

Deep Learning and Modern Natural Language Processing

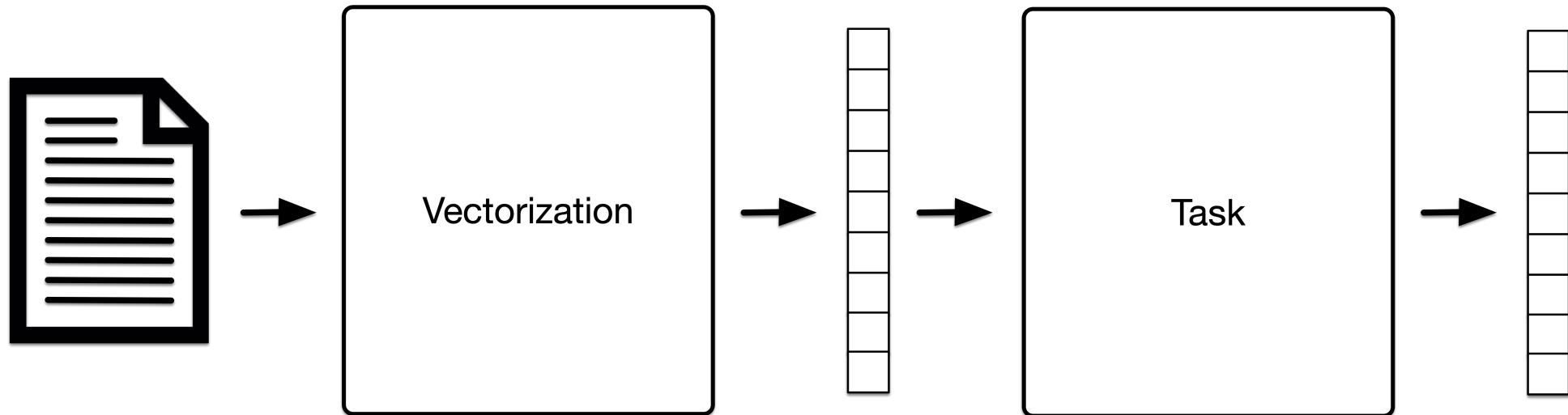
Zachary S. Brown

Outline

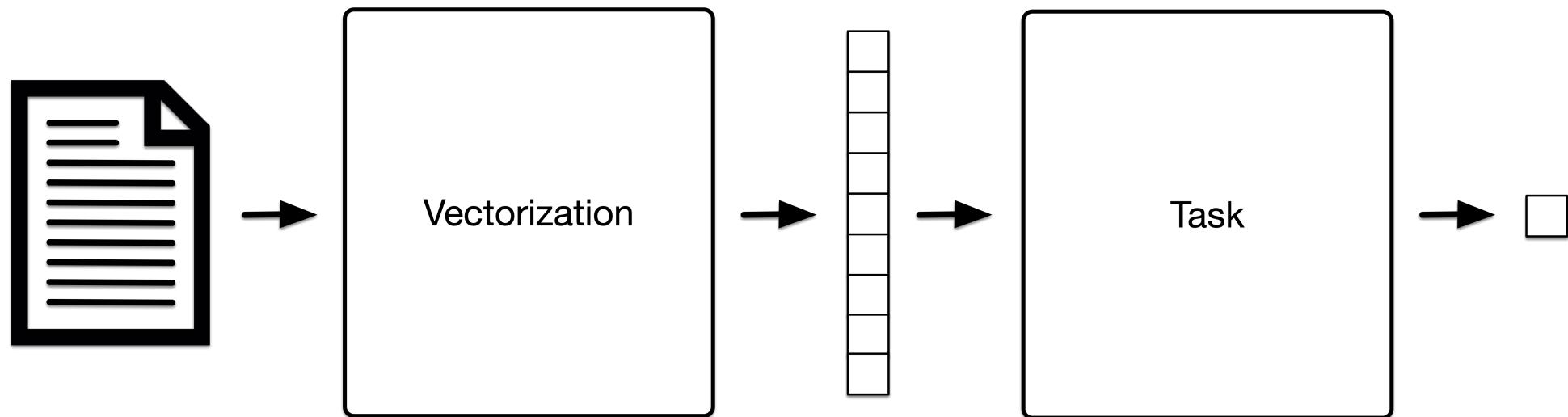
0. NLP Problem Structure
1. Text Classification and the Perceptron
2. Vectorization and Classification with RNNs
3. POS Tagging with RNNs
4. Sequence to Sequence Modeling

