

# Deep Learning

Abdelhak Mahmoudi  
[abdelhak.mahmoudi@um5.ac.ma](mailto:abdelhak.mahmoudi@um5.ac.ma)

INPT- 2020

# Content

1. Deep Artificial Neural Networks
2. Convolutional Neural Networks
- 3. Sequence Models**
4. Generative Models

# Sequence Models

- Applications
- Why not simple Deep NN ?
- Recurrent Neural Networks (RNN)
  - Architectures
  - Vanishing/Exploding gradients
- Long Short Term Memory Nets (LSTMs)
- Gated Recurrent Units (GRUs)
- Transformers

# Applications

Speech recognition



**Output  $y$**

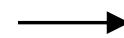
“The quick brown fox jumped over the lazy dog.”

Music generation



Sentiment classification

“There is nothing to like in  
this movie.”



DNA sequence analysis

AGCCCCTGTGAGGAAC TAG



AG~~CCCCTGTGAGGAAC~~ TAG

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.

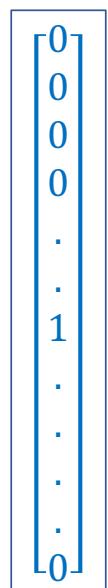


Yesterday, **Harry Potter** met  
**Hermione Granger**.

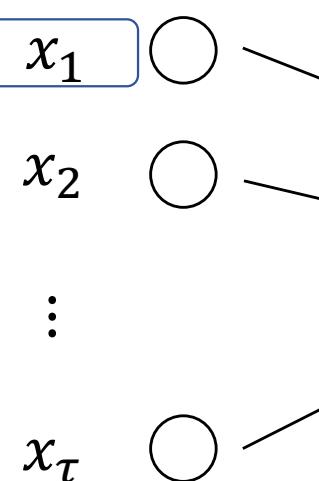
Abdelhak Mahmoudi

# Why not simple Deep NN ?

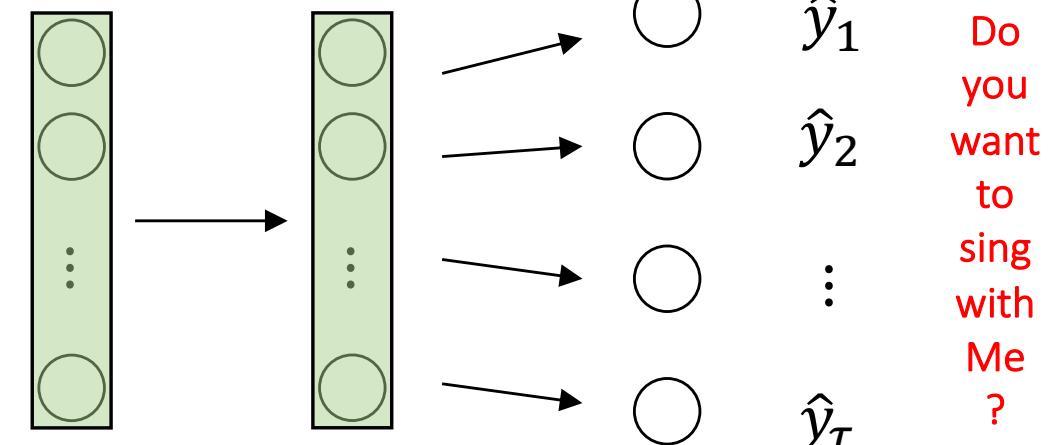
Text representation  
One hot Vector



Voulez  
vous  
chanter  
avec  
moi  
?



