Sequential Data

Patryk Utkała
Gradient Science Club 2023

Plan for Today

- Importance of order in data
- Sequential Data
- Recurrent Neural Network
- Exploding and vanishing gradients
- Long Short-Term Memory
- Introduction to Transformer

Resources

Python. Machine learning i deep learning. Biblioteki scikit-learn i TensorFlow 2. Wydanie III

Autorzy: **Sebastian Raschka, Vahid Mirjalili**

Importance of order in data

Bag-of-words

- Classic NLP model formed around counting words in data
- Ignores order of words
- Can somewhat keep track of order by using n-gram analysis
- Fixed vector length equal to vocabulary size
- Informes about word frequency

"The movie was not long and interesting"

"The movie was long and not interesting"

Sequential Data

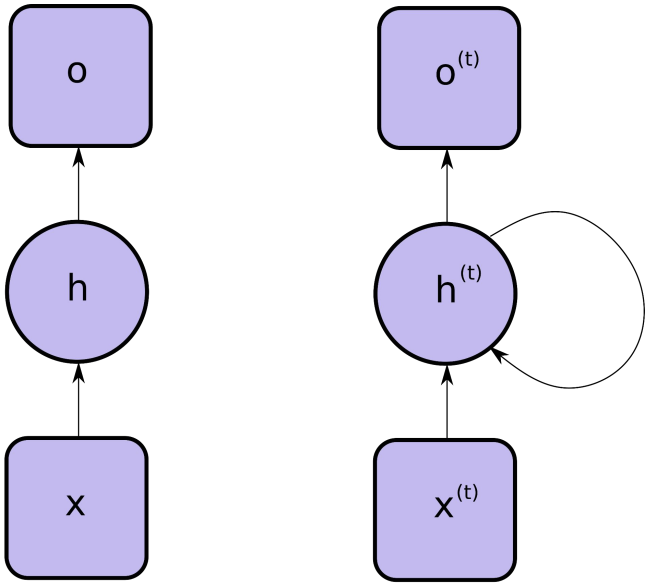
- In typical ML data elements are independent and identically distributed
- Sequential data means that elements are ordered into sequence

Sequence model categories

- many-to-one
- one-to-many
- many-to-many synchronized
- many-to-many delayed

Recurrent Neural Network

RNN

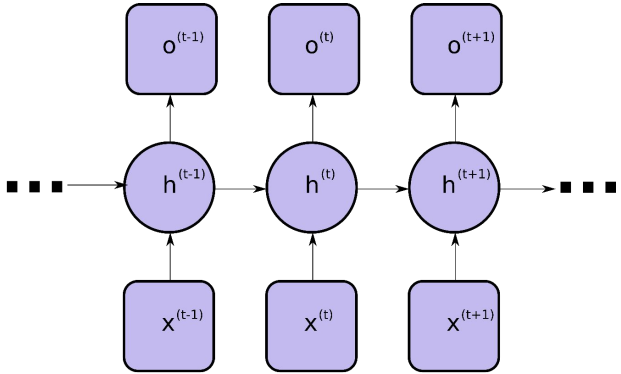


Standard single direction NN

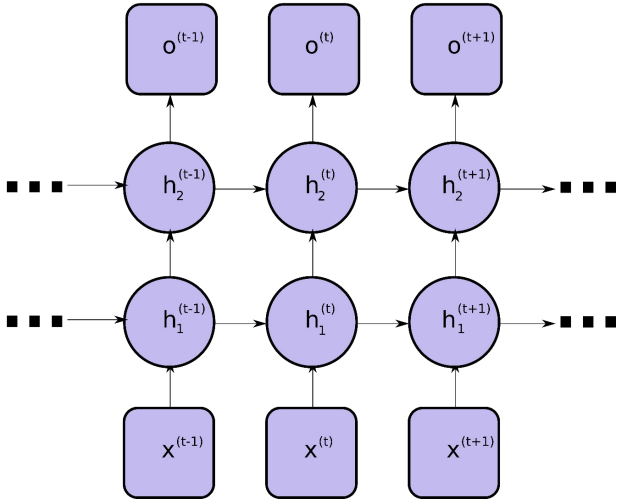
Recurrent NN

RNN

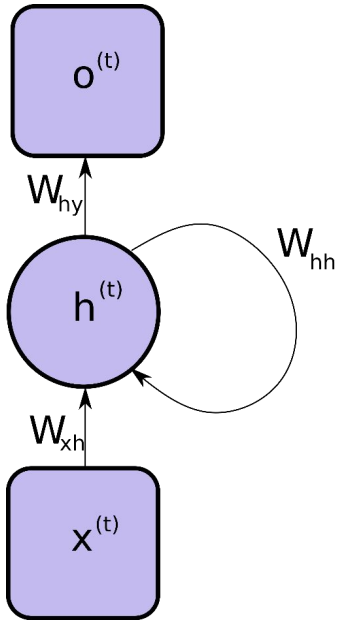
Single Layer



Multi Layer



RNN



$$h^{(t)} = F_h(W_{xh}x^{(t)} + W_{hh}h^{(t-1)} + b_h)$$

$$o^{(t)} = F_o(W_{ho}h^{(t)} + b_o)$$

Exploding and vanishing gradients

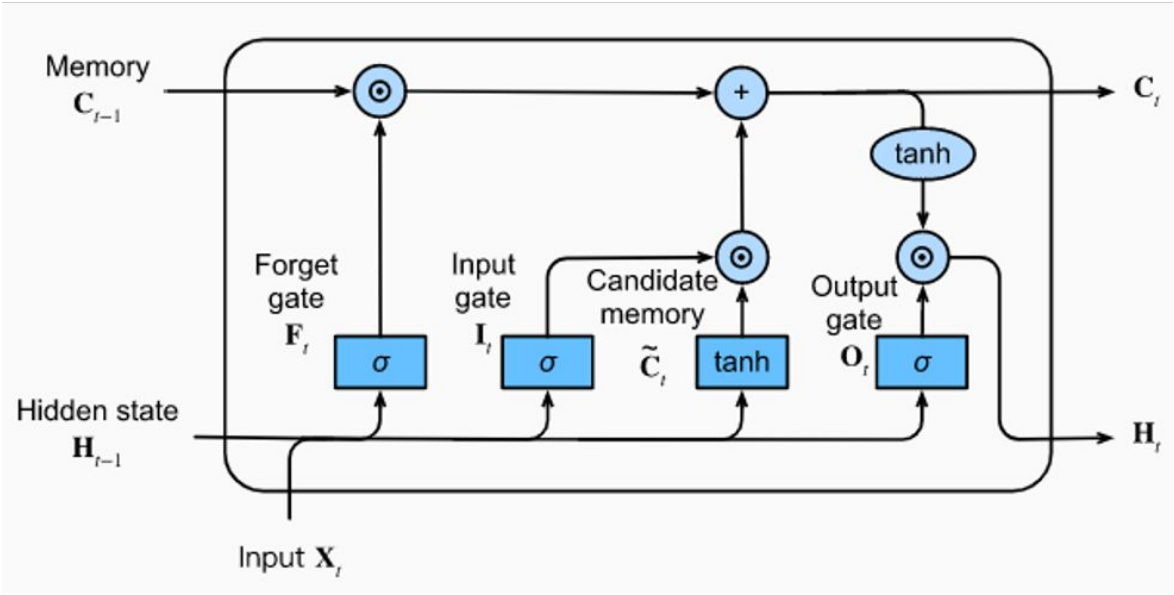
- dh^t/dh^k in loss function gradient calculation which comes to calculating $W_{hh}^{(t-k)}$
- Exploding gradient - $|W_{hh}| > 1$ => 1.1 ** 100 ~= 13780
- Vanishing gradient - $|W_{hh}| < 1$ => 0.9 ** 100 ~= 2.65e-5
- Gradients tend to vanish or explode for long sequences

Methods for dealing with exploding and vanishing gradients

- Gradient Clipping - set max value for gradients
- Truncated Backpropagation Through Time - propagates gradient through fixed steps
- Long Short-Term Memory

LSTM

LSTM



https://medium.com/@ottaviocalzone/an-intuitive-explanation-of-lstm-a035eb6ab42c

LSTM problems

- Difficult to train
- Can’t use transfer learning
- Parallelization is impossible

Transformer

- Proposed in 2017 in paper called “Attention is all you need”
- Based on attention mechanism
- Good with parallelization
- Transfer learning is possible
- Still state of the art

Questions
&
Discussion

Hands-on

Hands-on Title

All hands-on materials available at
github.com/Gradient-PG/gradient-live-session

Thank you!
See you next week on Reinforcement Learning.