



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Neural Classification of Linguistic Coherence using Long Short-Term Memories

Pashutan Modaresi, Matthias Liebeck and Stefan Conrad



Order of Sentences

Is what makes a text semantically meaningful

- **Hi!**
- **My name is Alan.**
- **I am a computer scientist!**

n!

Hi!
My name is Alan.
I am a computer scientist!

Hi!
I am a computer scientist!
My name is Alan.

I am a computer scientist!
Hi!
My name is Alan.

My name is Alan.
Hi!
I am a computer scientist!

I am a computer scientist!
My name is Alan.
Hi!

My name is Alan.
I am a computer scientist!
Hi!

Humans vs. Machines

Discourse Coherence

Linguistic Contradiction

Linguistic Redundancy

Pragmatics



Question

Is there a need to teach all these abilities to a machine?

Sentence Ordering

$$\Phi : \mathbb{R}^m \times \mathbb{R}^{m'} \rightarrow \mathbb{Z}$$

0, 1

-1, 0, 1

Hi!

My name is Alan.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$\begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_{m'} \end{bmatrix}$$

Question?

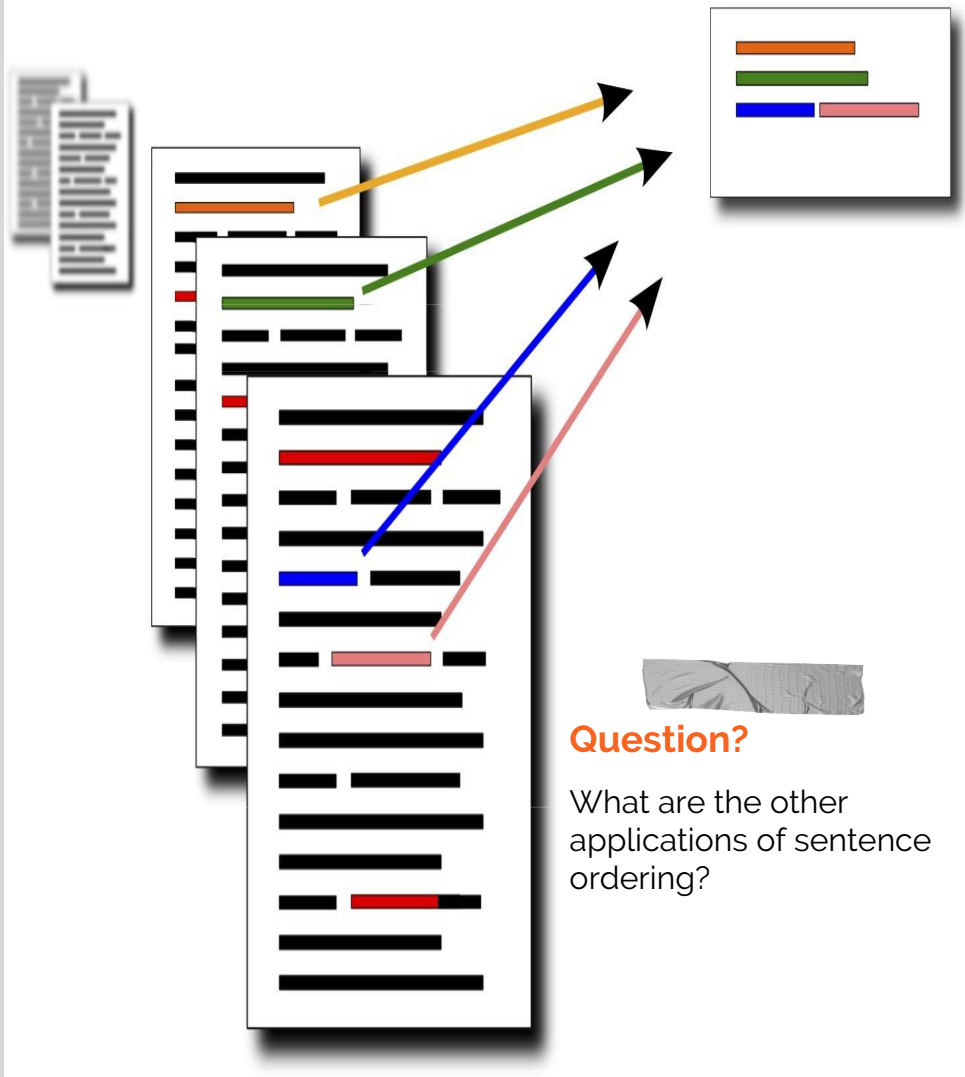
What about the sizes of m and m' ? Should they be equal?

