Recurrent neural networks: An Introduction

Pranav Kant Gaur

Graphics and Visualization Section

January 24, 2020

Problem: Top view

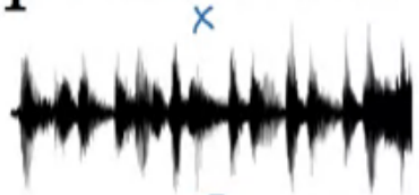
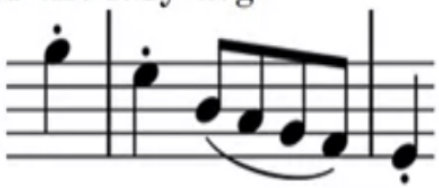
- Given examples of tuples (x, y)
- Estimate f, such that, $y = f(x)$
- Types:
  - Classification: Range of y is discrete and finite
  - Prediction: Range of y is continuous

Solution: Top view

- f is approximated by a Directed Graph, $G(V, E, W)$
- Solving Deep learning problem then becomes estimating $G(V, E, W)$, like designing electronic circuits.
- Neural networks: Adapative Weighted Directed Graphs
- Determining (V, E): Identifying NN architecture
- Determining W: Training NN

Motivation: Sequential X

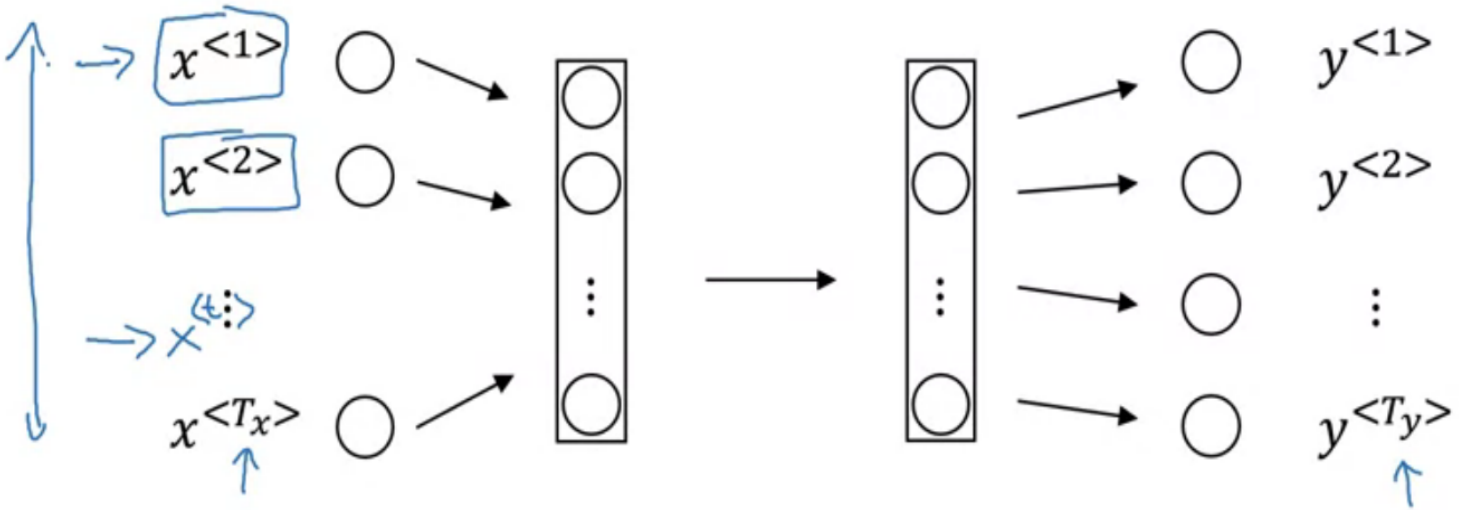
Examples of sequence data

Speech recognition		→	"The quick brown fox jumped over the lazy dog."
Music generation		→	
Sentiment classification	"There is nothing to like in this movie."	→	★☆☆☆☆
DNA sequence analysis	AGCCCCTGTGAGGAACTAG	→	AGCCCCTGTGAGGAACTAG
Machine translation	Voulez-vous chanter avec moi?	→	Do you want to sing with me?
Video activity recognition		→	Running
Name entity recognition	Yesterday, Harry Potter met Hermione Granger.	→	Yesterday, Harry Potter met Hermione Granger.

