Sequence Tagging with LSTMs

CS 759/859 Natural Language Processing Lecture 14

Samuel Carton, University of New Hampshire

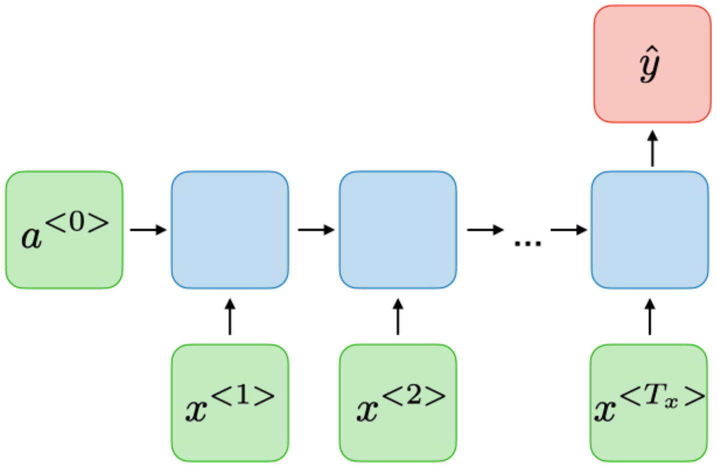
Last lecture

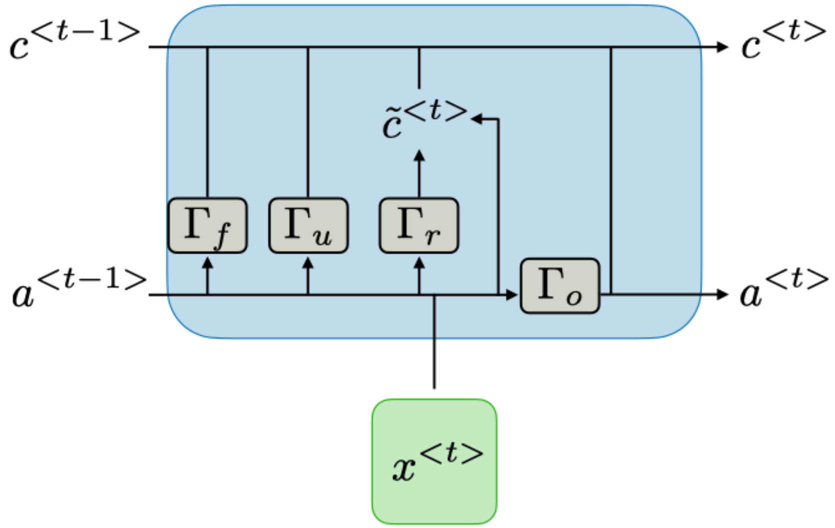
RNNs
- One-to-one
- **Many-to-one**
- Many-to-many