NLP Problem Structure

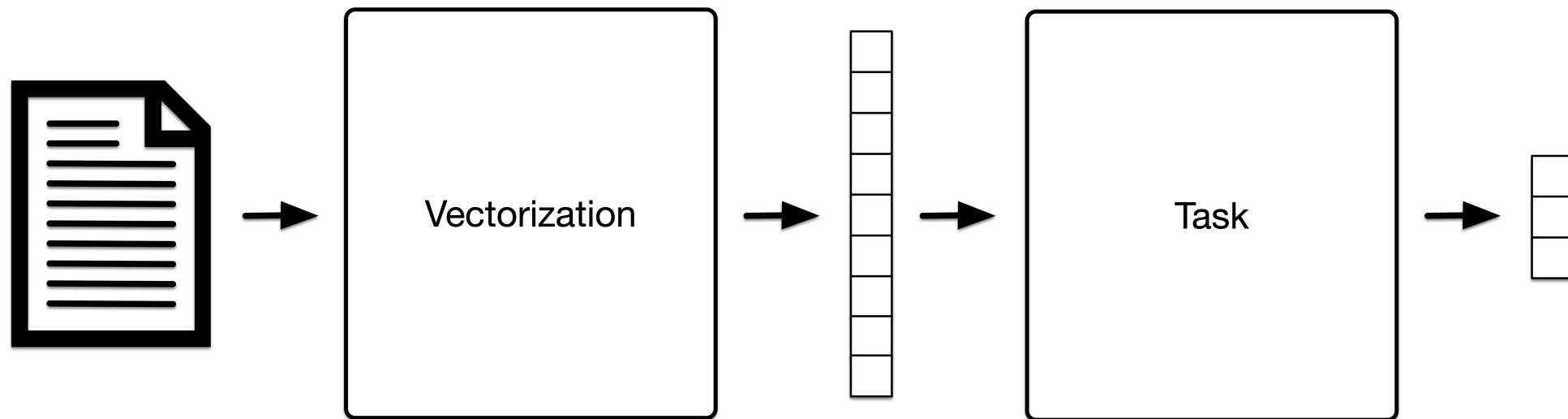
General Problem Structure



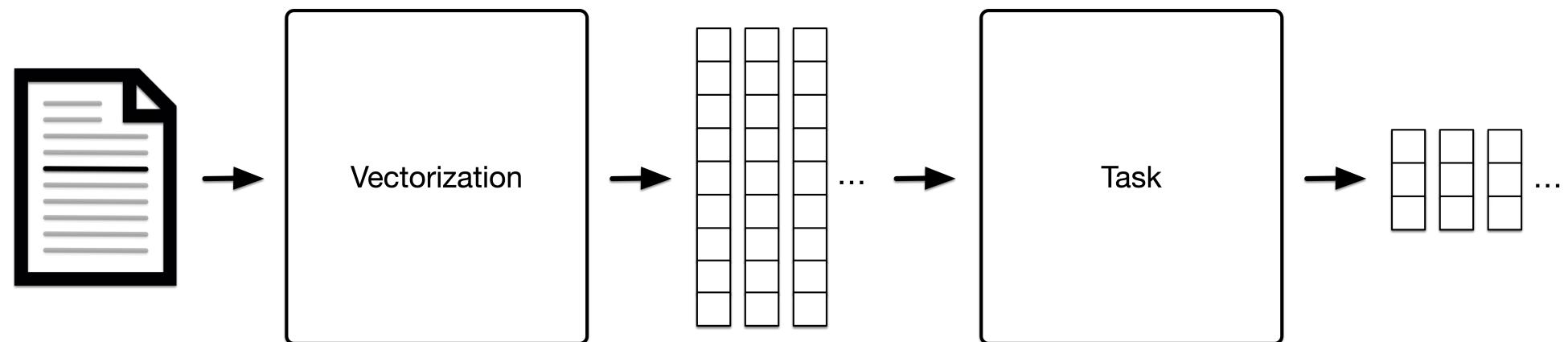
Binary Document Classification



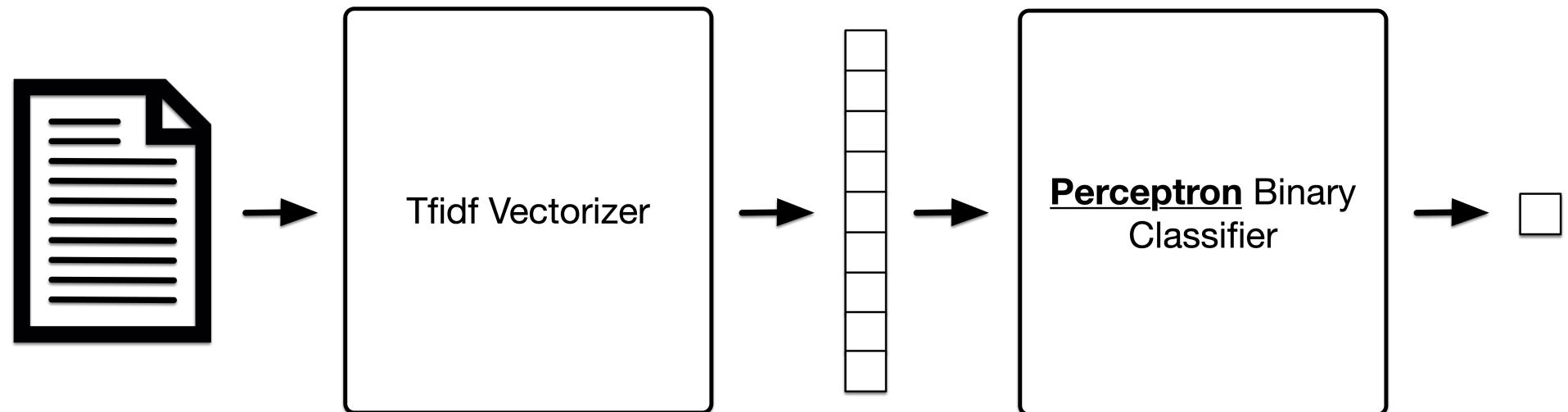
Multi-class Document Classification



Multi-class *Sequence* Classification



Starting Easy: Neural Net with Traditional Vectorization

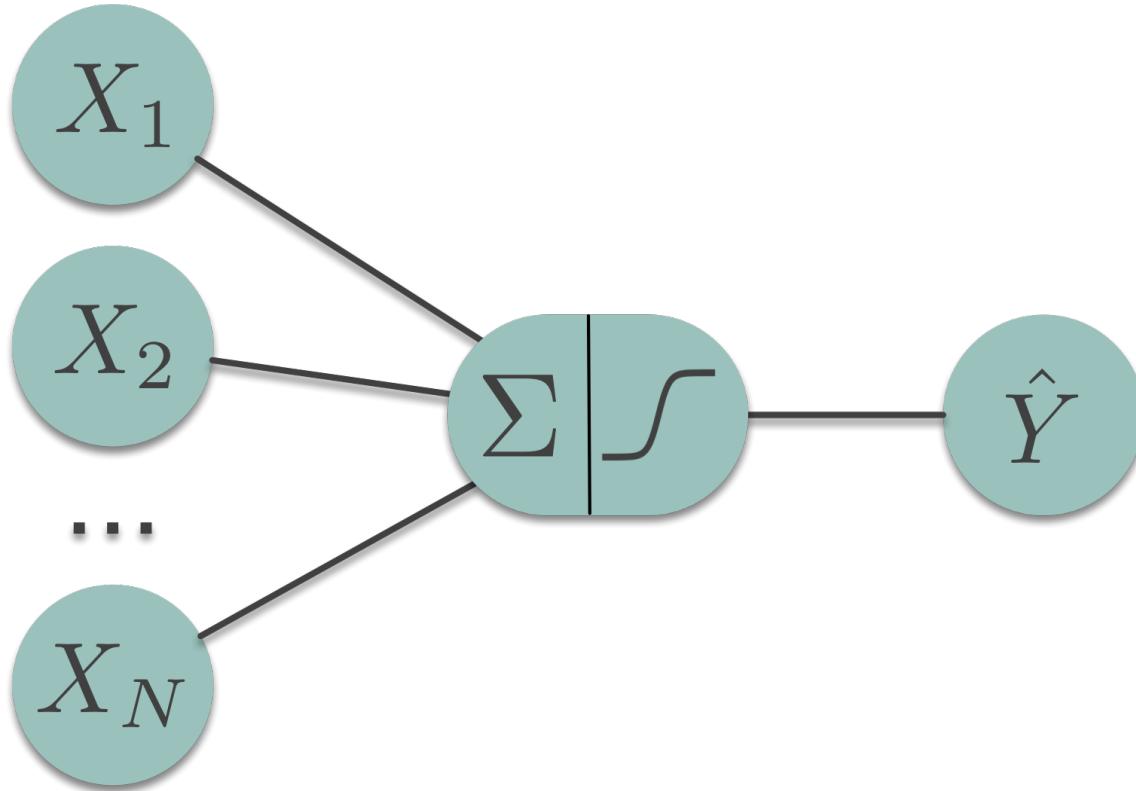


Text Classification and the Perceptron

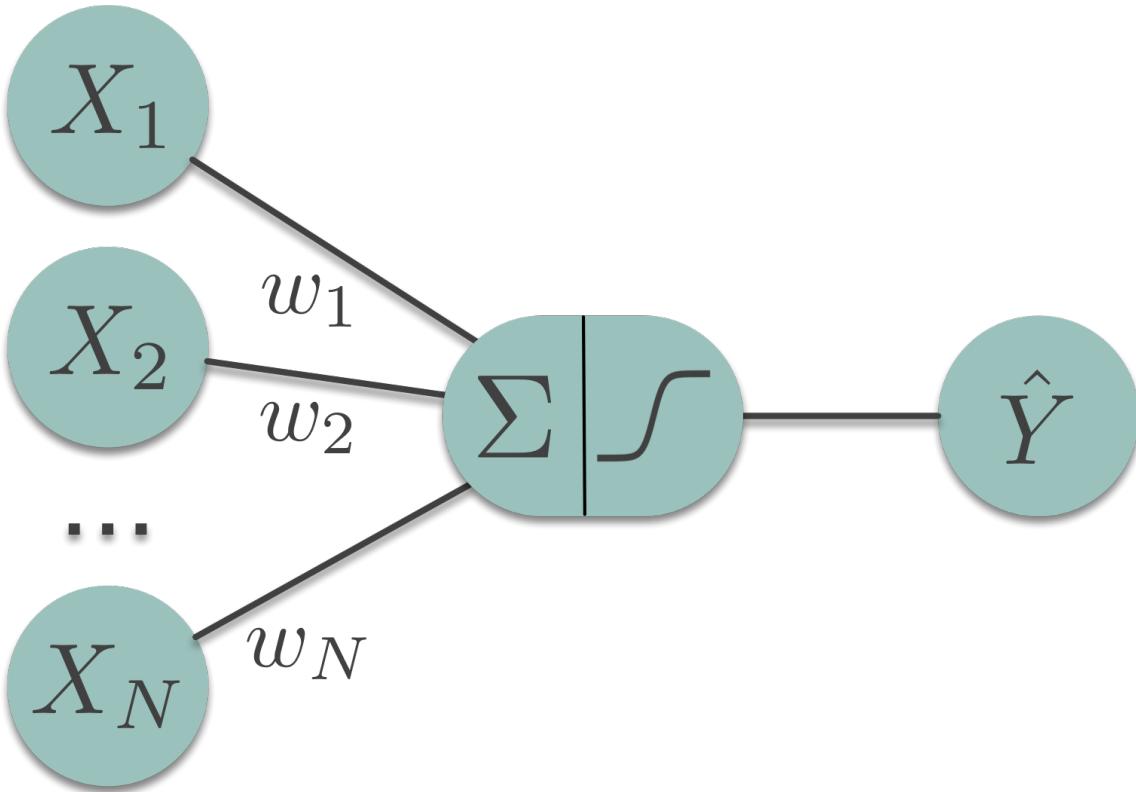
