Deep Learning

How Data Scientists become magicians

Sequence Data

Data that includes an order representation of its components is called *sequence data*.

Denver is nicer than Boulder.

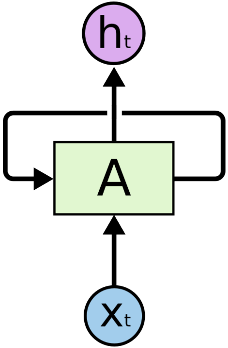
Boulder is nicer than Denver.

Sequence Data

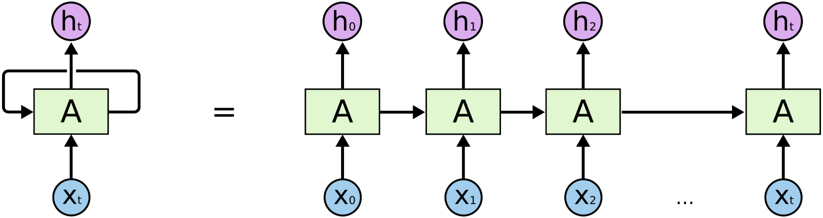
With a vocabulary of ”ATCG”, we can represent a string ”AACG” as,

$$x_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, x_2 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, x_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

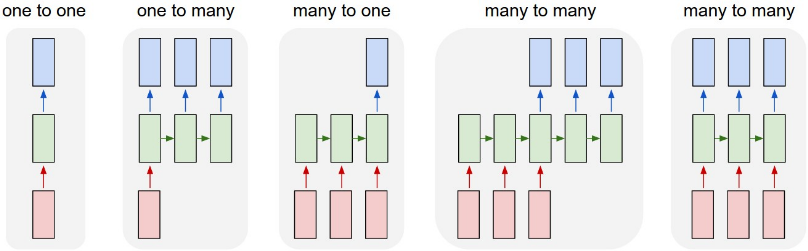
Recurrent Neural Networks



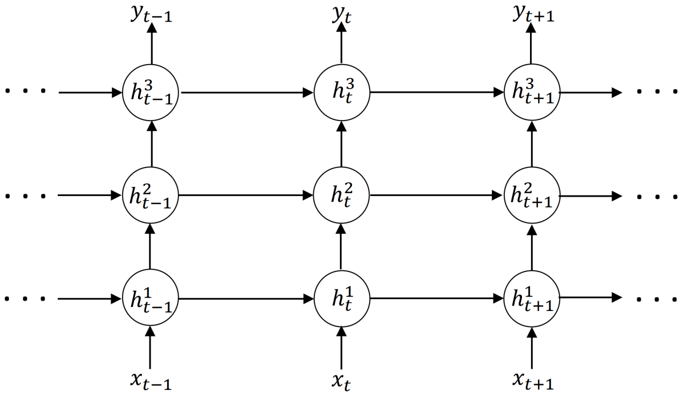
Recurrent Neural Networks



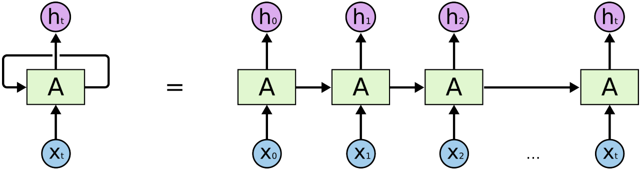
Types of Recurrent Neural Networks



Deep Recurrent Neural Networks



Recurrent Neural Network - Calculations

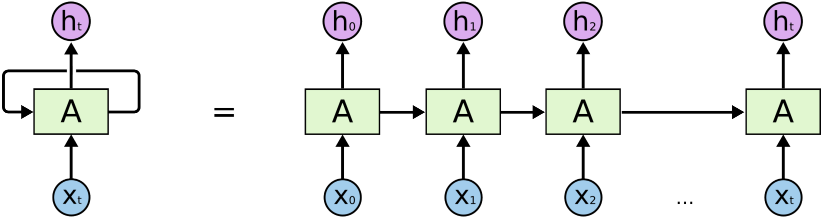


$$
\begin{aligned}
\mathbf{a}_t &= f(\mathbf{W}_{ax}\mathbf{x}_t + \mathbf{W}_{aa}\mathbf{a}_{t-1} + \mathbf{b}_a), \\
\mathbf{h}_t &= f(\mathbf{W}_{ha}\mathbf{a}_t + \mathbf{b}_h).
\end{aligned}
$$

Vanishing Gradient Problem

Deeper networks result in minimal changes in layers closer to input layer.