Andrew Ng

Why not a Feed-forward Neural Network?

Why not a standard network?

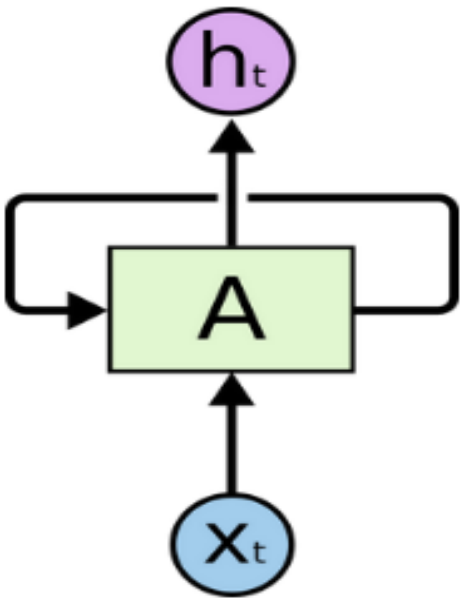


Problems:

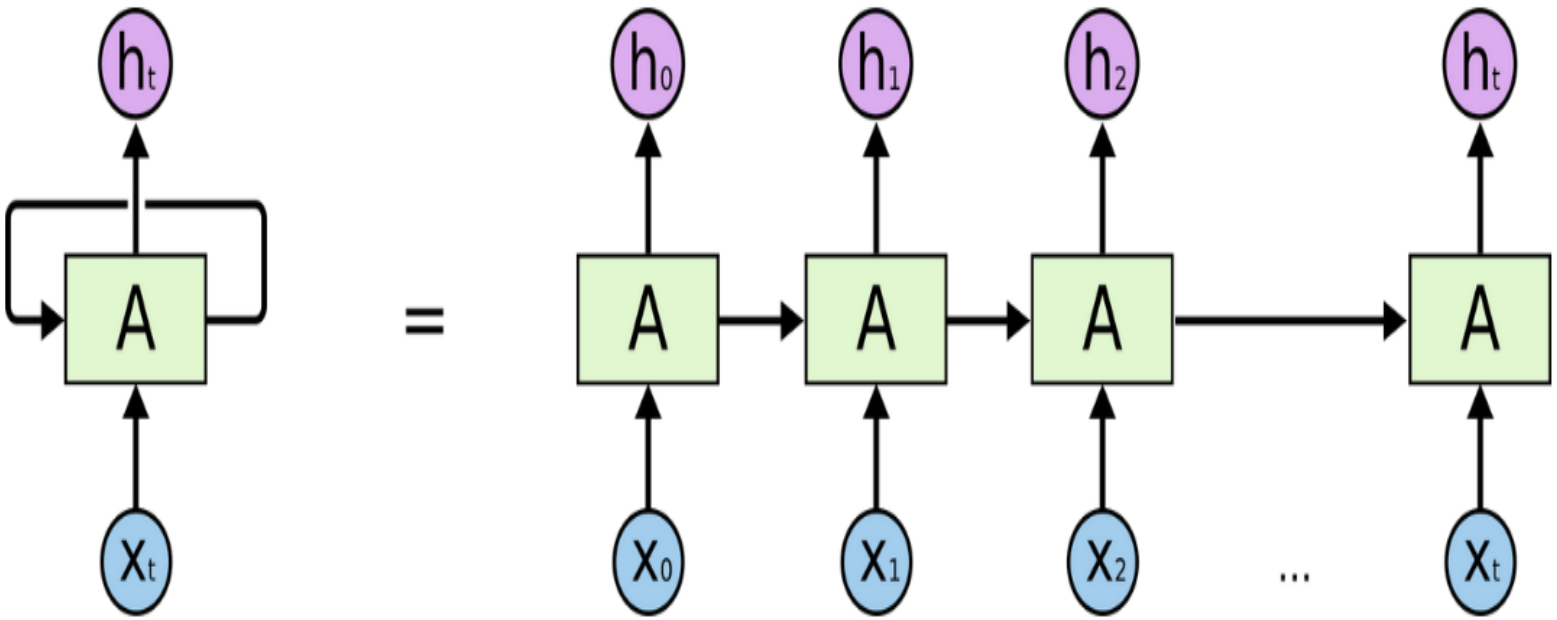
- Inputs, outputs can be different lengths in different examples.
- Doesn't share features learned across different positions of text.

