Recurrent Neural Networks (RNN)

Artificial Intelligence @ Allegheny College

Janyl Jumadinova

October 27, 2021

Alex Graves, "Supervised Sequence Labelling with Recurrent Neural Networks"

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

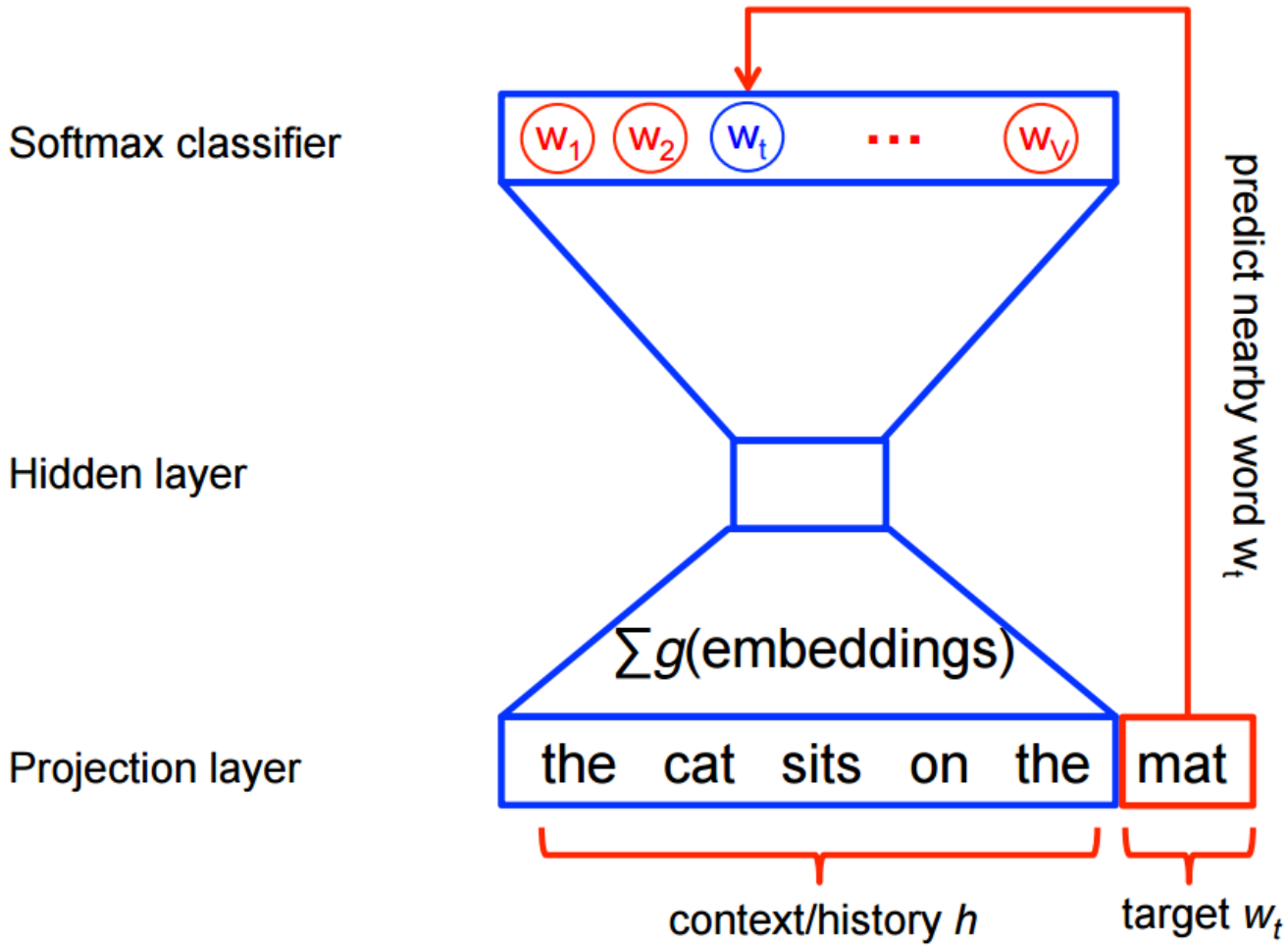
Word2Vec Model

- **Word2Vec** is used to learn vector representations of words, "word embeddings".
- This is typically a preprocessing step, where the learned vectors are fed into a discriminative model (such as RNN).
- Word2vec is a computationally-efficient predictive model for learning word embeddings from raw text.