Machine translation



$\tau$  Could be different

# Why not simple Deep NN ?

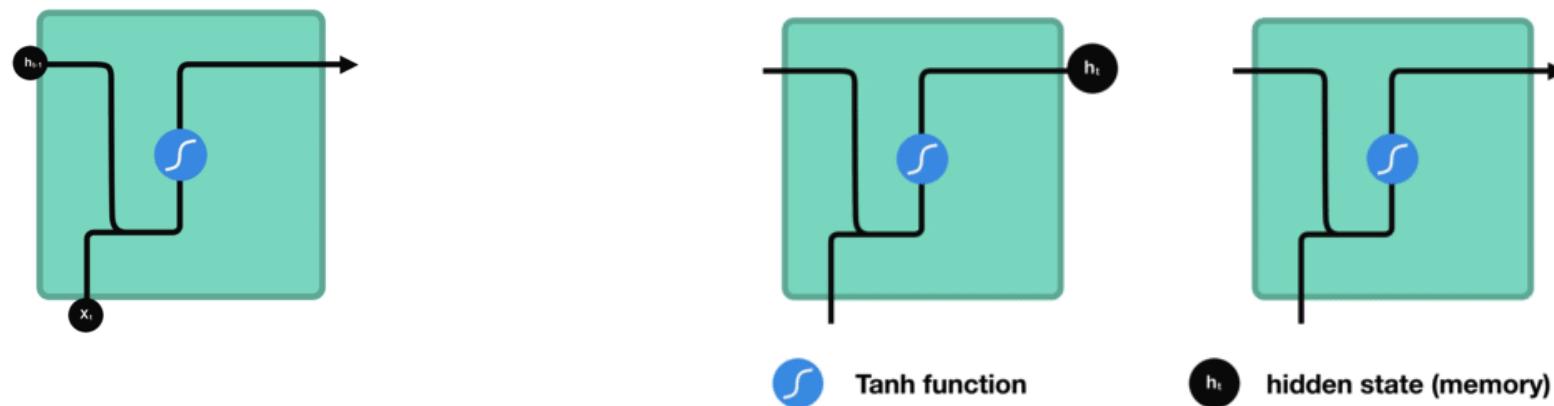
- Process sequences of **variable length**
  - Use a fixed window?
  - What about dependencies?
- Handle **long-term** dependencies
  - Still not taking into account the order in the sequence !
- Maintain **order's** information
  - How?
- **Sharing** parameters

“I'm Moroccan, I speak fluent.....”



“All people love Moroccan Couscous.”