Andrew Ng

RNN Cell



Unrolling



Short-term dependencies: No issues

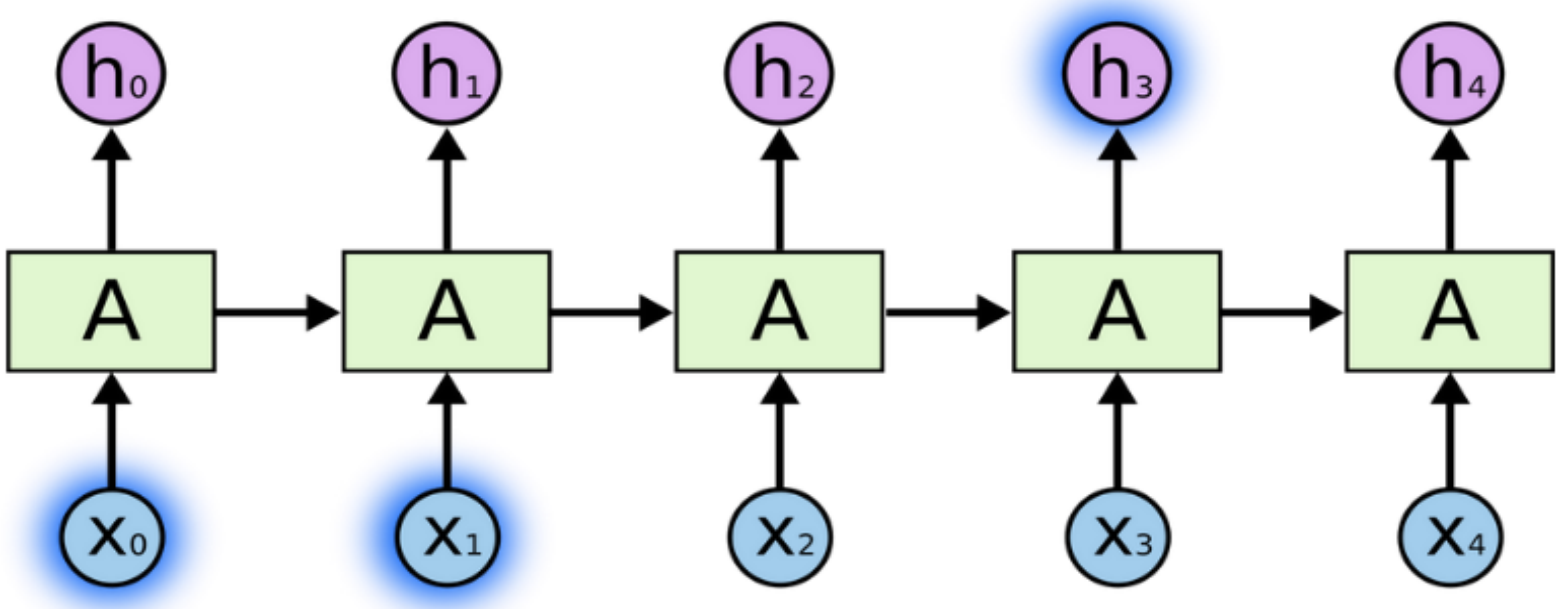


Figure: Consider text auto-completion: *The boy liked* $\langle$ *NEXT WORD?? his/her/etc* $\rangle$ *bicycle...*

