Long short-term memory (LSTMs)

LSTMs

Long Short-Term Memory

Today:

Deep Learning A-Z © SuperDataScience

Long Short-Term Memory

Today:

- A bit of history

Long Short-Term Memory

Today:

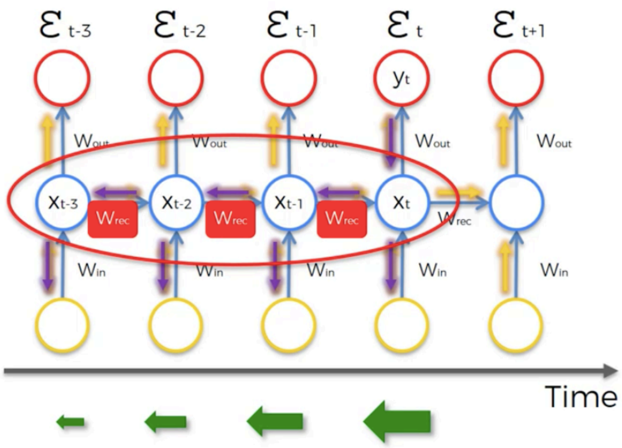
- A bit of history
- LSTM Architecture

Long Short-Term Memory

Today:

- A bit of history
- LSTM Architecture
- Example walkthrough

Long Short-Term Memory



$$\frac{\partial \mathcal{E}}{\partial \theta} = \sum_{1 \le t \le T} \frac{\partial \mathcal{E}_t}{\partial \theta} \quad (3)$$

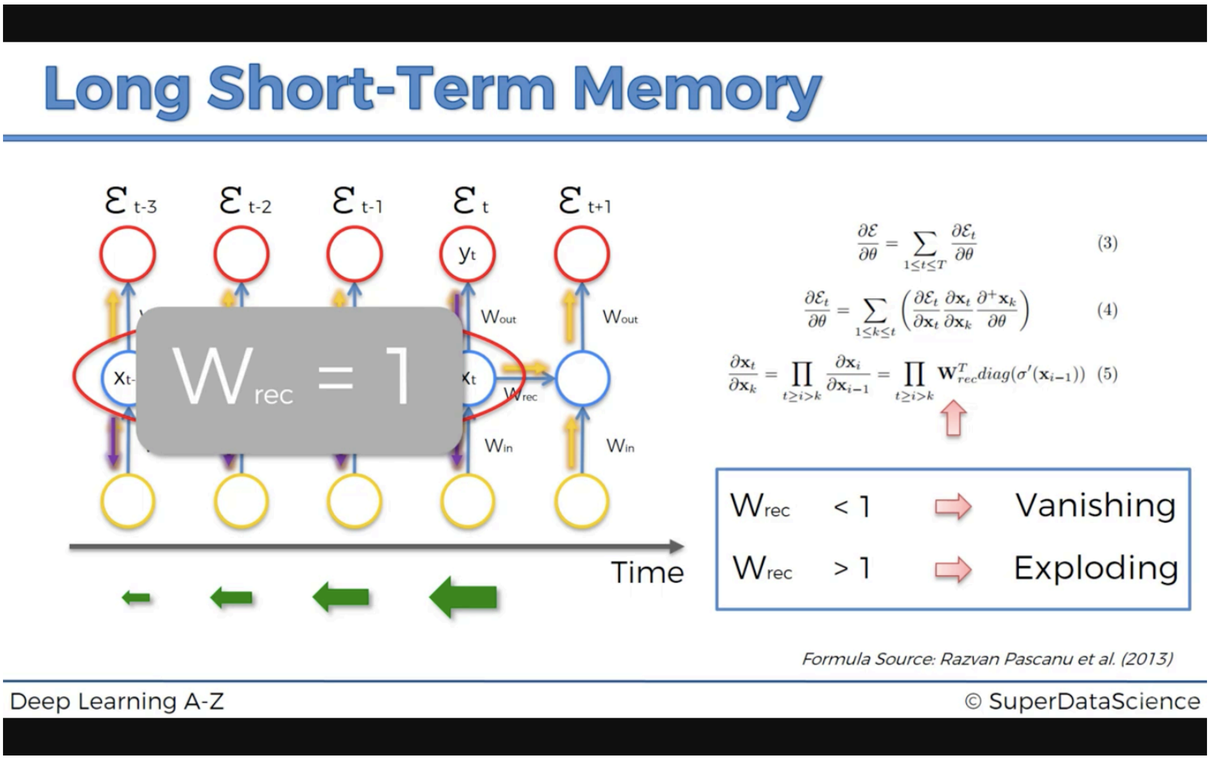
$$\frac{\partial \mathcal{E}_t}{\partial \theta} = \sum_{1 \le k \le t} \left(\frac{\partial \mathcal{E}_t}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} \frac{\partial^+ \mathbf{x}_k}{\partial \theta} \right) \quad (4)$$

$$\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} = \prod_{t \ge i > k} \frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_{i-1}} = \prod_{t \ge i > k} \mathbf{W}_{rec}^T diag(\sigma'(\mathbf{x}_{i-1})) \quad (5)$$

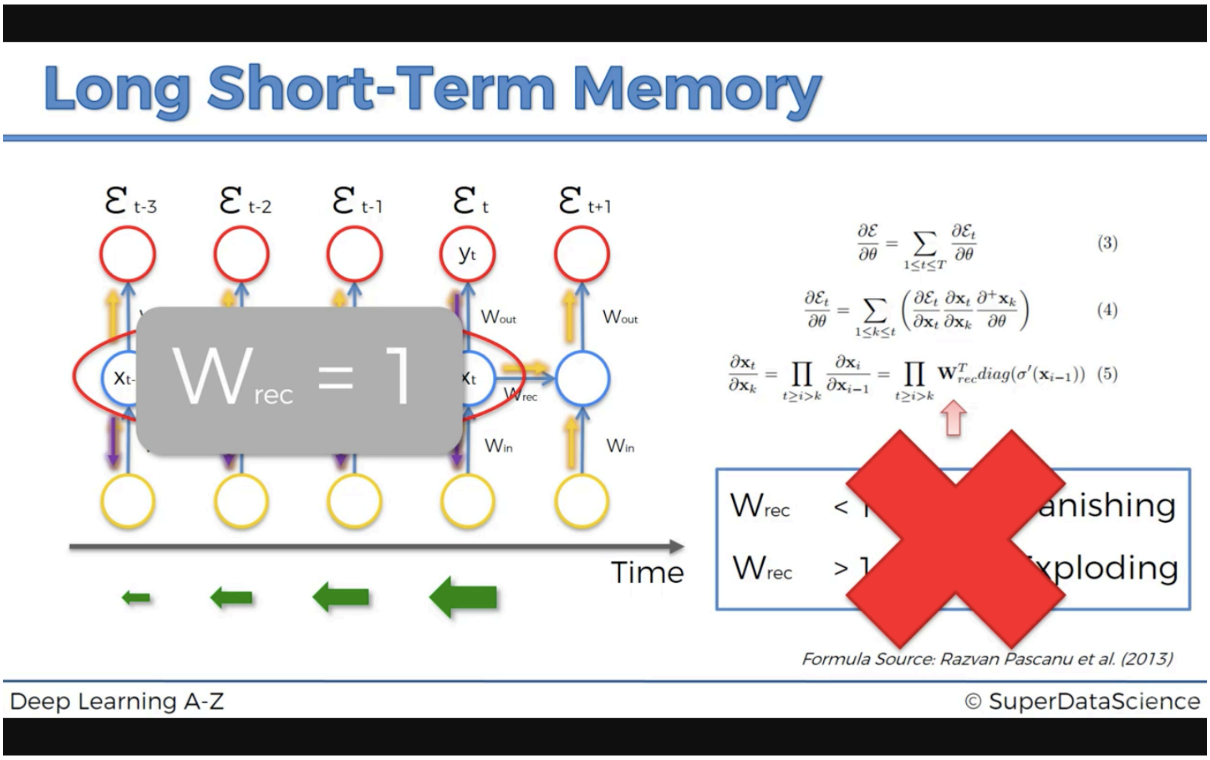
W_{rec} ~ small ⇨ Vanishing

W_{rec} ~ large ⇨ Exploding

Formula Source: Razvan Pascanu et al. (2013)



The first thing that comes to mind for solving this problem and that's exactly what has been done in LSTM.