Many Applications!
Focus was

TEXT

SUMMARIZATION

in the news domain



—
Treat the problem as a **classification** task

Number of Instances

$$- \sum_{n=1}^N \log p(z_n | x_1^{(n)}; x_2^{(n)})$$

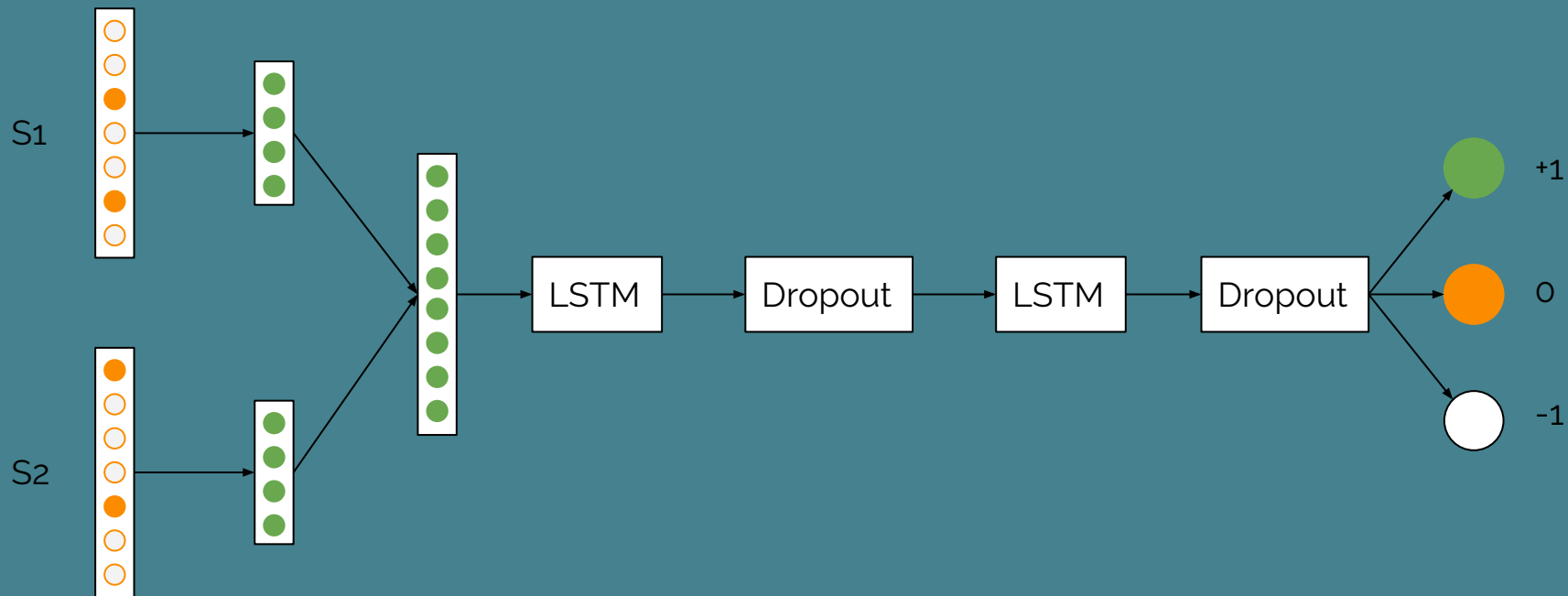
Class probability of the n-th pair



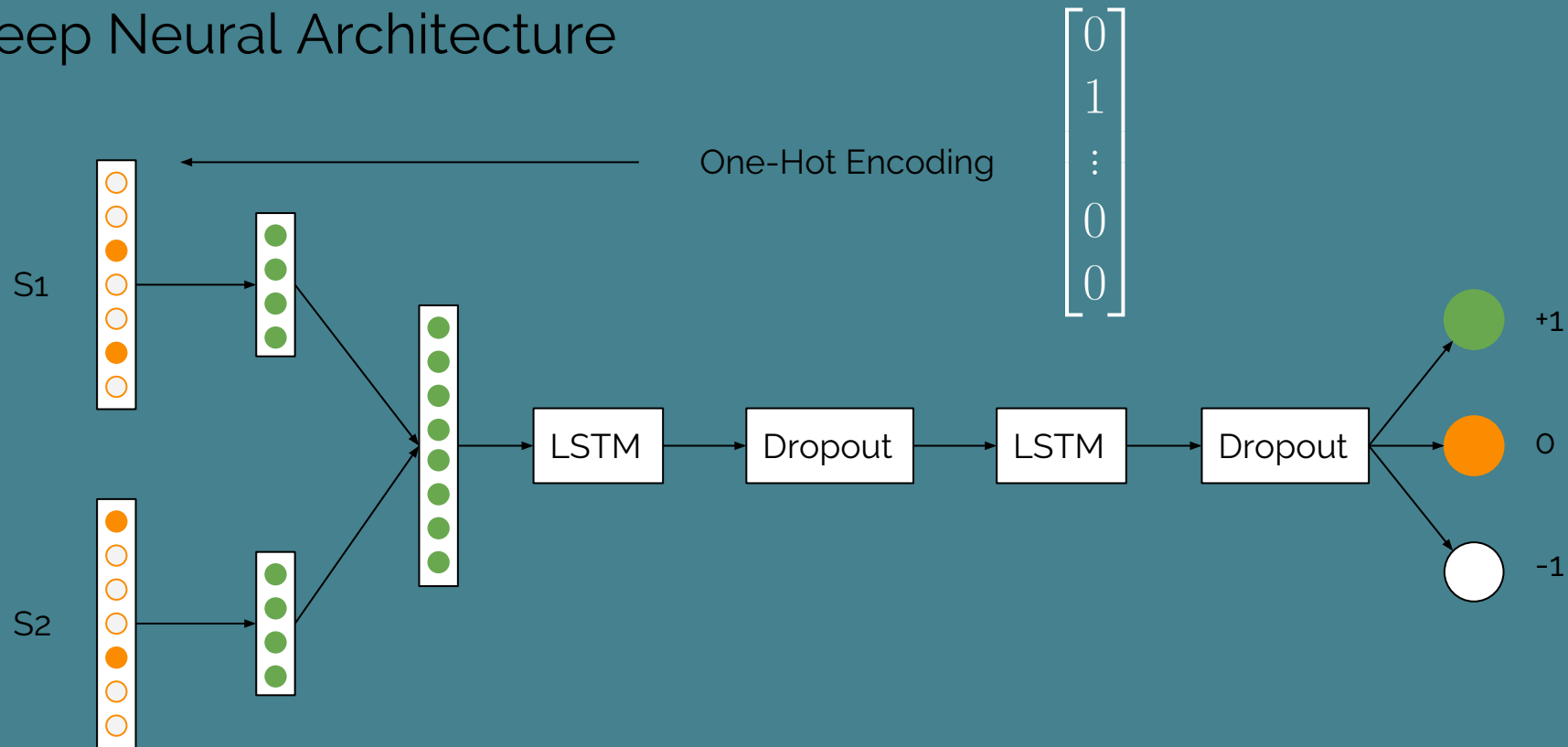
Question

Why do we use the negative log-likelihood and not the log-likelihood?