LSTMs

Increasing RNN capacity
- Depth
- Bidirectionality

Dropout





Types of RNN

Vanilla RNN: ht-1 and xt undergo one operation into ot and ht

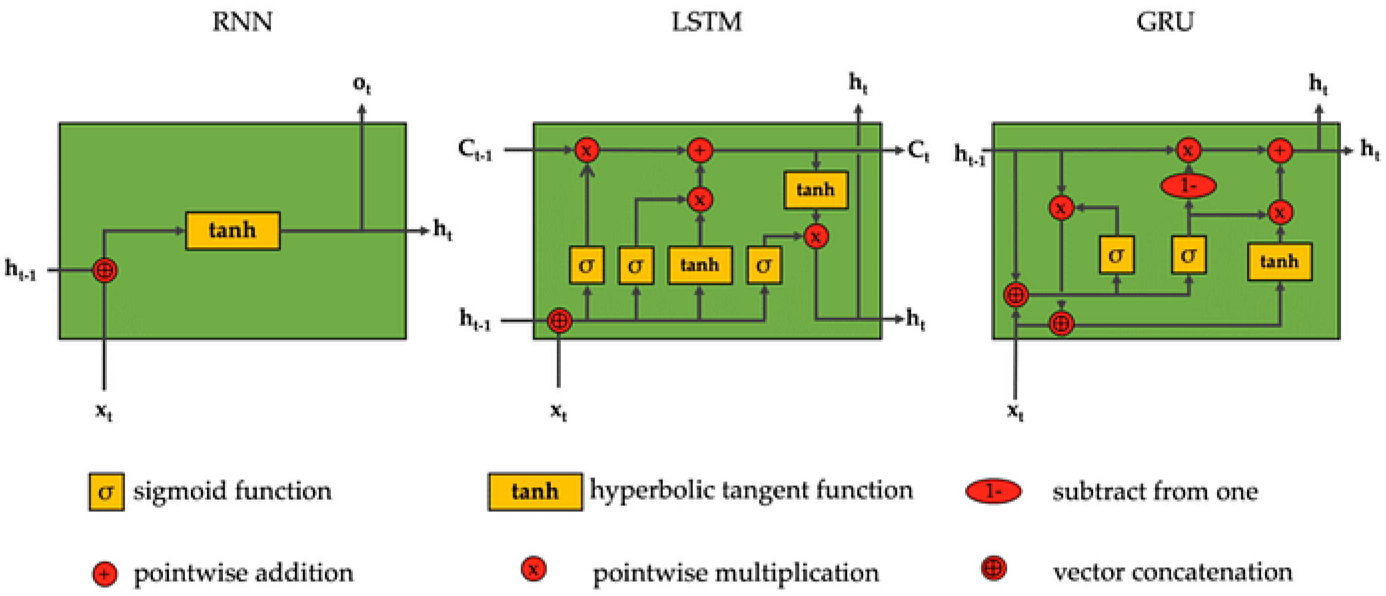
- No internal memory

LSTM:

- Memory cell that maintains information over time
- Separate layers (“gates”) control wha goes into/out of, what gets deleted out of memory cell

GRU:

- Simplified: update gate and reset gate

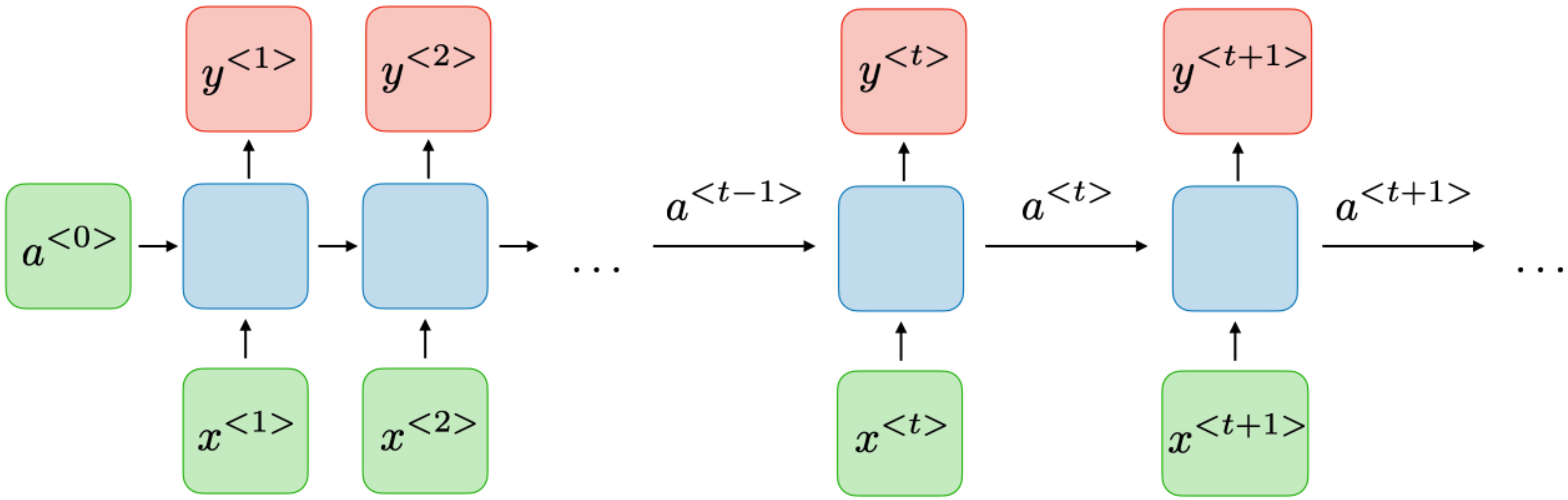


https://medium.com/@hassaanidrees7/rnn-vs-lstm-vs-gru-a-comprehensive-guide-to-sequential-data-modeling-03aab16647bb

LSTMs: NLP Swiss army knife

LSTMs are exciting for us because they are the Swiss army knife of NLP models.

