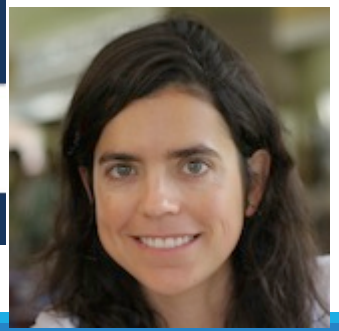
#DLUPC

Day 3 Lecture 3

Gated Units



Marta R. Costa-jussà
marta.ruiz@upc.edu

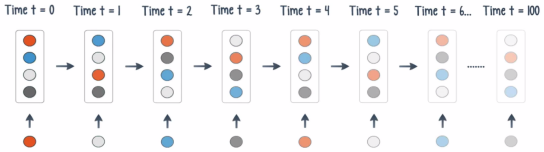
Ramón y Cajal Researcher
Universitat Politecnica de Catalunya
Technical University of Catalonia

telecom BCN

UPC

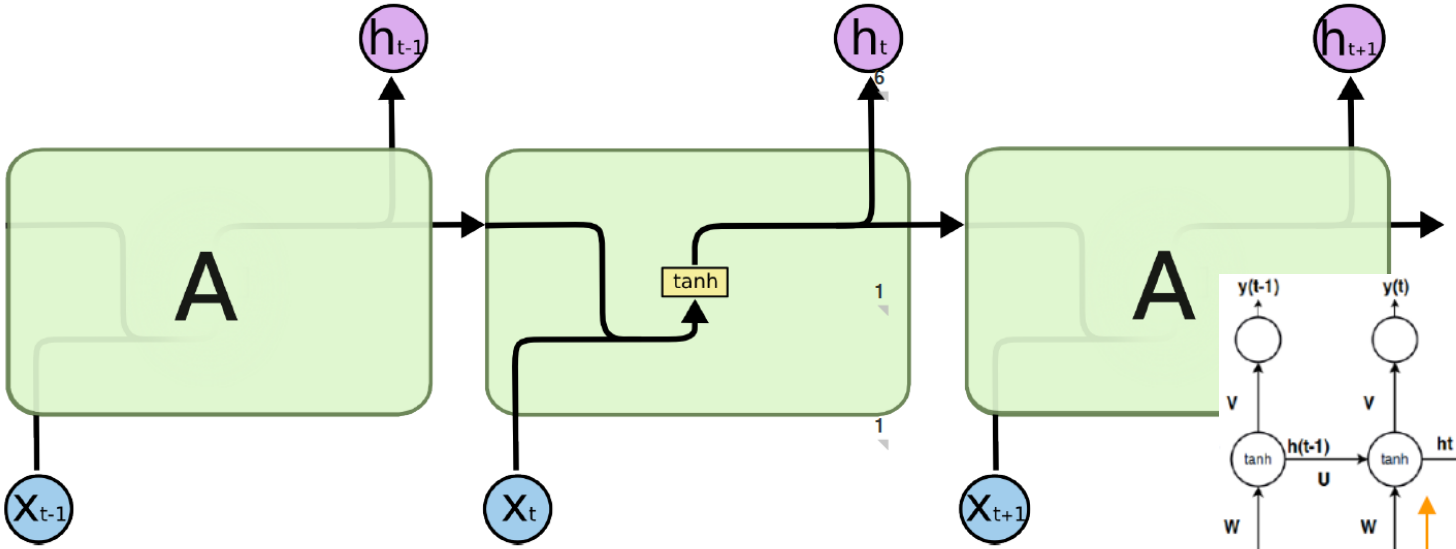
[course site]