Long-term dependencies: Vanishing(or Exploding) Gradients

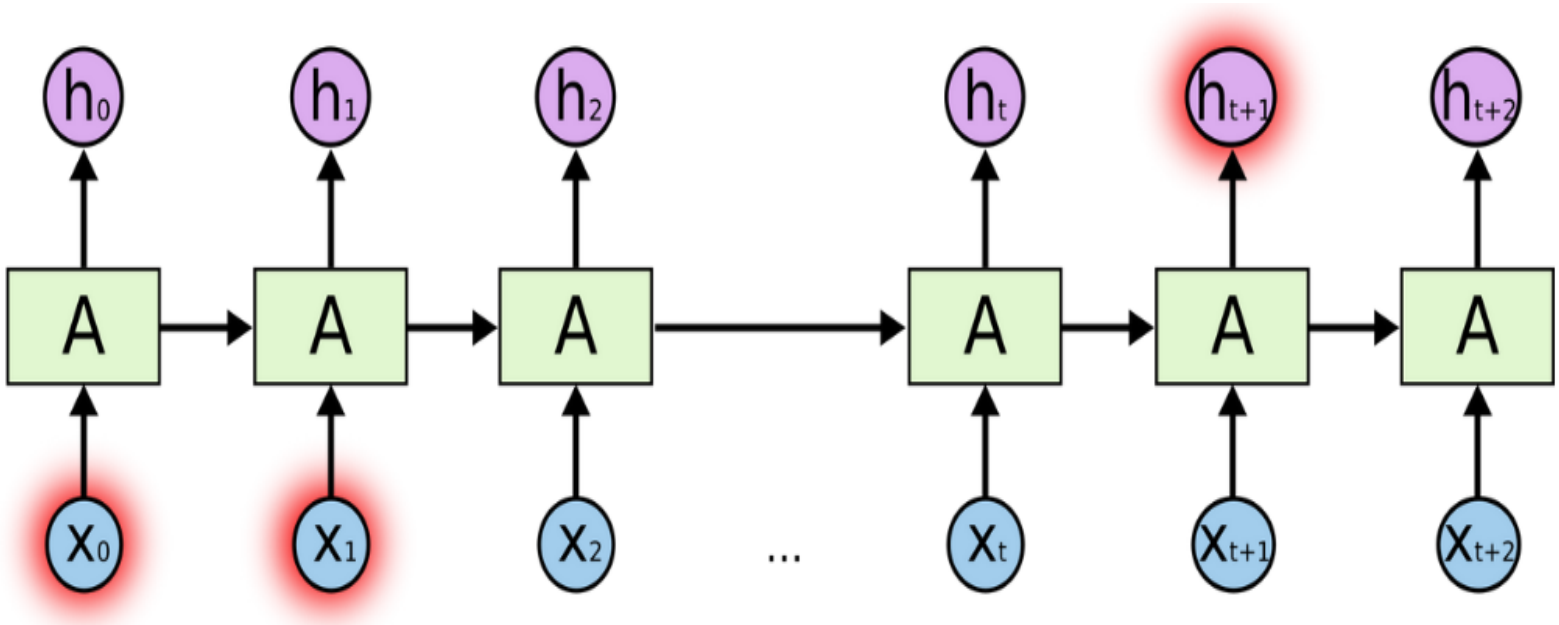


Figure: Consider: *A BARC employee, working on ...., published ⟨NEXT WORD, (his or their??)⟩....* **Solution:** Need to cherrypick info.

