

Natural Language Processing and Deep Learning





Table of content

- ◉ Natural Language Processing + Deep Learning
- ◉ Feed-forward Network
- ◉ Word2vec (Mikolov et al. 2013)
- ◉ Recurrent Neural Network
- ◉ Long short-term memory (Schmidhuber 1997)
- ◉ Neural Machine Translations (Wu et al. 2016)
- ◉ Neural Programmer (Neelakantan et al. 2016)
- ◉ Course of Action



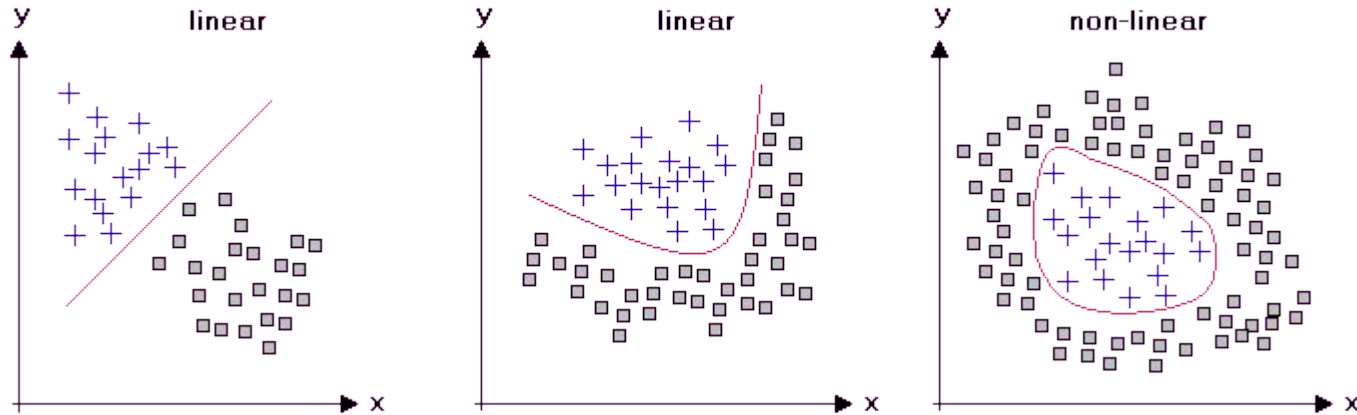
Natural Language Processing

Goal: Understand and represent meaning of languages to do useful tasks.

- AI Complete (Human level cognition, Multimodal)
- Semantics: Machine translations, Sentiment Analysis, Question-Answering



What is Deep Learning?



Example for a classification problem. This applies generally for distributions as well.