Decay of information through time

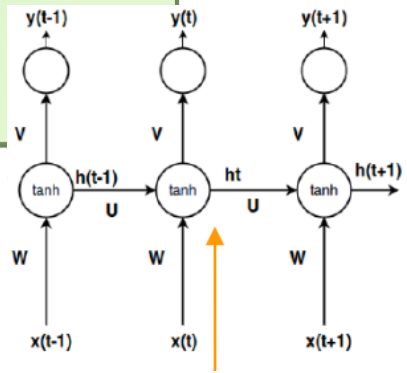


https://www.nextbigfuture.com/2016/03/recurrent-neural-nets.html

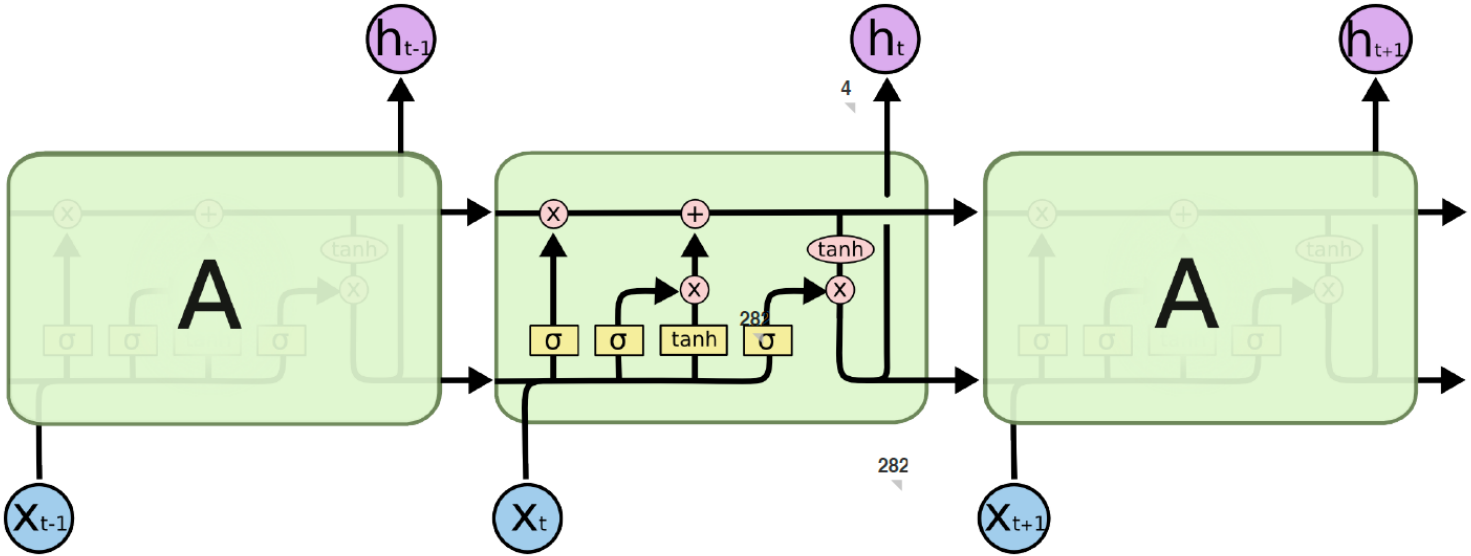
Standard RNN



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

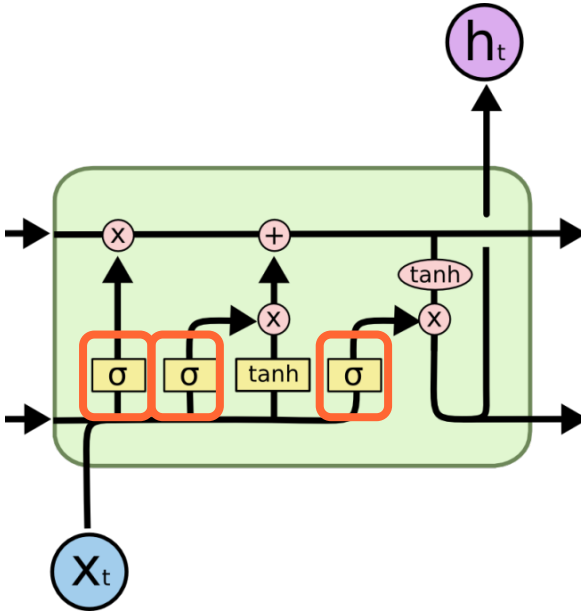


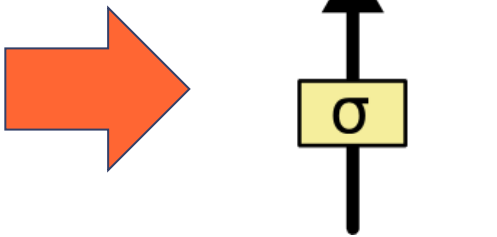
Long-Short Term Memory (LSTM)



Long Short-Term Memory (LSTM)

Three **gates** are governed by *sigmoid* units (btw [0,1]) define the control of in & out information..