Word2Vec Model

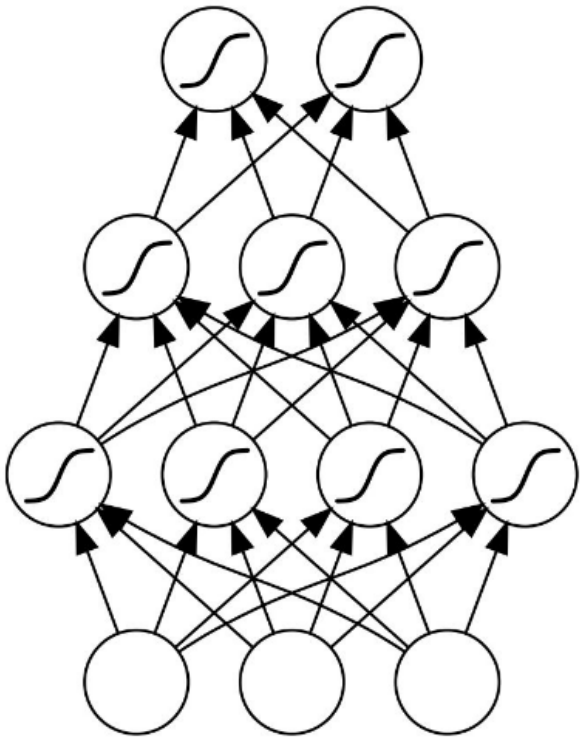
- **Word2Vec** is used to learn vector representations of words, "word embeddings".
- This is typically a preprocessing step, where the learned vectors are fed into a discriminative model (such as RNN).
- Word2vec is a computationally-efficient predictive model for learning word embeddings from raw text.
  (1) *Continuous Bag-of-Words model* (CBOW): predicts target words from context words.
  (2) *Skip-Gram model*: predicts source context words from target words.