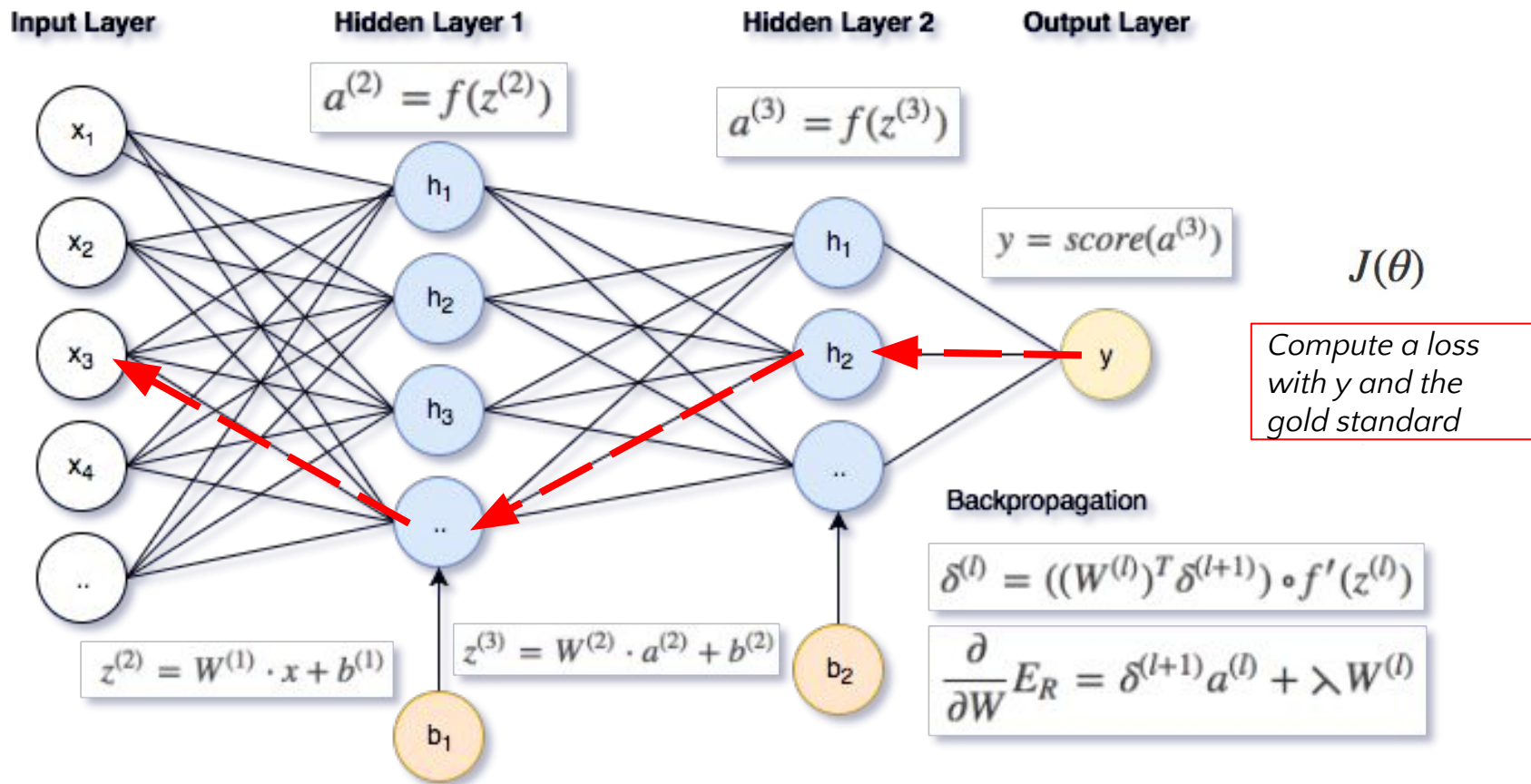
What is Deep Learning?

- ◉ $N > 1$ layers of nonlinear processing units that extract features in an unsupervised fashion
- ◉ Minimize or maximize a cost/loss function for a given model

Learning an implicit algorithm that largely outperforms classical methods on specific tasks (not all!)

