Jan Snoeijs 18/03/2018

Gated Recurrent Unit Networks

Report

Mathematical derivations applied to the problem:

Forward pass:

The choice of the reLU function for the output layer is to enable larger values than one at the output because the network is supposed to compute the sum of the numbers in the sequence.

Backward pass:

Where the following expression is the non-recursive part of

Which results in this expression:

Similarly, for the other coefficients:

For the remaining coefficients, only the partial non-recursive part is derived here because the rest is the same expression as for the derivatives computed above.

Finally, for the 3 bias coefficients we have:

The formulas derived above represents the error for one coefficient at a given time step and the total error should sum them all up over the whole sequence. (or part of it if using TBPTT instead of classical BPTT)

Algorithmically this translates into a double *for* loop.