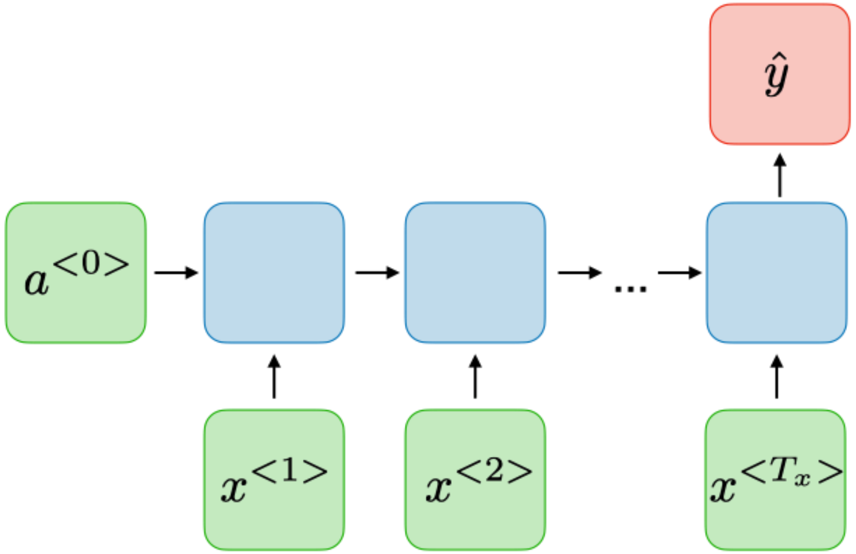
- Sequence classification
- Sequence tagging
- Language modeling
- Text-to-text (e.g. translation)



LSTMs: NLP Swiss army knife

LSTMs are exciting for us because they are the Swiss army knife of NLP models.

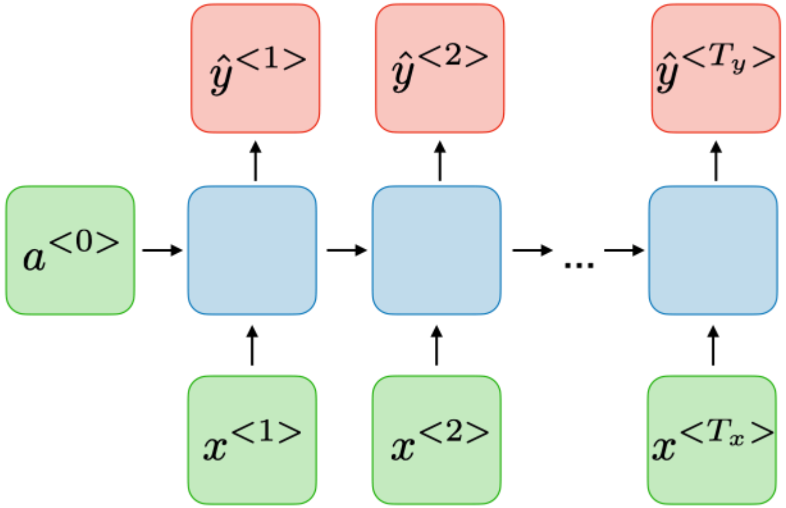
- **Sequence classification**
- Sequence tagging
- Language modeling
- Text-to-text (e.g. translation)



LSTMs: NLP Swiss army knife

LSTMs are exciting for us because they are the Swiss army knife of NLP models.