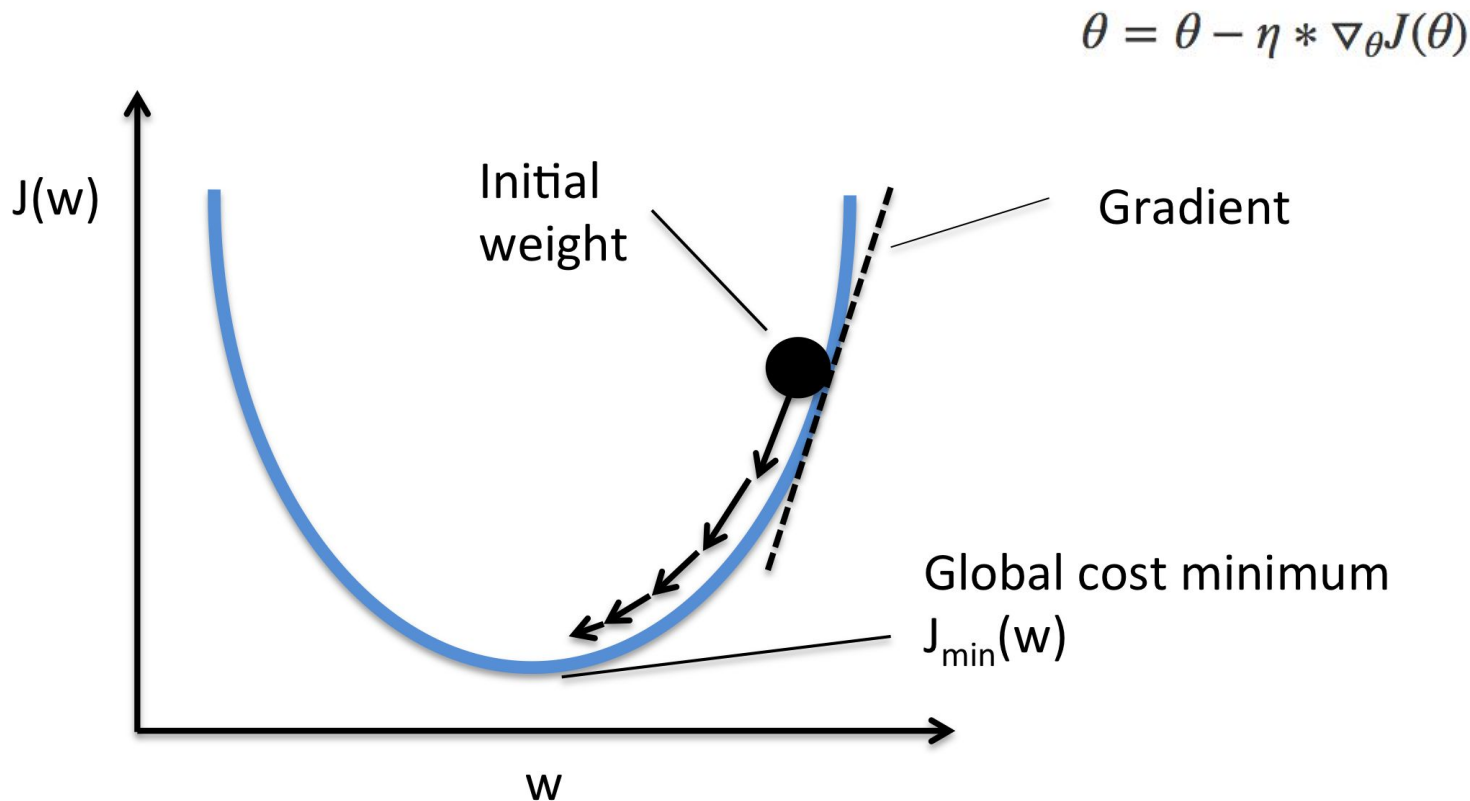


Feed-forward Networks





Gradient Descent





Word2Vec (Mikolov et al. 2013)

Unsupervised Word embedding method

Sliding window learn with context of center word.

Example: “I am a software developer from Hamburg”

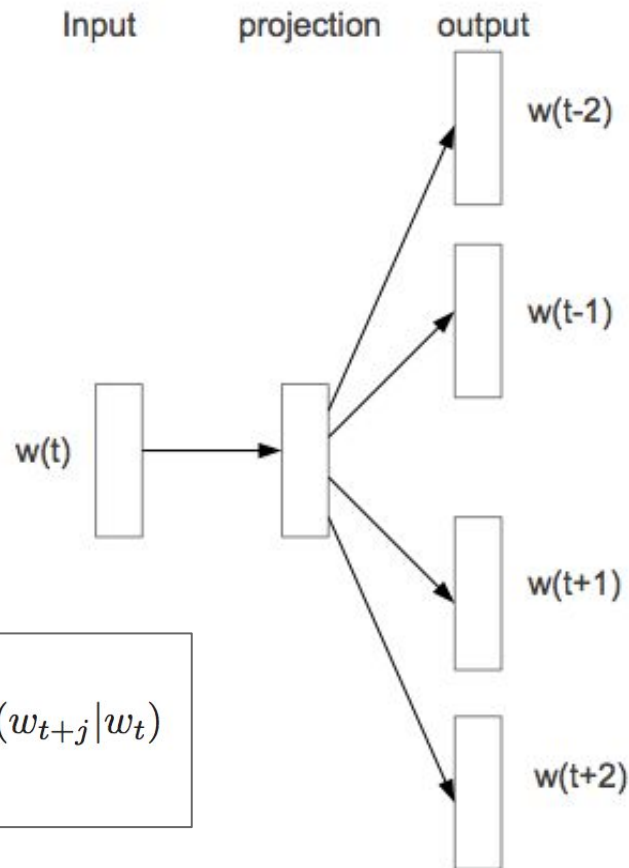
Window: 2

Center: “Software”

Context: “am a developer from”

