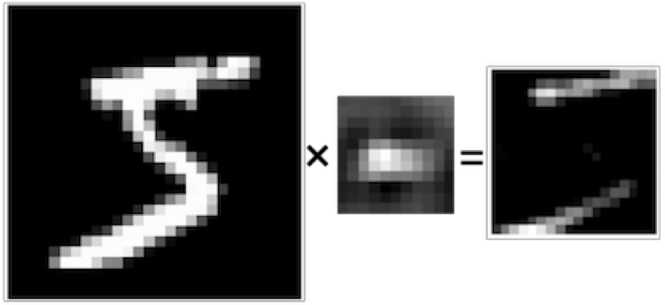
Universität Potsdam

Introduction to Deep Learning

Revision (selected topics)



Andreas Krug, M.Sc.

ankrug@uni-potsdam.de

28. January 2019

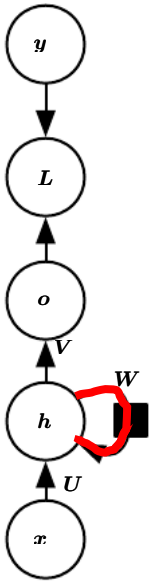
Orga

- Oral exam
  - 10 people on Moodle
  - 8 people picked a slot
  - You may change your 2 days in advance
  - Picking a time slot is mandatory
- PEP evaluation
  - 3 people participated
  - Please participate – I want to (briefly) discuss the results next week

Question 1

What makes RNNs stand out
from the other network architectures
you learned about so far (MLPs, CNNs)?

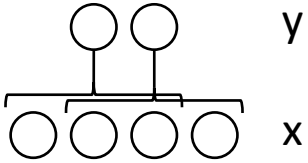
- Recurrent connections
- Special training method:
  back-propagation through time (BPTT)
- Weights shared across time

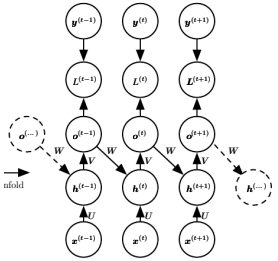


Question 2

What is the difference between
(a) applying (1D-)convolution
along the sequence dimension and
(b) using an RNN to process the sequence?

- Convolution processes a window of information (state-less)
- RNNs process the sequence with a hidden state per time step

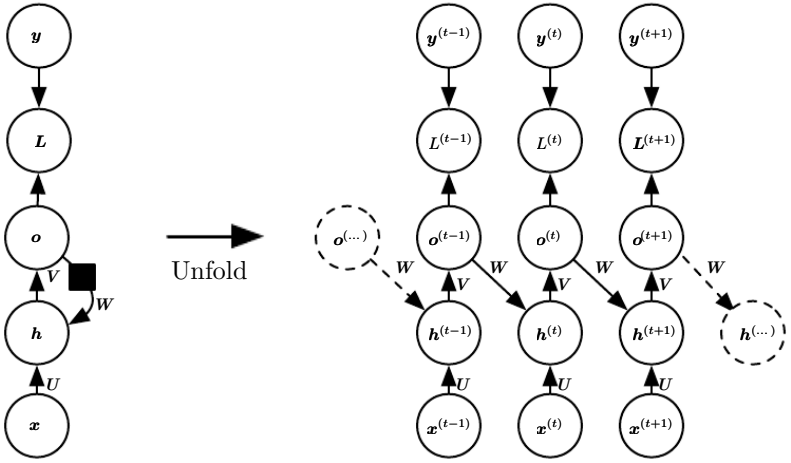




Question 3

What is "back-propagation through time"?

- Unroll the recurrent computation graph
- apply back-propagation



Question 4

Which problems can typically occur during RNN training and why?
Bonus: Outline possible remedies!

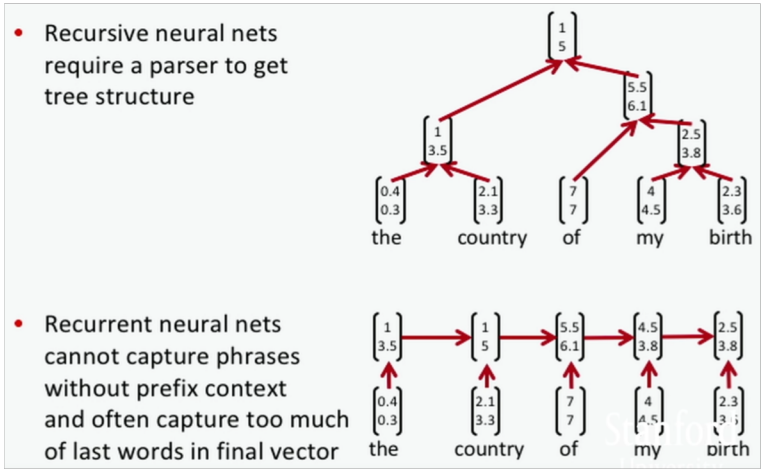
- Exploding or vanishing gradients
- $(W > 1)^t$ or $(W < 1)^t$
  extremely non-linear behavior
- long-term-dependencies are hard to capture
- gradient clipping (exploding)
  skip connections, LSTMs (vanishing)