Topics

- The perceptron and neural network optimization
- Example

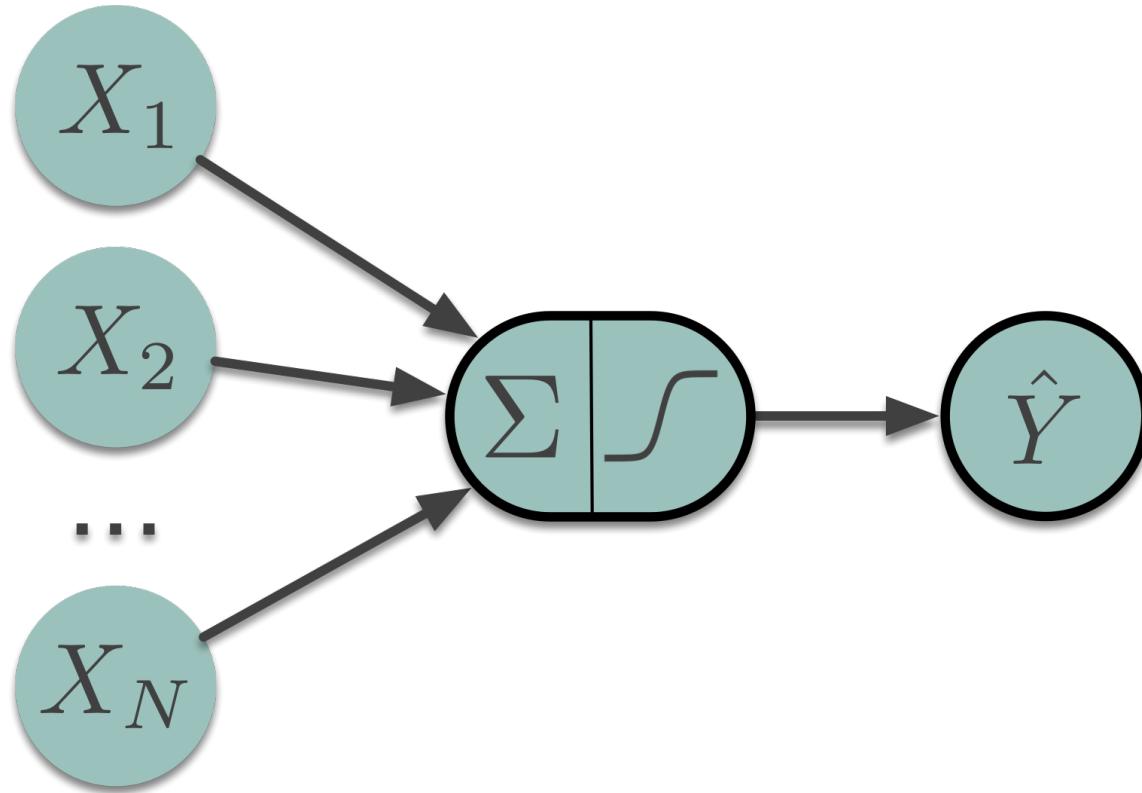
The Perceptron



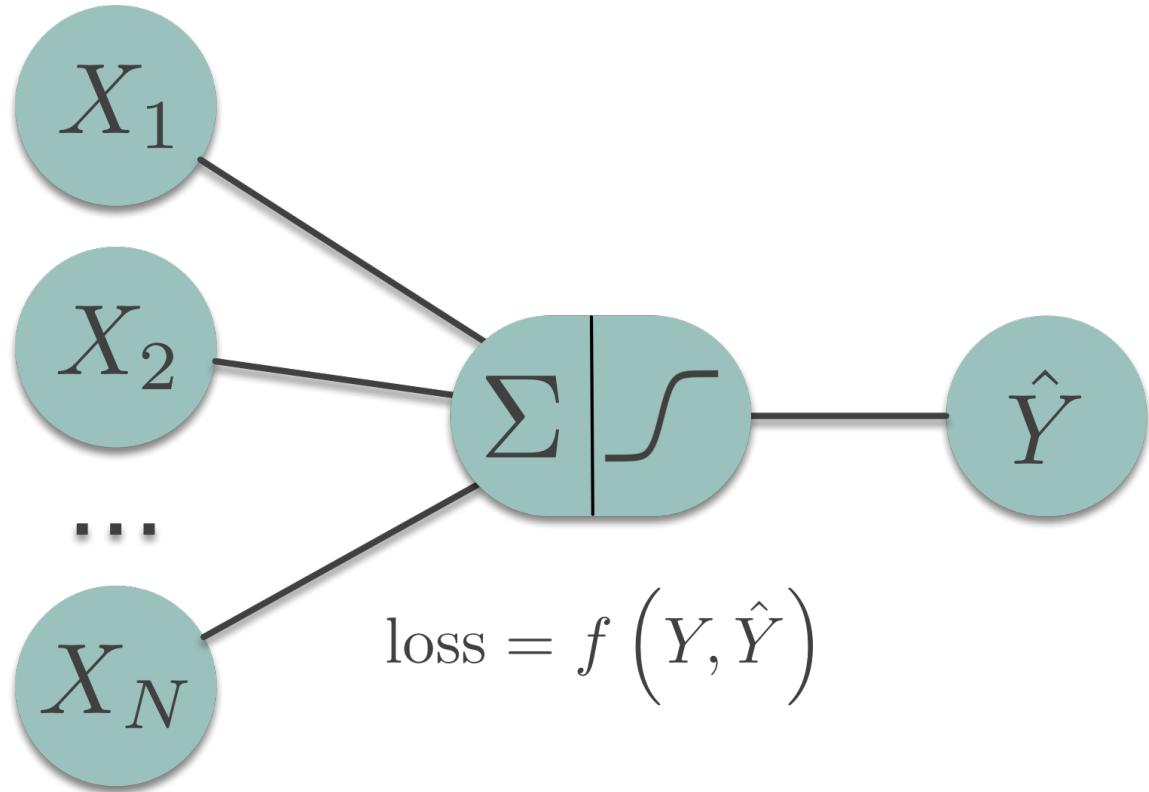
Weights



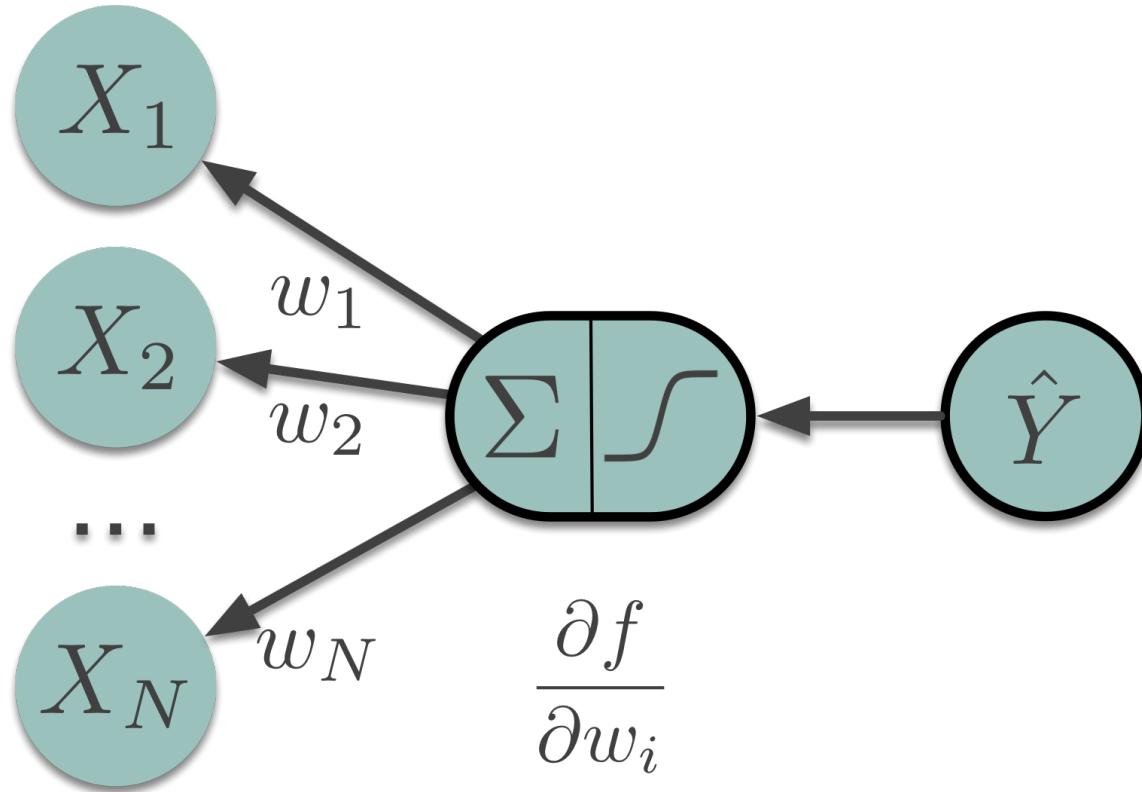
Forward Pass



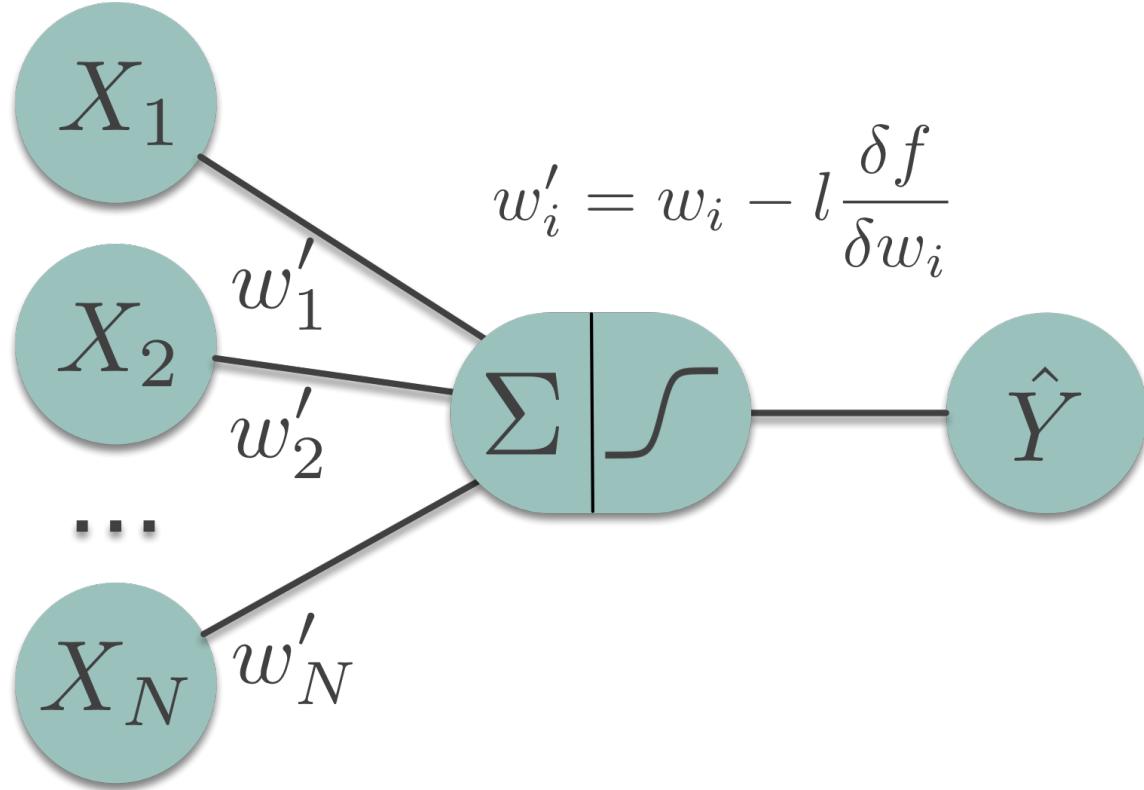
Loss



Calculate Gradients



Update Weights



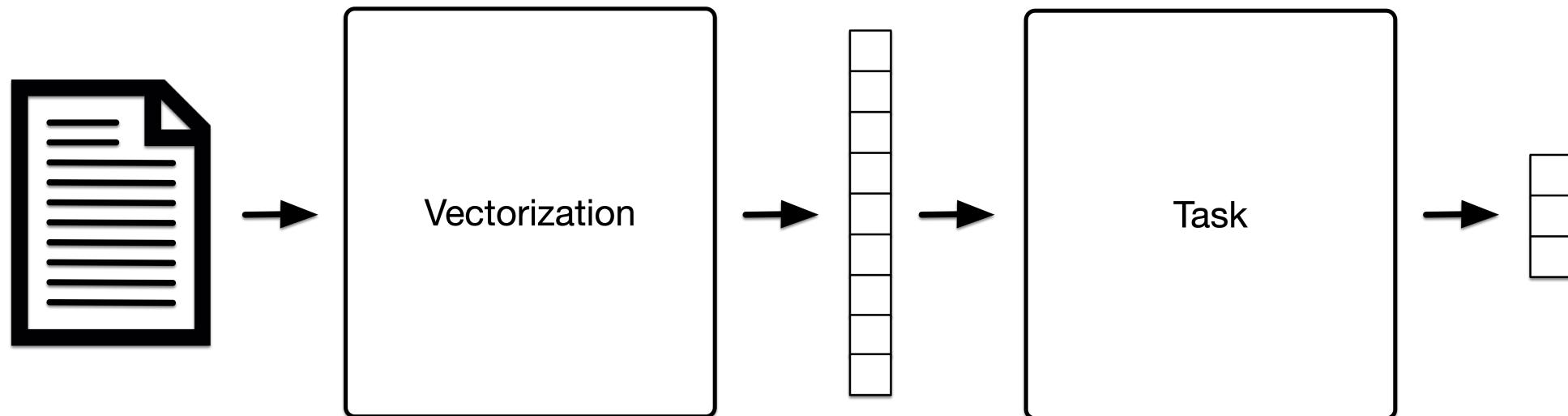
Perceptron Example

Vectorization and Classification with RNNs