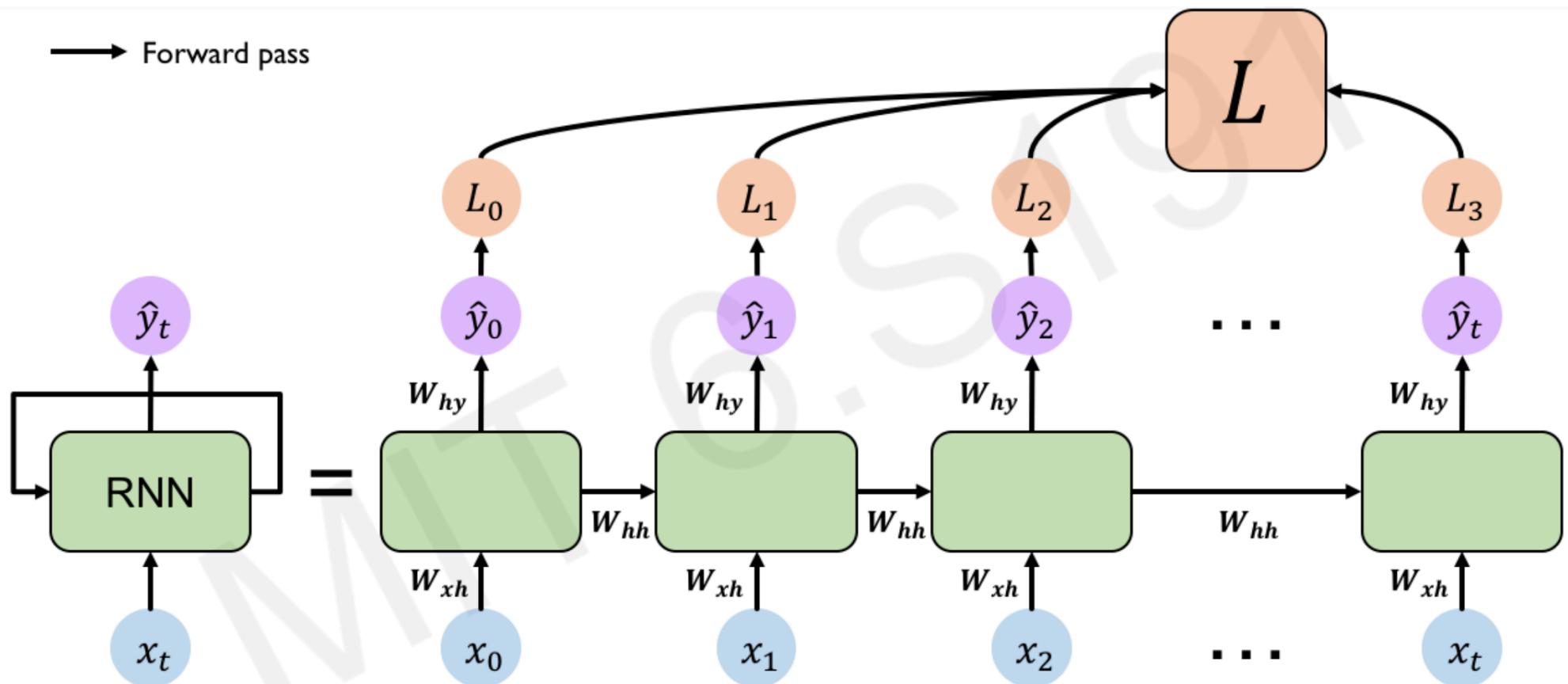
# RNN intuition