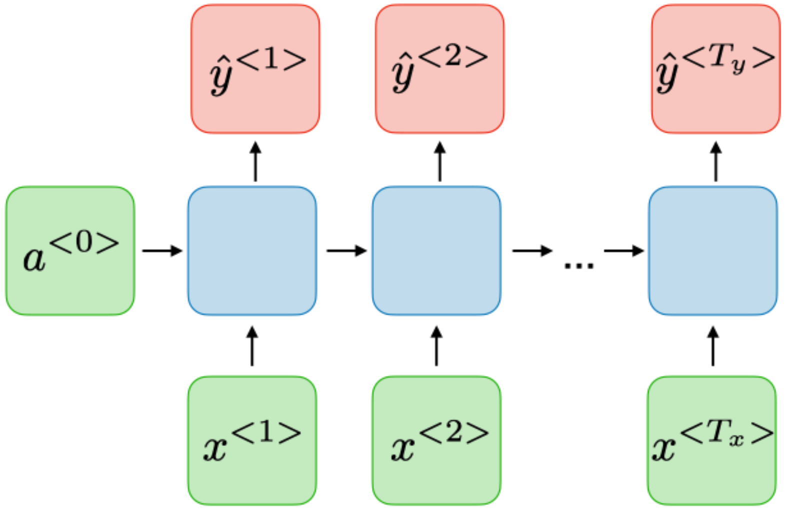
- Sequence classification
- **Sequence tagging**
- Language modeling
- Text-to-text (e.g. translation)



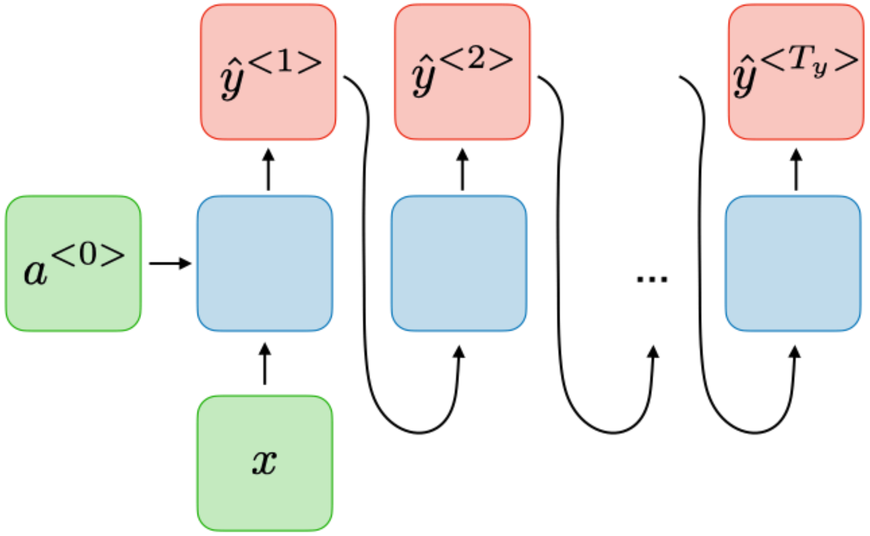
LSTMs: NLP Swiss army knife

LSTMs are exciting for us because they are the Swiss army knife of NLP models.

- Sequence classification
- Sequence tagging
- **Language modeling**
- Text-to-text (e.g. translation)