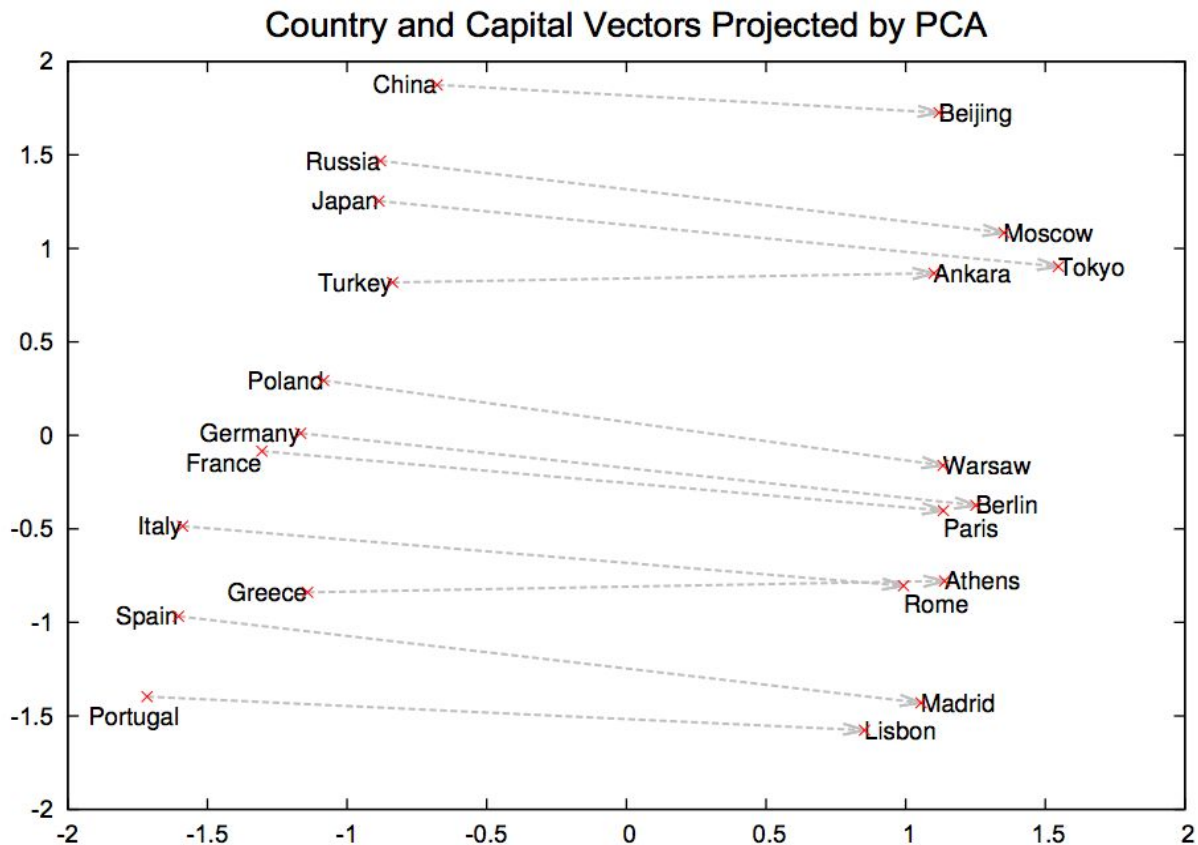
Other Models: Glove, doc2vec

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t)$$





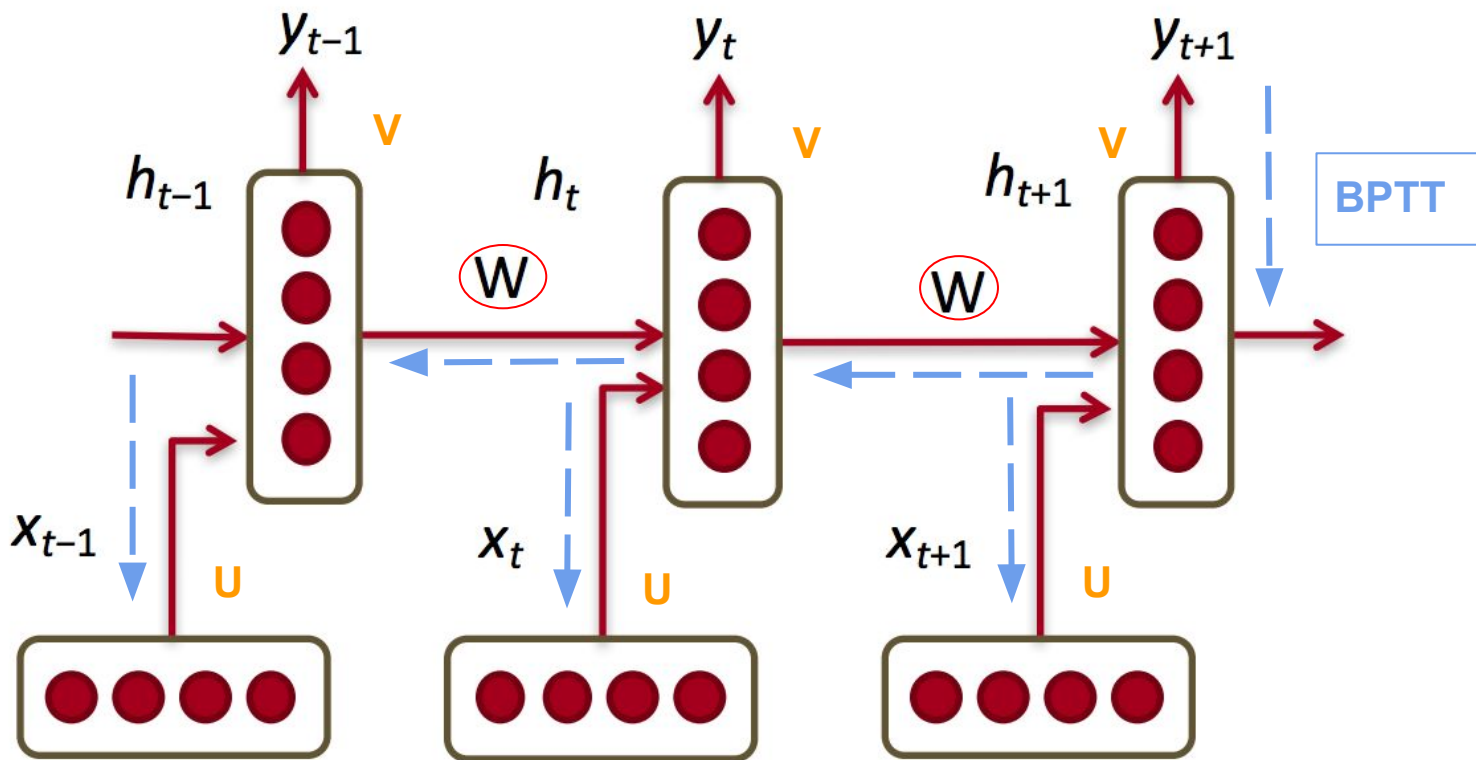
Word2Vec (Mikolov et al. 2013)