Deep Neural Architecture



Deep Neural Architecture

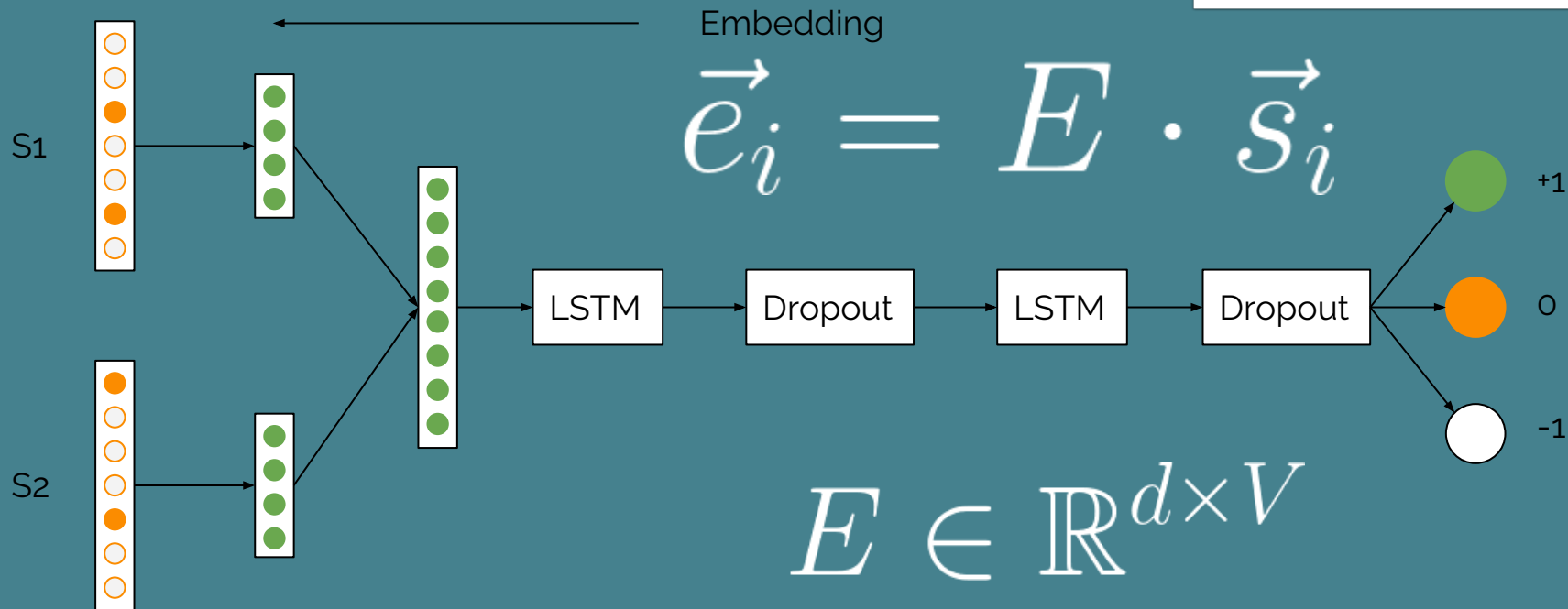


Deep Neural Architecture

Tip

Embedding: Simple matrix multiplication with input vector