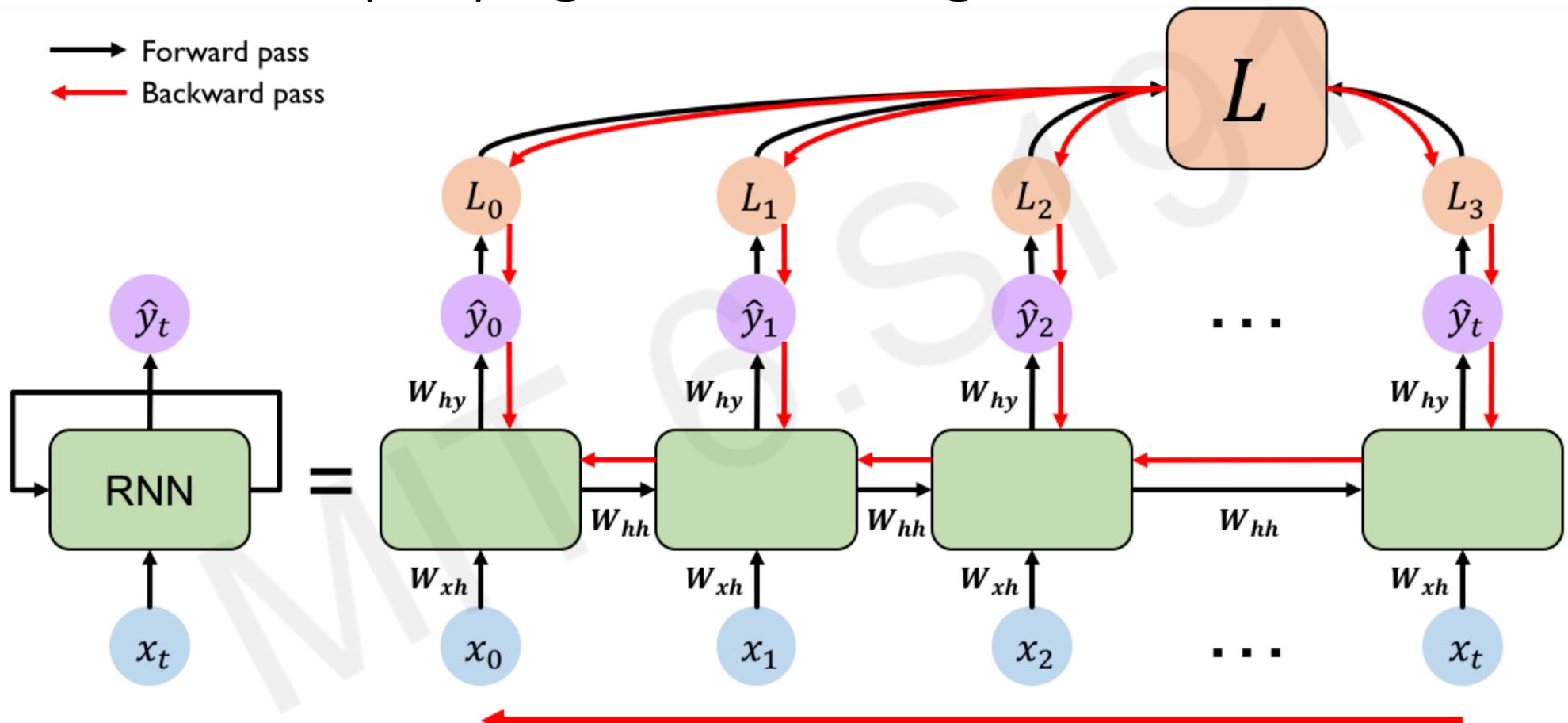
<https://towardsdatascience.com/@learnedvector>

