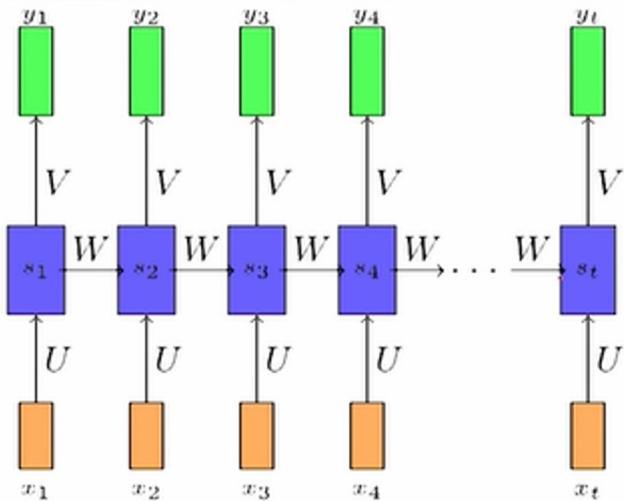
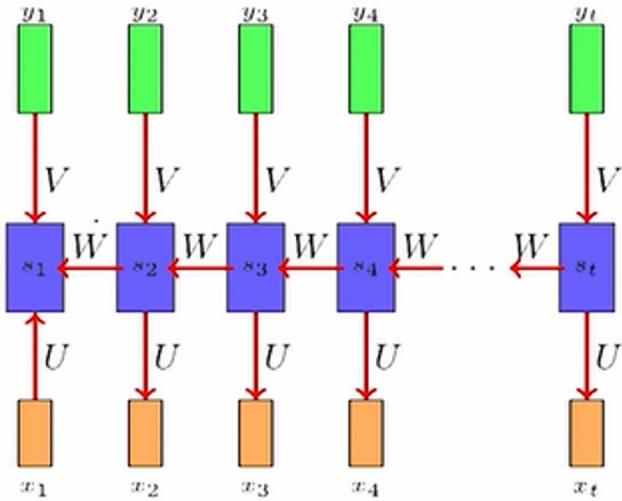


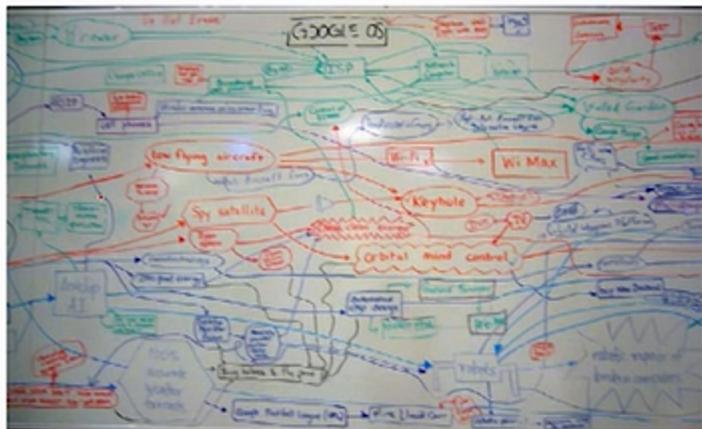
Selective Read, Selective Write, Selective Forget



- The state (s_i) of an RNN records information from all previous time steps
- At each new timestep the old information gets morphed by the current input
- One could imagine that after t steps the information stored at time step $t - k$ (for some $k < t$) gets completely morphed
so much that it would be impossible to extract the original information stored at time step $t - k$



- A similar problem occurs when the information flows backwards (backpropagation)
- It is very hard to assign the responsibility of the error caused at time step t to the events that occurred at time step $t - k$
- This responsibility is of course in the form of gradients and we studied the problem in backward flow of gradients
- We saw a formal argument for this while discussing vanishing gradients