Inside a RNN cell

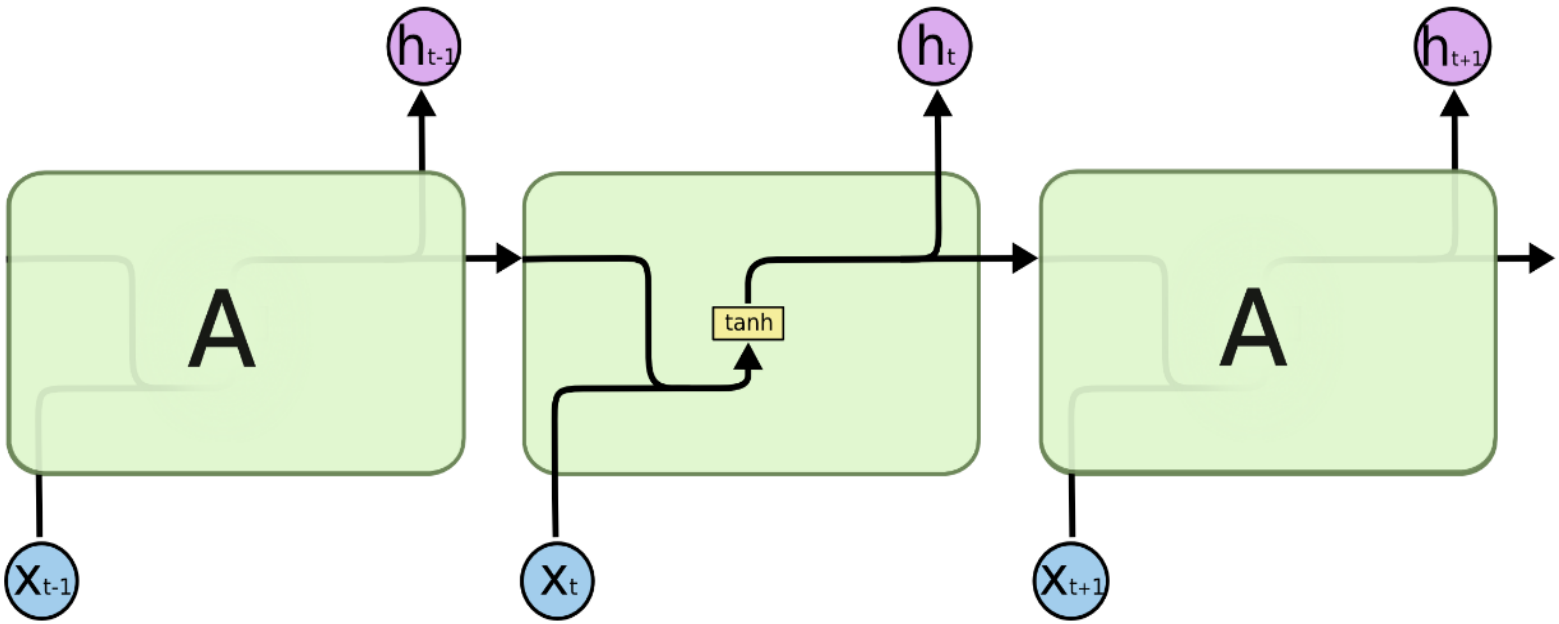


Figure: No mechanism to cherry-pick, **A BARC employee, working on ...., published** $\langle$*NEXT WORD, (his or their??)*$\rangle$ ....