Recurrent Neural Networks

Parameter
V
U
W



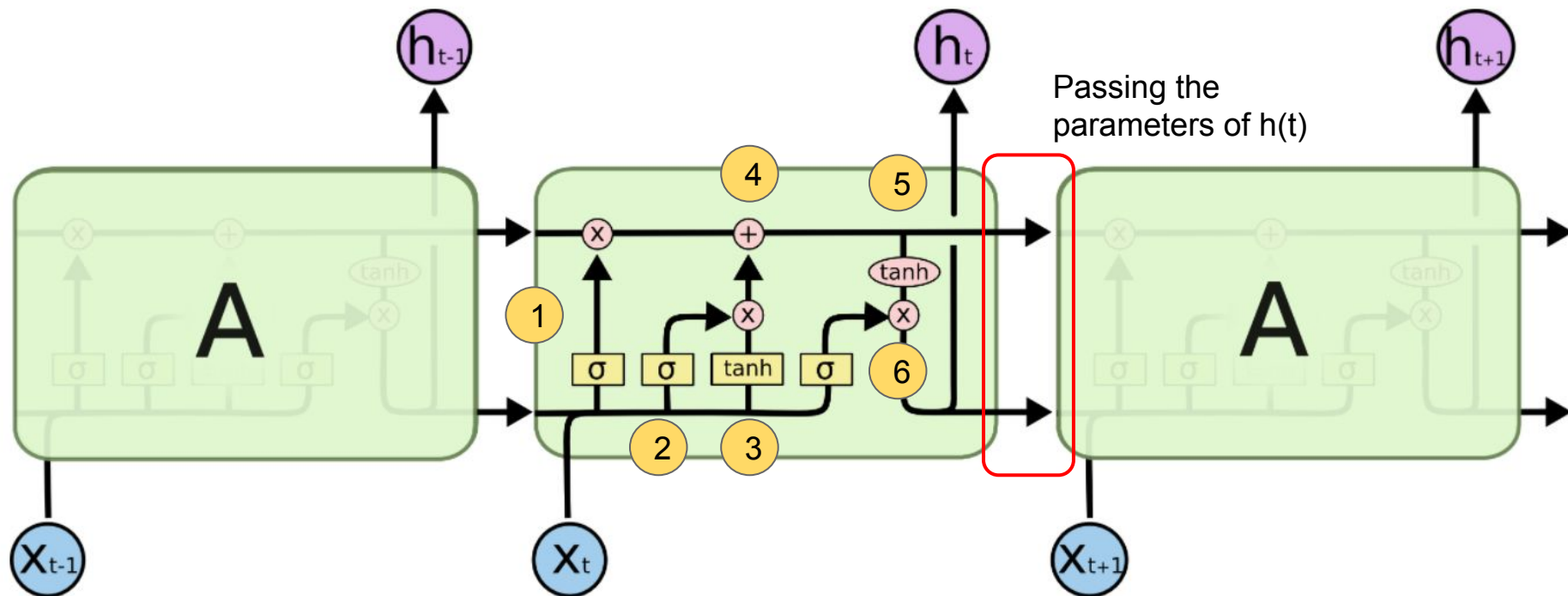


Long short-term memory

- Idea from Schmidhuber (1997) and is still state of the art
- Instead of RNNs use LSTMs!
- vanishing gradient for very large sequences
- More complex learning structure