- Let us see an analogy for this
 - We can think of the state as a fixed size memory
 - Compare this to a fixed size white board that you use to record information
 - At each time step (periodic intervals) we keep writing something to the board
 - Effectively at each time step we morph the information recorded till that time point
 - After many timesteps it would be impossible to see how the information at time step $t - k$ contributed to the state at timestep t



- Continuing our whiteboard analogy, suppose we are interested in deriving an expression on the whiteboard
 - We follow the following strategy at each time step
 - Selectively write on the board
 - Selectively read the already written content
 - Selectively forget (erase) some content
 - Let us look at each of these in detail

$$a = 1 \quad b = 3 \quad c = 5 \quad d = 11$$

Compute $ac(bd + a) + ad$

Say “board” can have only 3 statements at a time.

- ❶ ac
- ❷ bd
- ❸ $bd + a$
- ❹ $ac(bd + a)$
- ❺ ad
- ❻ $ac(bd + a) + ad$

$$ac = 5$$

$$bd = 33$$

Selective write

- There may be many steps in the derivation but we may just skip a few
- In other words we select what to write

$$a = 1 \quad b = 3 \quad c = 5 \quad d = 11$$

Compute $ac(bd + a) + ad$

Say “board” can have only 3 statements at a time.

- ① ac
- ② bd
- ③ $bd + a$
-
- ④ $ac(bd + a)$
- ⑤ ad
- ⑥ $ac(bd + a) + ad$

Selective read

- While writing one step we typically read some of the previous steps we have already written and then decide what to write next
- For example at Step 3, information from Step 2 is important
- In other words we select what to **read**

$$ac = 5$$

$$bd = 33$$

$$bd + a = 34$$

$$a = 1 \quad b = 3 \quad c = 5 \quad d = 11$$

Compute $ac(bd + a) + ad$

Say “board” can have only 3 statements at a time.

- ➊ ac
- ➋ bd
- ➌ $bd + a$
- ➍ $ac(bd + a)$
- ➎ ad
- ➏ $ac(bd + a) + ad$

Selective forget

- Once the board is full, we need to delete some obsolete information
- But how do we decide what to delete? We will typically delete the least useful information
- In other words we select what to **forget**

$$ac = 5$$

$$ac(bd + a) = 170$$

$$bd + a = 34$$

$$a = 1 \quad b = 3 \quad c = 5 \quad d = 11$$

Compute $ac(bd + a) + ad$

Say “board” can have only 3 statements at a time.

- ➊ ac
- ➋ bd
- ➌ $bd + a$
- ➍ $ac(bd + a)$
- ➎ ad
- ➏ $ac(bd + a) + ad$

$$ad + ac(bd + a) = 181$$

$$ac(bd + a) = 170$$

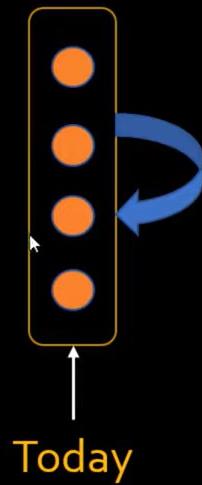
$$ad = 11$$

- There are various other scenarios where we can motivate the need for selective write, read and forget
- For example, you could think of our brain as something which can store only a finite number of facts
- At different time steps we selectively read, write and forget some of these facts
- Since the RNN also has a finite state size, we need to figure out a way to allow it to selectively read, write and forget

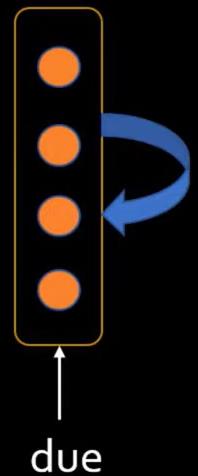
Today, due to my current job situation and family conditions, I need to take a loan.

Last year, due to my current job situation and family conditions, I had to take a loan.

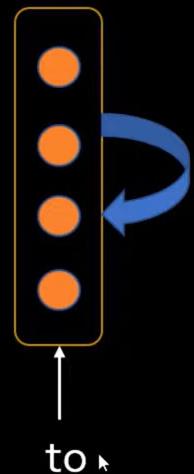
Today, due to my current job situation and family conditions, I need to take a loan.