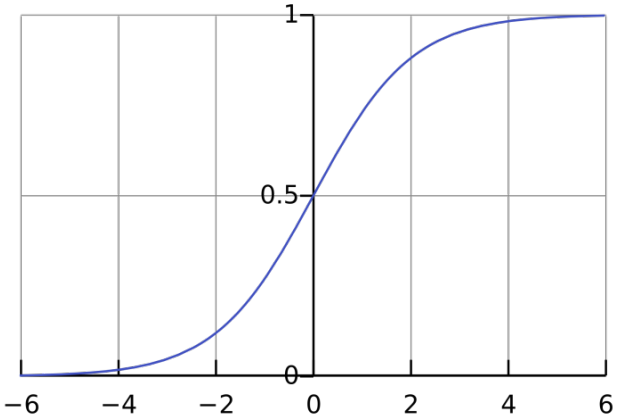
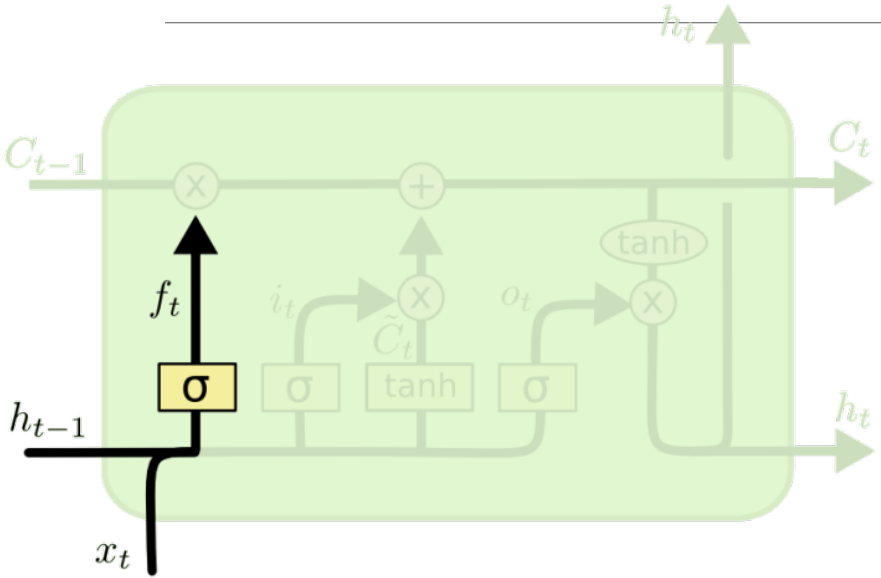


Figure: Cristopher Olah, "Understanding LSTM Networks" (2015)

slide credit: Xavi Giro

Forget Gate



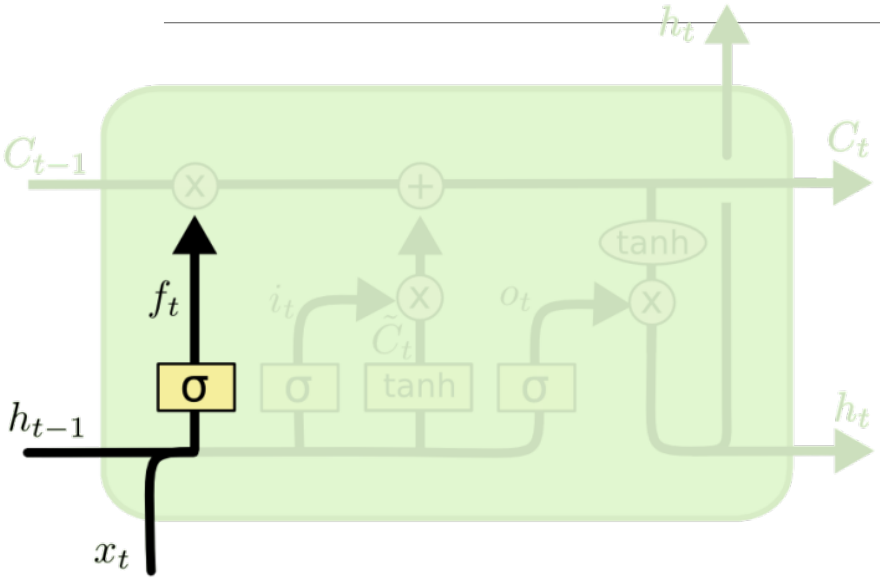
Forget Gate:

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

Concatenate

Figure: Cristopher Olah, "Understanding LSTM Networks" (2015) / Slide: Alberto Montes