Word2Vec Model



https://www.tensorflow.org/tutorials/representation/word2vec

Recurrent Neural Networks



Output Layer

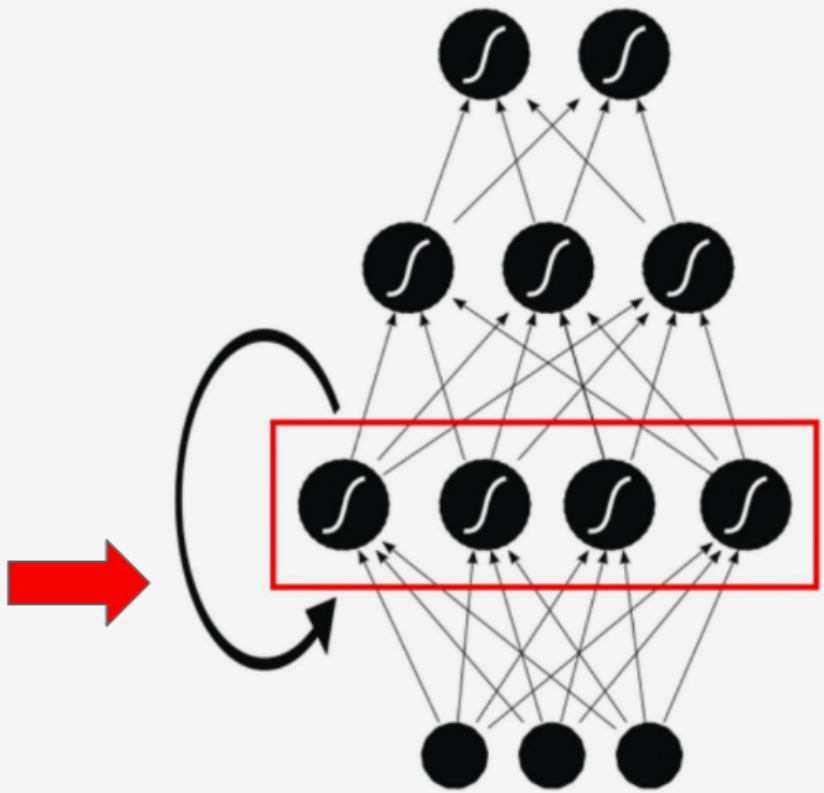
Hidden Layers

Input Layer

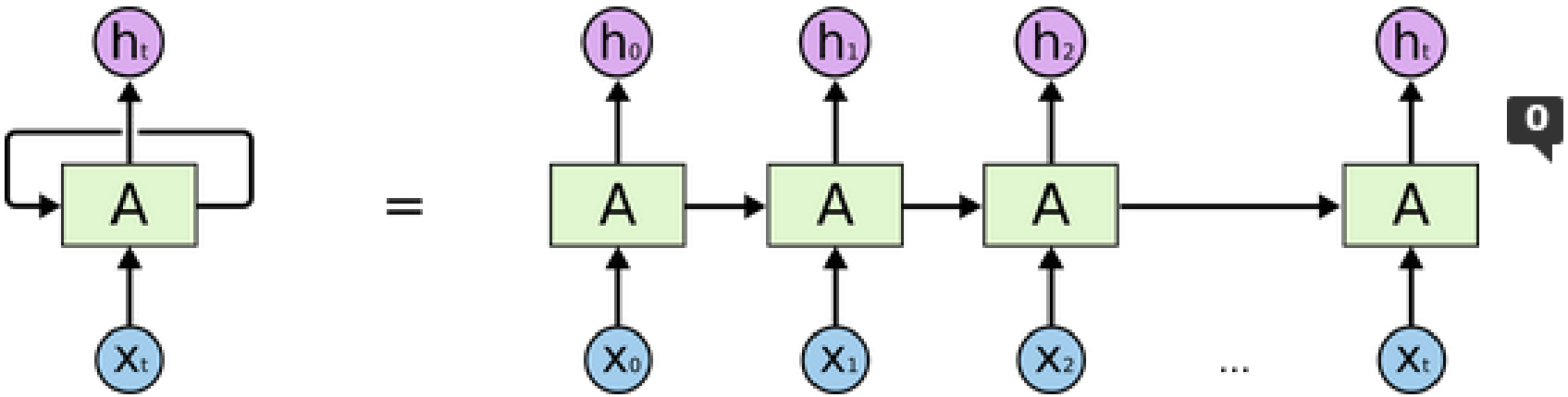
The output depends ONLY on the current input.

Recurrent Neural Networks

The hidden layers and the output depend from previous states of the hidden layers