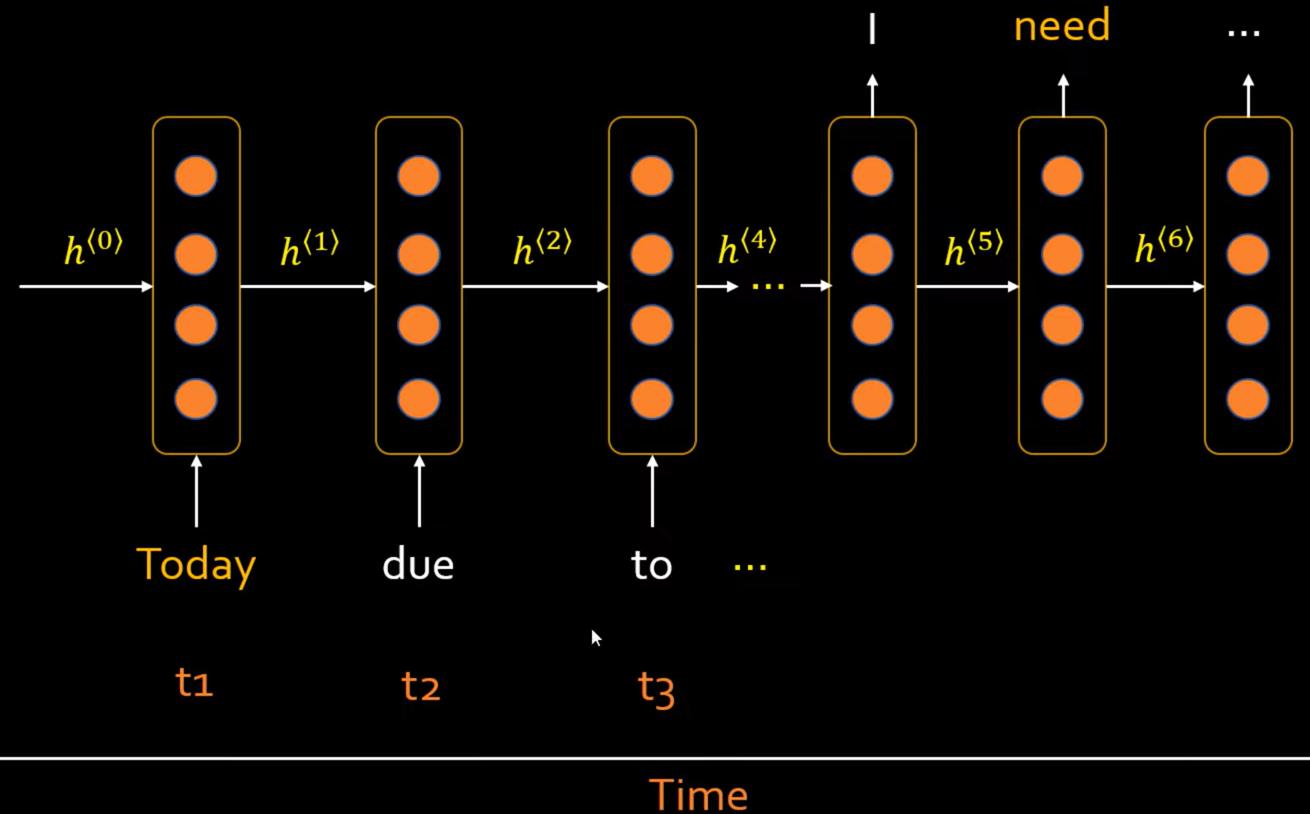
Today, due to my current job situation and family conditions, I need to take a loan.