Abdelhak Mahmoudi

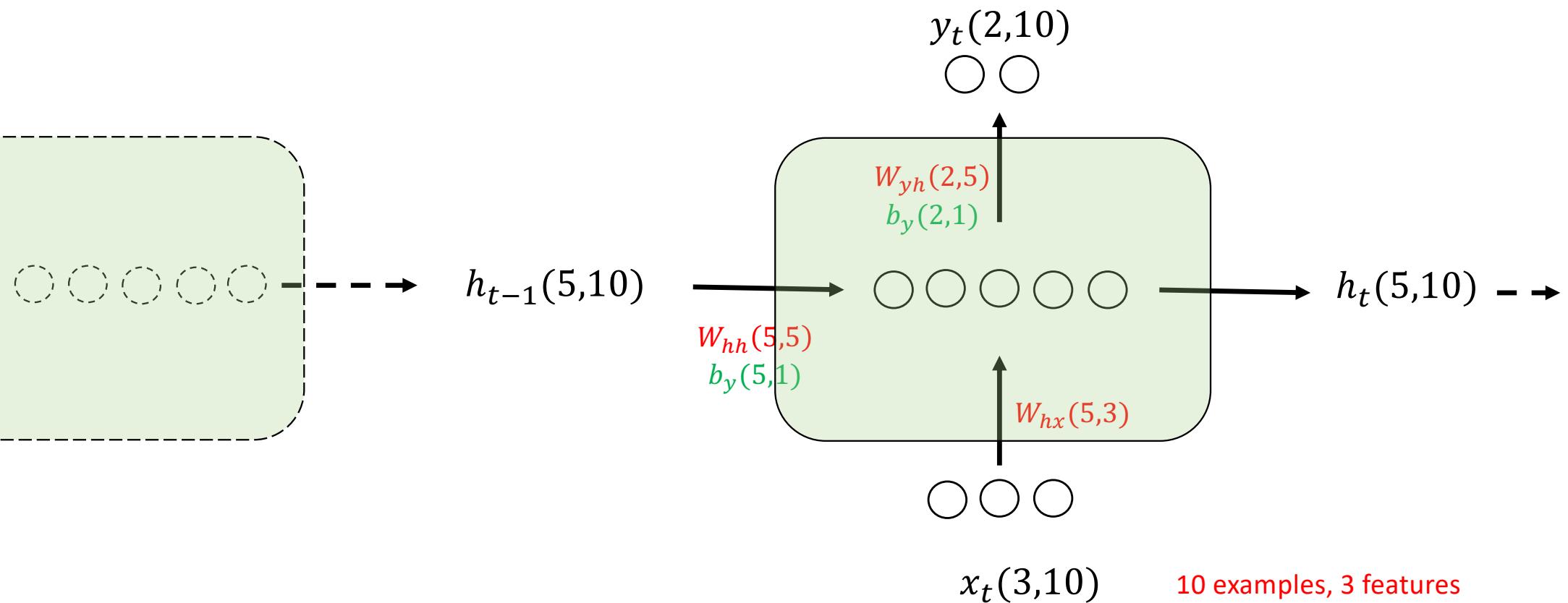
# RNN Architecture



# RNN : Backpropagation through time

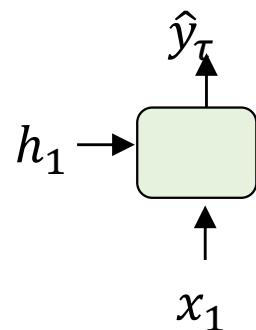


# RNN - Dimensions

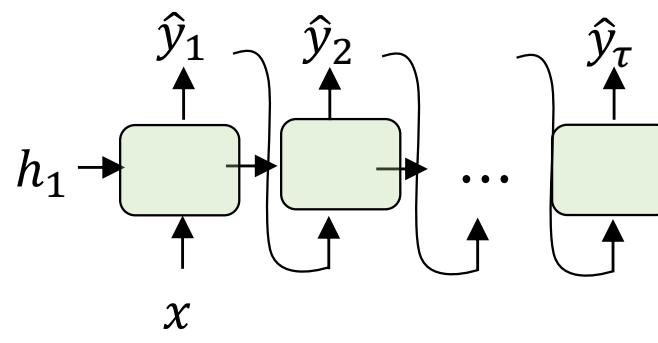


# RNN – Different Architectures

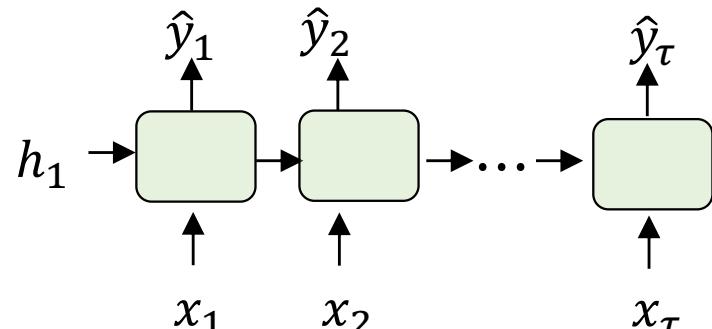
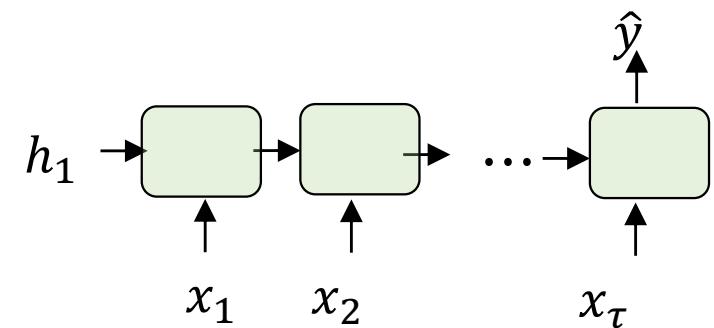
One to one



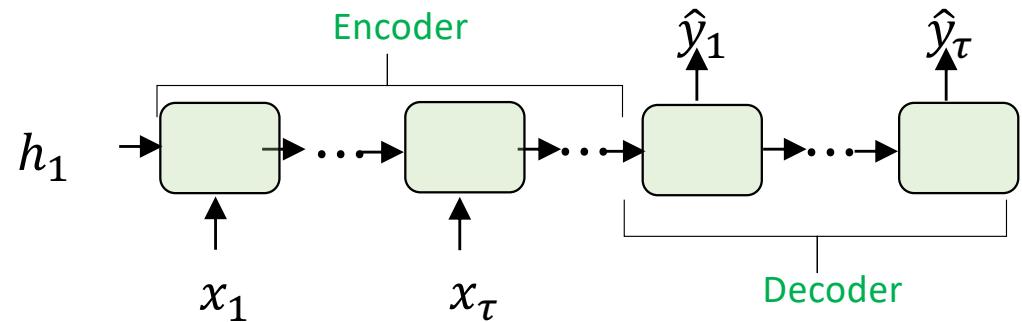
One to many  
(Music/ Text Generation)



Many to one  
(Sentiment)