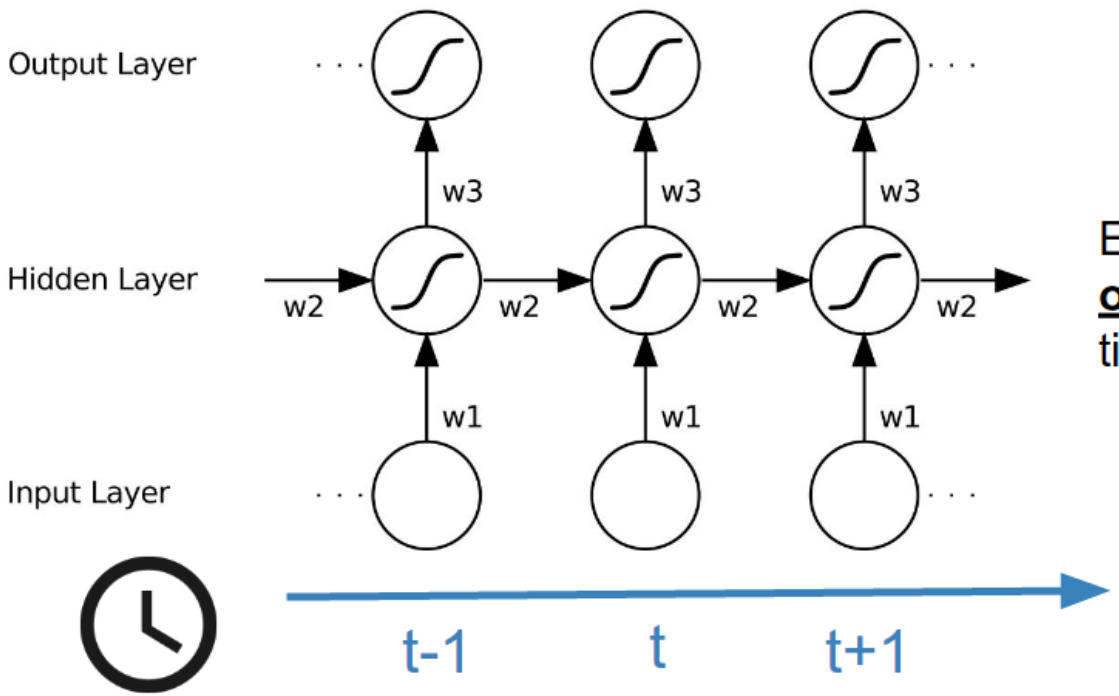
Recurrent Neural Networks



An unrolled recurrent neural network.

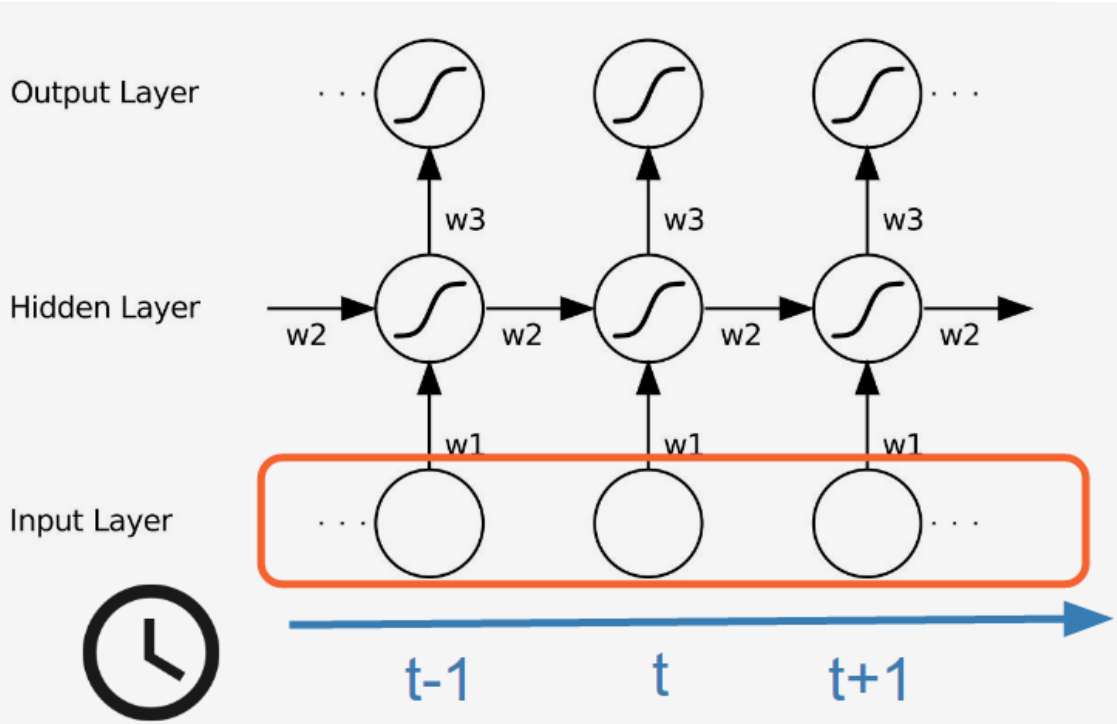
Based on an encoder-decoder scheme, using Seq2Seq model.