Vanishing Gradient Problem



Gated Recurrent Units

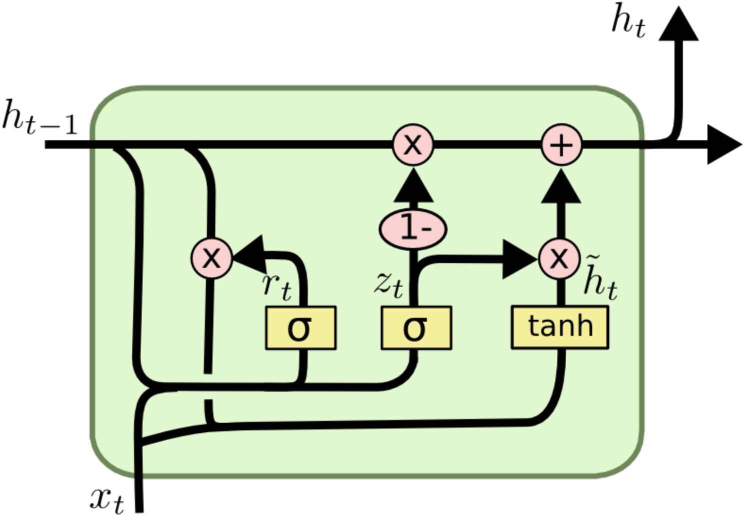
- Update gate - controls how much to update the passed through values
- Reset gate - how to reset the passed through value based on the current input

$$\mathbf{z}_t = f(\mathbf{W}_{zh}\mathbf{h}_{t-1} + \mathbf{W}_{zx}\mathbf{x}_t + \mathbf{b}_z),$$
$$\mathbf{r_t} = f(\mathbf{W}_{rh}\mathbf{h}_{t-1} + \mathbf{W}_{rx}\mathbf{x}_t + \mathbf{b}_r).$$

Gated Recurrent Units

$$\tilde{\mathbf{h}}_t = g(\mathbf{W}_{hh}(\mathbf{r}_t * \mathbf{h}_{t-1}) + \mathbf{W}_{hx}\mathbf{x}_t + \mathbf{b}_h)$$
$$\mathbf{h}_t = (1 - \mathbf{z}_t) * \mathbf{h}_{t-1} + \mathbf{z}_t * \tilde{\mathbf{h}}_t$$

Gated Recurrent Units



Long Short Term Memory Cells

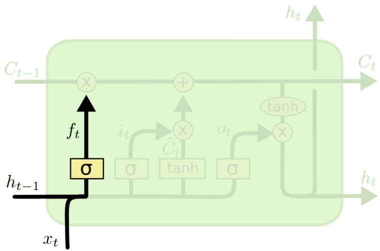
- Forget gate - controls how much to forget from past memory
- Input gate - controls how much to update future memory based on current input
- Output gate - controls how to update the cell output

$$
\begin{aligned}
\mathbf{f}_t &= f(\mathbf{W}_{fh}\mathbf{h}_{t-1} + \mathbf{W}_{fx}\mathbf{x}_t + \mathbf{b}_f), \\
\mathbf{i}_t &= f(\mathbf{W}_{ih}\mathbf{h}_{t-1} + \mathbf{W}_{ix}\mathbf{x}_t + \mathbf{b}_i), \\
\mathbf{o}_t &= f(\mathbf{W}_{oh}\mathbf{h}_{t-1} + \mathbf{W}_{ox}\mathbf{x}_t + \mathbf{b}_o).
\end{aligned}
$$