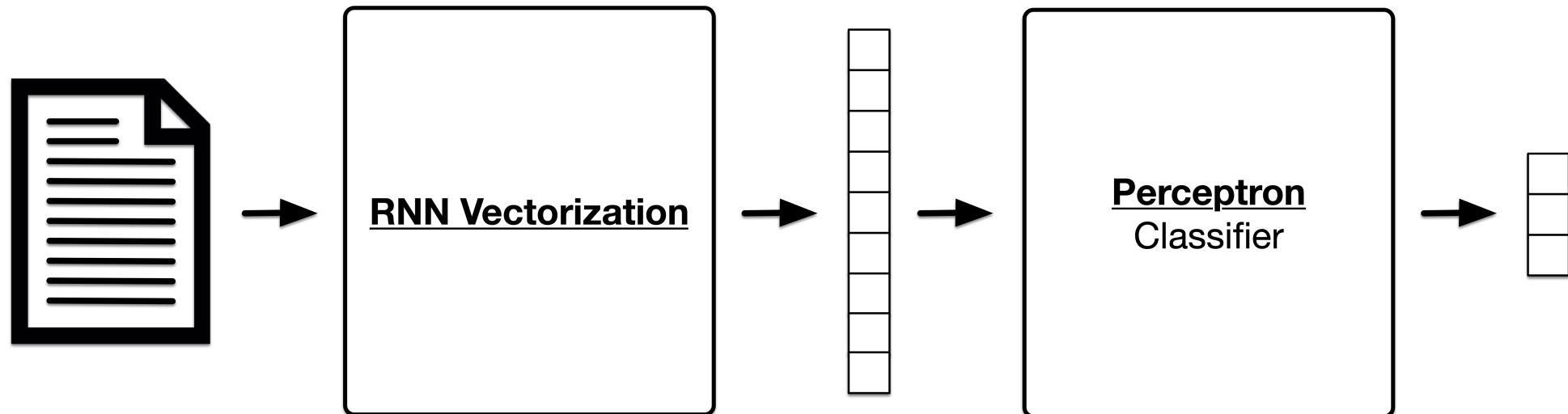
Topics

- RNN Encoding for Text Classification
- Recurrent Neural Networks
- Word Embeddings
- Example

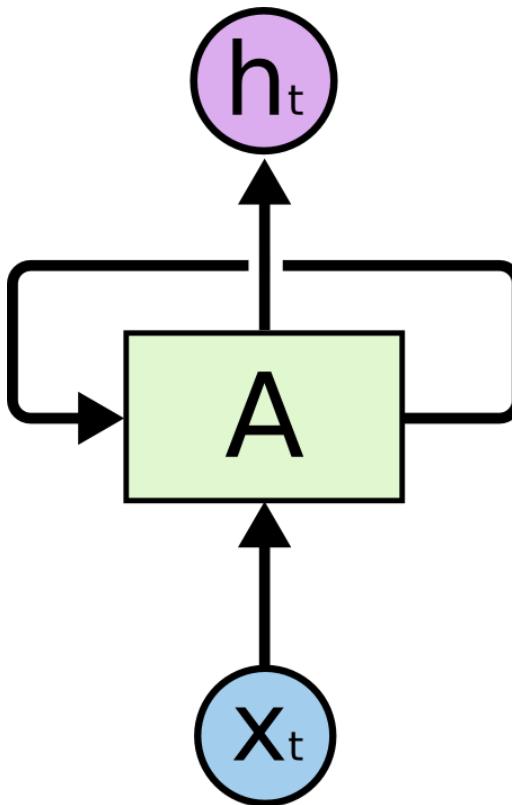
Multi-class Document Classification



Text Vectorization with Recurrent Neural Networks

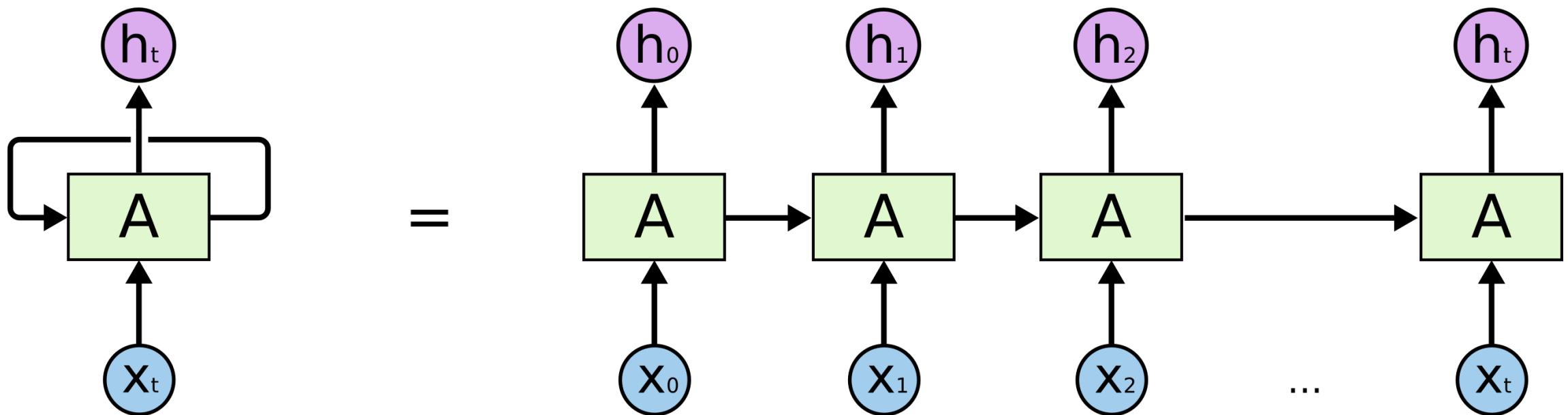


Recurrent Neural Networks

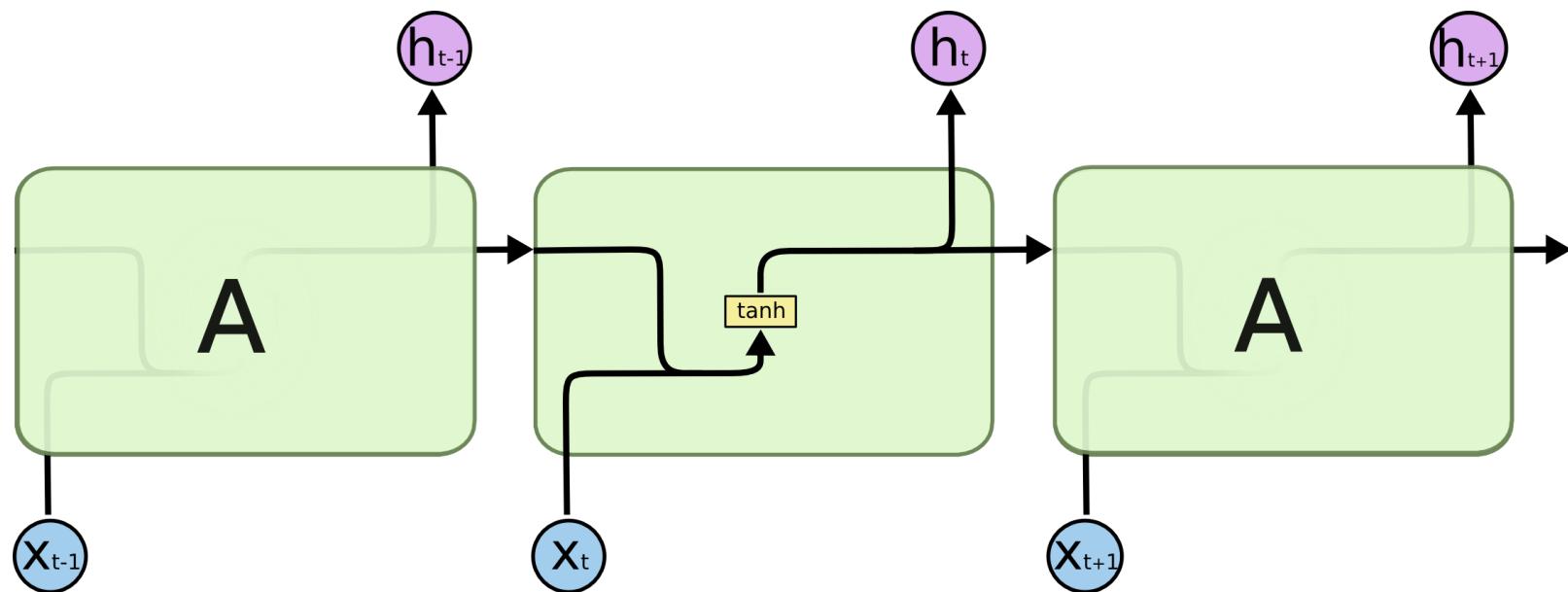


Images graciously sourced from [Understanding LSTM Networks](#) by Christopher Olah