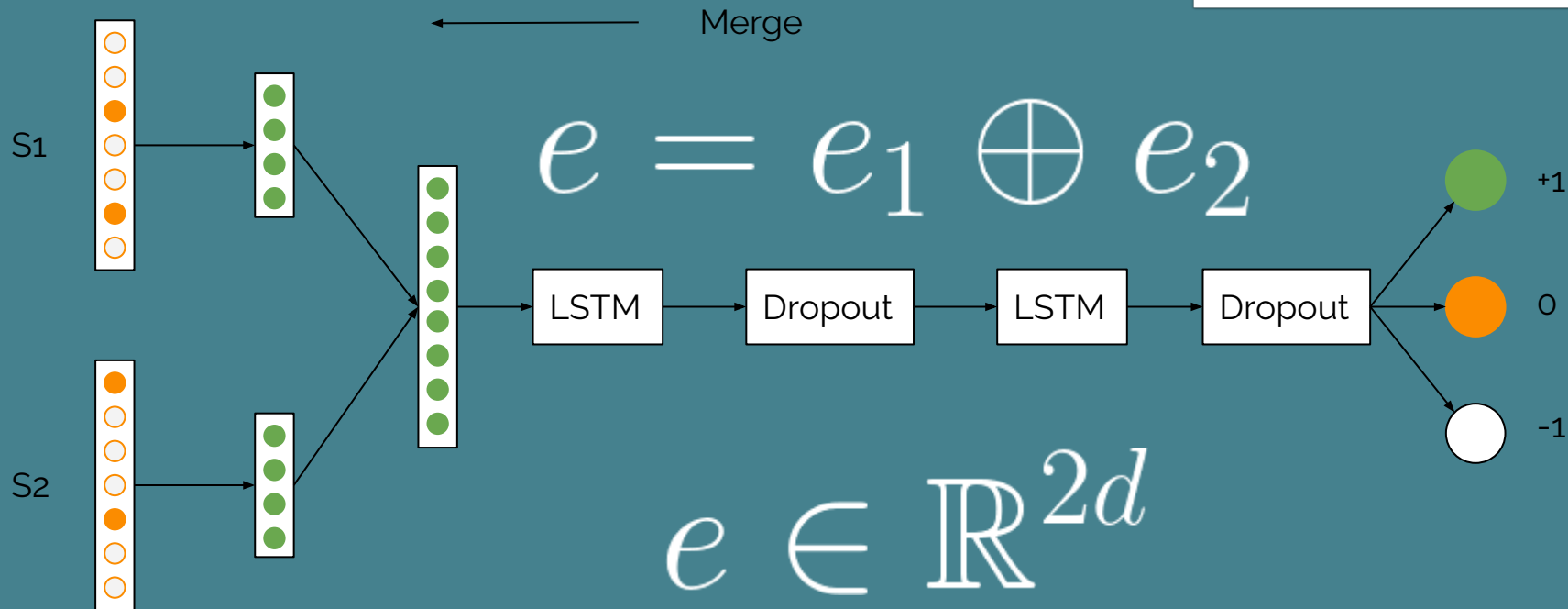
Init the matrix E



Deep Neural Architecture

Tip

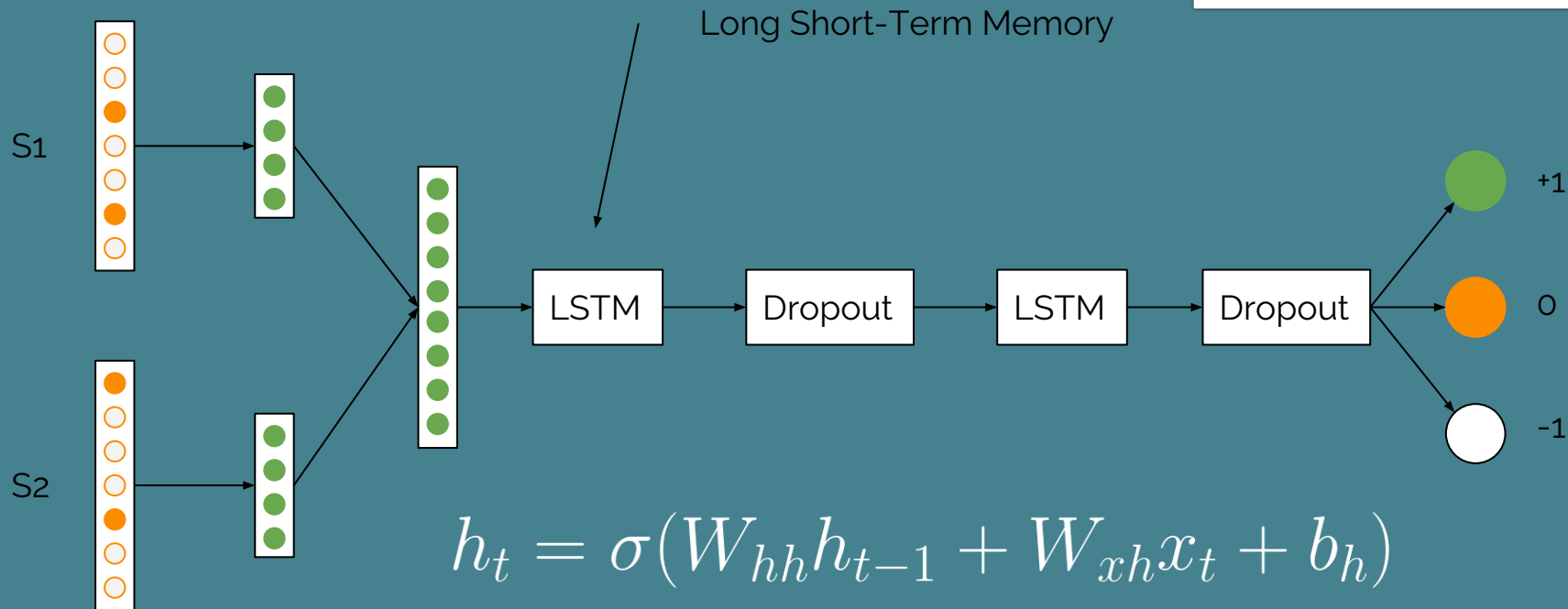