Forget Gate: Example



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

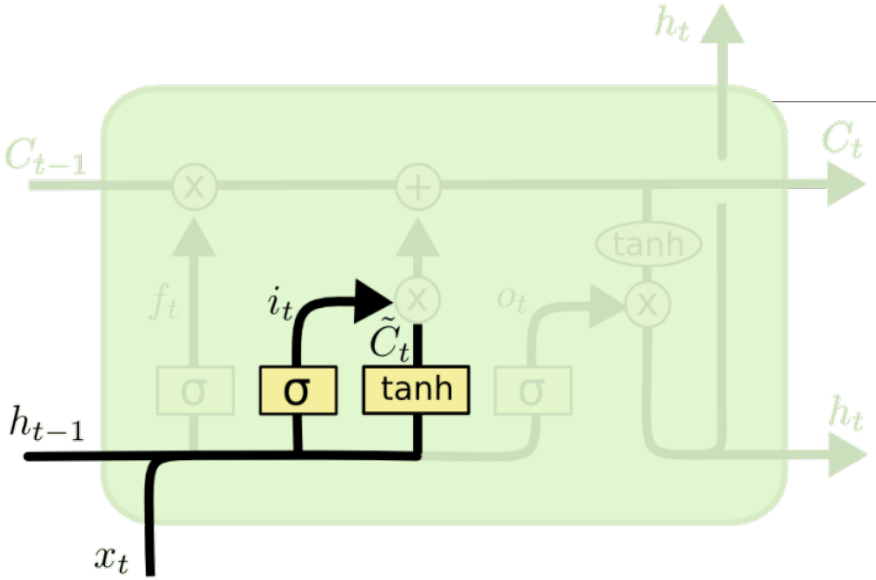
LANGUAGE MODELING

Joan es un chico activo y Anna es una chica calmada

Forget about ”male” gender

Figure: Cristopher Olah, “Understanding LSTM Networks” (2015) / Slide: Alberto Montes

Input Gate



Input Gate Layer

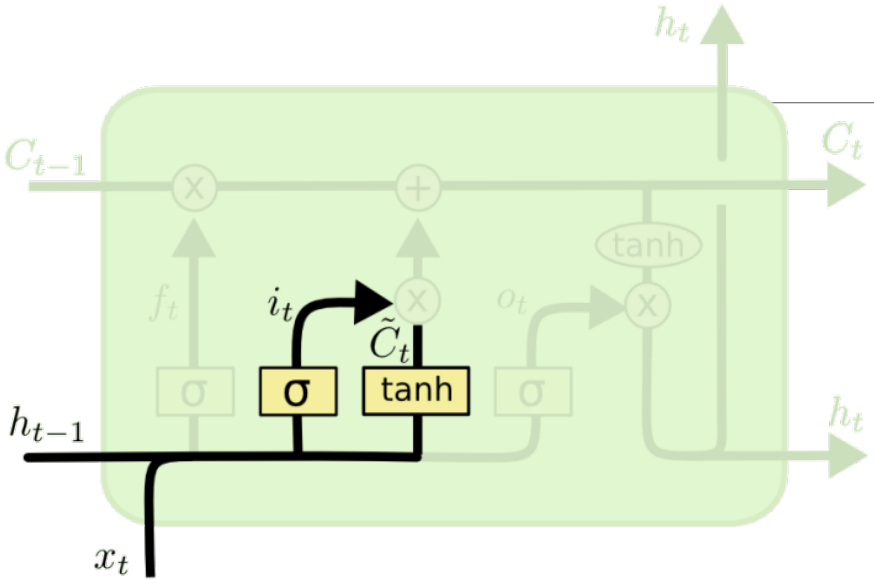
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

New contribution to cell state

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

Classic neuron

Input Gate: Example



Input Gate Layer

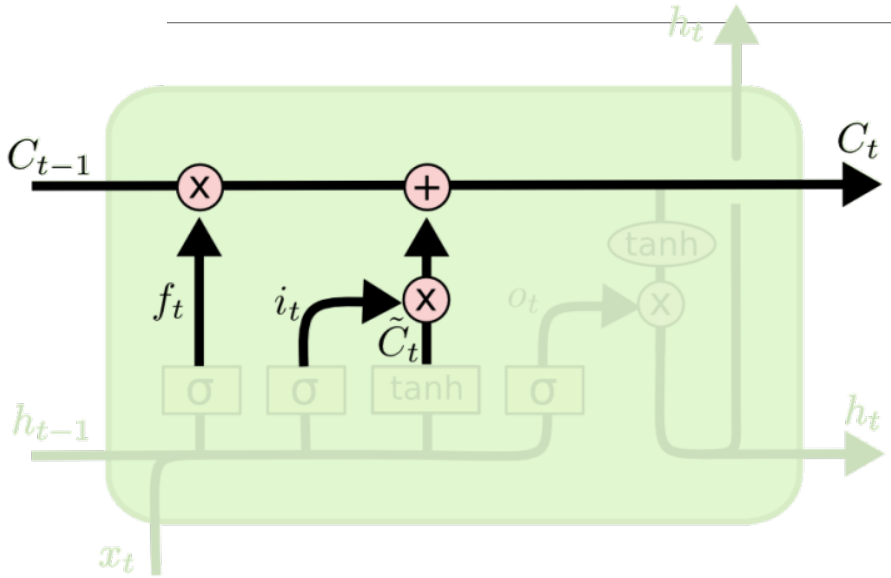
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