Long Short-Term Memory

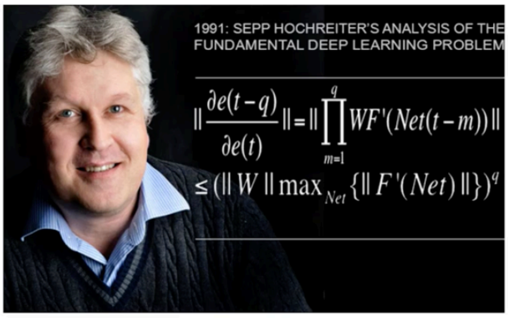
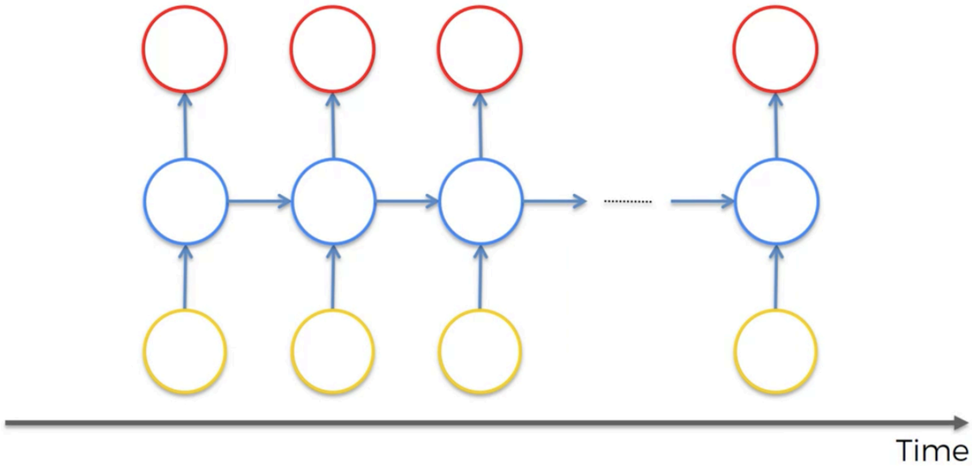


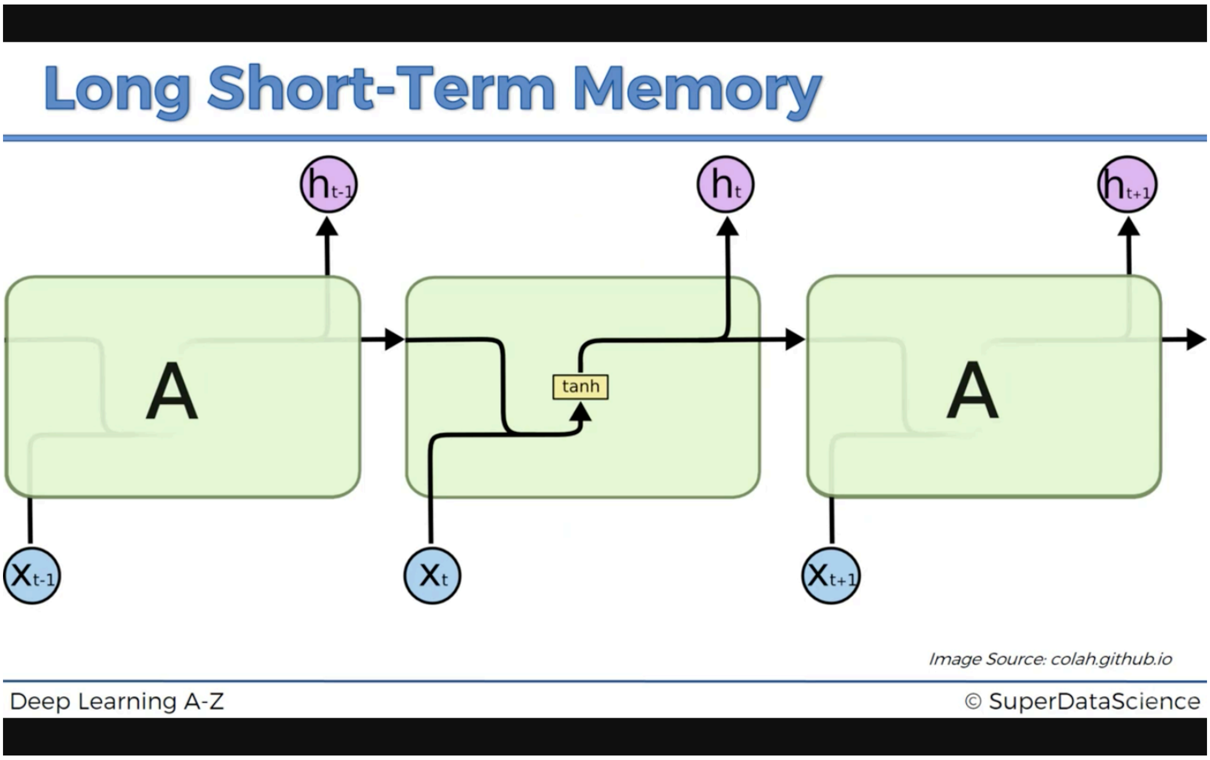
Image Sources: people.idsia.ch, ics.usi.ch