- Some extensions to the original (attention, memory, multilayers)



Long short-term memory



The repeating module in an LSTM contains four interacting layers.



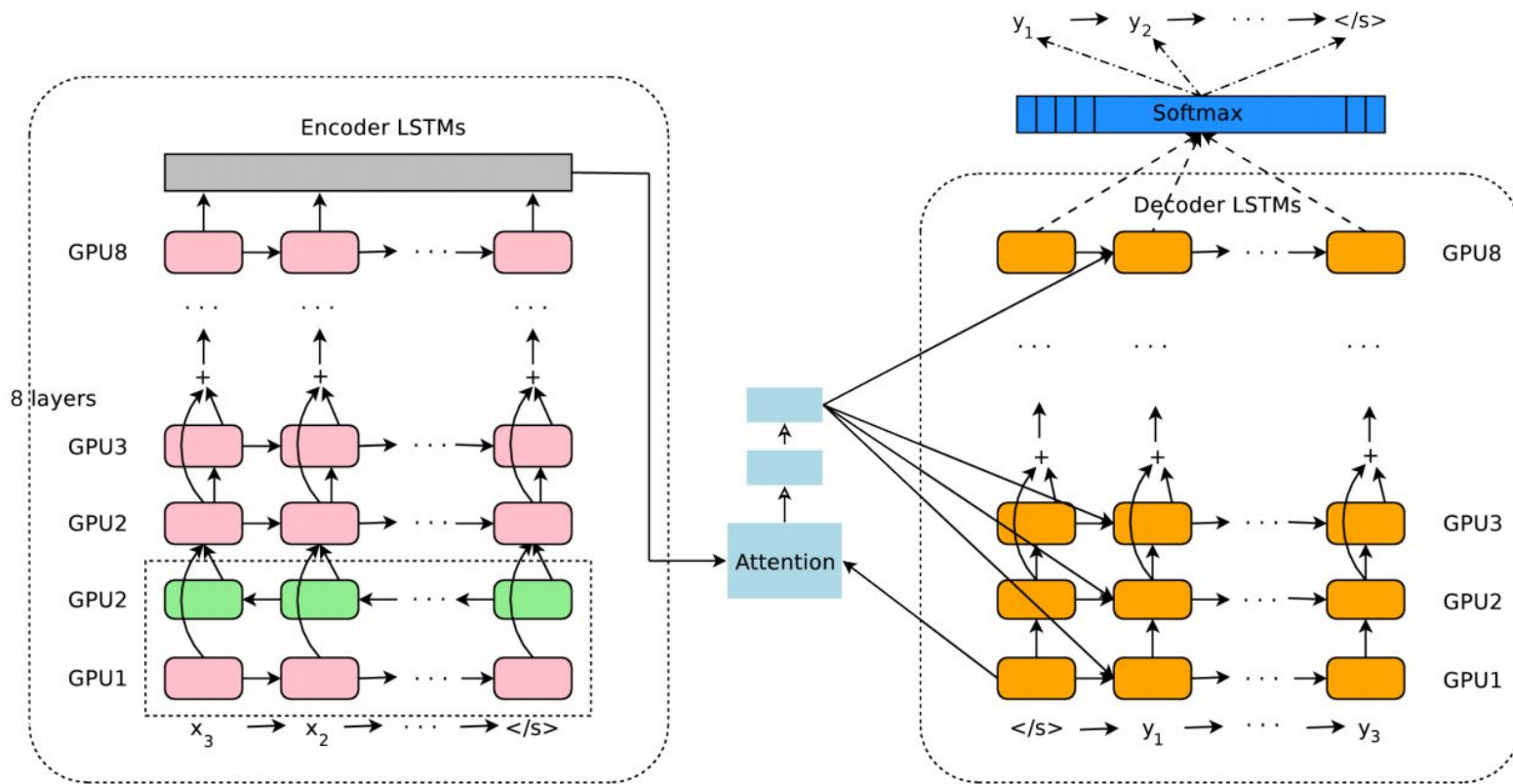
Neural Machine Translations

$$= \prod_{i=1}^N P(y_i | y_0, y_1, y_2, \dots, y_{i-1}; \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_M)$$