LANGUAGE MODELING

Joan es un chico activo y Anna es una chica calmada

Input about ”female” gender

Update Cell State

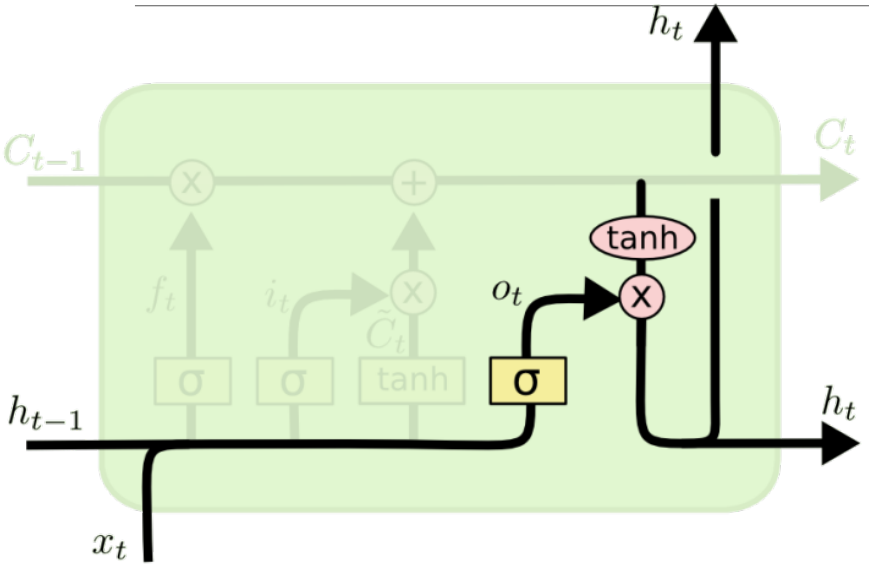


Update Cell State (memory):

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

Figure: Cristopher Olah, "Understanding LSTM Networks" (2015) / Slide: Alberto Montes

Output Gate



Output Gate Layer

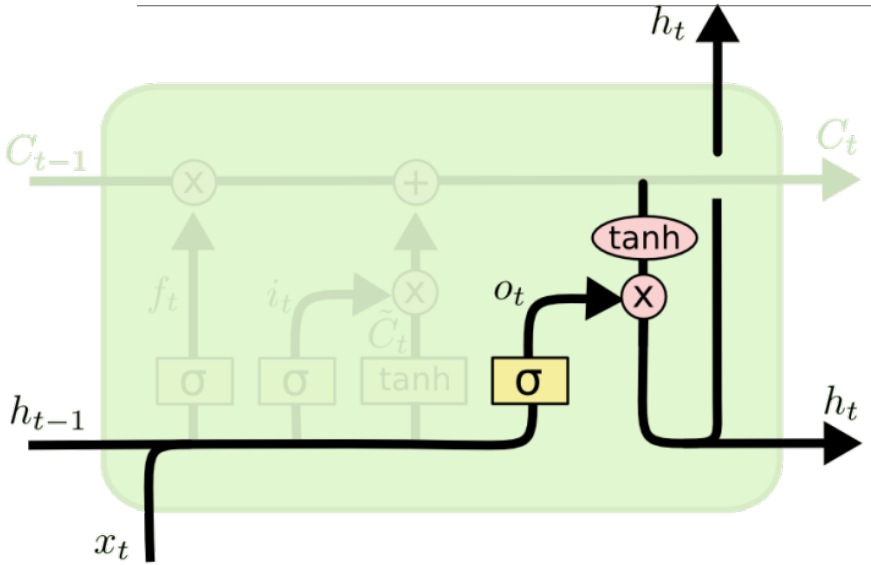
$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

Output to next layer

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

Figure: Cristopher Olah, "Understanding LSTM Networks" (2015) / Slide: Alberto Montes

Output Gate: Example



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

LANGUAGE MODELING

Joan es un chico activo y Anna es una chica calmada

Info relevant for a verb? : ”female”, “singular”, “3rd person”

Figure: Cristopher Olah, “Understanding LSTM Networks” (2015) / Slide: Alberto Montes

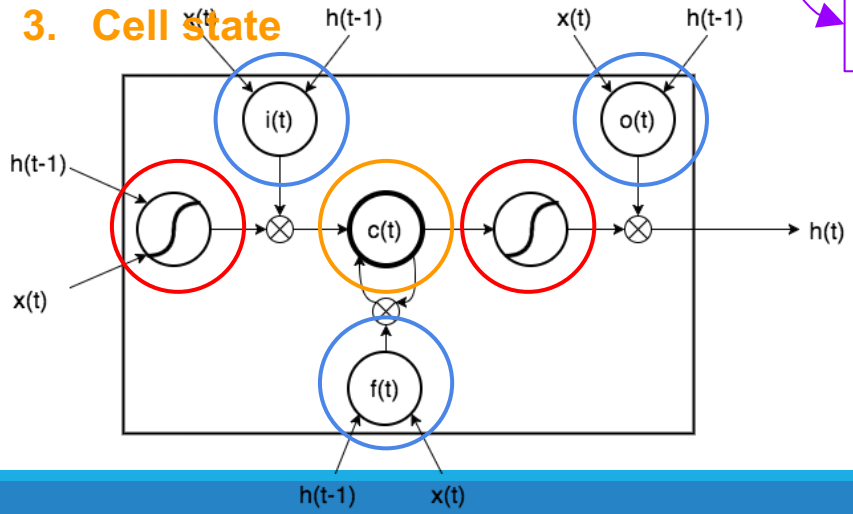
LSTM: parameters

An LSTM cell is defined by two groups of neurons plus the cell state (memory unit):