Long Short-Term Memory

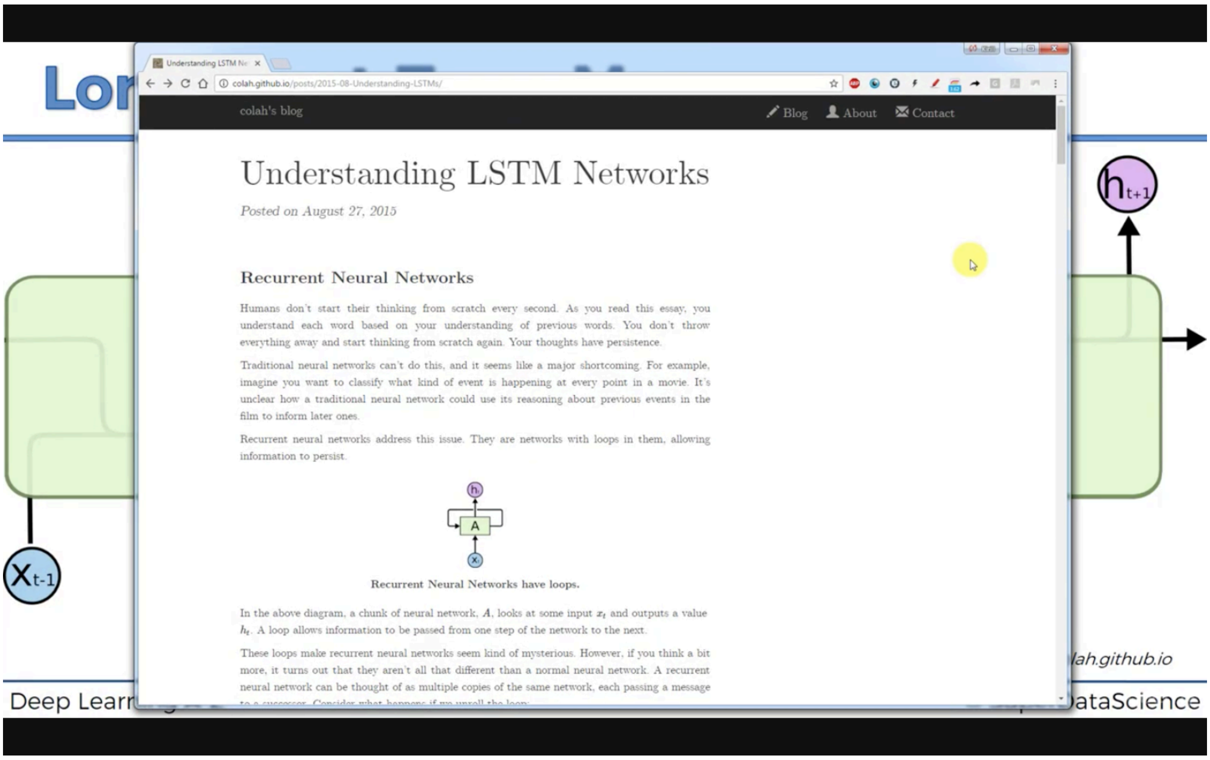


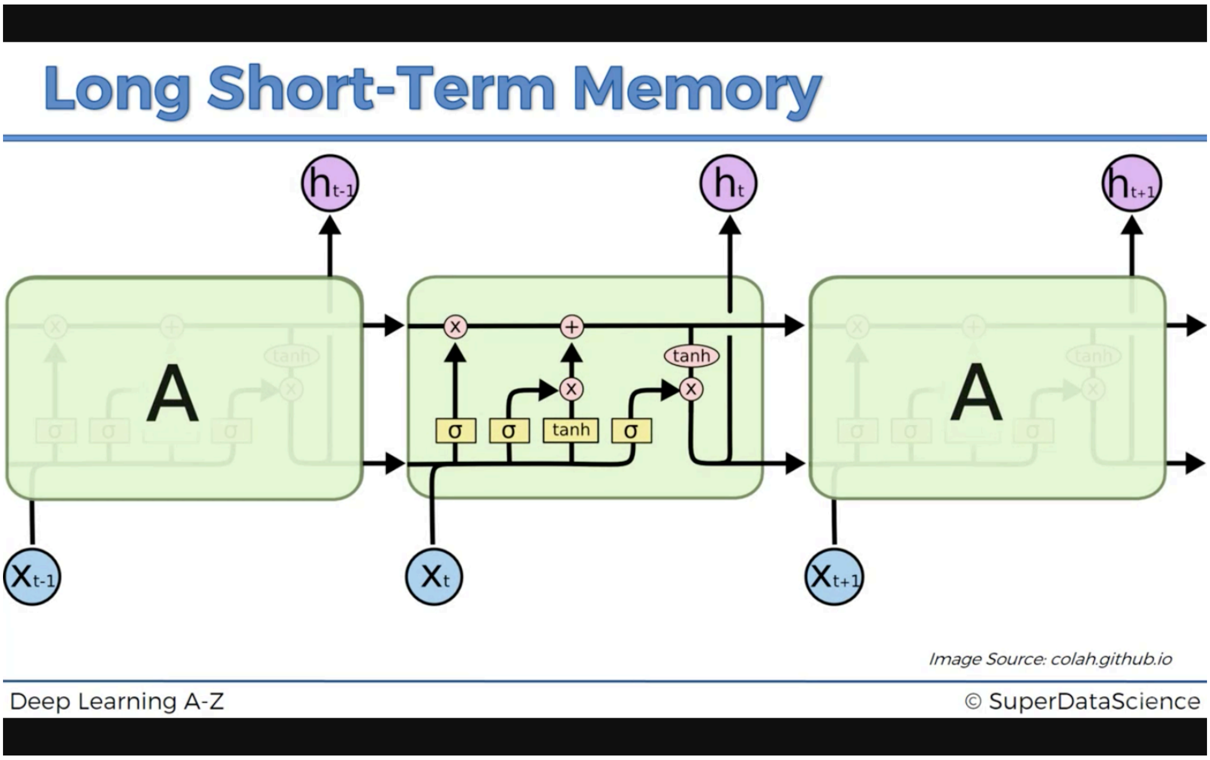
Deep Learning A-Z
© SuperDataScience

Recurrent neural network on ravelled temporal loop

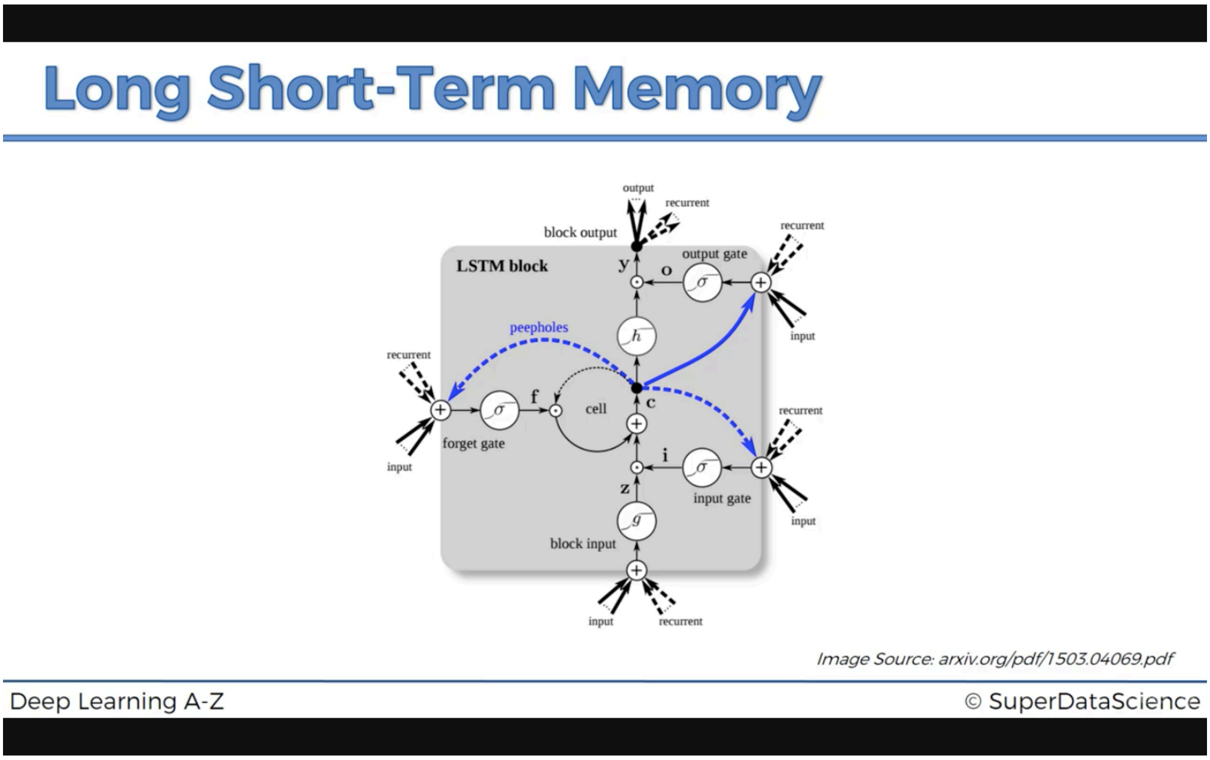


This is what it's look like if dig in inside the recurrent neural network.
This is standard RNN.