Concatenate the embeddings



Deep Neural Architecture

Tip

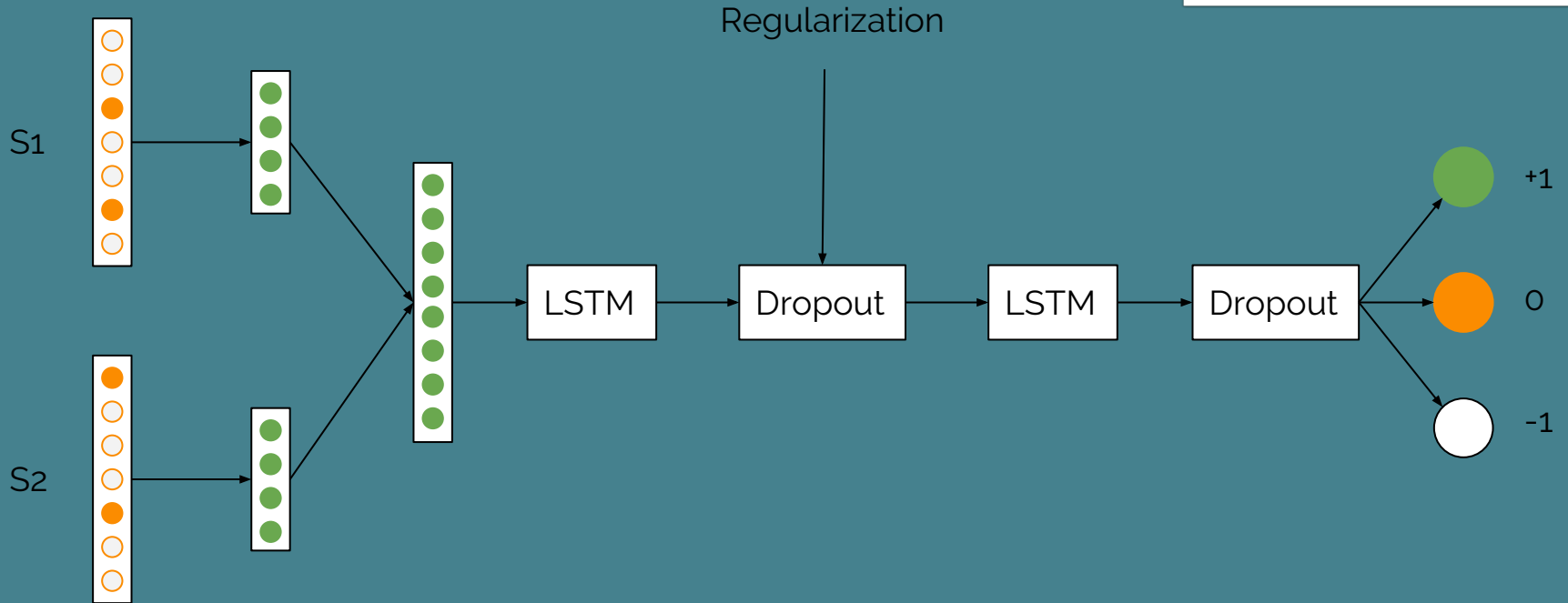
LSTM: Just a special kind of RNNs addressing their difficulties