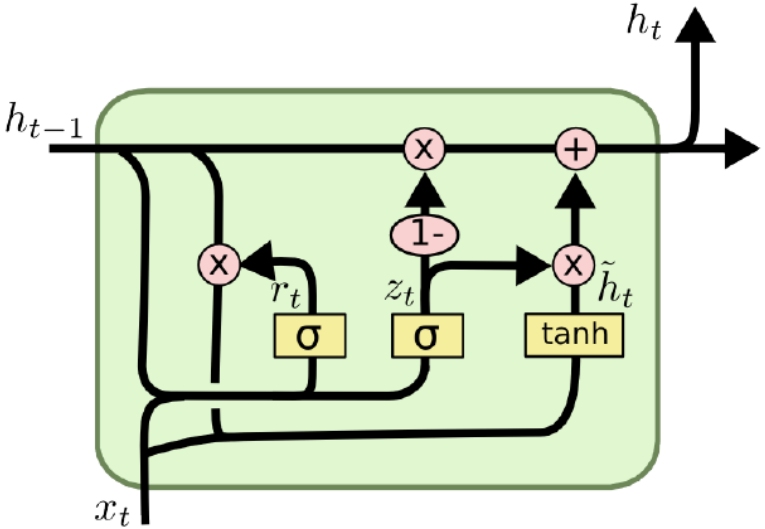
1. **Gates**
2. **Activation units**
3. **Cell state**

$$N^i_{params} = 4 \times (N^i_{inputs} \times N^i_{units} + N^i_{units} \times N^i_{units} + N^i_{units})$$

3 gates + input activation



Gated Recurrent Unit (GRU)



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

$$N^i_{params} = 3 \times (N^i_{inputs} \times N^i_{units} + N^i_{units} \times N^i_{units} + N^i_{units})$$

Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." AMNLP 2014.