- End to-end-learning + expensive
- Deep LSTM Networks (8 layers) - encoder (source language) - decoder (target language)
- Attention mechanism
- Parallelization + Low precision arithmetic units (Tensor Processing Units - TPUs)
- Word pieces instead of characters or words



Neural Machine Translations



Google's Neural Machine Translation System 2016



Neural Machine Translations

Table 4: Single model results on WMT En→Fr (newstest2014)

Model	BLEU	CPU decoding time per sentence (s)
Word	37.90	0.2226
Character	38.01	1.0530
WPM-8K	38.27	0.1919
WPM-16K	37.60	0.1874
WPM-32K	38.95	0.2118
Mixed Word/Character	38.39	0.2774
PBMT [15]	37.0	
LSTM (6 layers) [31]	31.5	
LSTM (6 layers + PosUnk) [31]	33.1	
Deep-Att [45]	37.7	
Deep-Att + PosUnk [45]	39.2	



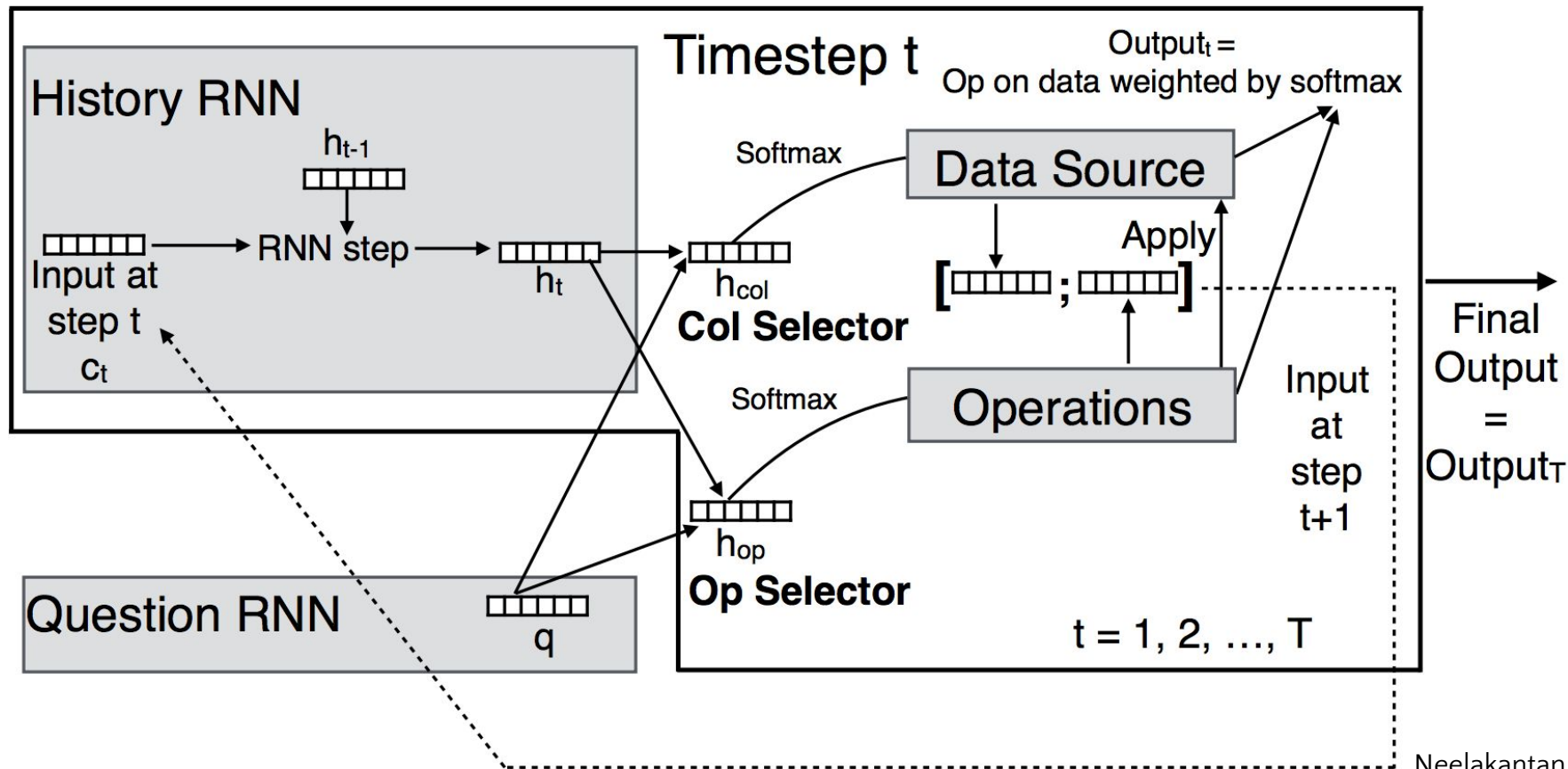
Neural Programmer

- Map natural language to logical form and execute discrete operations on a database
- 15 operations (e.g. count, min, max) are augmented to a Neural Network
- Input a question and output a program

ID	Question		Step 1	Step 2	Step 3	Step 4
1	what is the total number of teams?	Operation	-	-	-	count
		Column	-	-	-	-
2	how many games had more than 1,500 in attendance?	Operation	-	-	>=	count
		Column	-	-	attendance	-



Neural Programmer





Neural Programmer

Method	Dev Accuracy	Test Accuracy
Baselines from Pasupat & Liang (2015)		
Information Retrieval System	13.4	12.7
Simple Semantic Parser	23.6	24.3
Semantic Parser	37.0	37.1
Neural Programmer		
Neural Programmer	34.1	34.2
Ensemble of 15 Neural Programmer models	37.5	37.7

Test on WikiTableQuestions Dataset (not synthetic)



Course of Action

- Gain “deep” understanding of architectures and hyper parameters
 - CNNs + GANs as well
- Interested in multiple input sources (Multimodal) combining vision, text, speech
- Learn to use GPUs for computation
- Working knowledge of TensorFlow



Conferences

- Annual Meeting of the Association for Computational Linguistics (ACL) – <http://acl2018.org/>
 - Melbourne, Australia (July 15–20 2018)