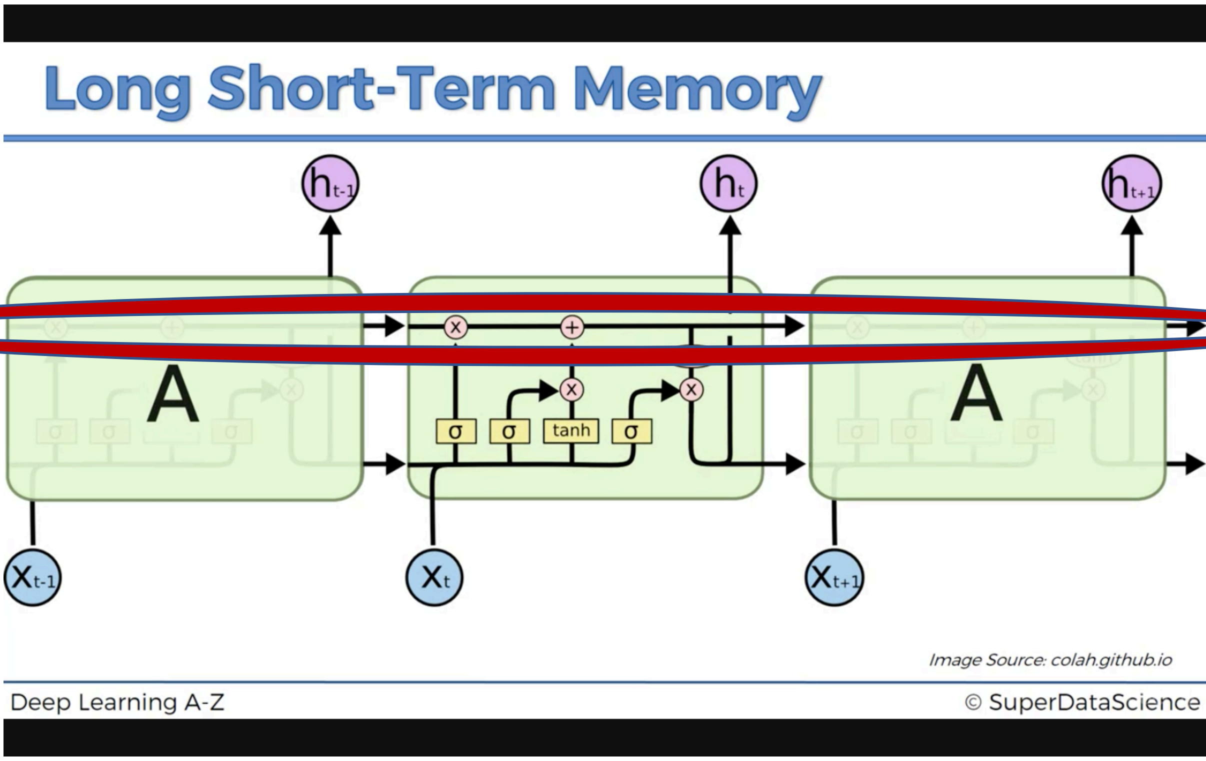




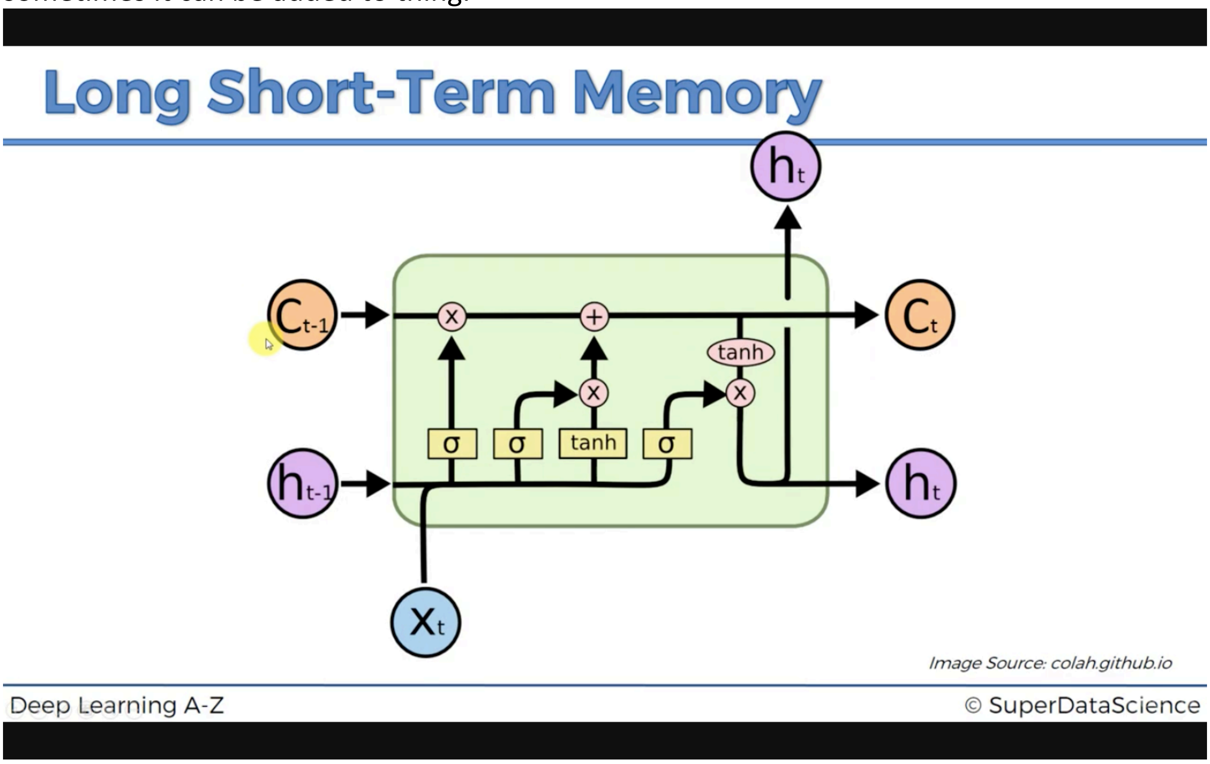
This LSRM version.



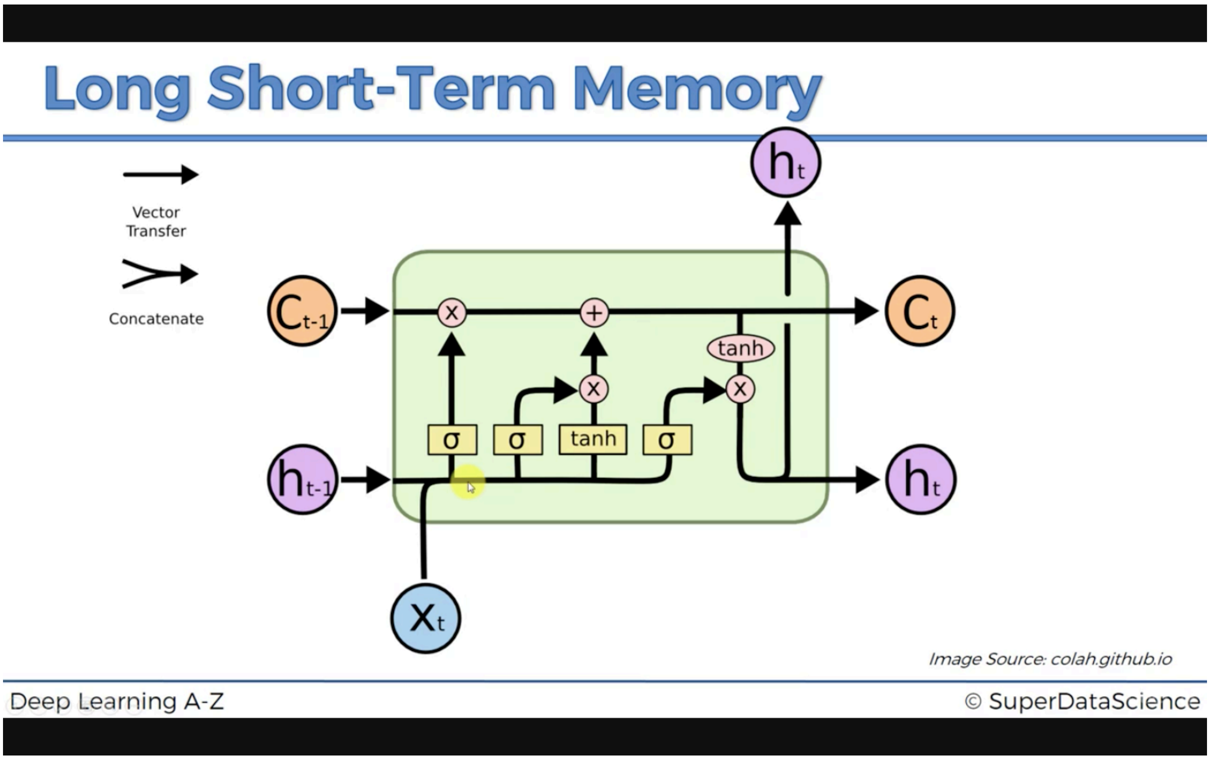
What usually shown in other sources.



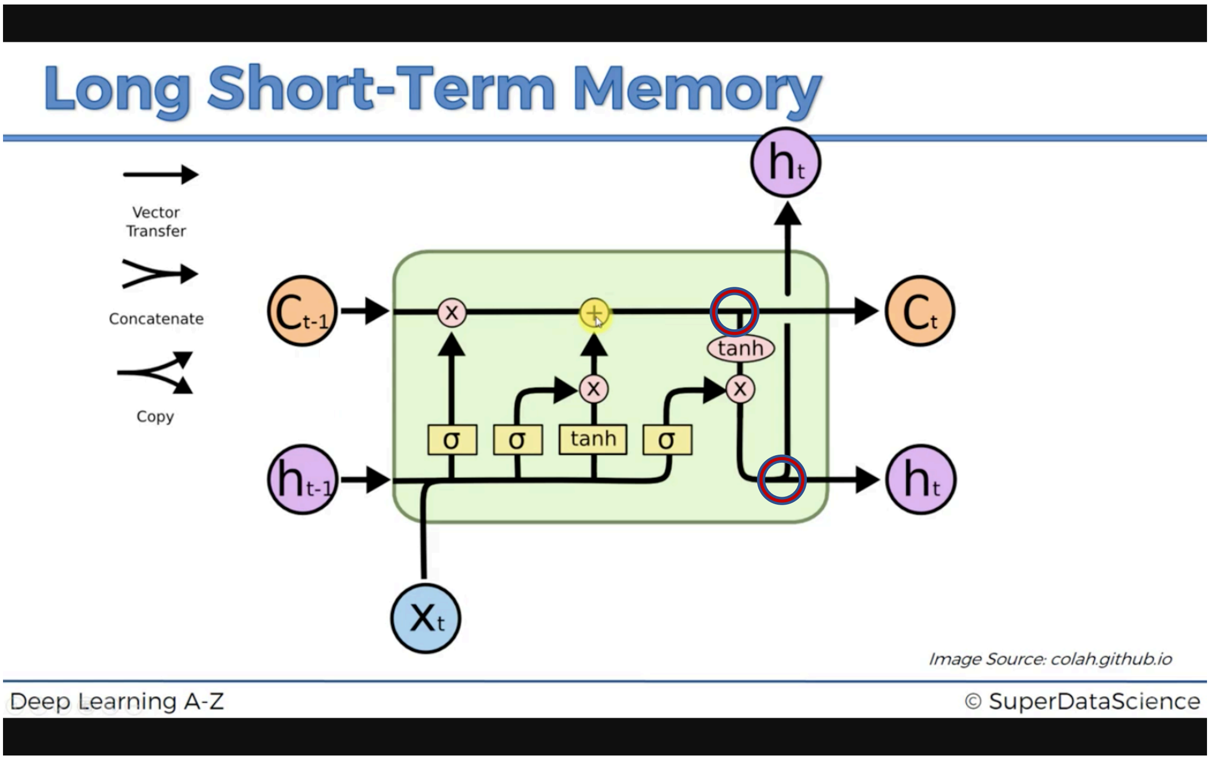
When we said W_{rec} is equal to 1, that this line over here. So LSRM has a memory cell (or unofficially memory pipeline) which as it goes through time, it can freely flow through time from left to right or in backpropagation right to left. Sometimes it can be removed (X sign) sometimes it can be added to thing.

