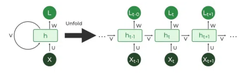
Day-9: LSTM & GRU Networks

Need for LSTM and GRU Networks

RNN Cell

Most important part of an RNN is the **Hidden State (a.k.a. Memory State)**

Remembers some information about a sequence



Vanishing Gradient Problem

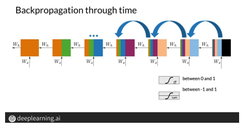
Suffer from short-term memory

Difficulty in carrying information from earlier steps to later ones

Eg: Trying to process a paragraph of text, RNN's may leave out important information

During backpropogation, gradient shrinks as it back propagates through time

So the gradient value become extremely small and doesn't contribute much to learning



Long Short-Term Memory (LSTM) Network

Designed to overcome the vanishing gradient problem faced by traditional RNNs

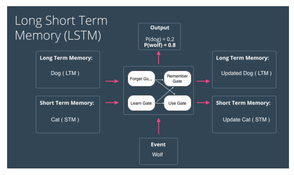
Movie example

It follows two mechanisms

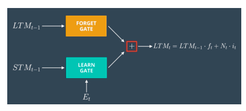
Forgetting Mechanism: Forget all scene information that is not worth remembering.

Saving Mechanism: Save information that is important and can help in the future

Architecture



Learn Gate: Necessary information that we have recently learned from STM can be applied to the current input.



Takes Event (Et) and Previous Short Term Memory (STMt-1) as input and keeps only relevant information for prediction.

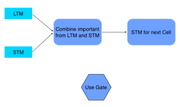
Forget Gate: forgets information that is not useful.

Takes Previous Long Term Memory (LTMt-1) as input and decides on which information should be kept and which to forget.

Remember Gate: LTM information that we haven't forgotten and STM and Event are combined together in Remember gate which works as updated LTM.

Combine Previous Short Term Memory (STMt-1) and Current Event (Et) to produce output

Use Gate: This gate also uses LTM, STM, and Event to predict the output of the current event which works as an updated STM.



Combine important information from Previous Long Term Memory and Previous Short Term Memory to create STM for next and cell and produce output for the current event.

Gated Recurrent Units (GRU) Network

Very similar to Long Short Term Memory(LSTM)

Just like LSTM, GRU uses gates to control the flow of information.

But it's architecture is simpler that LSTM, making it faster and better memory management

Architecture

Reset Gate (Short term memory)

The Reset Gate is responsible for the short-term memory of the network i.e the hidden state (Ht). Here is the equation of the Reset gate

Update Gate (Long Term memory)

Similarly, we have an Update gate for long-term memory and the equation of the gate is shown below.

Working

The GRU takes two inputs as vectors: the current input and the previous hidden state

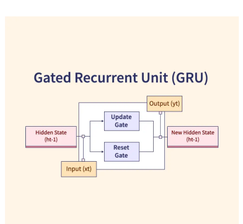
We'll calculate the values for each gate

We perform an element-wise multiplication between the current input and the previous hidden state vectors (Like dot-product)

Creating vectors of input specefic to each gate

Apply an activation function element-wise to each element in these vectors

Outputs values between 0 and 1, which will be used by the gates to control information flow



Hands-on