- Neural Information Processing Systems Conference (NIPS)
 - 2017, December 04–09.2017, California, Long Beach
- International Conference on Learning Representations (ICLR)
- SIGKDD Conference on Knowledge Discovery and Data Mining



Bibliography

- M. Abadi, et al.** (2016). 'TensorFlow: A system for large-scale machine learning'.
- D. Andor, et al.** (2016). 'Globally Normalized Transition-Based Neural Networks'.
- D. Chen & C. D. Manning** (2014). 'A Fast and Accurate Dependency Parser using Neural Networks'. In Conference on Empirical Methods in Natural Language Processing, pp. 740–750.
- I. Goodfellow, et al.** (2016). Deep Learning. MIT Press.
<http://www.deeplearningbook.org>.
- S. Hochreiter & J. Schmidhuber** (1997). 'Long Short-Term Memory'. Neural Computation 9(8):1735–1780.



Bibliography

- Q. V. Le & T. Mikolov** (2014). 'Distributed Representations of Sentences and Documents'.
- A. Neelakantan, et al.** (2016). 'Neural Programmer: Inducing Latent Programs with Gradient Descent'.
- A. Neelakantan, et al.** (2017). 'Learning a Natural Language Interface with Neural Programmer'.
- J. Pennington, et al.** (2014). 'GloVe: Global Vectors for Word Representation'.
- I. Sutskever, et al.** (2014). 'Sequence to Sequence Learning with Neural Networks'.
- Y. Wu, et al.** (2016). 'Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation'.



Useful additional resources

- **Stanford NLP with Deep Learning 2016** – Manning / Socher
https://www.youtube.com/watch?v=OQQ-W_63UgQ
- **spaCy** – <https://explosion.ai/blog/>
- **Intuition** about **DL + NLP** – <http://sebastianruder.com/#open>
- **TensorFlow** – https://www.tensorflow.org/get_started/get_started
- Great papers **overall** – <http://colah.github.io/>
- Great source for many **basics** – <http://www.wildml.com/>
- **Understandable DL** posts – <http://distill.pub/>
- **Stanford Researcher:** <http://karpathy.github.io/>



Thanks!

Any **questions** ?

• matthias.nitsche@haw-hamburg.de