Deep Neural Architecture

Tip

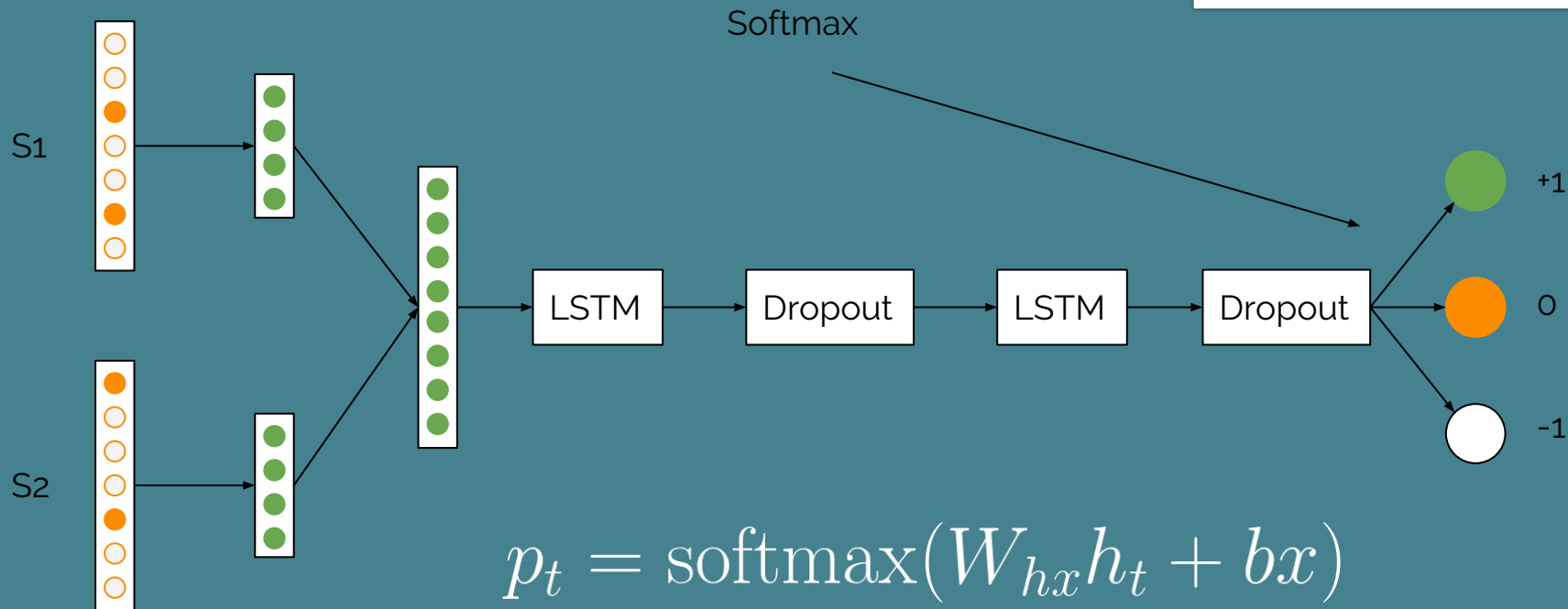
Dropout: Sets a random set of its arguments to zero.

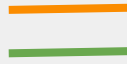
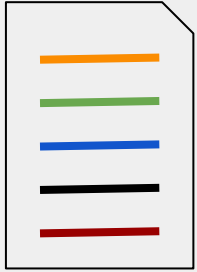


Deep Neural Architecture

Tip

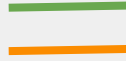
Dropout: Sets a random set of its arguments to zero.





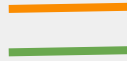
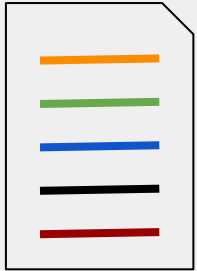
+1

Correct
Order



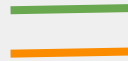
-1

Wrong
Order



+1

Correct
Order



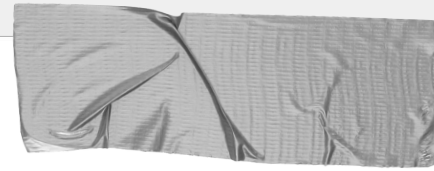
-1

Wrong
Order



0

Mising
Context



Data

How to collect the required data to train the network?

→ Binary

→ Ternary

Baseline - SVM

English

German

Binary

Ternary

Binary

Ternary

0.24

0.16

0.25

0.16

SVMs: Not really appropriate for sequential modelling

Macro-Averaged F1

	English		German	
	Binary	Ternary	Binary	Ternary
True	0.95	0.97	0.94	0.97
MC	—	0.98	—	0.97
False	0.95	0.98	0.94	0.98
Overall	0.95	0.98	0.94	0.97

Lessons Learned

- Use appropriate tools for sequence modeling
- RNNs are slow. First train on a subset of data
- Train deep models with lots of data points
- Find a way to automatically annotate data
- Use regularization (be generous)

**Thank You
For
Your
Attention**