Recurrent Neural Networks



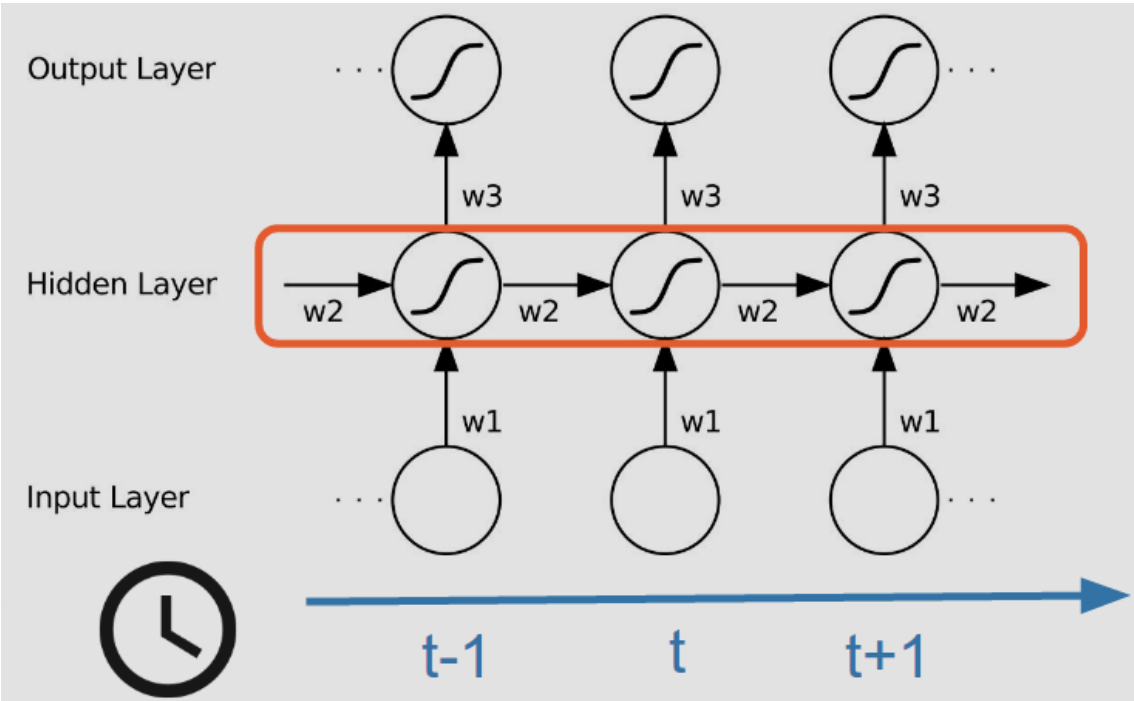
Each node represents **a layer of neurons** at a single timestep.