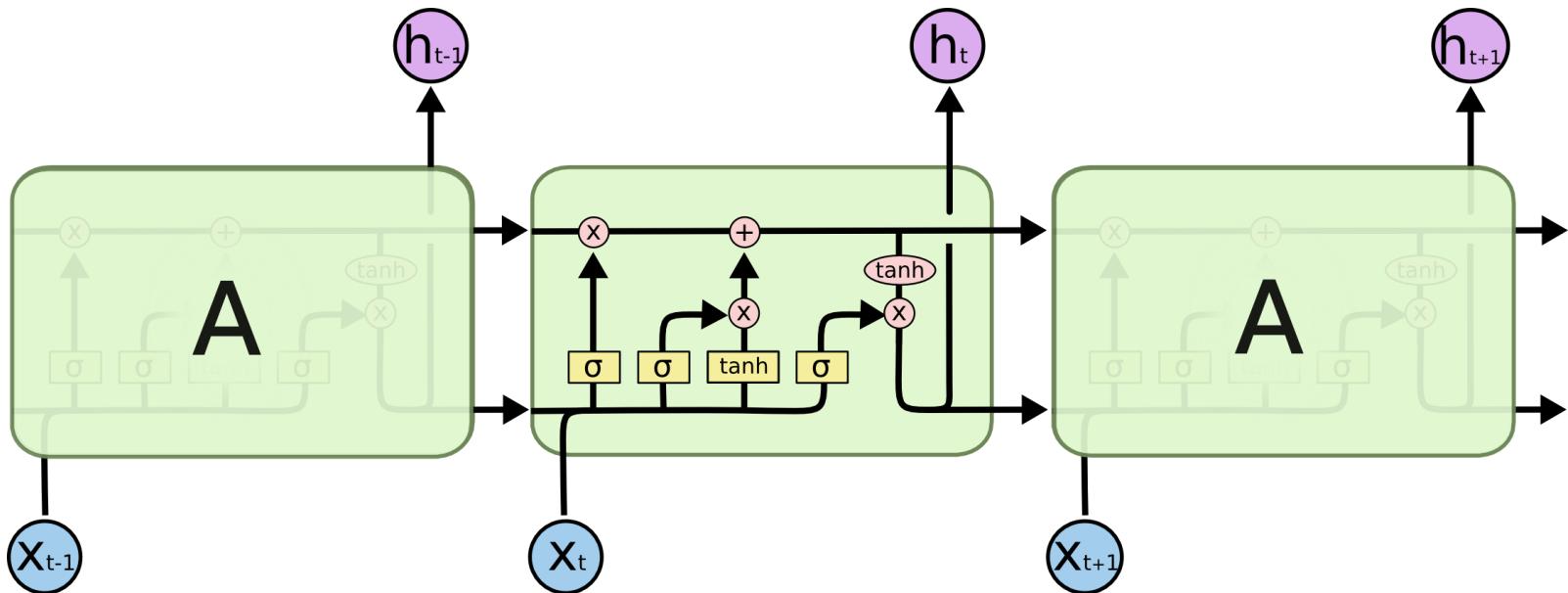
Recurrent Neural Networks (unrolled)



Vanilla Recurrent Neural Networks

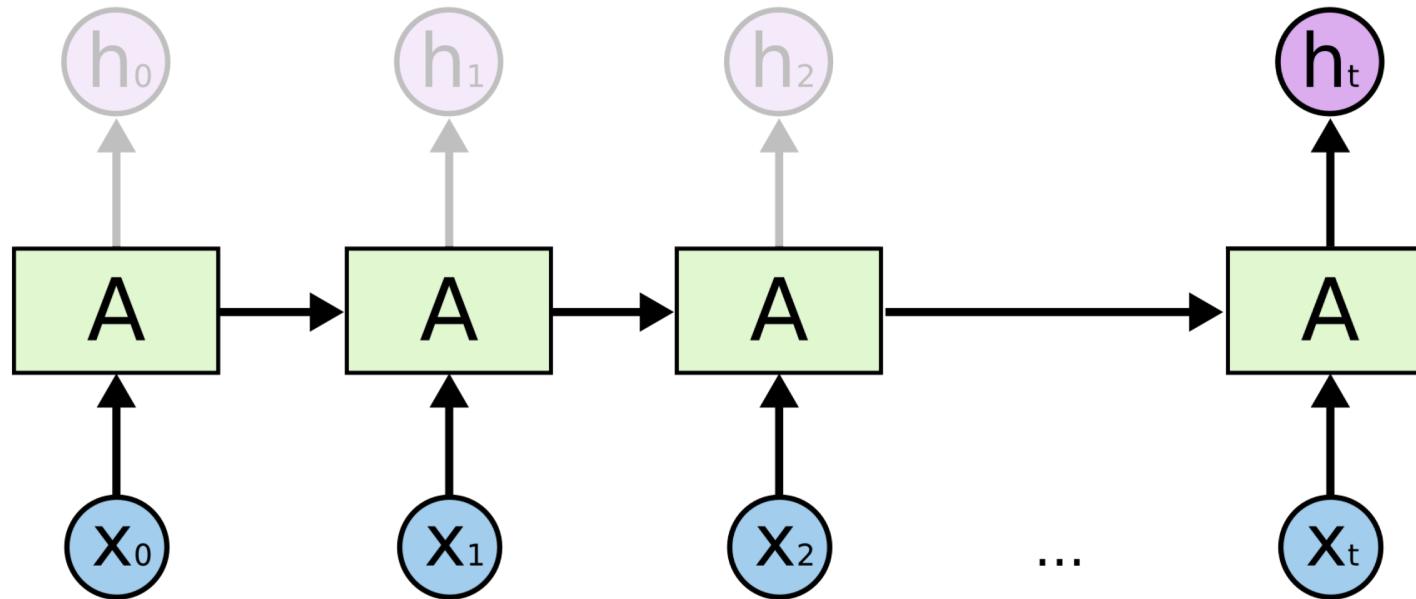


Long Short-term Memory (LSTM) Networks

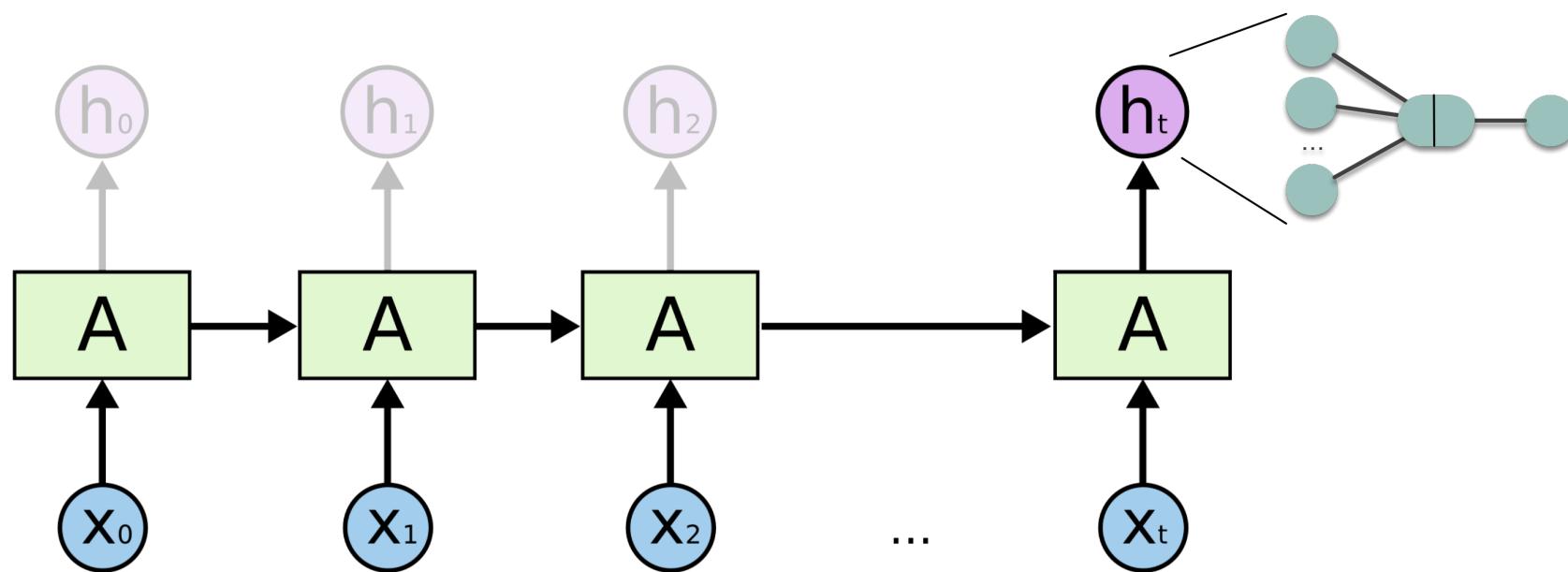


For a deeper dive into the necessity for and implementation of LSTM networks, see [Understanding LSTM Networks](#) by Christopher Olah

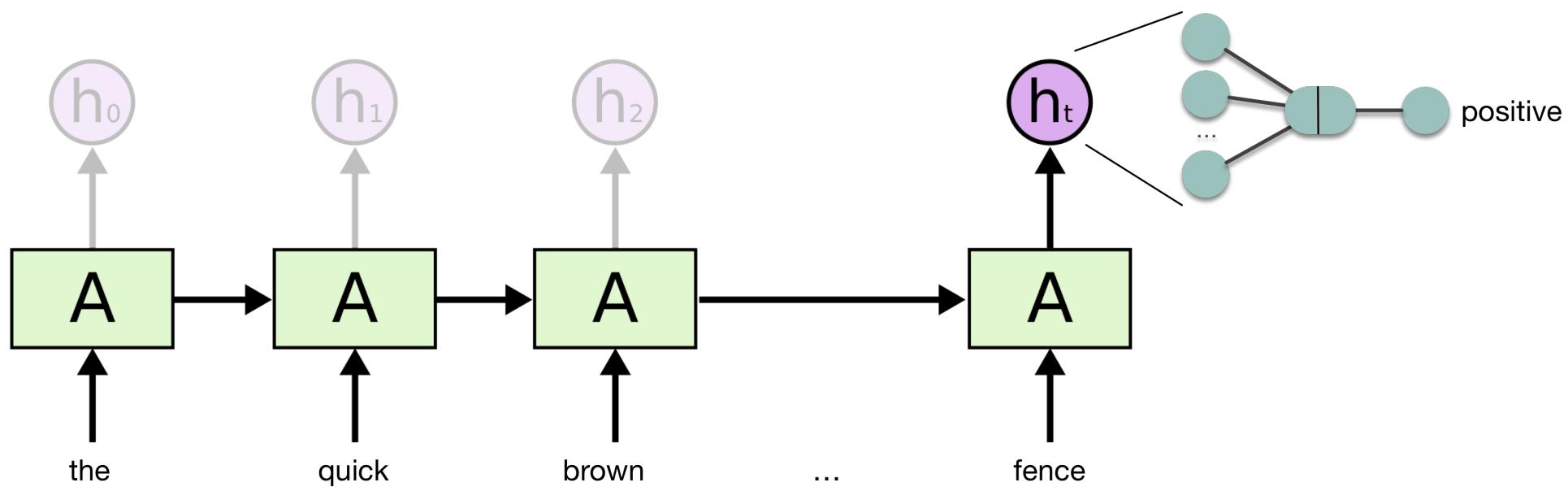
Encoding a Sequence to a single Vector



Sequence Classification



Sequence Classification



Wait, what about the inputs to the RNN??