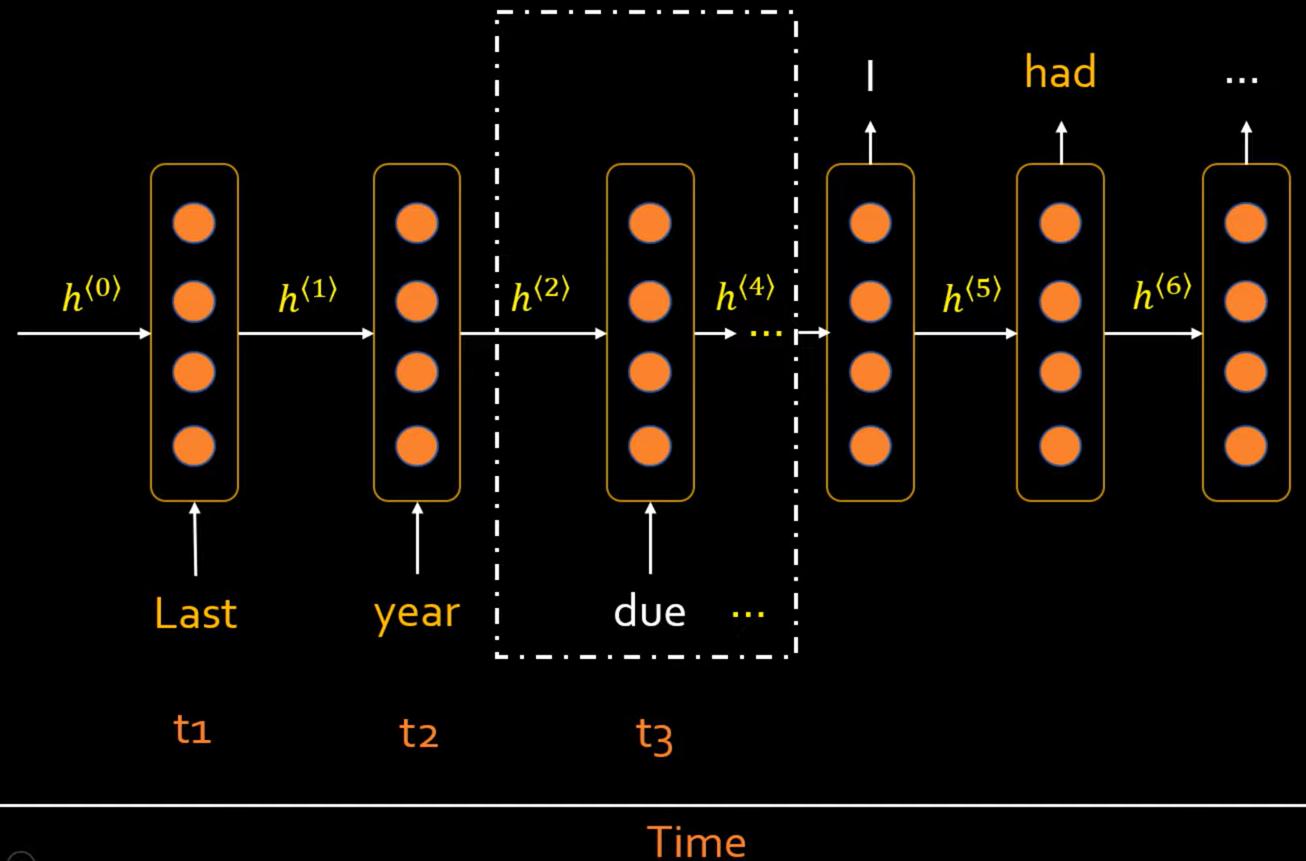
Today, due to my current job situation and family conditions, I need to take a loan.

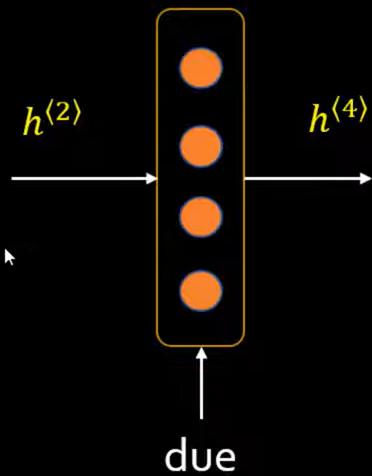


Today, due to my current job situation and family conditions, I need to take a loan.