Recurrent Neural Networks



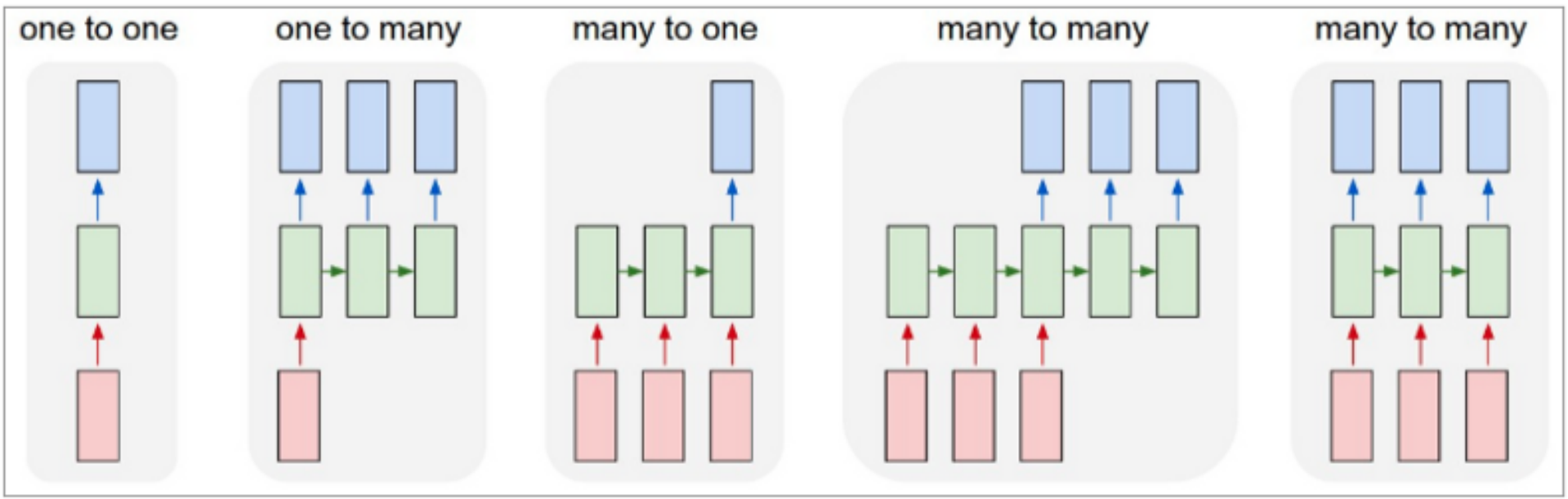
The input is a **SEQUENCE** **x(t)** of any length.

Recurrent Neural Networks



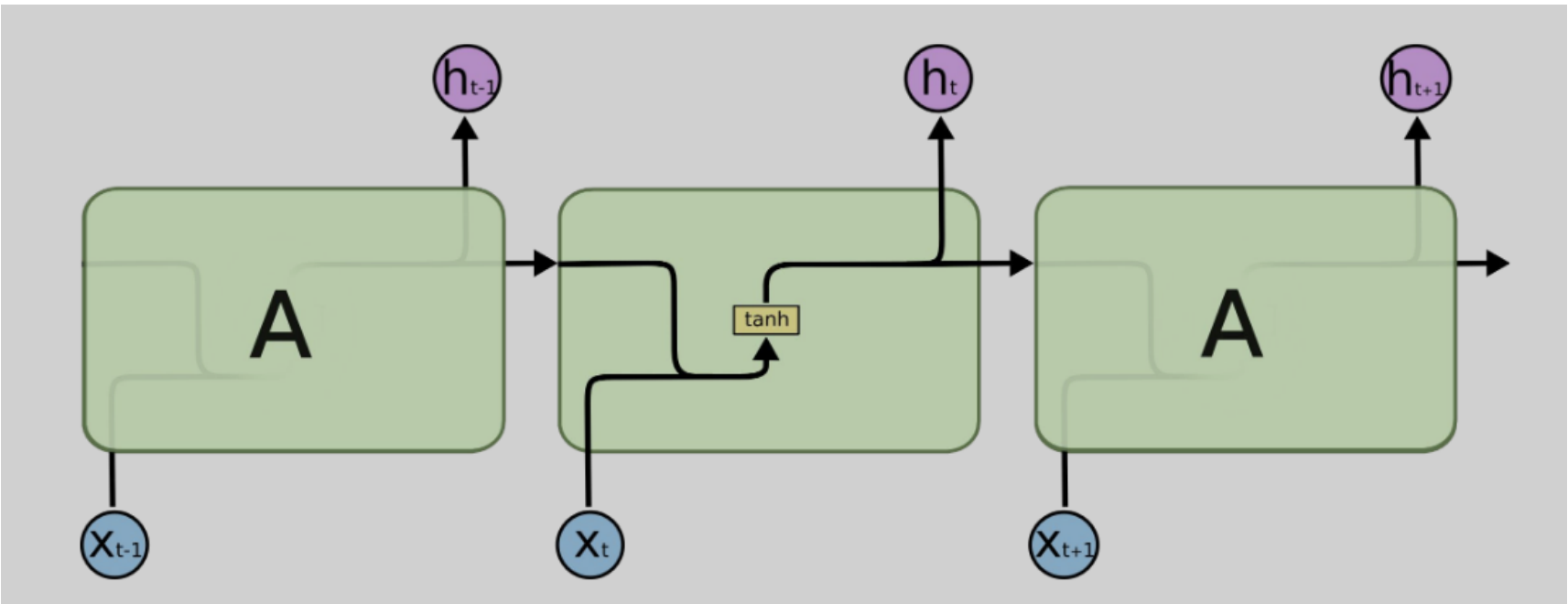
Must learn temporally shared weights w2; in addition to w1 & w3.