Word Embeddings: Bag-of-Words vs. Dense Representations

apple	1								
ask		1							
baby			1						
ball				1					
banana					1				
child						1			
computer							1		
dog								1	
...								...	
woman									1

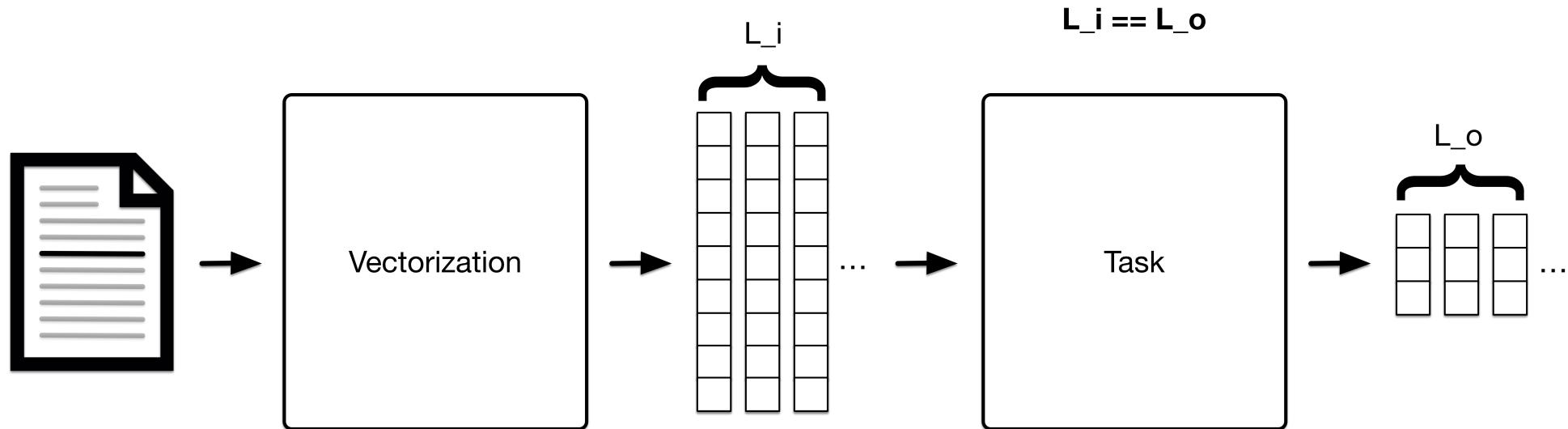


apple	0.23	0.00	0.74	0.78	0.01	0.7	...	0.4
ask	0.13	0.03	0.9	0.3	0.5	1.2	...	0.16
baby	0.46	0.74	0.56	0.7	0.06	0.75	...	0.65
ball	0.85	0.65	0.48	0.75	0.58	0.59	...	0.54
banana	0.56	0.84	0.24	0.84	0.27	0.11	...	0.
child	0.09	0.25	0.15	0.55	0.36	0.15	...	0.17
computer	0.44	0.17	0.16	0.81	0.66	0.33	...	0.24
dog	0.19	0.06	0.77	0.57	0.23	0.31	...	0.22
...
woman	0.49	0.76	0.11	0.64	0.81	0.67	...	0.13