slide credit: Xavi Giro

Visual Comparison FNN, Vanilla RNNs and LSTMs

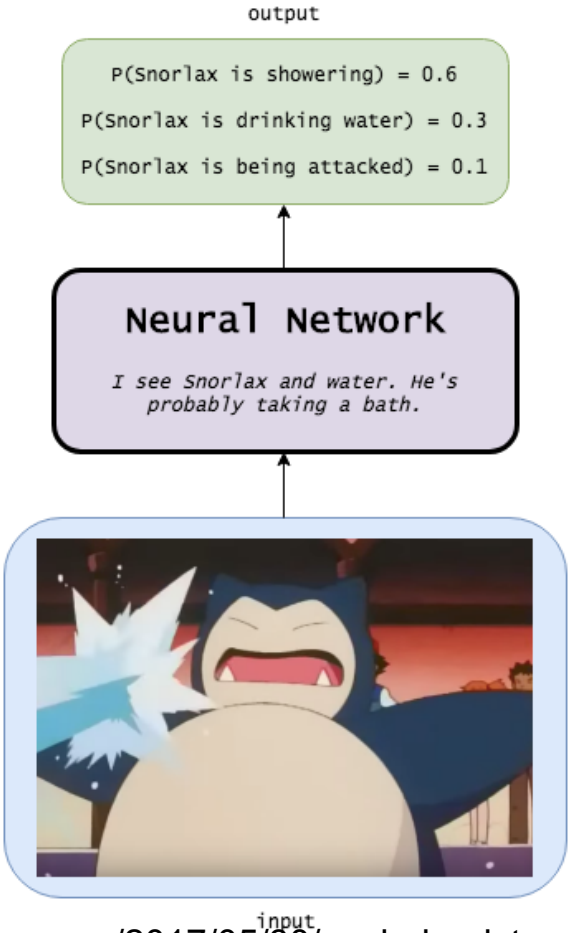


Image src http://blog.echen.me/2017/05/30/exploring-lstms/

Vanilla RNNs

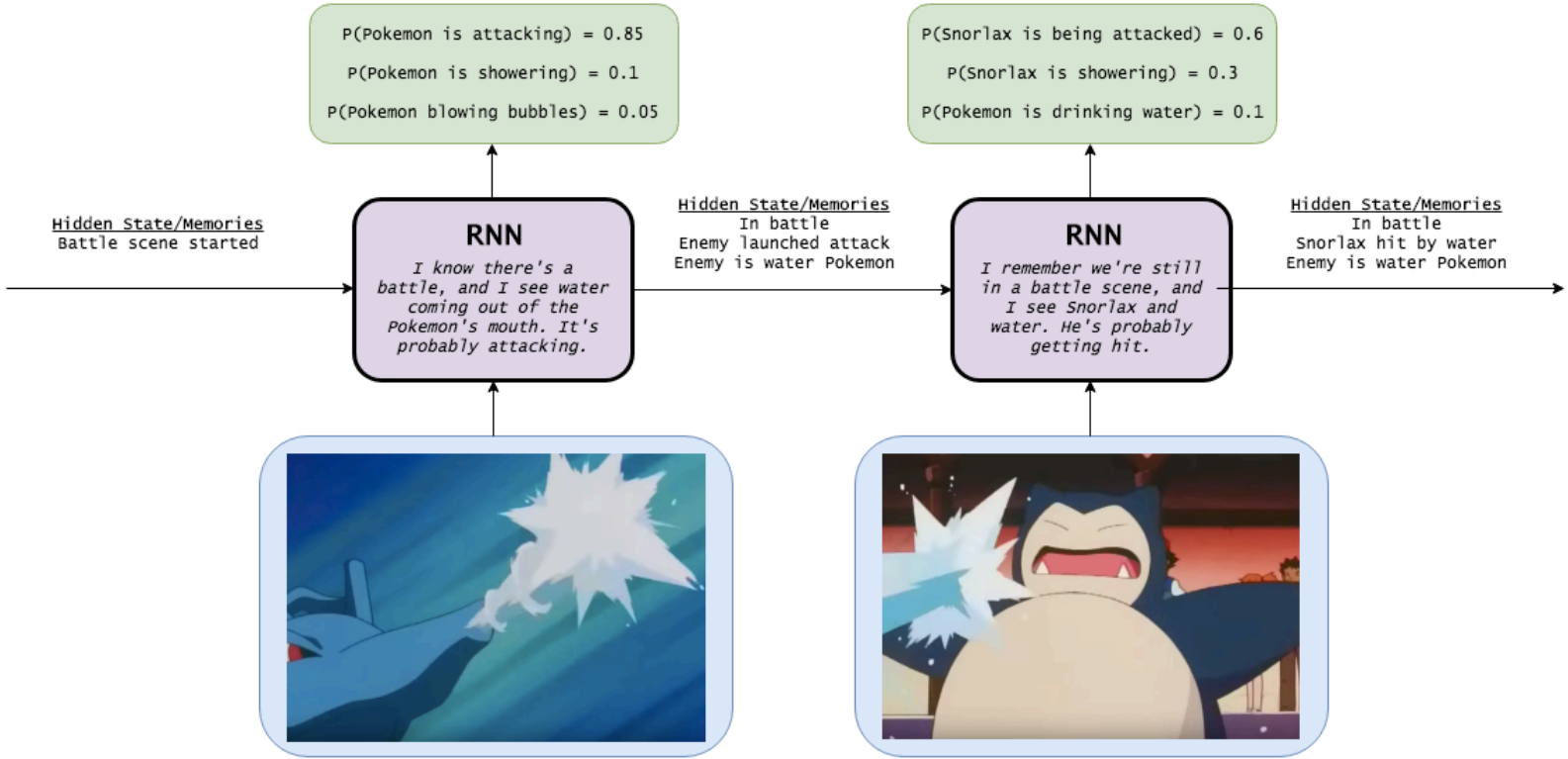
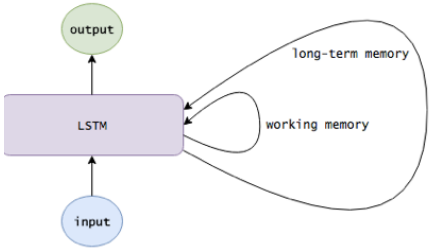


