RNN group - Artificial Intelligence Project

Members:

Martina Ceccon, Mirko Gaslini, Laura Gotra, Jonathan Meyer, Stan Schepers

First step: Understand what RNN and LTSM are

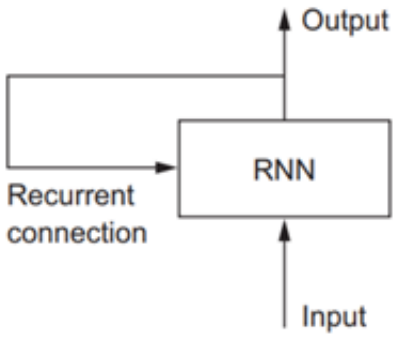
We shared papers and YouTube videos to gain an understanding of how RNNs works and what kinds there are, focussing on LSTMs and GRUs.

- **Youtube playlist:**
  https://www.youtube.com/playlist?list=PLpaiN9Sbao5lJb9N3F3PPoJSAp4rMB4vi

- **GitHub link:**
  https://github.com/MirkoGaslini/Artificial_Intelligence_20-21

What is a Recurrent Neural Network (RNN)?

"Recurrent means the output at the current time step becomes the input to the next time step. At each element of the sequence, the model considers not just the current input, but what it remembers about the preceding elements."[1]



Why use Long Short-Term Memory (LSTM)?

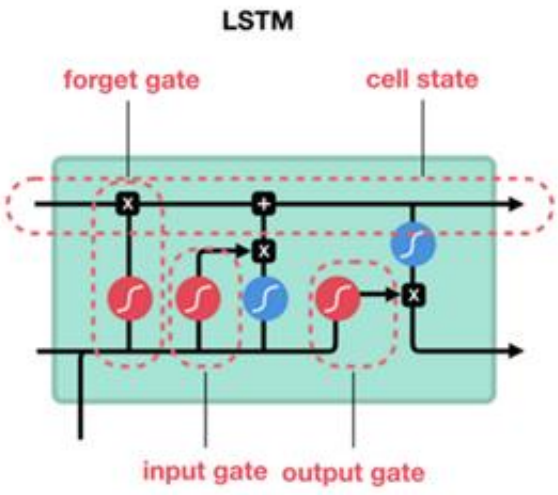
➔ Problem: *The Vanishing/Exploding Gradient Problem*

"In a network of n *hidden layers,* n *derivatives will be multiplied together. If the derivatives are large then the gradient will increase exponentially as we propagate down the model until they eventually explode, and this is what we call the problem of* **exploding gradient**. *Alternatively, if the derivatives are small then the gradient will decrease exponentially as we propagate through the model until it eventually vanishes, and this is the* **vanishing gradient** *problem"*[2]

[1] https://towardsdatascience.com/recurrent-neural-networks-by-example-in-python-ffd204f99470
[2] https://towardsdatascience.com/the-vanishing-exploding-gradient-problem-in-deep-neural-networks-191358470c11

How does Long Short-Term Memory (LSTM) work?

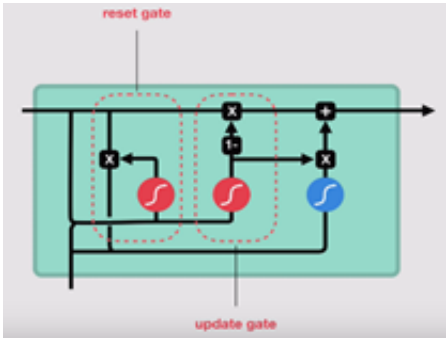
A common LSTM unit is composed of a *cell*, an *input gate*, an *output gate* and a *forget gate*. The advantage of an LSTM cell compared to a common recurrent unit is its cell memory unit. The cell vector has the ability to encapsulate the notion of forgetting part of its previously stored memory, as well as to add part of the new information. To illustrate this, one has to inspect the equations of the cell and the way it processes sequences under the hood.[3]



(https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21)

Gated Recurrent Units (GRUs)

Gated Recurrent Units have a similar architecture to LSTM networks, but only use two gates as opposed to the three shown above. The other main difference is that instead of a cell-state and a hidden state, GRUs only make use of a hidden state. The two gates are similar in function to the ones used in LTSMs; the update gate regulates which information is used from the past (memory) and current input (i.e. it has the same role as the forget and input gates of LTSM), and the reset gate controls how much past information is forgotten. The main advantage of the GRU is a lower computational complexity, which makes it faster to train than the LTSM.



[3] https://en.wikipedia.org/wiki/Long_short-term_memory

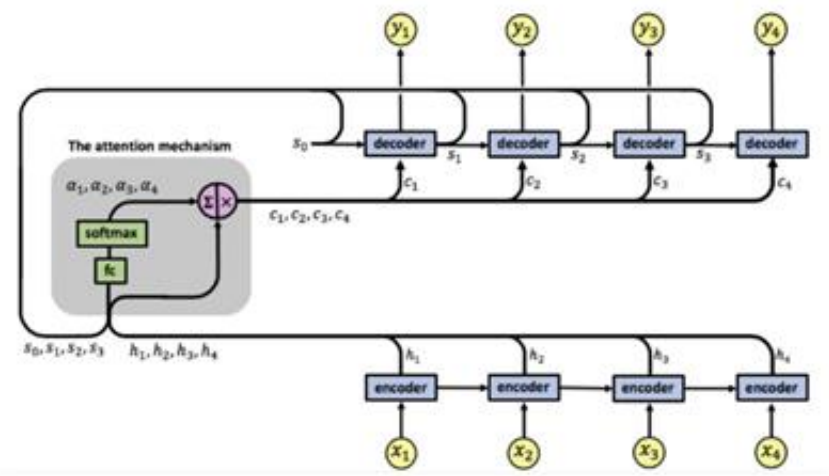
What about Attention?

Attention is a mechanism combined in the RNN allowing it to focus on certain parts of the input sequence when predicting a certain part of the output sequence, enabling easier learning and of higher quality. Combination of attention mechanisms enabled improved performance in many tasks making it an integral part of modern RNN networks.[4]

See the paper:

"Attention is all you need"

https://github.com/MirkoGaslini/Artificial_Intelligence_20-21/blob/main/attention-is-all-you-need.pdf



RNN with an attention mechanism

Useful libraries for python:

- Keras: https://keras.io/api/layers/recurrent_layers/
- Theano: https://pypi.org/project/theano-lstm/
- Pytorch: https://medium.com/intel-student-ambassadors/implementing-attention-models-in-pytorch-f947034b3e66

Second step: Decide what to do

We thought about to implement a LSTM recurrent neural network and test the results with and without **attention**.

[4] https://medium.com/datadriveninvestor/attention-in-rnns-321fbcd64f05#:~:text=Attention%20is%20a%20mechanism%20combined,learning%20and%20of%20higher%20quality.&text=The%20encoder%20outputs%20a%20single,as%20input%20to%20the%20decoder.https://duckduckgo.com/?ia=answer