LSTM cell

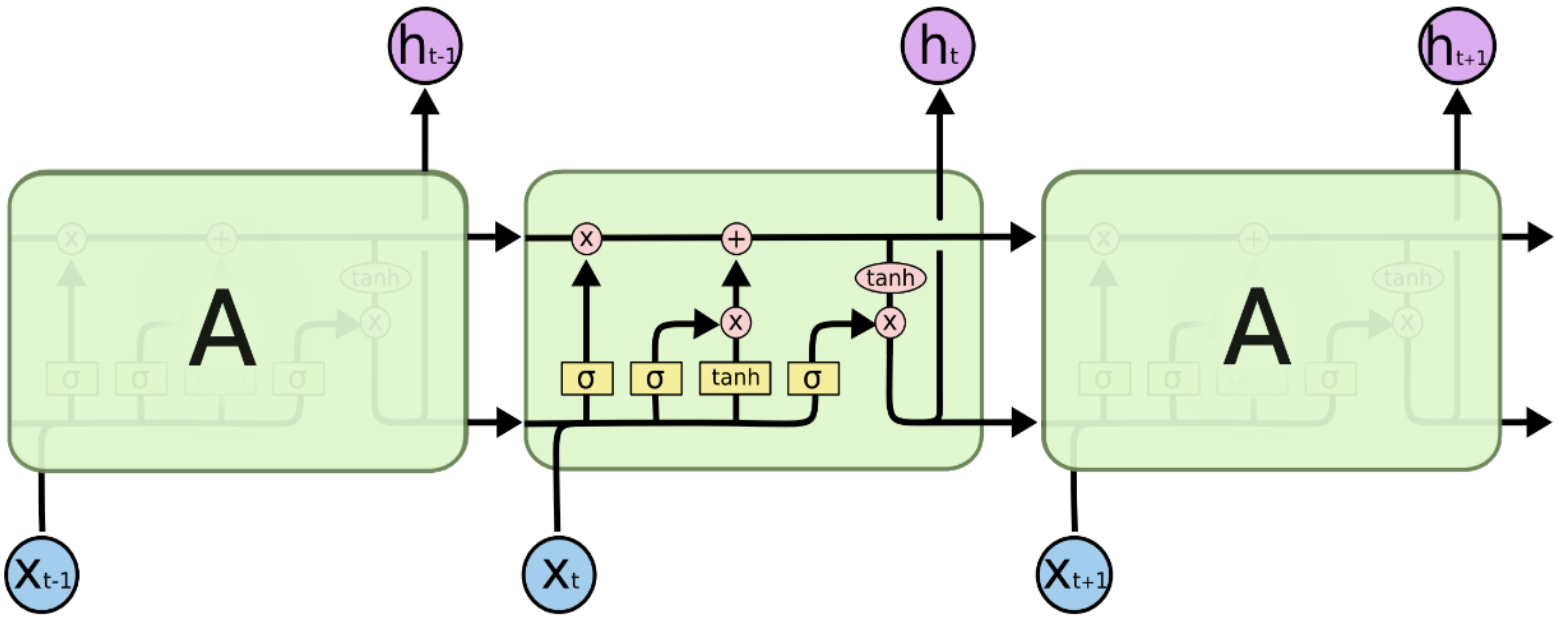


Figure: Gating mechanism for cherry-picking, **A** *BARC employee, working on ....,* *published* $\langle$*NEXT WORD, (his or their??)*$\rangle$....

Cell state

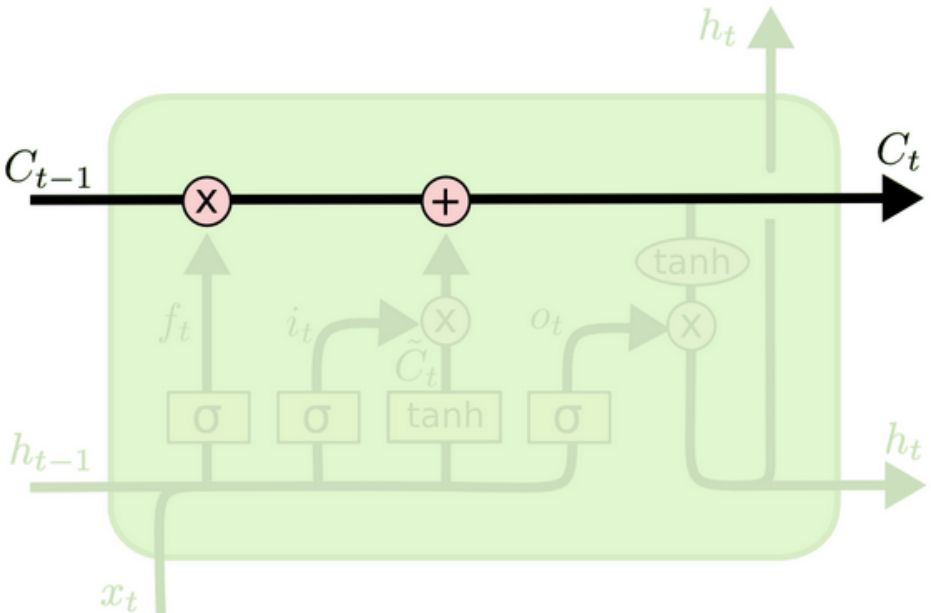


Figure: Stores relevant context, **A** *BARC employee, working on ...., published* $\langle$*NEXT WORD*$\rangle$ *...*, **A** implies Singular.

