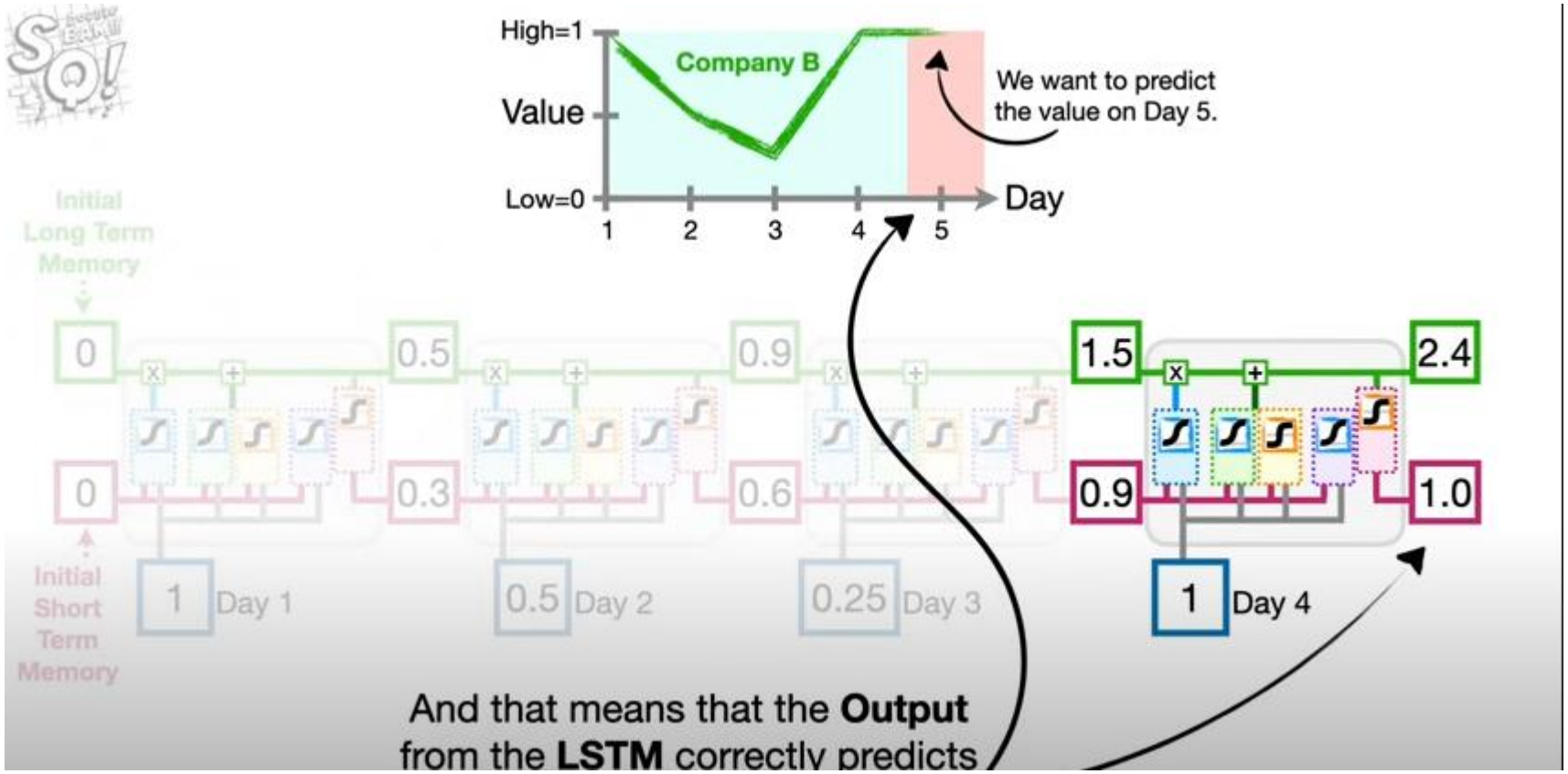


LSTM- Example

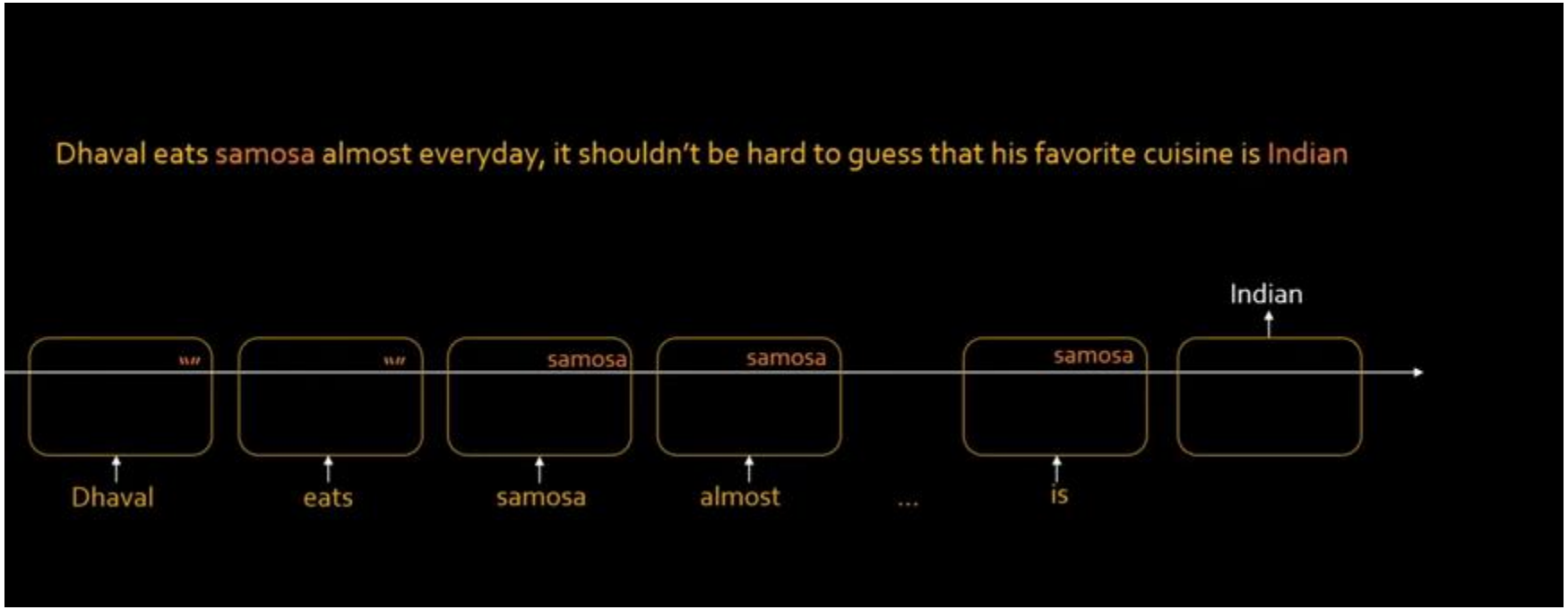
LSTM- Example(Company A)