Many to many

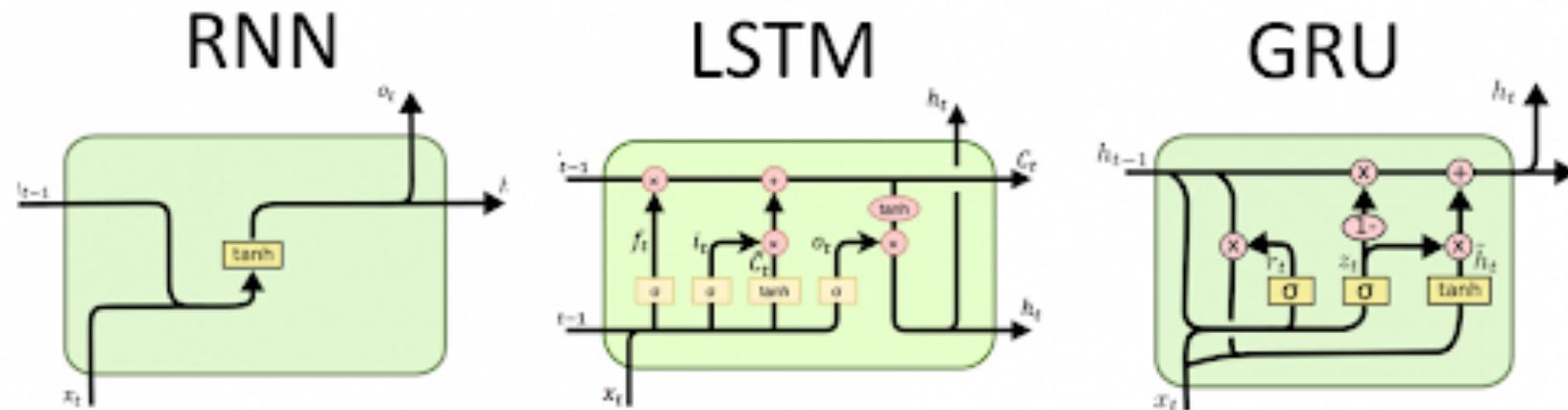


Many to many  
(Translation)

# Vanishing/Exploding gradients

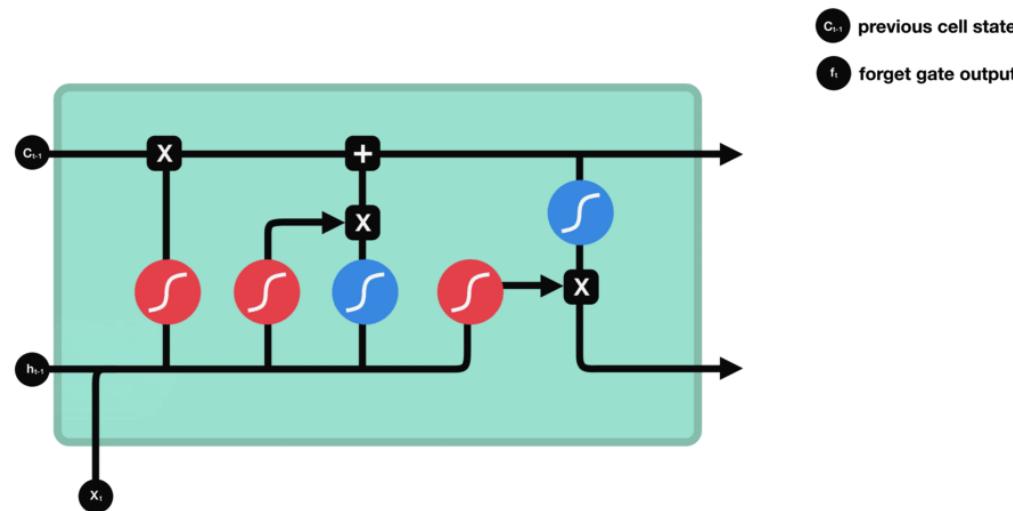
- $W^{[l]} =: W^{[l]} - \alpha \frac{\partial L}{\partial W^{[l]}}$
- $W^{[l]} < 1 \rightarrow \frac{\partial L}{\partial W^{[l]}} < 1 \rightarrow$  Vanishing  $\rightarrow$  slow down training
- $W^{[l]} > 1 \rightarrow \frac{\partial L}{\partial W^{[l]}} > 1 \rightarrow$  Exploding  $\rightarrow$  divergence
- Solution
  - Batch normalization
  - Random Weights Initialization
  - Use Gradient Clipping
  - **Use gated cells GRU, LSTM**