LSTM Gate

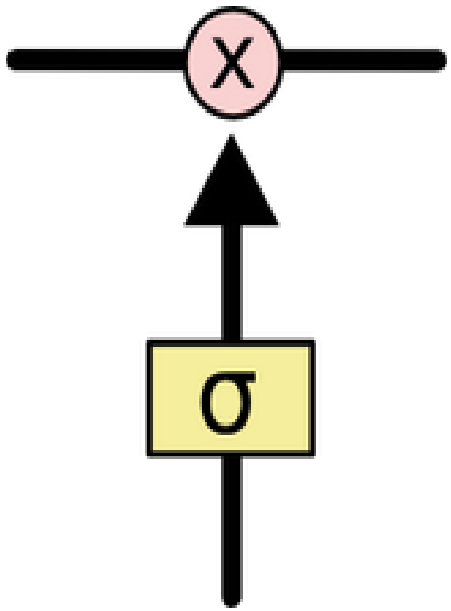
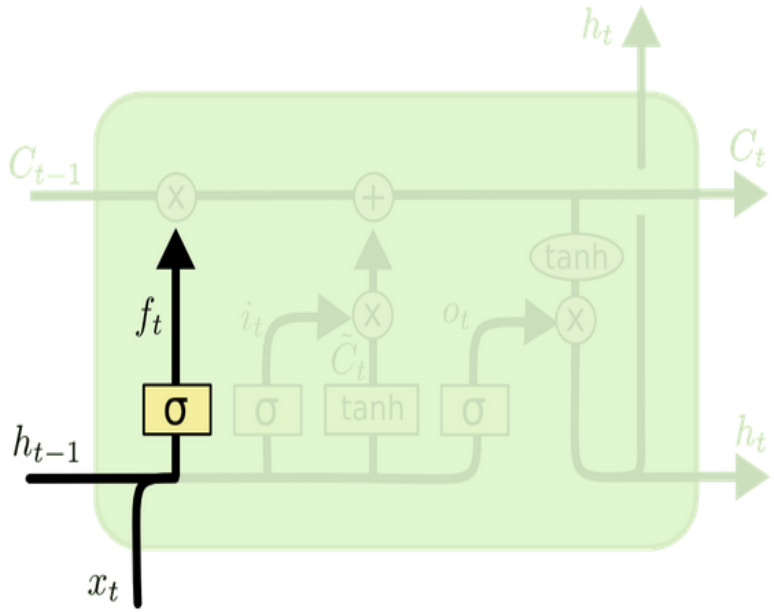


Figure: Controls the relevance of last state, input and candidate output in determining the final state and output of a LSTM cell

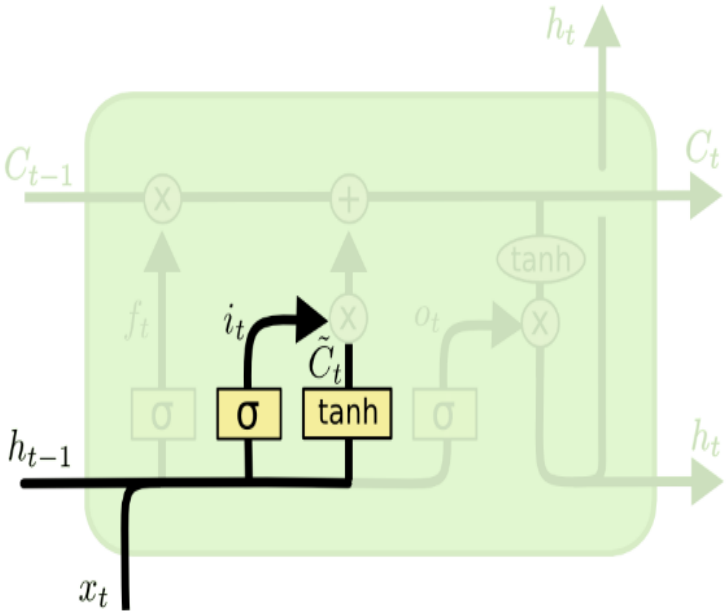
LSTM Walkthrough: Forget control signal



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

Figure: Controls relevance of the state of previous cell for the current cell, **A** *BARC employee, working on ...., published* $\langle$*NEXT WORD, (his or their??)*$\rangle$ ..., focus on state of cell representing **A** only and reject state updates from other cells.