LSTM- Example(Company B)

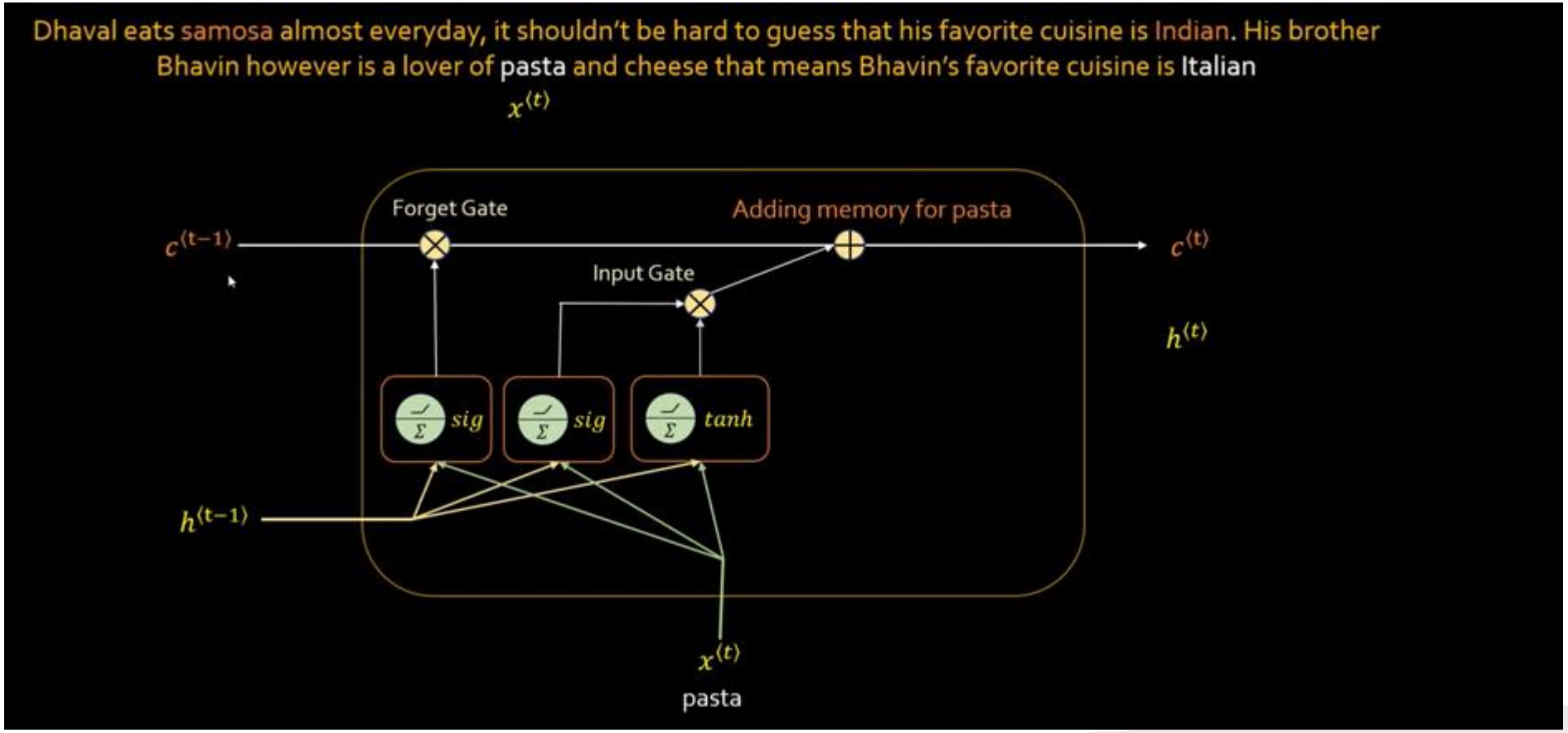


LSTM- Example 2



LSTM- Example 2

LSTM- Example 2



Reference

- https://colah.github.io/posts/2015-08-Understanding-LSTMs/
- https://www.youtube.com/watch?v=LfnrRPFhkuY
- https://www.youtube.com/watch?v=YCzL96nL7j0&t=5s