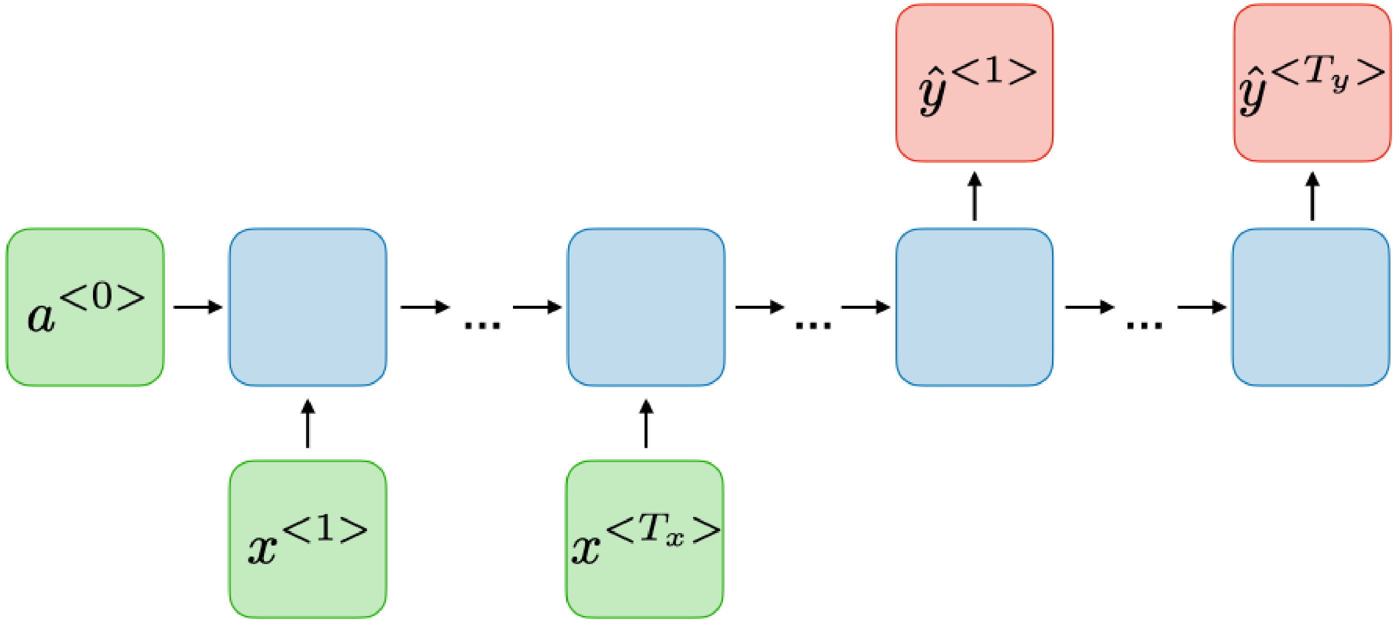
and



LSTMs: NLP Swiss army knife

LSTMs are exciting for us because they are the Swiss army knife of NLP models.

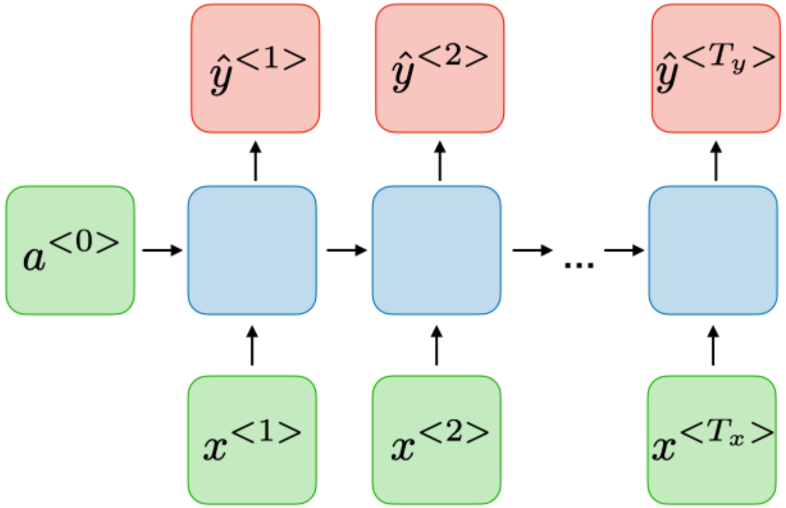
- Sequence classification
- Sequence tagging
- Language modeling
- **Text-to-text (e.g. translation)**



LSTMs: NLP Swiss army knife

LSTMs are exciting for us because they are the Swiss army knife of NLP models.

- Sequence classification
- **Sequence tagging**
- Language modeling
- Text-to-text (e.g. translation)



Sequence tagging

Basic idea: Given a corpus of text where each **word** has a label, learn to predict word labels for unseen texts

- Part-of-speech tagging
- Named entity recognition
  - "In his speech to the UN today, **George Bush** addressed the rising problems of…"
- Explanations
  - "You are a real **piece of garbage** human being." → Predicted toxic

Evaluation: Use the same metrics as for classification (Acc/P/R/F1)

- Two choices for aggregation:
  - **Calculate score for each text and then take mean**
  - Concatenate all texts together and calculate over one long sequence
- F1 preferable for very unbalanced tasks

Named-entity recognition (NER)

Goal: Identify the **named entities** (people, corporations, etc) in a piece of text.