Image src http://blog.echen.me/2017/05/30/exploring-lstms/

LSTMs



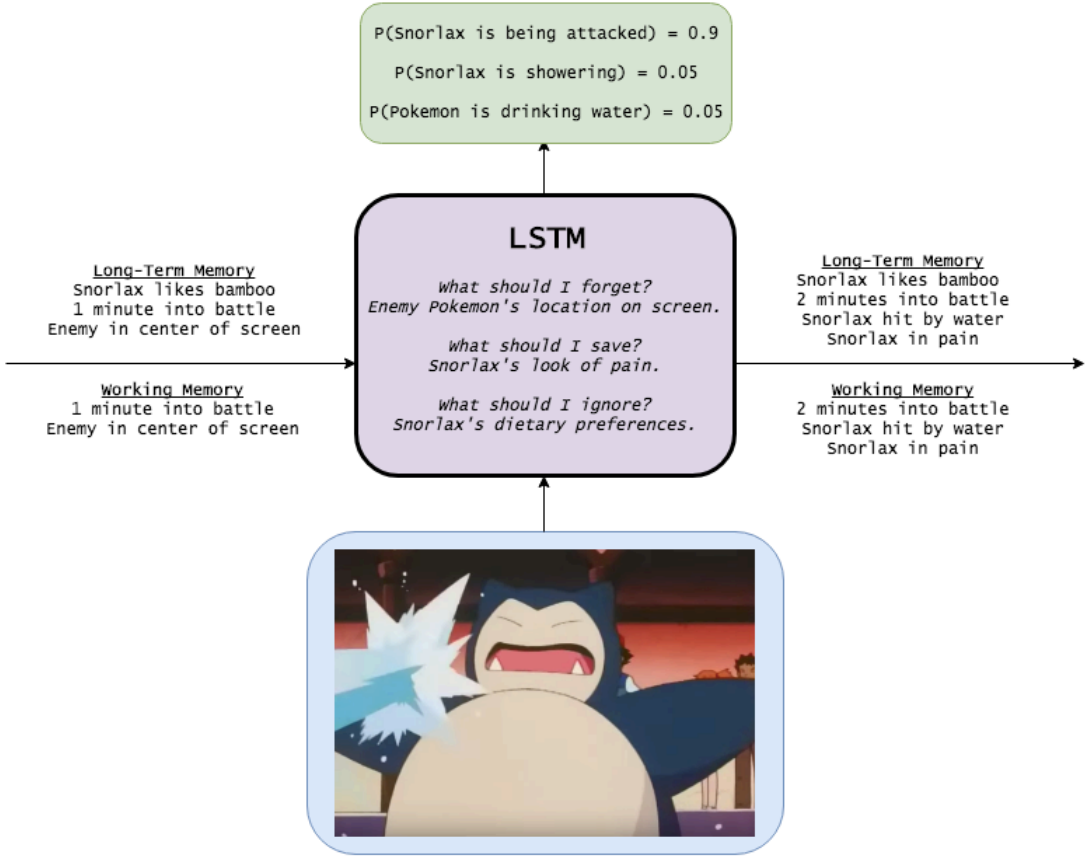


Image src http://blog.echen.me/2017/05/30/exploring-lstms/

Other RNN extensions

Bidirectional RNNs

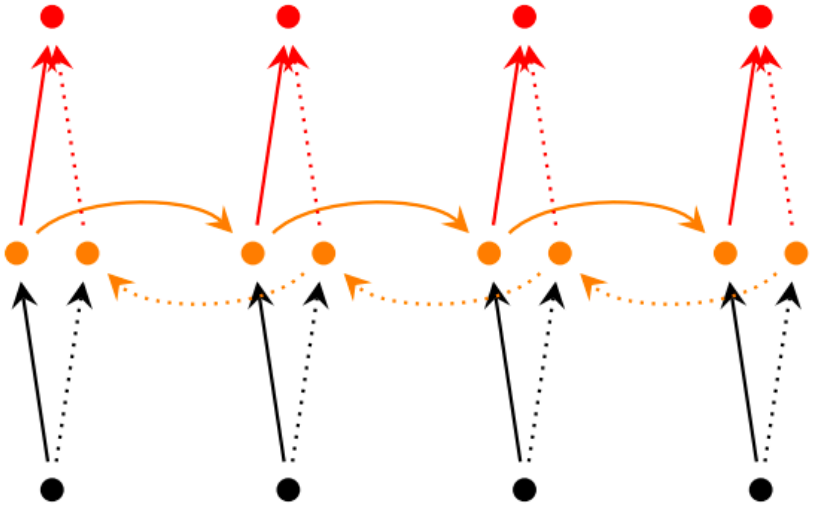


Image src: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

Deep RNNs

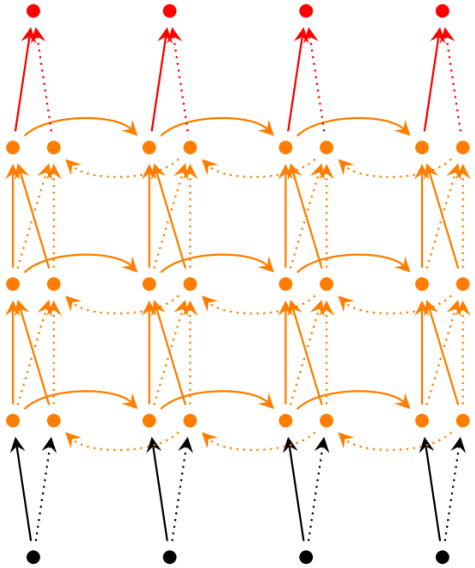


Image src: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

Still ISSUES with RNNs??

Thanks ! Q&A ?