# RNN, LSTM, GRU

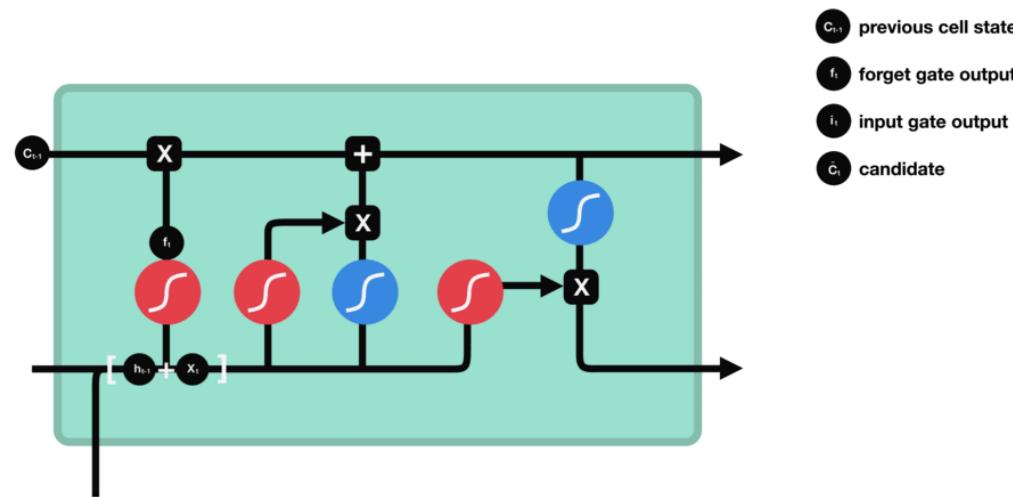


# Long Short Term Memory (LSTM)

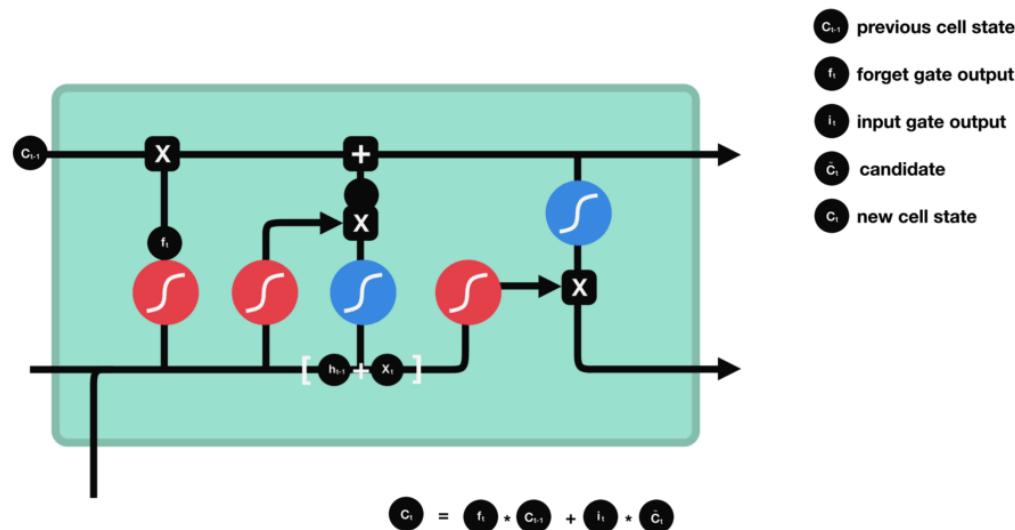
Sepp Hochreiter, 1991 !



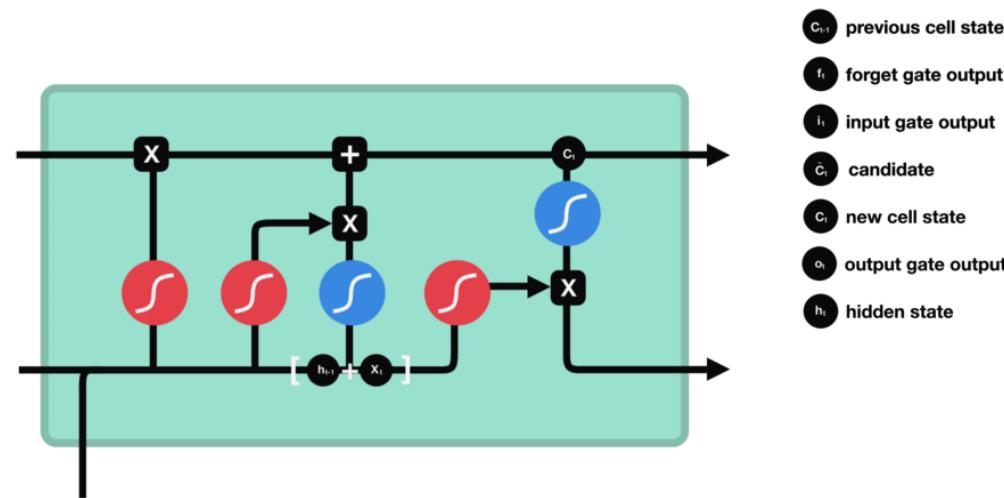
# Long Short Term Memory (LSTM)