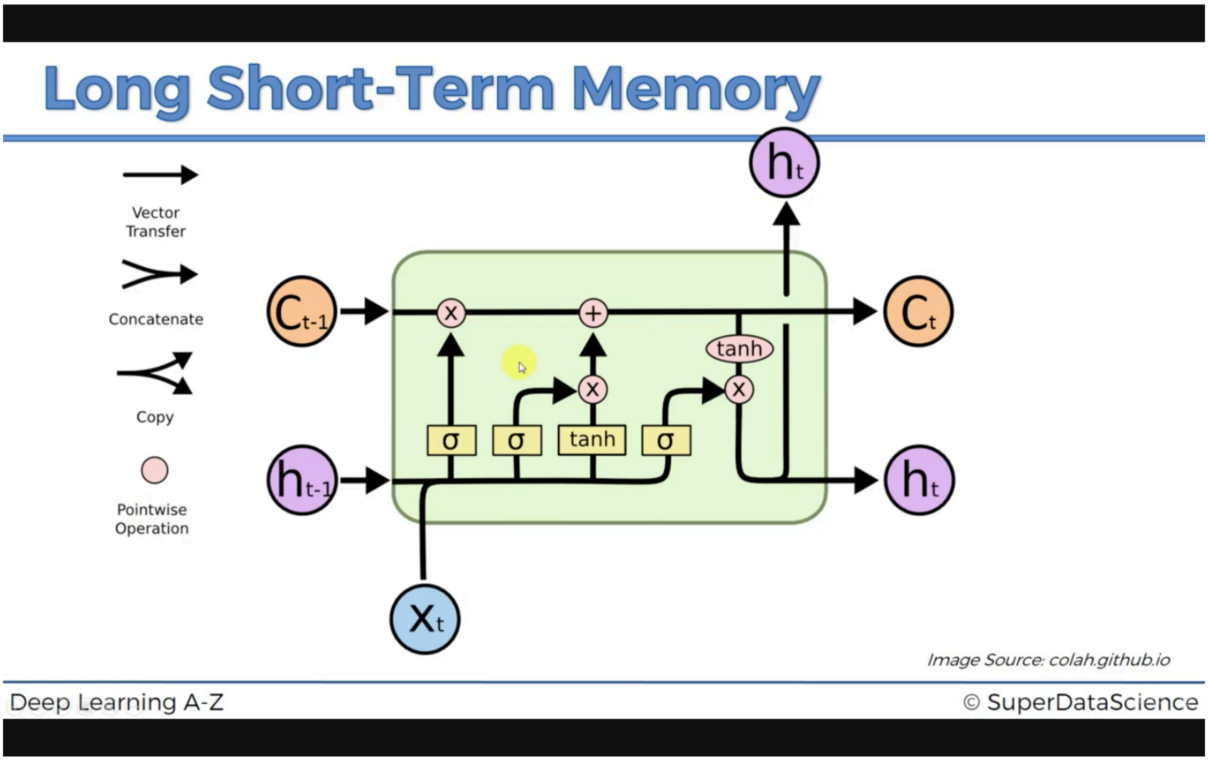


C stands for memory (cell). h is our output which is going to world (the upward) or it's going to the next modulo or block (the horizontals). X_t is our input.
This has 3 inputs and 2 outputs (because the h's are essentially the same).
Reminder: there is not just 5 neurons in here, there is lots of values behind. We refrene them as vectors