Long Short Term Memory Cells

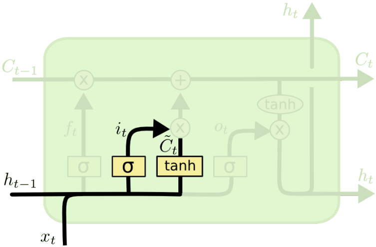
$$
\begin{aligned}
\tilde{\mathbf{c}}_t &= g(\mathbf{W}_{ah}(\mathbf{h}_{t-1}) + \mathbf{W}_{ax}\mathbf{x}_t + \mathbf{b}_a),\\
\mathbf{c}_t &= \mathbf{f}_t * \mathbf{c}_{t-1} + \mathbf{i}_t * \tilde{\mathbf{c}}_t,\\
\mathbf{h}_t &= \mathbf{o}_t * g(\mathbf{c}_t).
\end{aligned}
$$

LSTM - Forget Gate



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

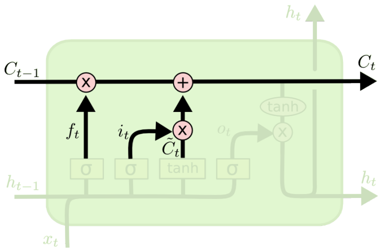
LSTM - Input Gate



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

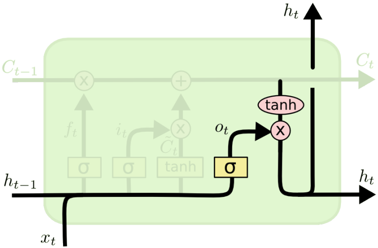
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

LSTM - Cell Memory Update



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

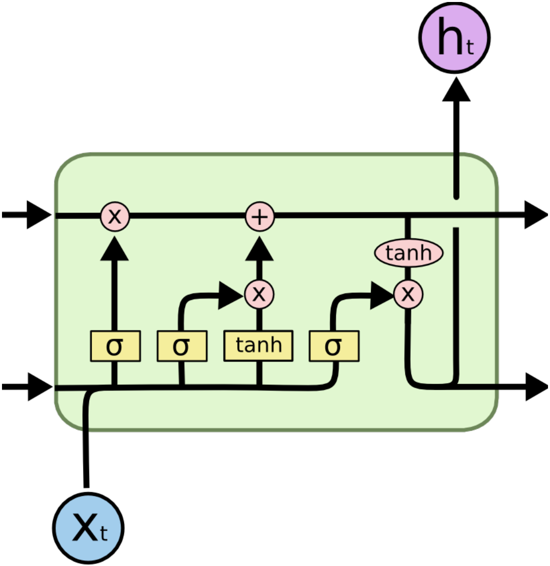
LSTM - Output Gate



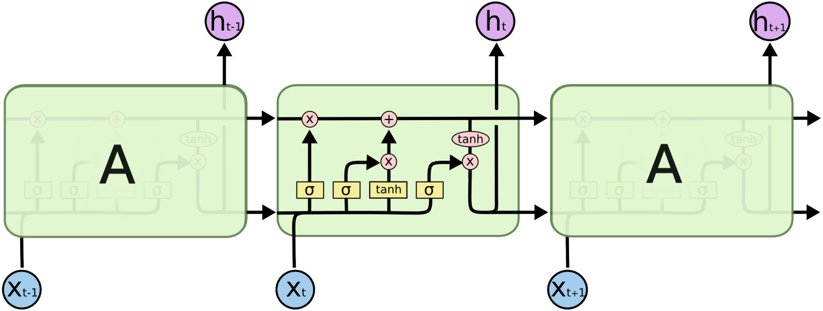
$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

LSTM



LSTM



Bidirectional RNNs

Send values forwards and backwards to previous cells.

- In text, allows us to extract more context based on words before and after the current one.
- Requires the full text to be available

Questions

These slides are designed for educational purposes, specifically the CSCI-470 Introduction to Machine Learning course at the Colorado School of Mines as part of the Department of Computer Science.

Some content in these slides are obtained from external sources and may be copyright sensitive. Copyright and all rights therein are retained by the respective authors or by other copyright holders. Distributing or reposting the whole or part of these slides not for academic use is HIGHLY prohibited, unless explicit permission from all copyright holders is granted.