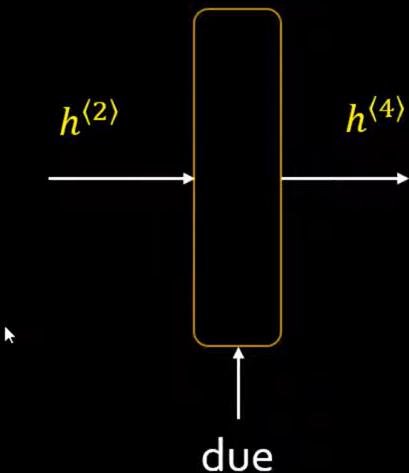


Last year, due to my current job situation and family conditions, I had to take a loan.

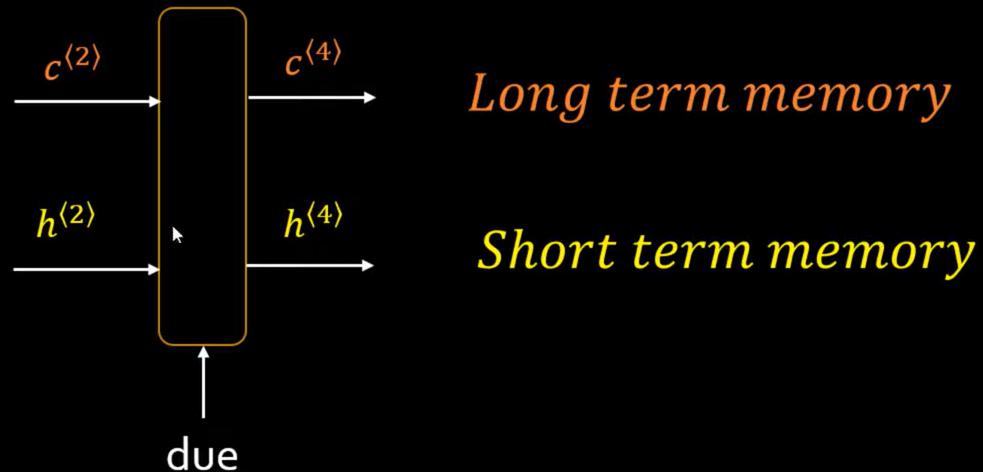




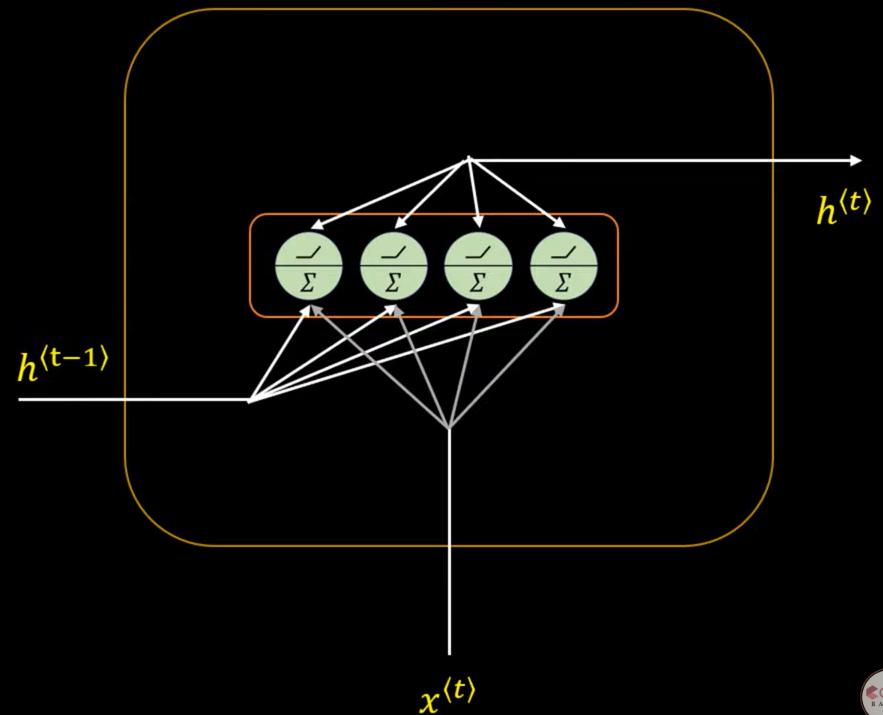
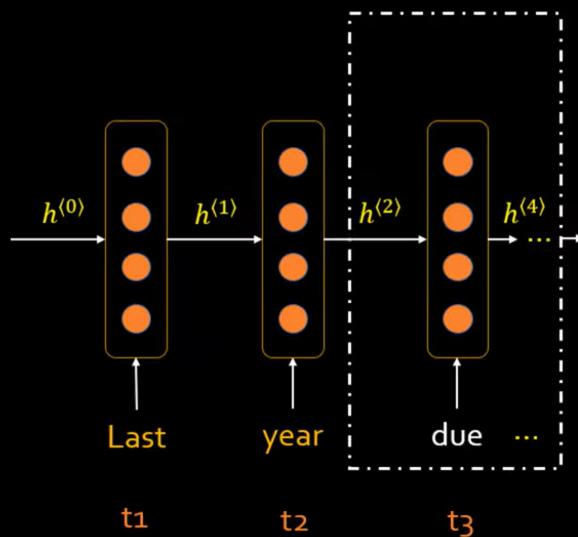
Memory cell