Question 5

How do RNNs generalize to recursive NNs?

won‘t be asked in the exam

- Weight sharing in trees (instead of chains)
- Tree has to be given e.g. by a parser

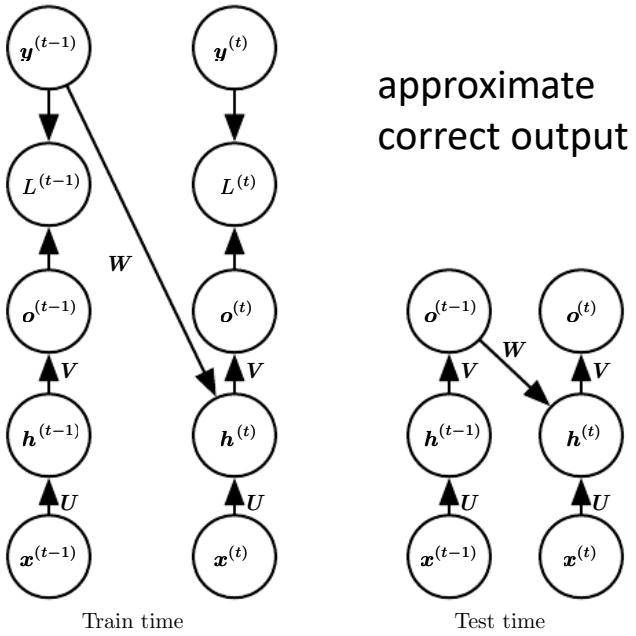


From Stanford “NLP with Deep Learning“ Lecture 14
https://youtu.be/RfwgqPkWZ1w

Question 6

What is "Teacher Forcing"?
Bonus: Discuss advantages and problems!

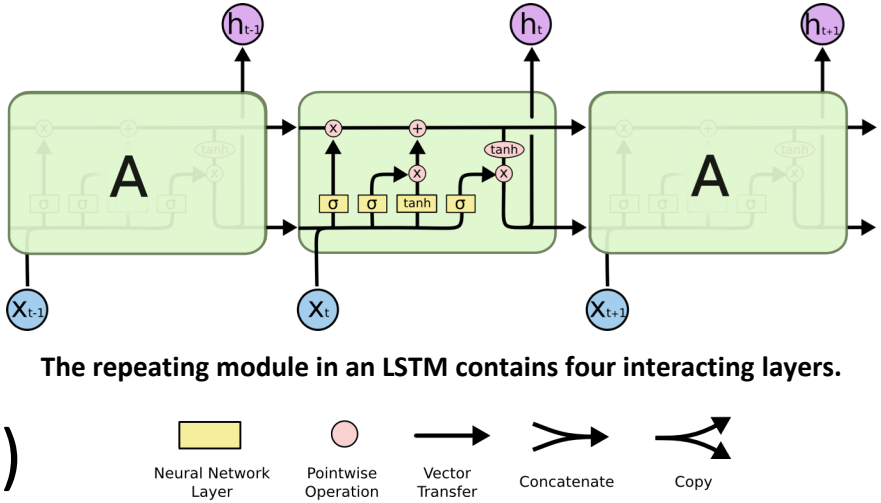
- Only for models with output-to-hidden connections
- During training: ground truth y(t) is used as o(t)
- Pro: parallelized training (without h-h connections)
- Con: o(t) in training can be different from o(t) during test time → mixed training



Question 7

What is an LSTM
and how does it address the challenge of learning long-term dependencies?

- Self-loops to produce paths where gradient can flow for long durations
- Weight on self-loop conditioned on context (gates)



The repeating module in an LSTM contains four interacting layers.

https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Question 8

The forget gate in an LSTM uses a sigmoid function on the linear transformation of the hidden layer and a new input.

Could other functions be used as well and why (not)?

- Obtain values between 0 and 1 (how much of the information goes through the gate)

Your exam questions

1 MLPs, Gradient Descent & Backpropagation
2 CNNs
3 RNNs, LSTMs
4 Attention & Memory
5 Practical Methodology/Good Practice
6 Regularization
7 Optimization
8 Autoencoders
9 Introspection

- For each topic: Write down 1 or 2 questions, which you would ask as the examiner (or you would like to be asked), individually - *if possible digitally* 30‘
- Try your favorite questions on your neighbor 15‘
- Afterwards, I‘ll collect the questions 5‘
  I will have a look all questions
  Those which I find suitable for the exam,
  I will share with you – and also use some of them

Assignments

- Reading on Model Compression & Transfer Learning (no exercise next week, but Q&A)
- Participate in the PEP evaluation **until Sunday Feb 3**
- Time for your project
- prepare your final project presentation
  - write me on MM if you want to present and how much time you will need **until Friday Feb 1**

Slides & assignments on: https://mlcogup.github.io/idl_ws18/schedule