Important for large-scale text analysis

- E.g. Extracting structured information from scientific literature
- E.g. Performing market research over social media

Usually treated as sequence tagging task, where each word is tagged as (1) part of an entity or (2) not part of an entity

F1 preferable as a metric because usually unbalanced

P2- Na2/3Ni1/4TixMn3/4-xO2 was prepared through a simple solid state method. The precursor solution was prepared by mixing desirable amount of Ni(CH3COO)2*4H2O, Mn(CH3COO)2*4H2O and CH3COONa and titanium citrate solution. The obtained mixture was heated at 400 degC for 12 h. The ground powder was ball-milled for 1 h and was subsequently calcinated at 900 degC in air for 12 h to synthesize Na2/3Ni1/4TixMn3/4-xO2 (x=0, 0.05, 0.10, 0.15, 0.20, 0.30).

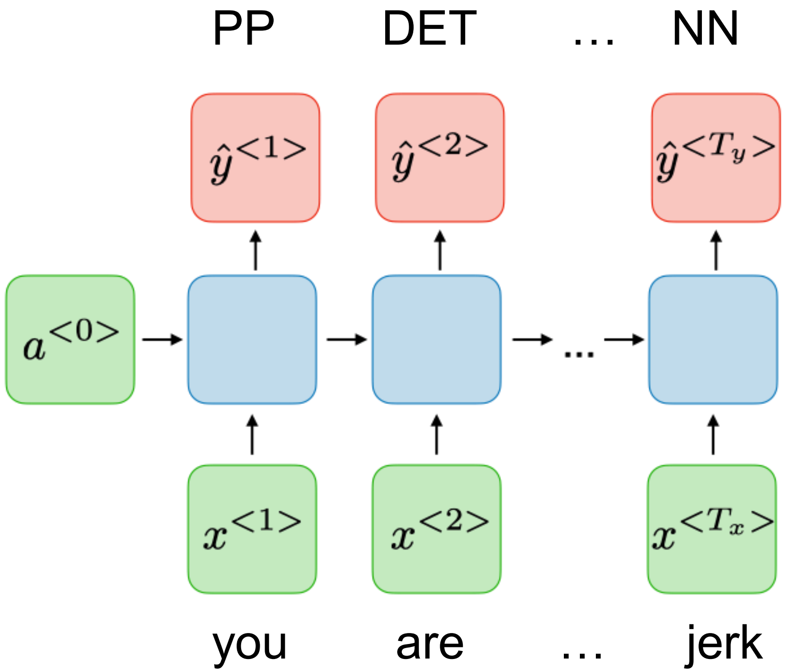
Figure 1: Part of an example synthesis procedure included in the dataset with entity annotations from Zhao et al. (2015). Colors represent entity types and underlines represent span boundaries. Colors: Target, Nonrecipe-operation, Unspecified-Material, Operation, Material, Condition-Unit, Number.

Tim O'Gorman, Zach Jensen, Sheshera Mysore, Kevin Huang, Rubayyat Mahbub, Elsa Olivetti, and Andrew McCallum. 2021. MS-Mentions: Consistently Annotating Entity Mentions in Materials Science Procedural Text. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*

Sequence tagging

Context sensitive.

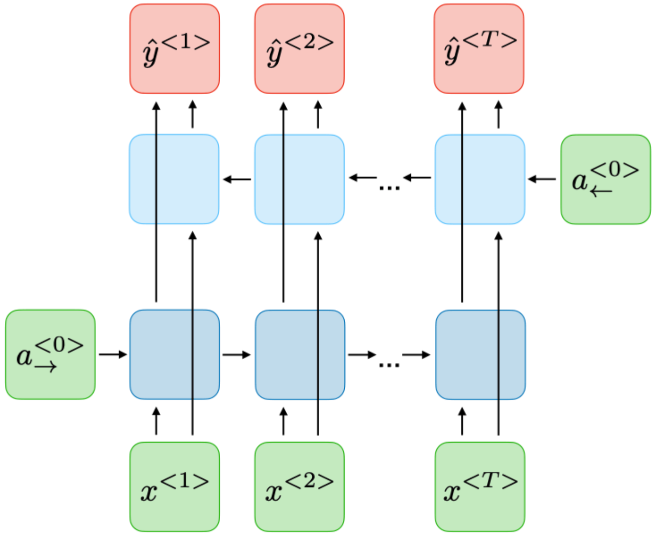
- “You are a real jerk!”
- “I am really craving some Jamaican jerk chicken right now.”