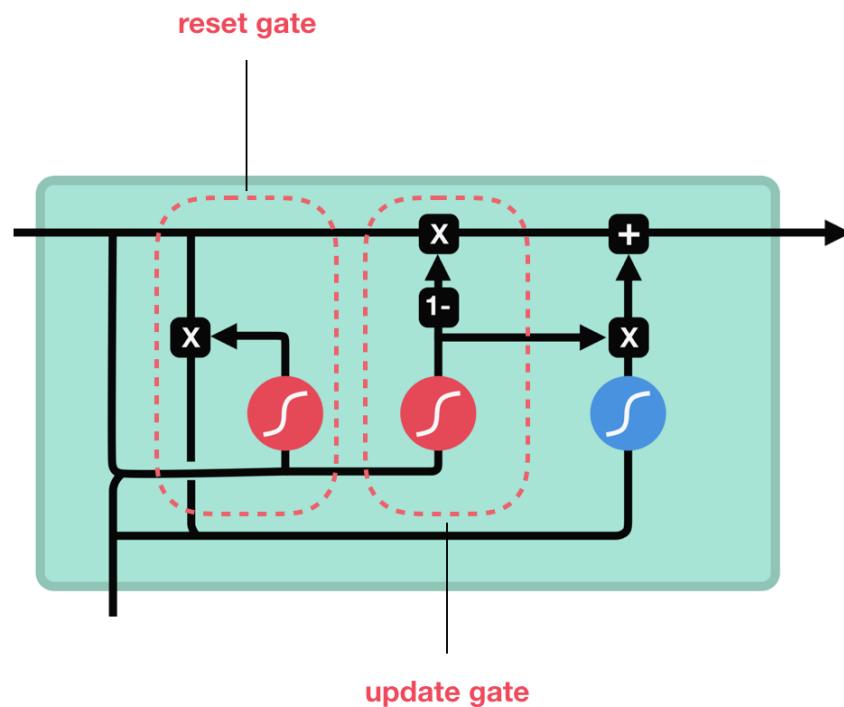
# Long Short Term Memory (LSTM)



# Long Short Term Memory (LSTM)

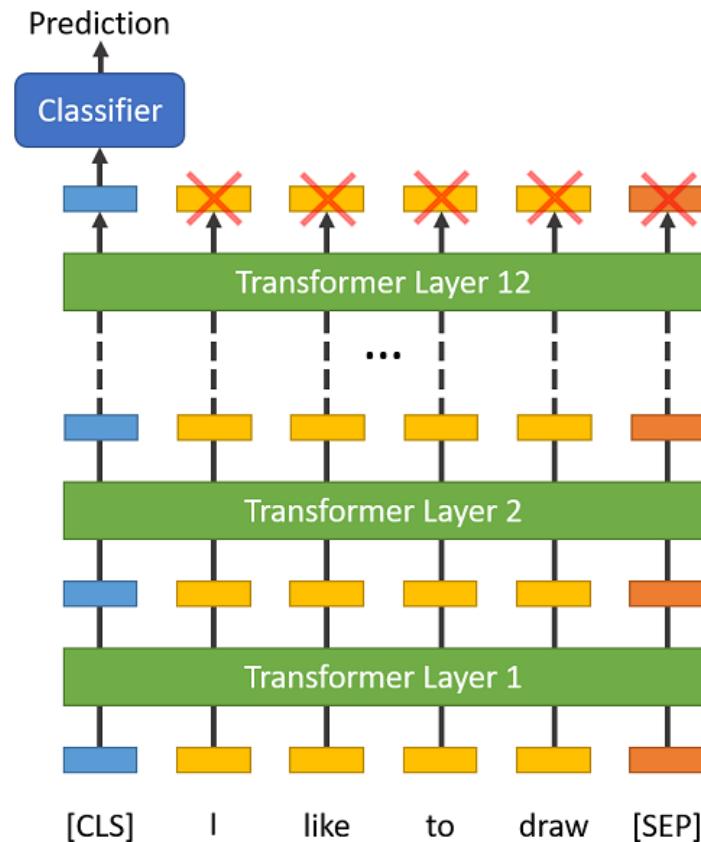


# Gated Recurrent Units



# Seq2Seq for Natural Language

- RNN
- LSTM
- Word2Vec
- GRU
- Attention
- Bidirectional LSTM
- **Transformer** (Attention is all what you need!)



# Transformers

- BERT (Bidirectional Encoder Representations from Transformers)
  - A Lite BERT (ALBERT)
- Generative Pre-Training (GPT)
  - GPT2
- ELMo (Embeddings from Language Models)