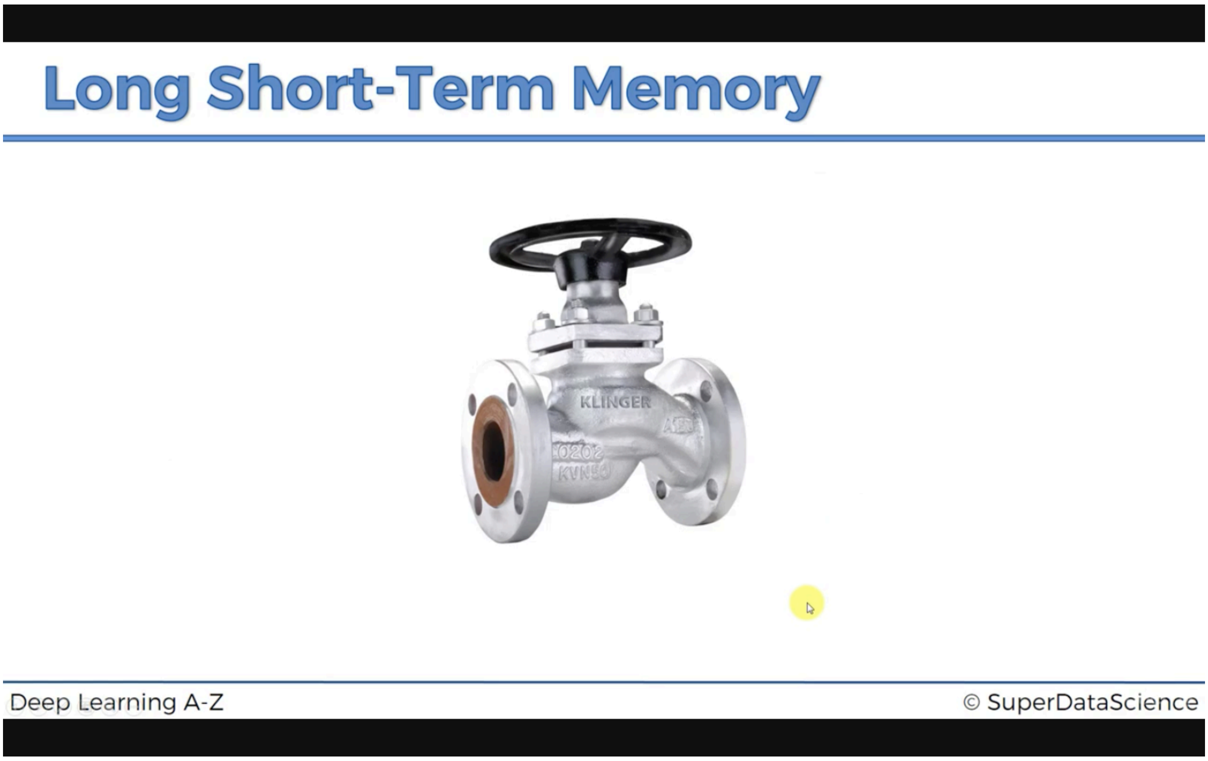


The concatenate: combining two lines in parallel

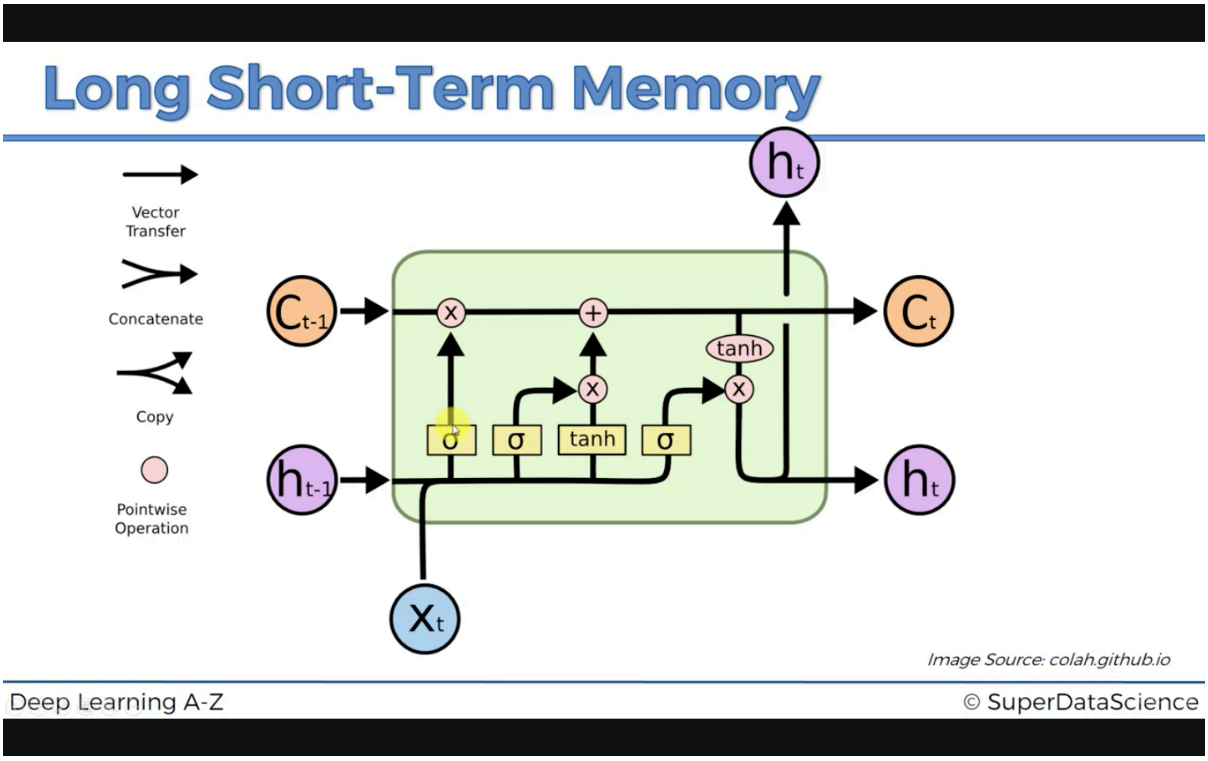




X's are Valves. X's from left to right: 1. Forget valve (f) 2. Memory valve (v) 3. Output valve (o)



This is valve



If you notice valves are controlled by sigma which it stands for sigmoid activation function. The reason for using sigmoid is because it is between 0 and 1. (+) is like a T-shaped pipe.

T-shaped pipe

Long Short-Term Memory

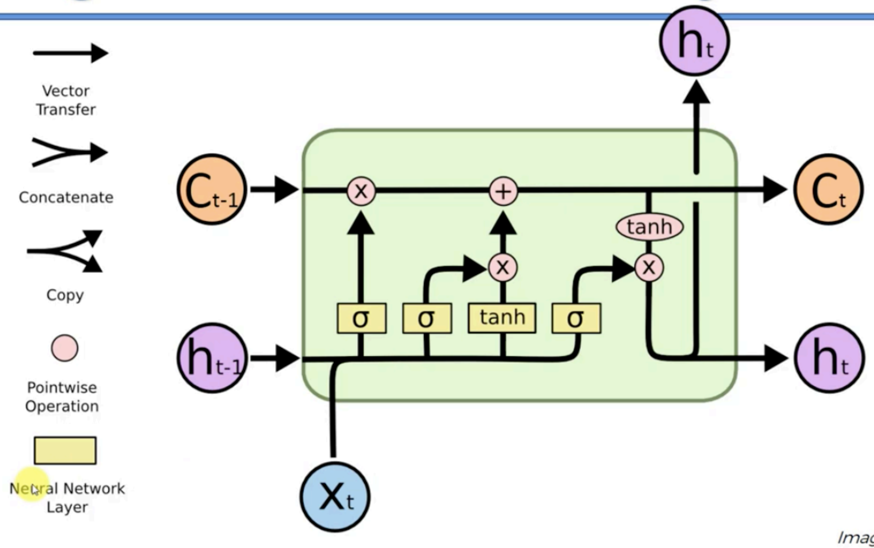


Image Source: colah.github.io

Long Short-Term Memory

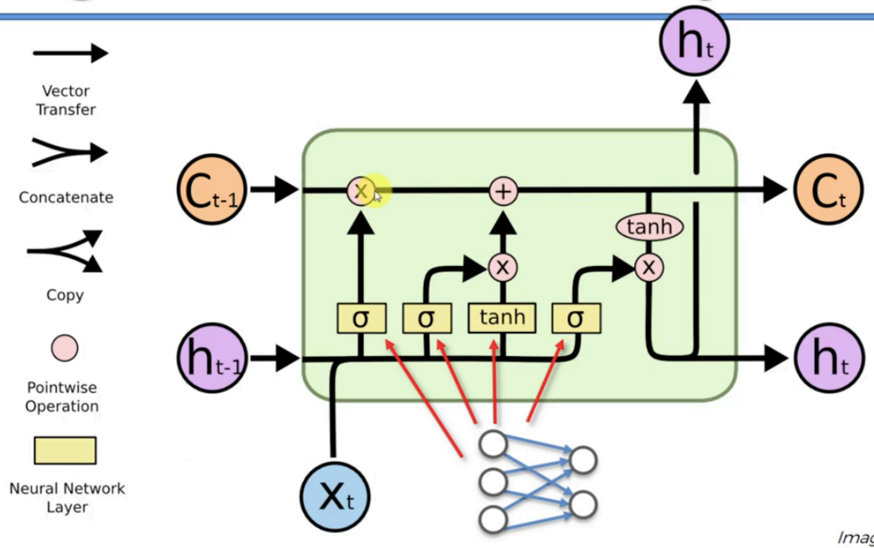
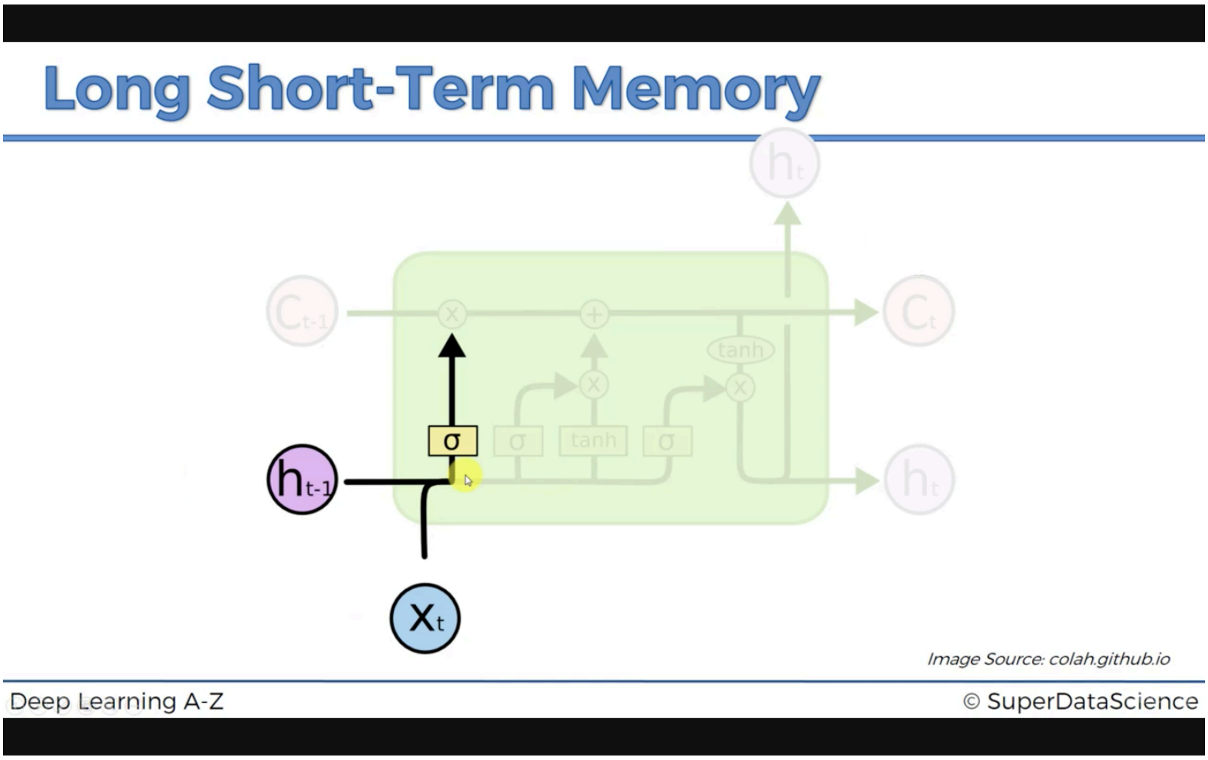
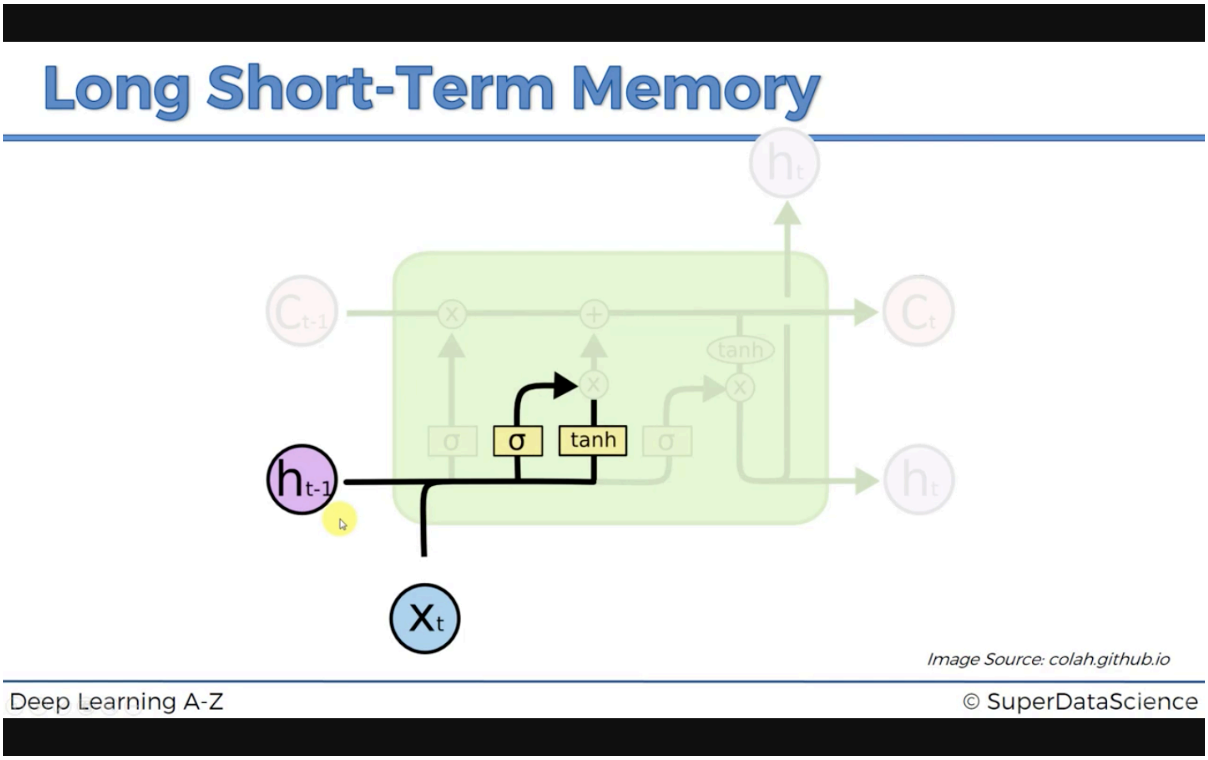