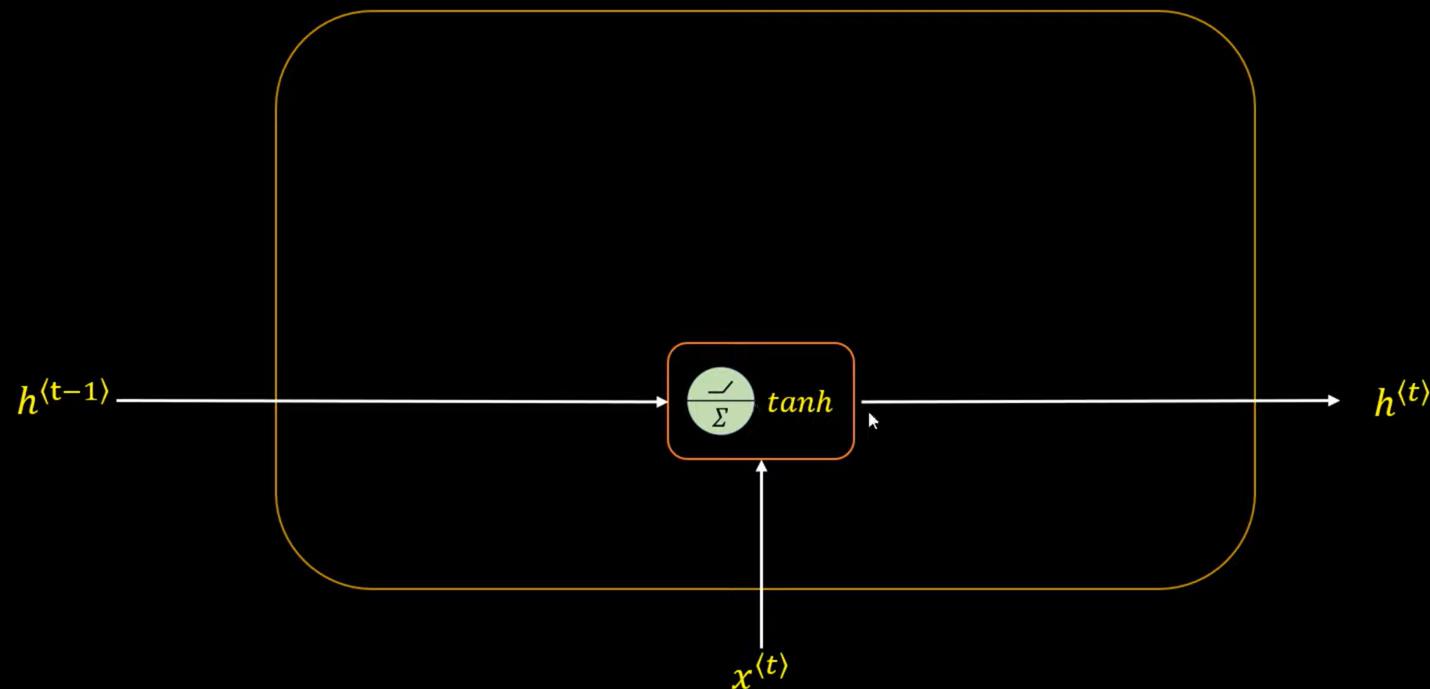
Memory cell