Recurrent Neural Networks



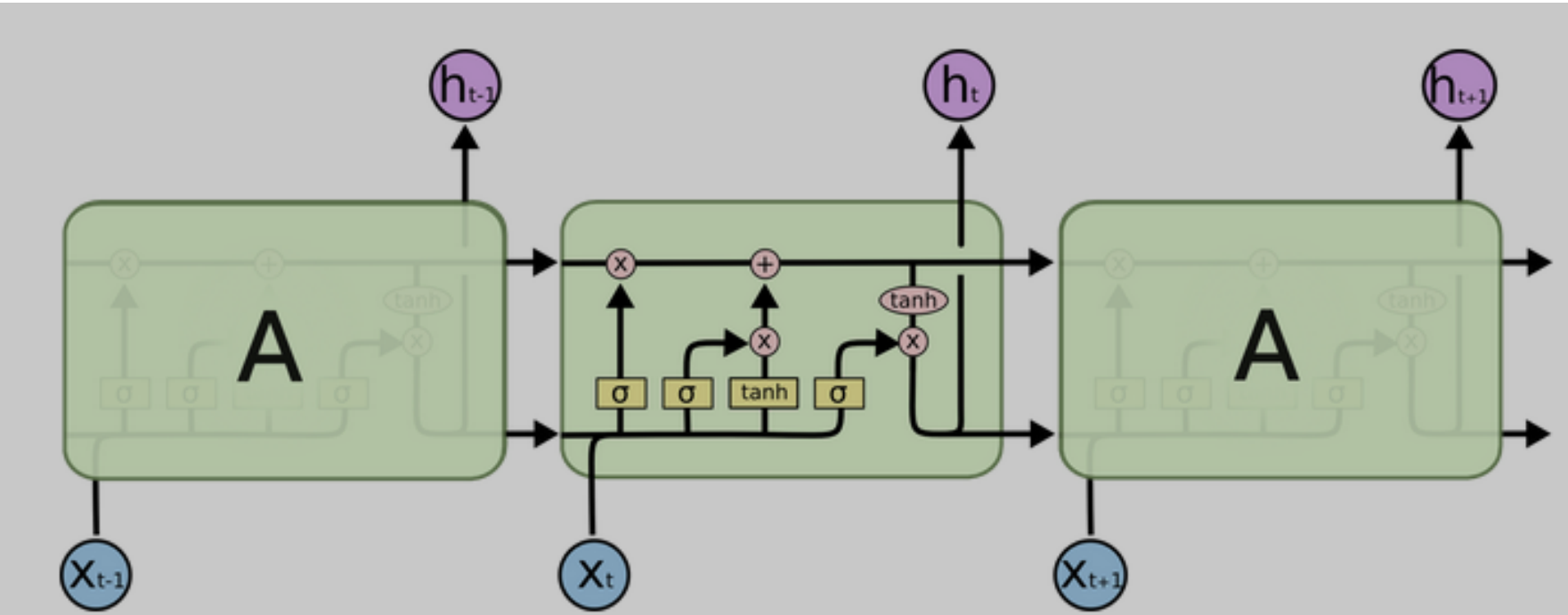
Long Short-Term Memory (LSTM)