LSTM Walkthrough: Input control signal

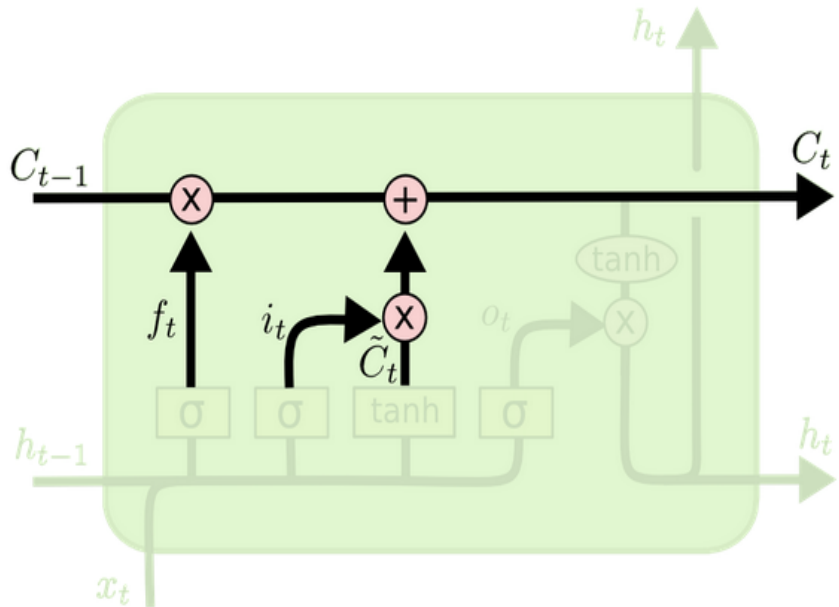


$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Figure: Controls relevance of the input for the current cell, **A** *BARC employee, working on ...., published* $\langle$*NEXT WORD, (his or their??)*$\rangle$ ..., words other than **A** will not be allowed to alter the state of their cells

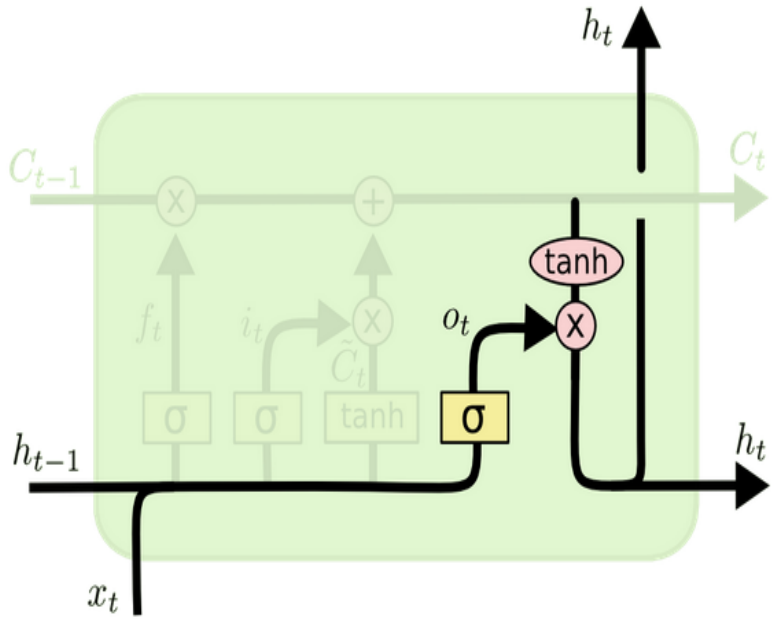
LSTM Walkthrough: Forget and Input gate



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

Figure: Cells state is an additive combination of input and last state controlled by respective gates. For above application, any cell between **A** and ⟨WORD PREDICTION⟩ cell will have *singular* as the state update and no contributions from their word input.

LSTM Walkthrough: Output gate



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

$$h_t = o_t * \tanh\left(C_t\right)$$

Figure: Controls the relevance of current cell output depending on target application.In our auto-completion application, next word prediction is generated via output gate from one cell only, output signal from rest of the cells stays shut via output gate.

References

- http://colah.github.io/posts/2015-08-Understanding-LSTMs/
- https://www.coursera.org/learn/nlp-sequence-models
- https://arxiv.org/abs/1507.05717
- https://distill.pub/2019/memorization-in-rnns/

Thank you!!