LSTM Text Encoding and Classification Example

LSTMs for Sequence to Sequence Modeling

Fixed Length Sequence to Sequence



Use Cases

Use Cases

- POS Tagging

Use Cases

- POS Tagging
- Named Entity Recognition

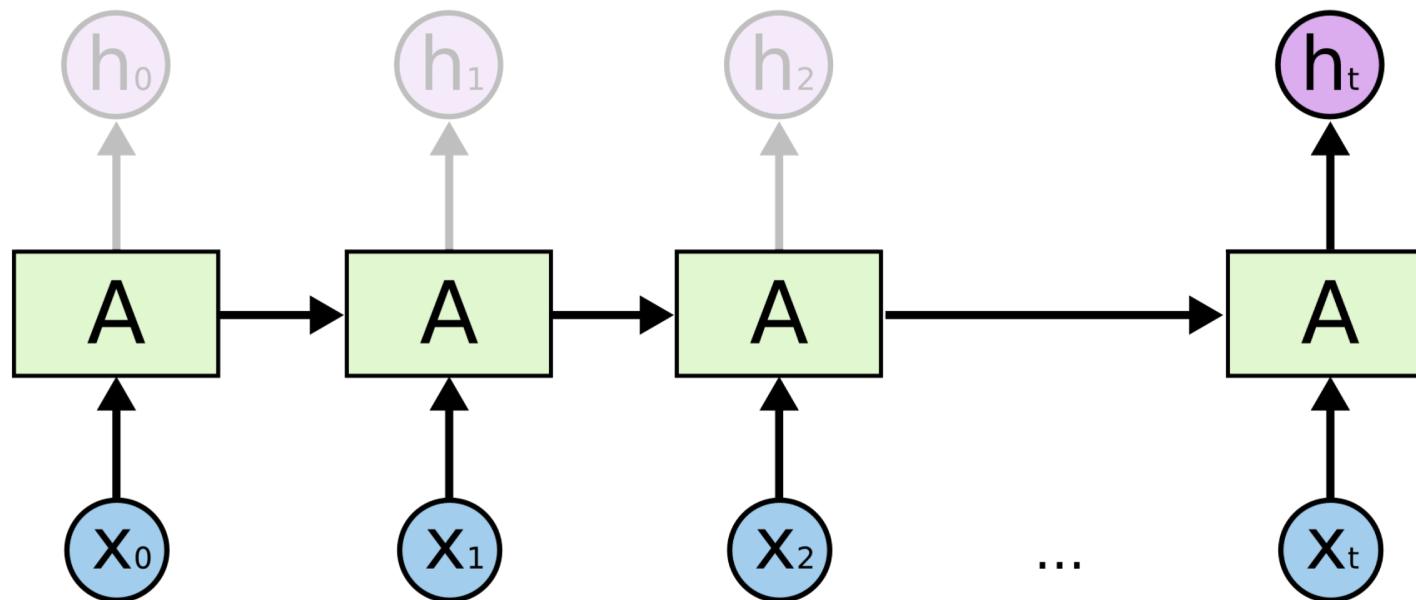
Use Cases

- POS Tagging
- Named Entity Recognition
- Extractive Summarization

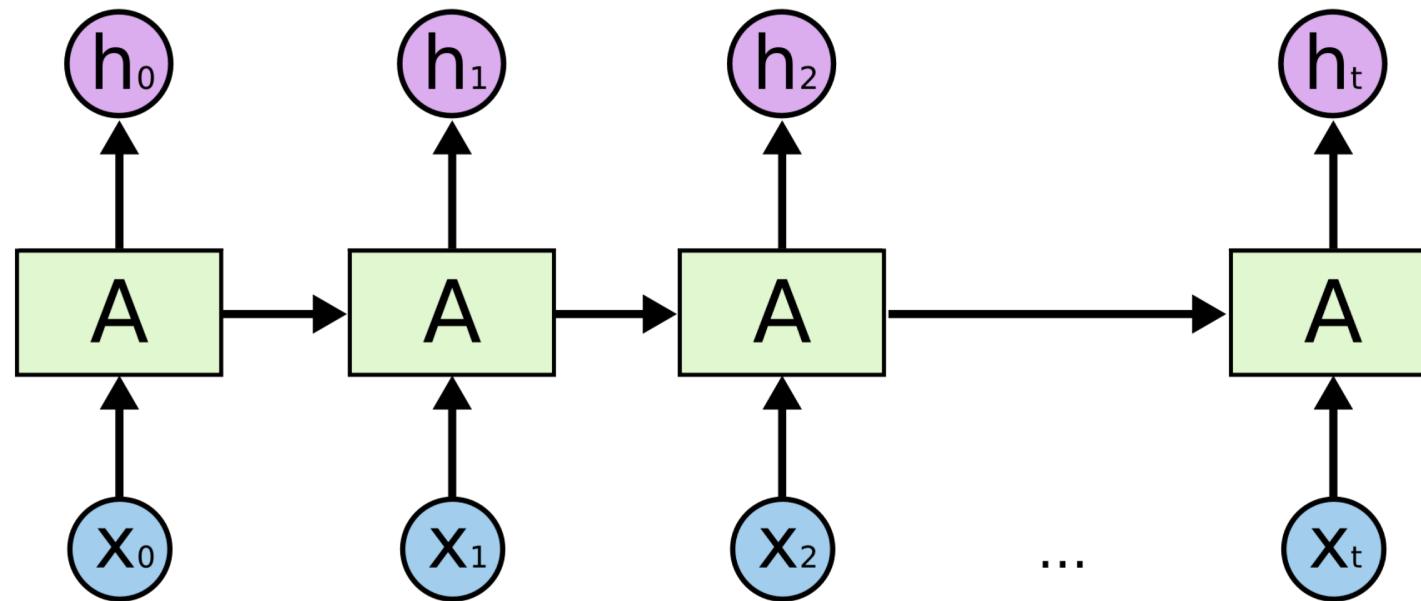
Use Cases

- POS Tagging
- Named Entity Recognition
- Extractive Summarization

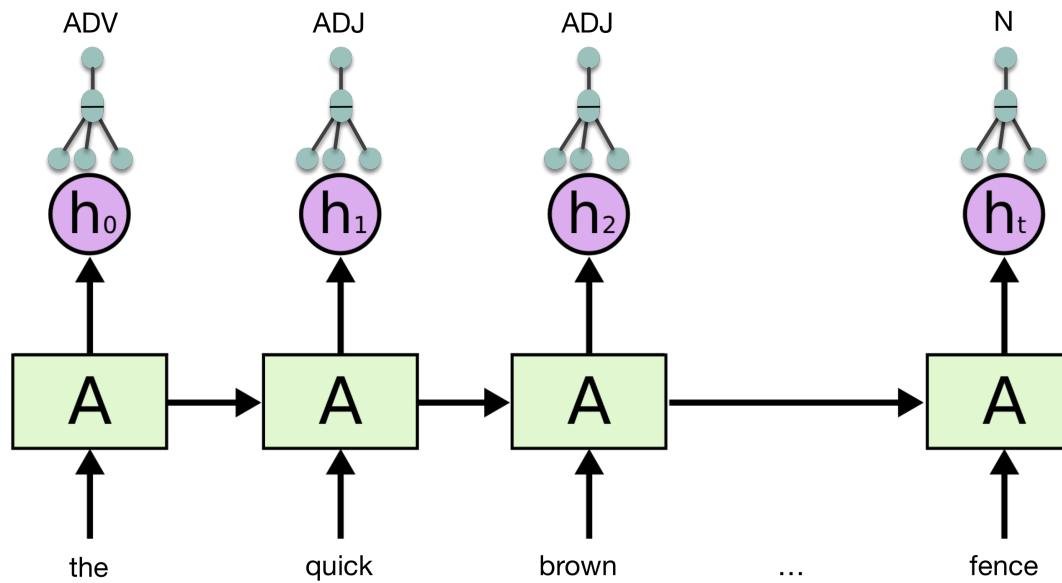
Before: Encoding a Sequence to a single Vector



