**Paper Link**

**Video Summary**

N/A

**Pre-requisite papers (Roots)**

This stuff is pretty self-explanatory with all the other stuff in this folder given since more complex models are based on this, but these notes are meant to make the differences more apparent

**Notes**

Equations for standard RNN

The RNN works when the alignment between inputs and outputs is known ahead of time

Map the input sequence to a fixed size vector using 1 RNN and map the target vector to the target sequence with another RNN, possibly using an LSTM for long term dependencies



LSTM is given an entire input sequence represented by the fixed vector v (given by the final LSTM's hidden state)

Deep LSTMs (stacked) outperform regular LSTMs

This paper reversed the order of the input sequence (c, b, a) and maintained the order of the output sequence so that the first words for the prediction and the input sequence would be close to each other / related

Reversing the input sequence is different than using a bidirectional architecture

**Branches**

**Applications/Insights**

Reversing the order of an input sequence is useful (even when using a bidirectional architecture)