Based on a standard RNN whose neuron activates with tanh



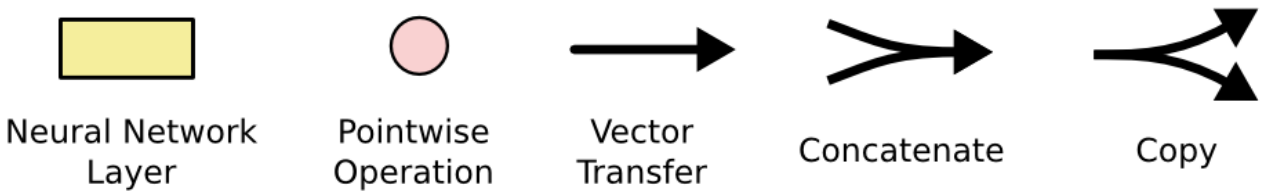
Cristopher Olah, "Understanding LSTM Networks" (2015)

Long Short-Term Memory (LSTM)



The repeating module in an LSTM contains four interacting layers.

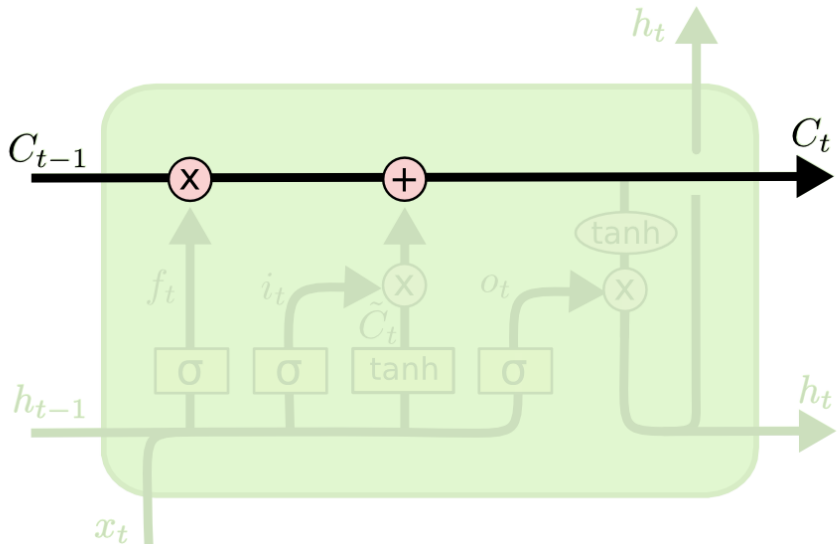
Long Short-Term Memory (LSTM)



- Each line carries an entire vector from the output of one node to the inputs of others.
- Pointwise operations are operations such as vector addition.
- Yellow boxes are learned neural network layers.
- A "Copy" line denote its content being copied and the copies going to different locations.

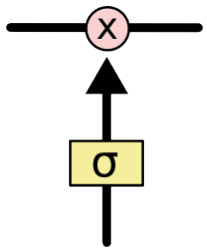
Long Short-Term Memory (LSTM)

The **cell state** runs through the entire chain, with only some minor linear interactions.



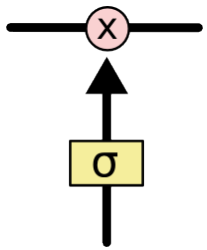
Long Short-Term Memory (LSTM)

The **gate** structures allow to remove or add information to the cell state.



Long Short-Term Memory (LSTM)

The **gate** structures allow to remove or add information to the cell state.



Disadvantage of RNN/LSTM

- Suffer from memory-bandwidth limited problems.
- Alternative? Transformer architecture (replace recurrence/convolution with attention).

TensorFlow Tutorial

- TensorFlow Recurrent Neural Networks
- Text Generation with Recurrent Neural Networks