Now: Encoding a Sequence to a *Sequence* of Vectors



Implementing Part-of-Speech Tagging

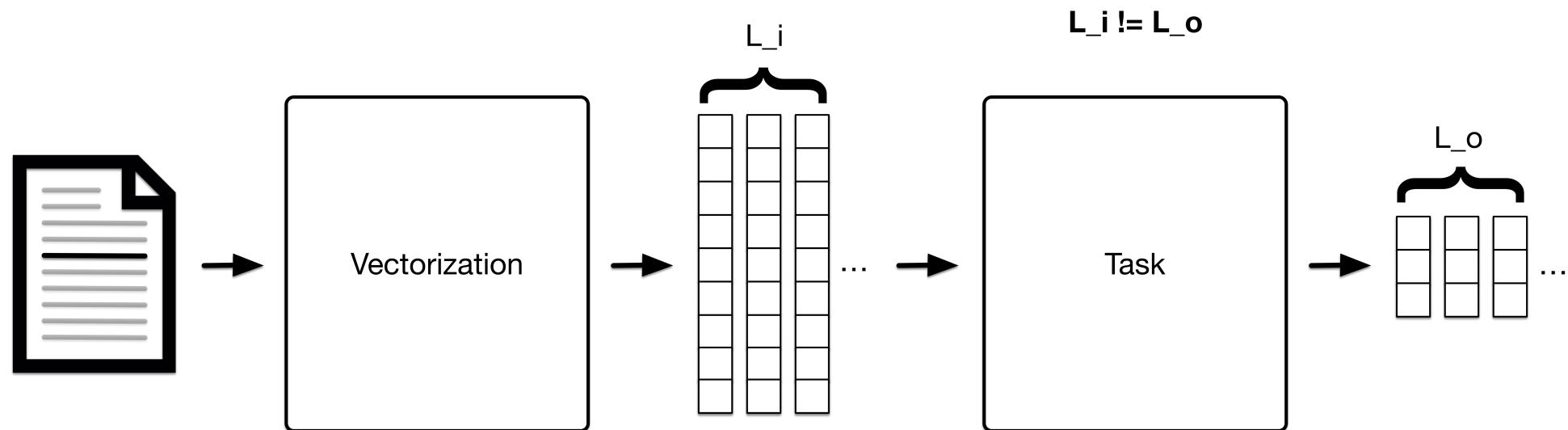


Sequence to Sequence Modeling

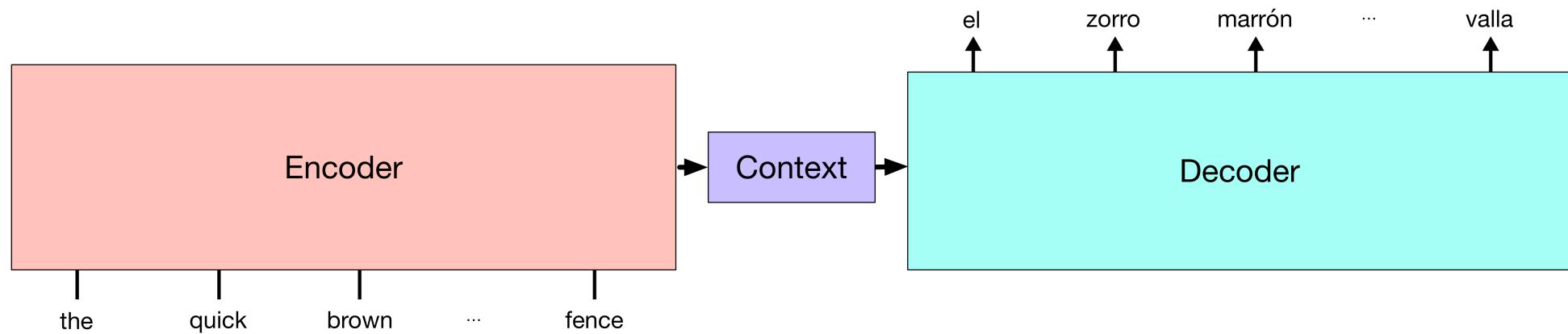
Topics

- Encoder-Decoder Architecture
- Neural Machine Translation

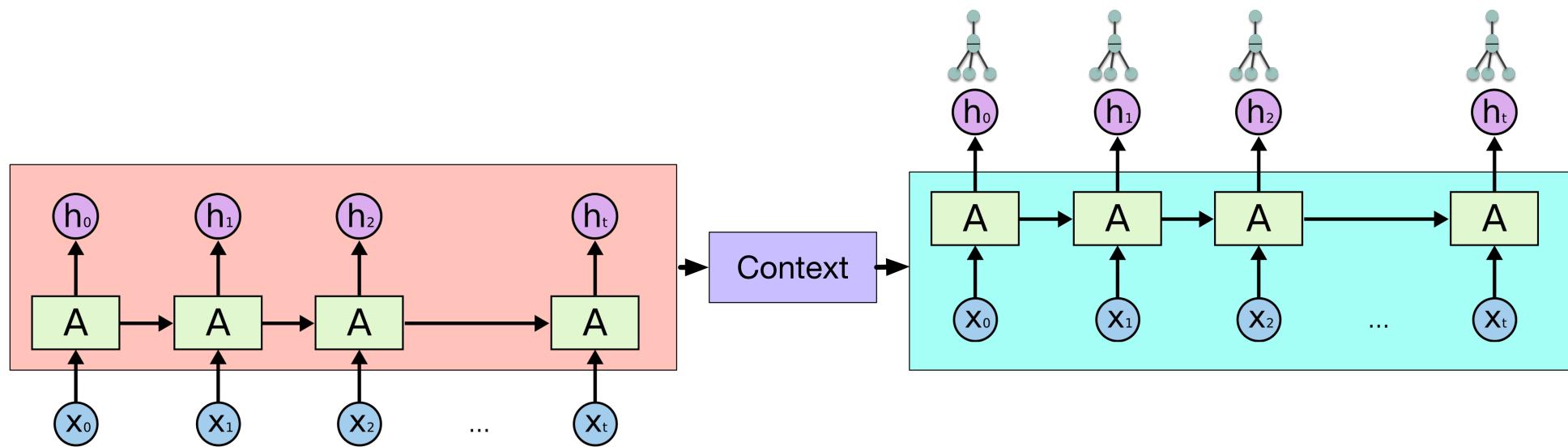
Variable Length Sequence to Sequence



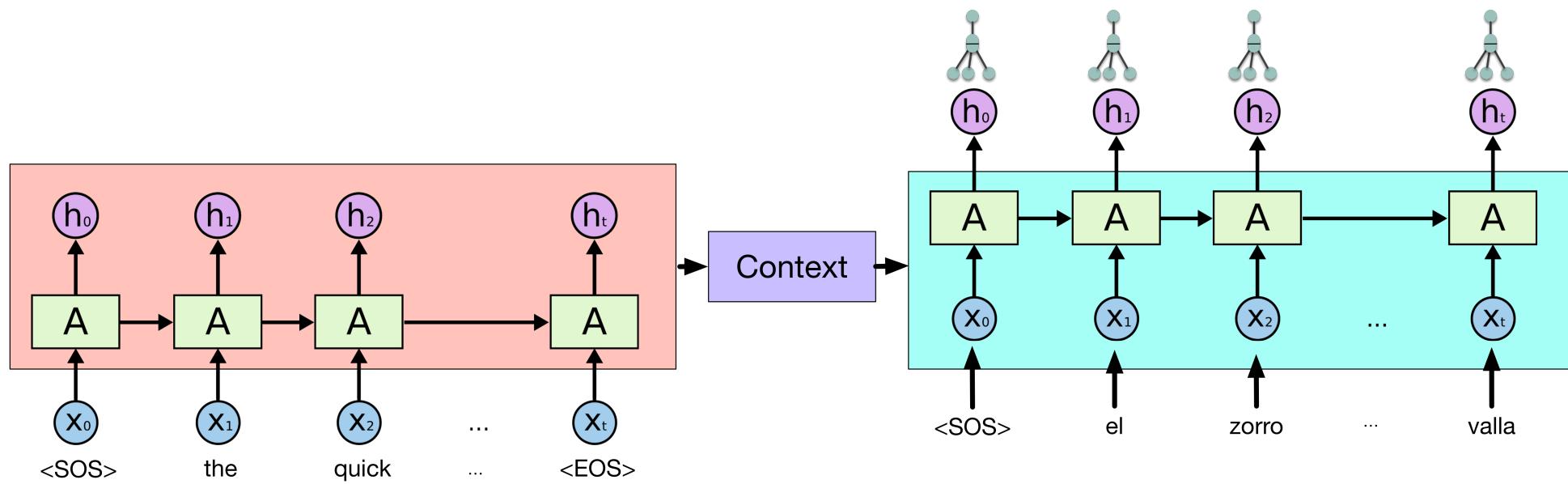
Sequence-to-Sequence Overview



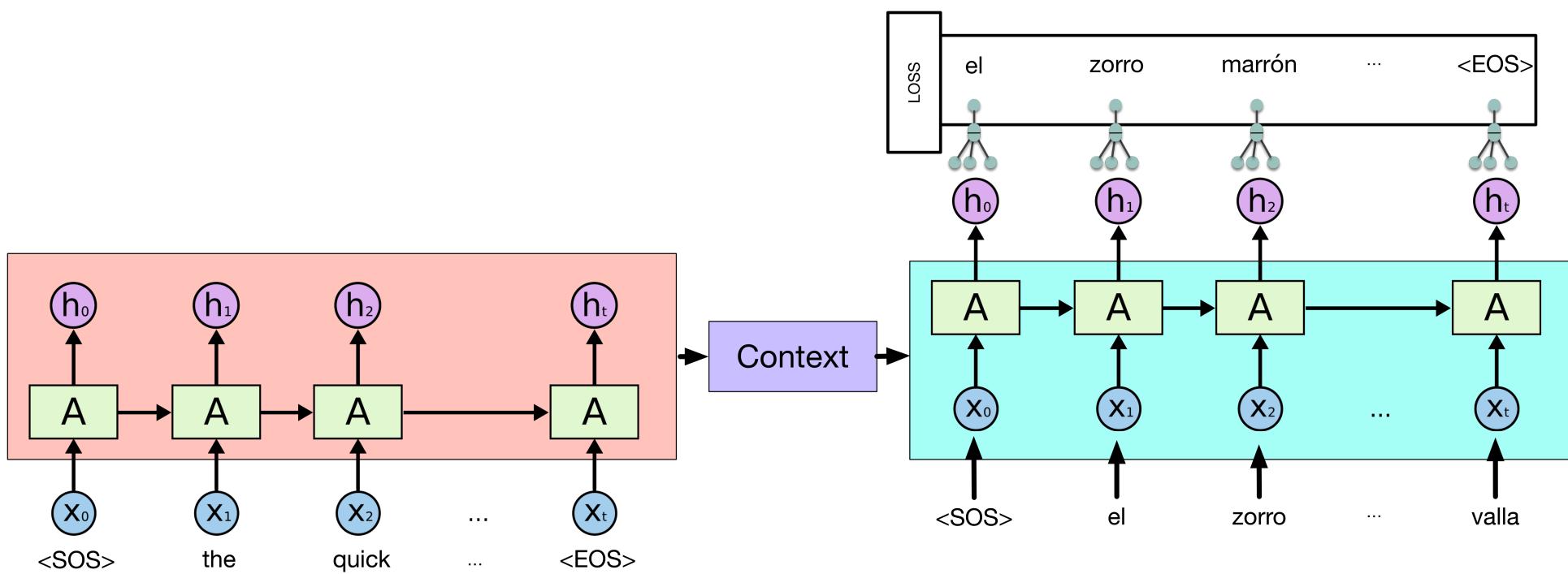
Encoder-Decoder Architecture



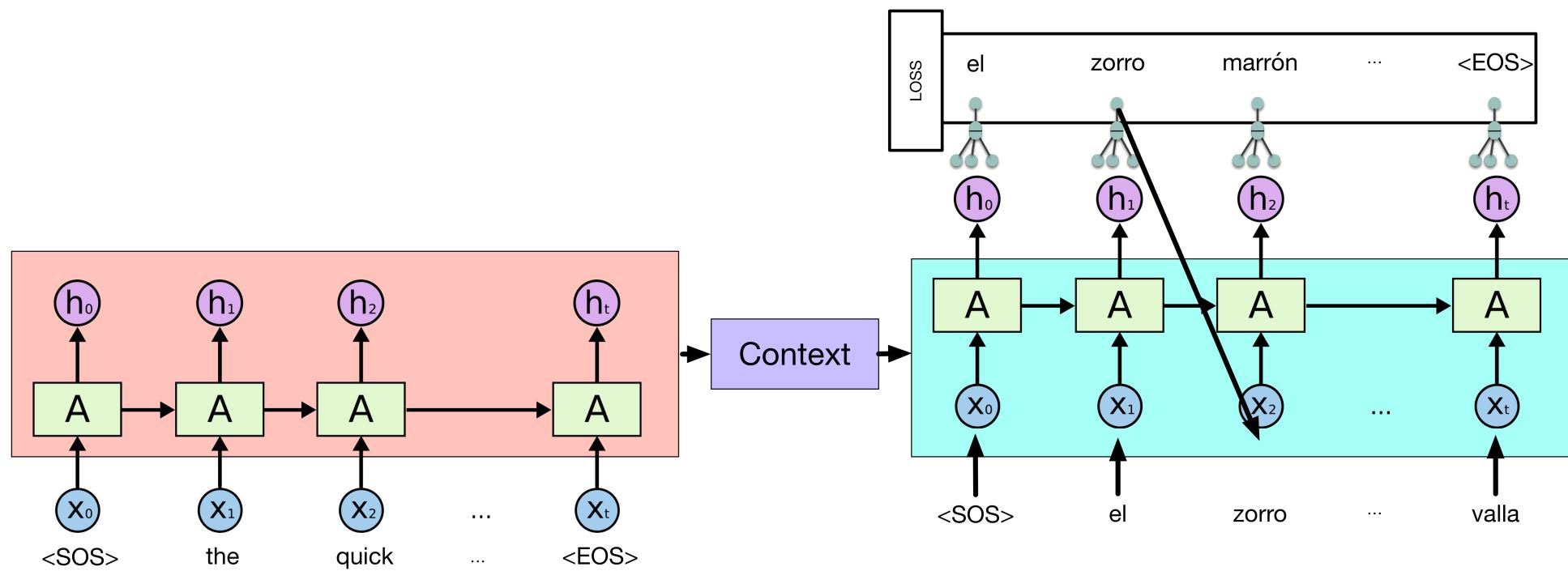
Sequence-to-Sequence Training: Input



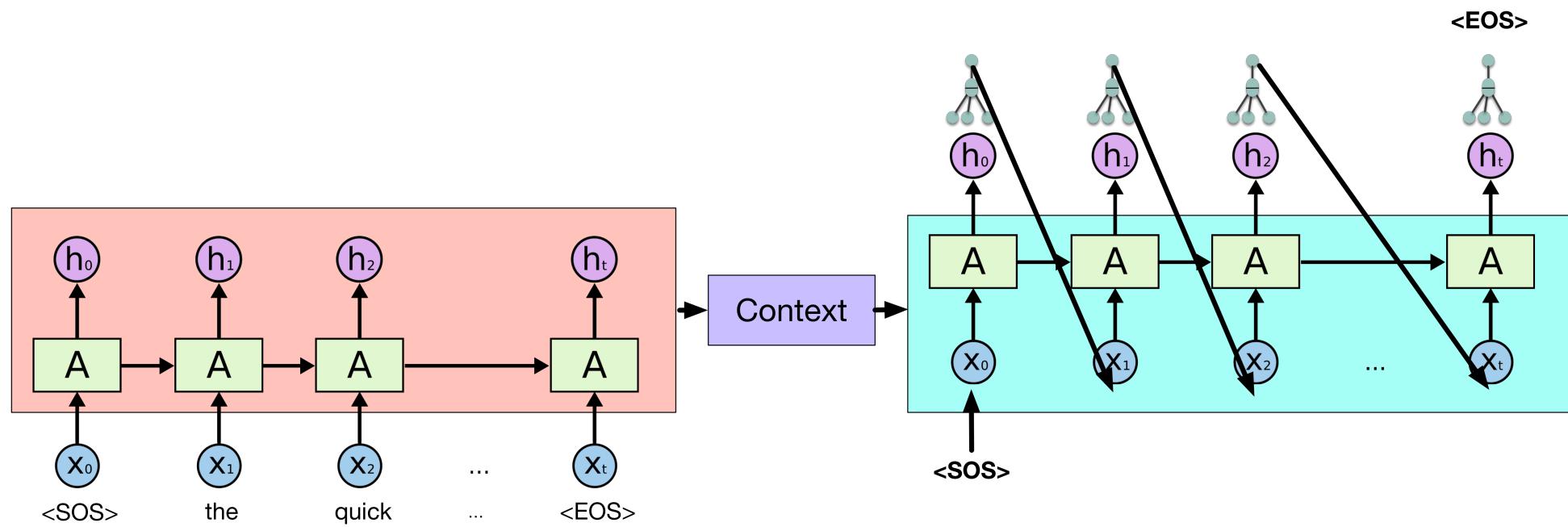
Sequence-to-Sequence Training: Loss



Sequence-to-Sequence Training: Teacher Forcing



Sequence-to-Sequence Inference



In []: