Important Facts Regarding Deep Learning

Randil Pushpananda

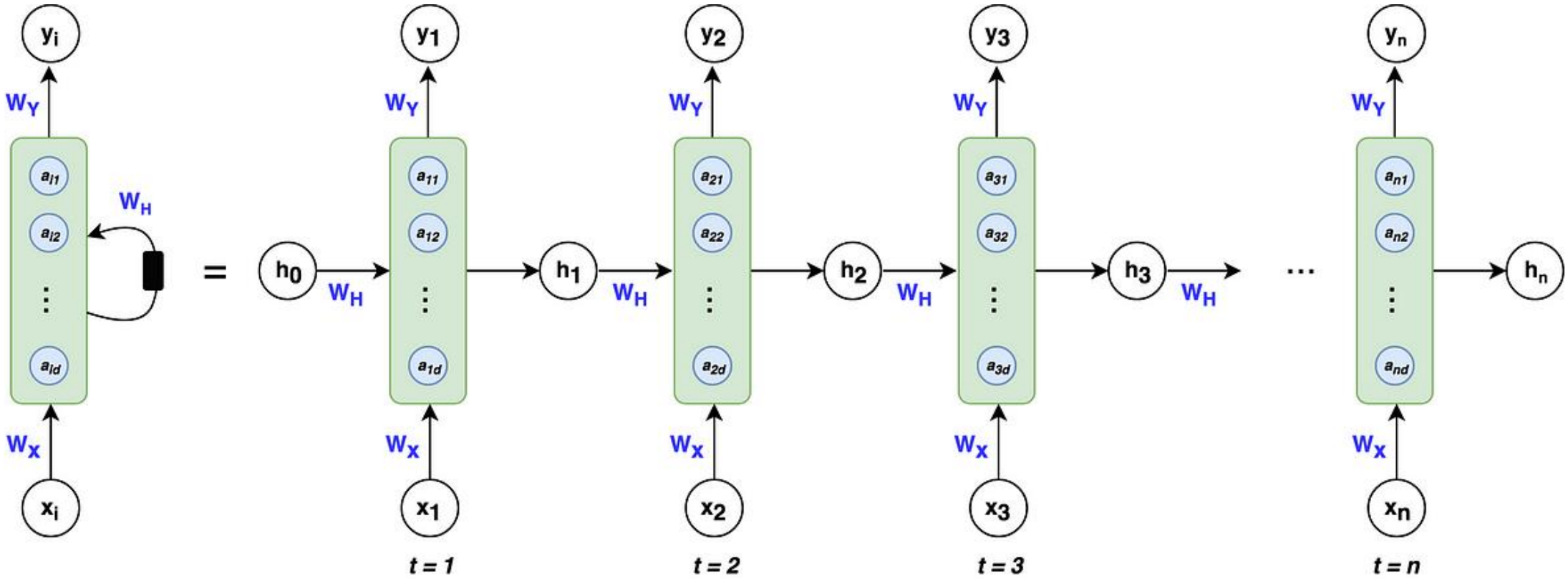
University of Colombo School of Computing

No 35, Reid Avenue, Colombo 07

rpn@ucsc.cmb.ac.lk

Natural Language Processing

RNN Revisited



Problems of RNN

- RNNs are not able to keep track of Long Term Dependencies
- Vulnerable to vanishing and exploding gradient problems

Problems of RNN

- RNNs are not able to keep track of Long Term Dependencies
- Vulnerable to vanishing and exploding gradient problems

LSTM And GRU

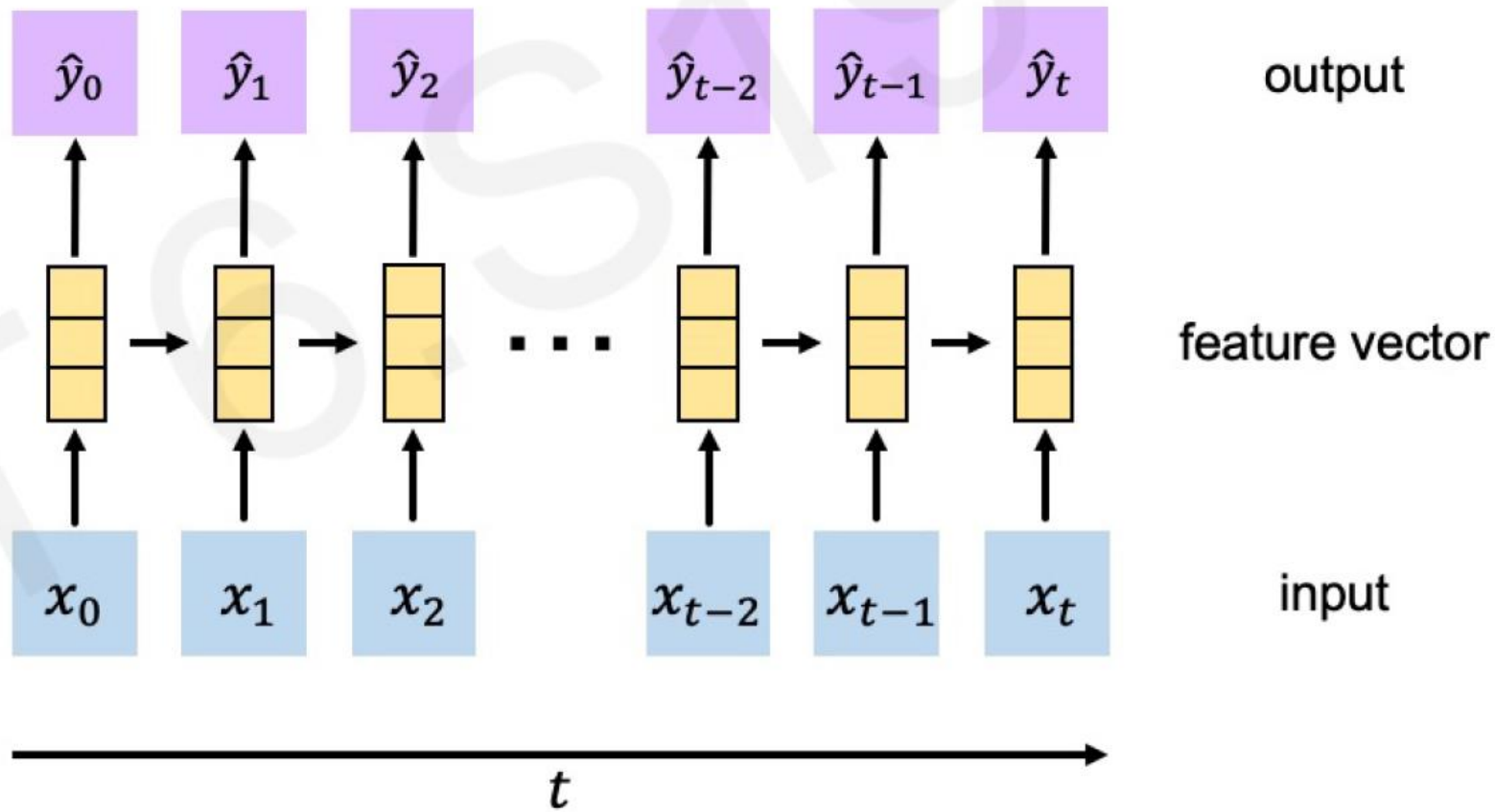
- Computationally Expensive
- Sill have the problems for very large sequences

Goal of Sequence Modeling

RNNs: recurrence to model sequence dependencies

Limitations of RNNs

- Encoding bottleneck
- Slow, no parallelization
- Not long memory



Attention Is All You Need (Transformer)

- Intuition Behind Self Attention
  - Attending to the most important parts of an input.

- Attention in computer vision (2014)
  - Attention used to highlight important parts of an image that contribute to a desired output.

- Attention in NLP:
  - 2015: Machine Translation
  - 2017: Language modeling with Transformer Networks

- Sequence Modeling
  - Challenges with RNNs:
    - Long Range Dependencies
    - Gradient Vanishing and Explosion
    - Large # of Training Steps
    - Recurrence Prevents Parallel Computation

- Transformer Networks
  - Facilitate Long Range Dependencies
  - No gradient vanishing and explosion
  - Fewer Training Steps
  - No recurrence that facilitate parallel computation

What is Self-Attention

- In the context of Transformers, self-attention refers to how the model assigns importance or attention to different positions or elements within an input sequence when making predictions or processing that sequence.

- Read the following Paper:

https://arxiv.org/abs/1706.03762

NOTE: one or two questions are based on the above paper