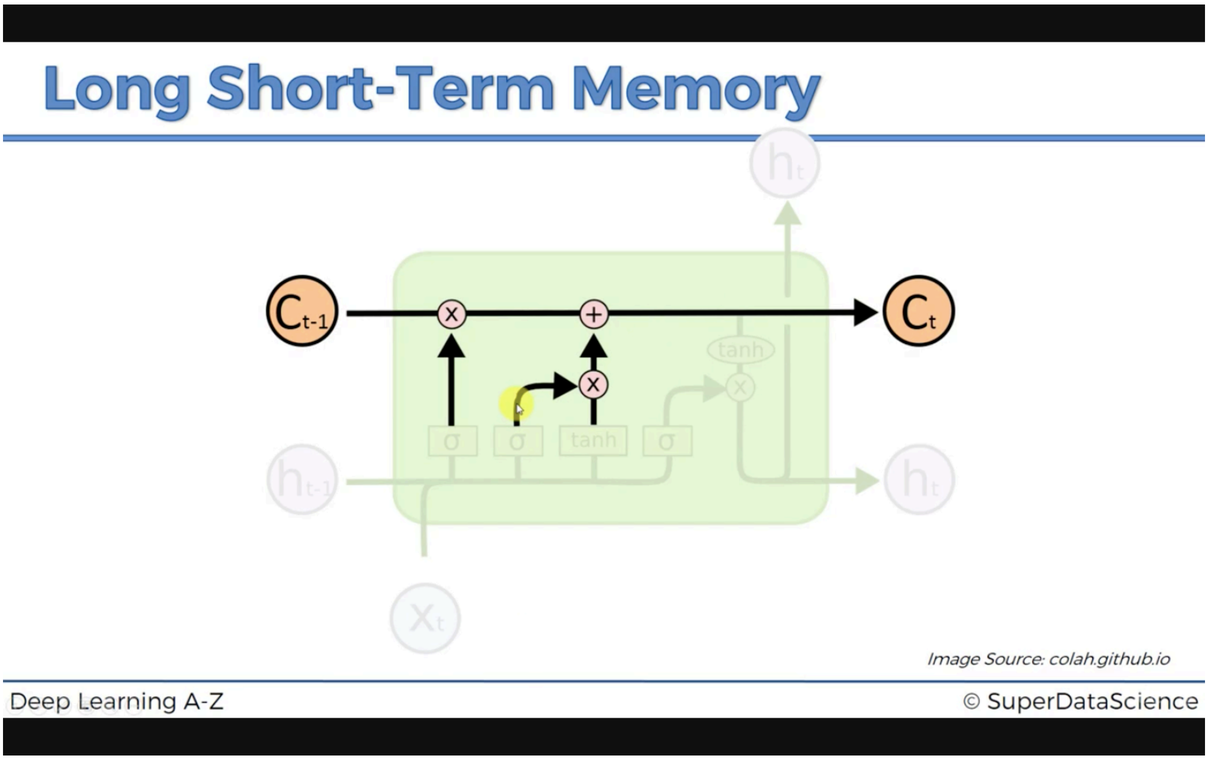
Image Source: colah.github.io

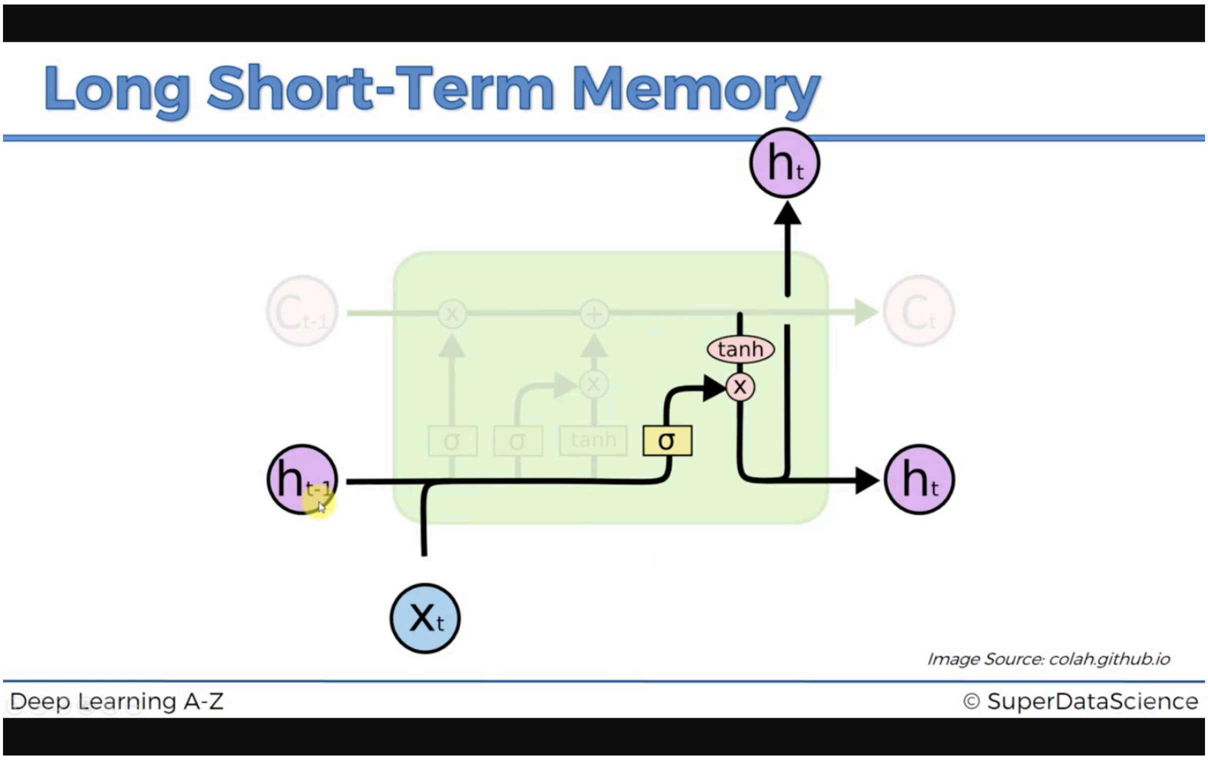


In here we got a value coming in and also a value coming from a previous node which together are combined to decide whether the valve (X) should go ahead or not.