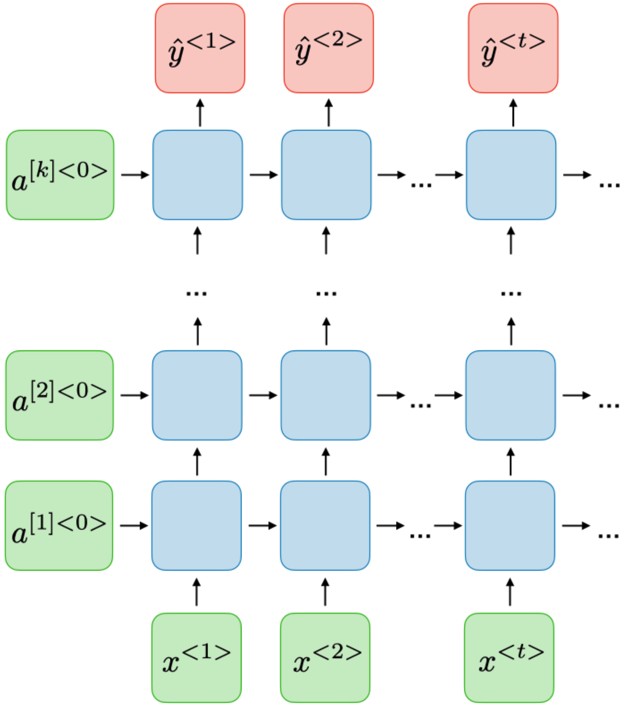


Sequence tagging

Context sensitive.

- “You are a real jerk!”
- “I am really craving some Jamaican jerk chicken right now.”



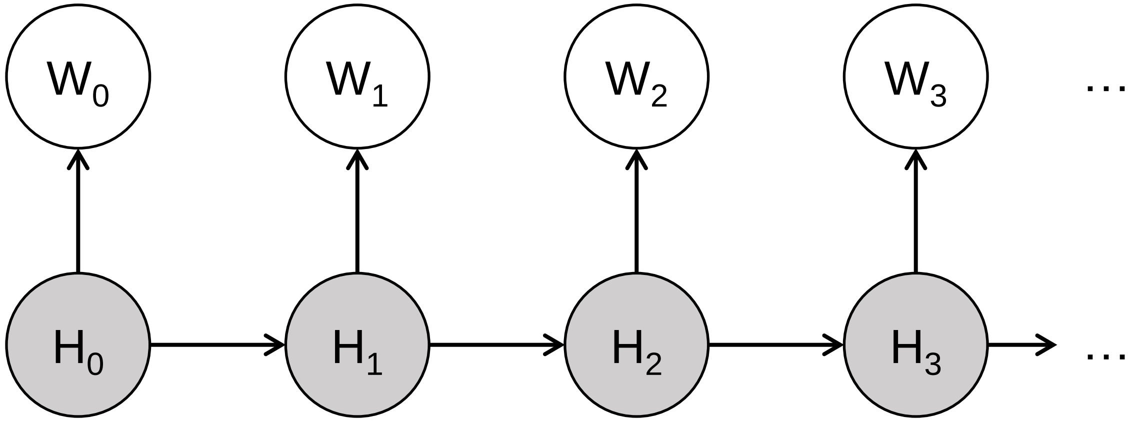


POS tagging with HMM

A popular application of HMMs in NLP is part-of-speech tagging

We imagine a generative story where parts-of-speech occur in a Markov chain, and then they emit words conditioned on their value.