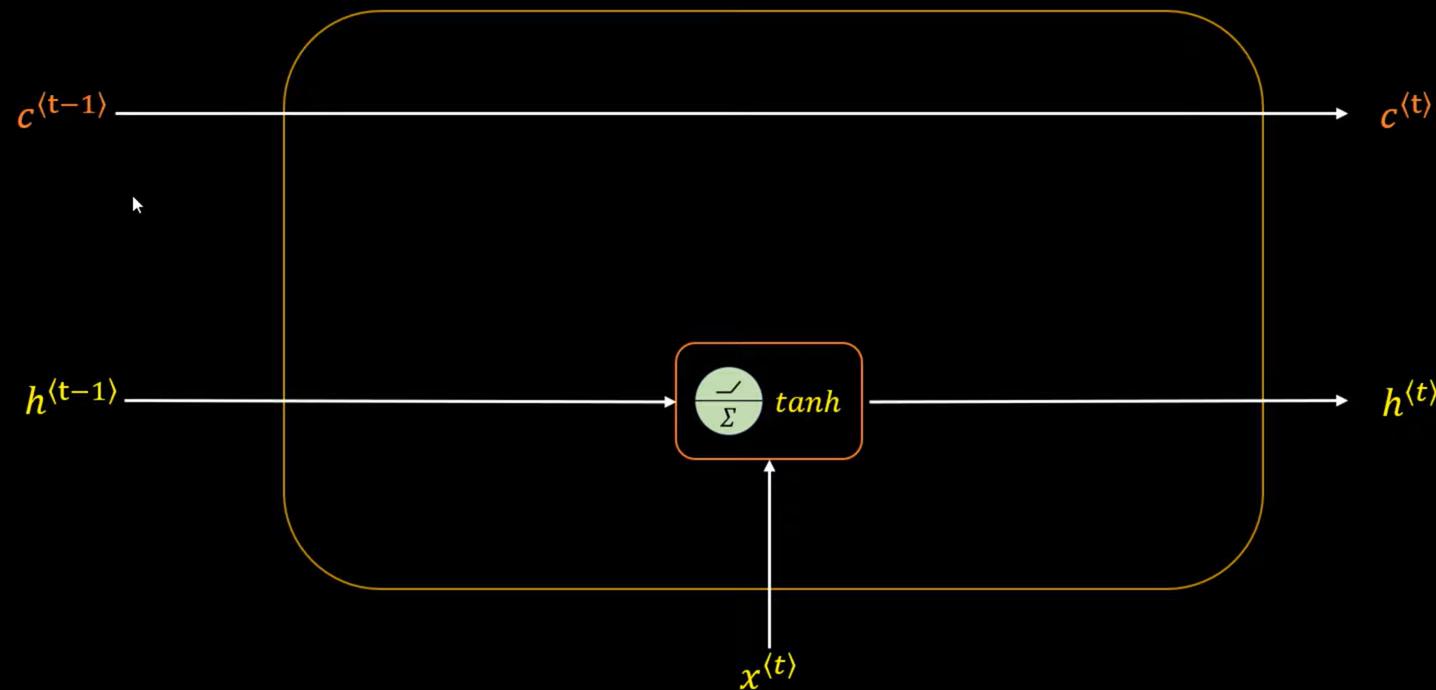
Short term memory cell in traditional RNN



Short term memory cell in traditional RNN



Short term memory and long term memory



Dhaval eats samosa almost everyday, it shouldn't be hard to guess
that his favorite cuisine is ...

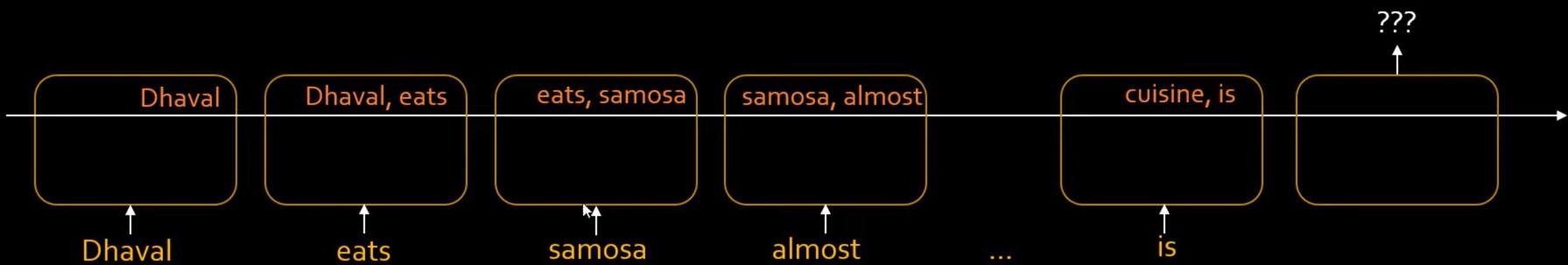


Dhaval eats samosa almost everyday, it shouldn't be hard to guess
that his favorite cuisine is Indian

Dhaval eats samosa almost everyday, it shouldn't be hard to guess
that his favorite cuisine is Indian

Dhaval eats samosa almost everyday, it shouldn't be hard to guess that his favorite cuisine is Indian

Traditional RNN

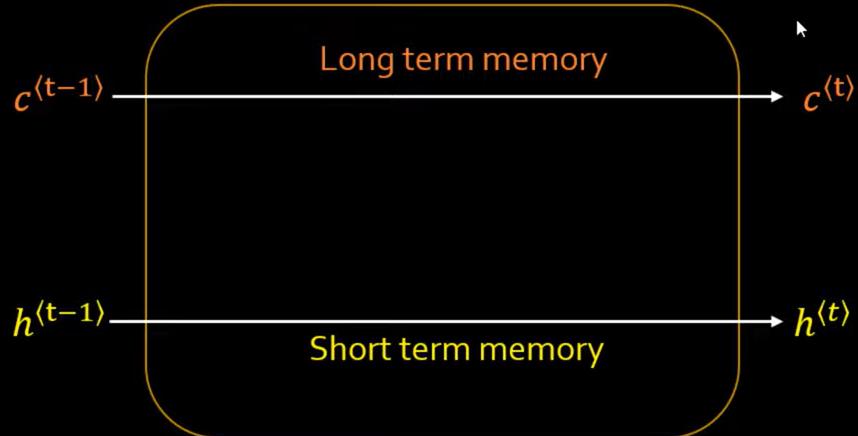


Dhaval eats samosa almost everyday, it shouldn't be hard to guess that his favorite cuisine is Indian. His brother Bhavin however is a lover of pastas and cheese that means Bhavin's favorite cuisine is ...



Pause (k)

LSTM



GRU