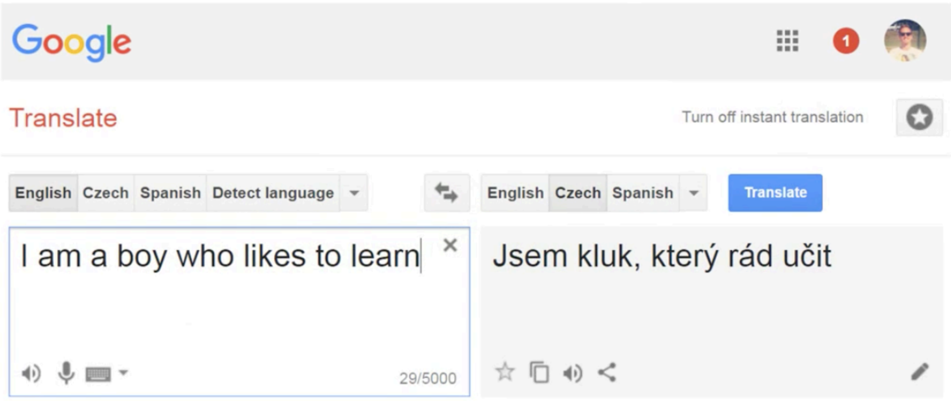
These two are combine again to decide first which value pass through and also if that value should pass through or not.





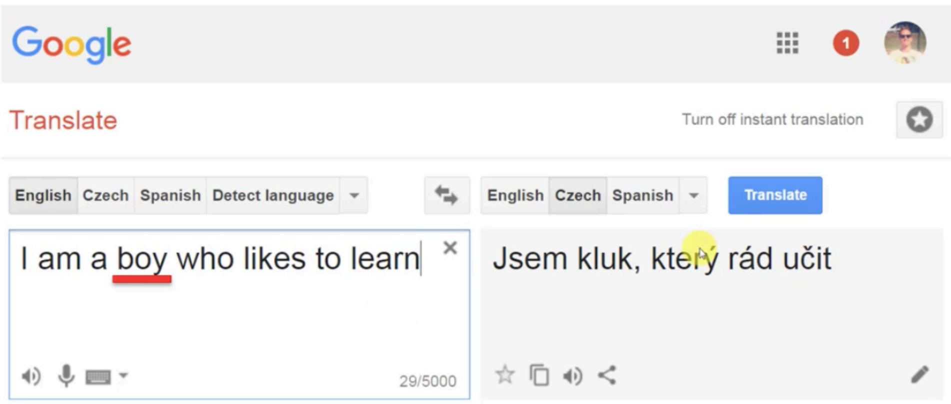
And finally, these two are combines to decide what part of memory cell or pipeline is going to be output.

Long Short-Term Memory



Deep Learning A-Z © SuperDataScience

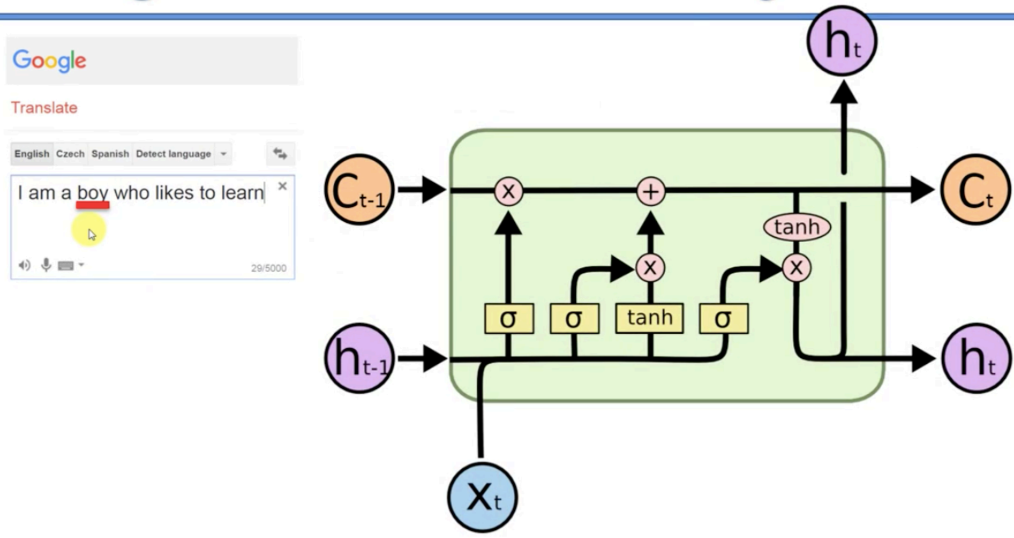
Long Short-Term Memory



Deep Learning A-Z © SuperDataScience

In here if we change the boy to girl, the structure of sentence will change.

Long Short-Term Memory



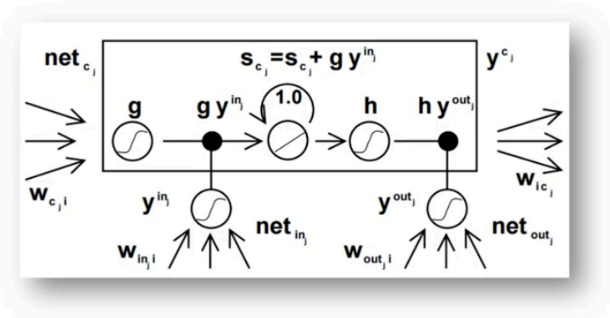
Let’s say boy go through the C freely. If we add a girl or a name like Amanda, it comes from X. then we close the first valve we had. Then open the second valve then put the subject on memory cell and also, we extract more information out of it (for example if it’s feminine, singular, capitalized, etc.). then it flows through the memory cell and on the third valve, it will extract the information of its gender so it goes as a input to the next modulo to make the necessary changes in there.

Long Short-Term Memory

Additional Reading:

Long Short-Term Memory

By Sepp Hochreiter & Jurgen Schmidhuber (1997)



Link:

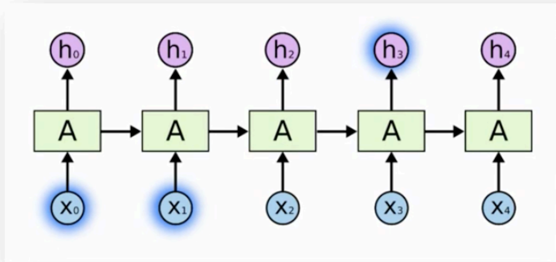
http://www.bioinf.jku.at/publications/older/2604.pdf

Long Short-Term Memory

Additional Reading:

Understanding LSTM Networks

By Christopher Olah (2015)



Link:

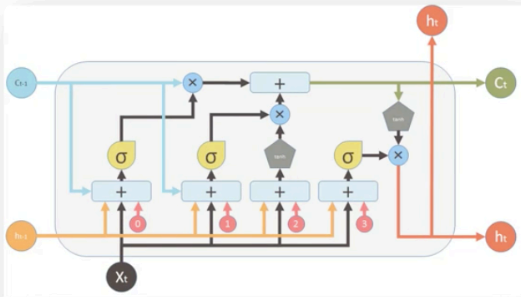
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Long Short-Term Memory

Additional Reading:

Understanding LSTM and its diagrams

By Shi Yan (2016)



Link:

https://medium.com/@shiyan/understanding-lstm-and-its-diagrams-37e2f46f1714

Deep Learning A-Z © SuperDataScience