i	sentence	you	to	read	this	sentence	.
PP	V	PP	PREP	V	DET	NN	PUNCT



Transition matrix

$P(H_t|H_{t-1})$

		h_0	h_1	…
	h_0	…	…	…
H_{t-1}	h_1	…	…	..
	…	...	…	…

Emission matrix

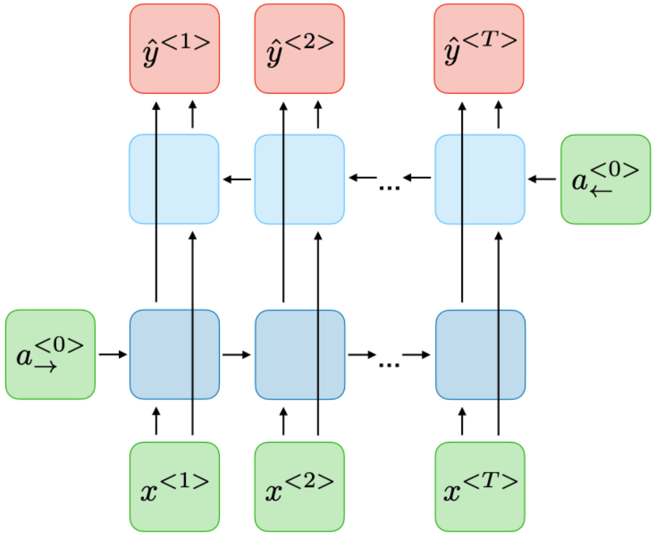
$P(W_t|H_t)$

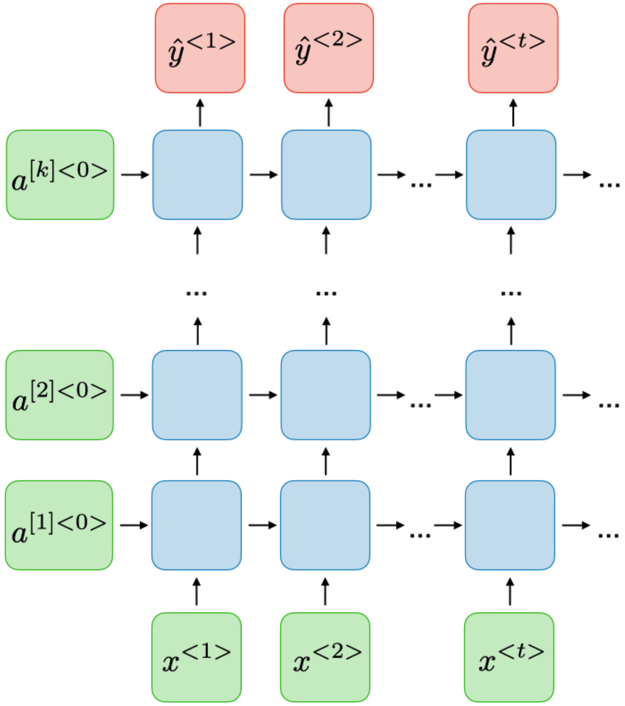
		w_0	w_1	…
	h_0	…	…	…
H_t	h_1	…	…	..
	…	...	…	…

Sequence tagging

Context sensitive.

- "You are a real jerk!"
- "I am really craving some Jamaican jerk chicken right now."





Twitter POS tagging dataset

Named Entity Recognition in Tweets: An Experimental Study (Ritter et al., 2011)

https://raw.githubusercontent.com/aritter/twitter_nlp/master/data/annotated/pos.txt

```
@paulwalk USR
It PRP
's VBZ
the DT
view NN
from IN
where WRB
I PRP
'm VBP
living VBG
for IN
two CD
weeks NNS
. .
Empire NNP
State NNP
Building NNP
= SYM
ESB NNP
. .
Pretty RB
bad JJ
storm NN
here RB
last JJ
evening NN
. .
```

```
@MiSS_SOTO USR
I PRP
think VBP
that DT
's VBZ
when WRB
I PRP
'm VBP
gonna VBG
be VB
there RB

On IN
Thanksgiving NNP
after IN
you PRP
done VBN
eating VBG
its PRP
#TimeToGetOut HT
unless IN
you PRP
wanna VBP
help VB
with IN
the DT
dishes NNS
```

POS tagging with PyTorch

Code description

- Downloading Twitter-specific Glove Vectors
- Downloading a Twitter POS tagging dataset
- Preprocessing it
- Training and evaluating a LSTM POS tagger

Notebook headings

- **Loading GloVe vectors with Gensim**
- Reading and preprocessing POS data
- Training a LSTM POS tagger
  - Dataset
  - DataLoader
  - Model
  - Trainer

Concluding thoughts

Sequence tagging

- POS tagging

LSTMs as a NLP Swiss army knife

Domain-specific word embeddings

Masked loss