Introduction to Recurrent Neural Networks

Komal Teru
Mila, McGill University

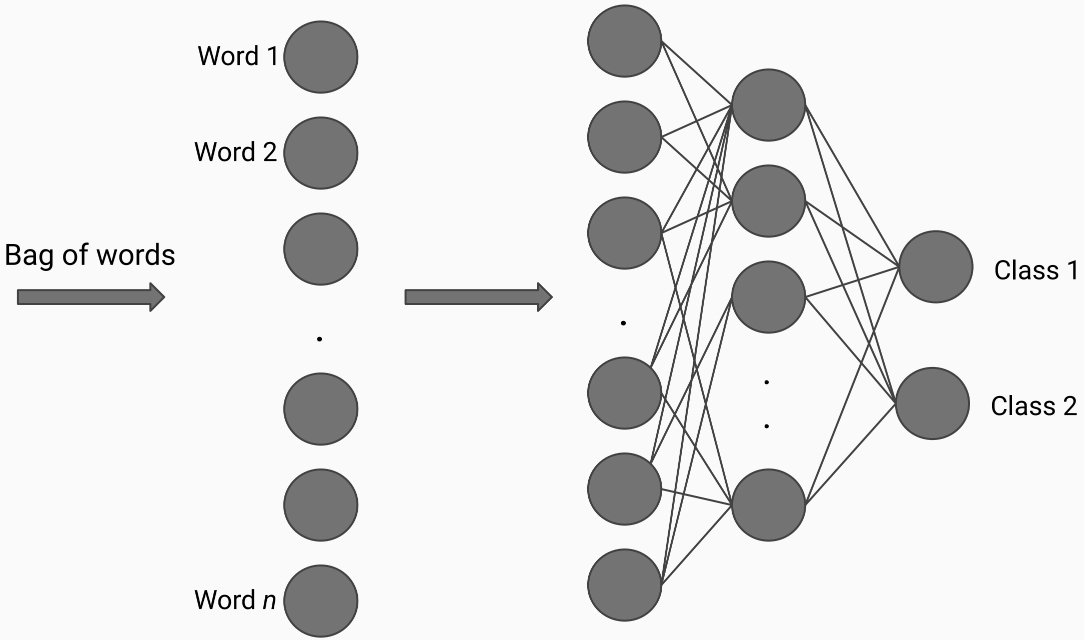
Topics to cover today

- **Intro to computations of Recurrent Neural Networks**

Dealing with text

Sentiment analysis

42 is the answer to life, universe and everything.



Dealing with text

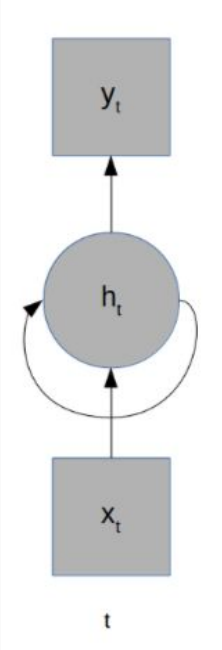
Disadvantages of using MLP with texts

- We lose the inherent temporal nature of text
- Huge number of parameters for large vocabularies

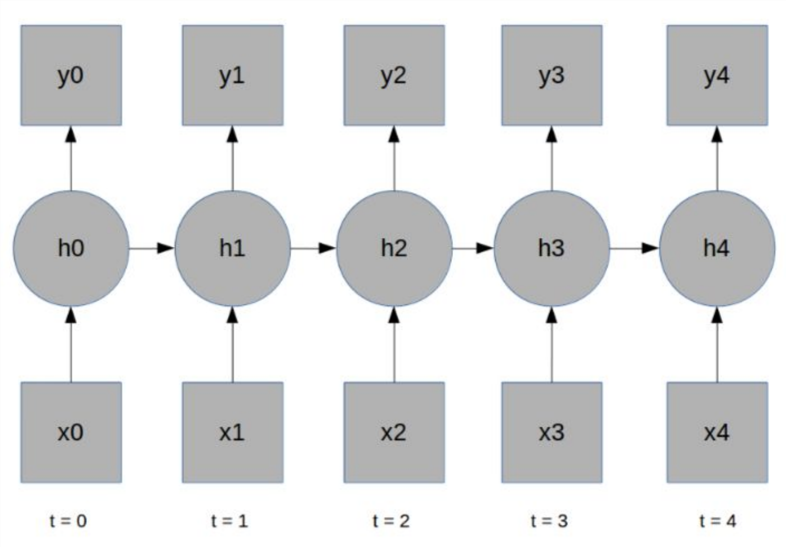
Recurrent Neural Network

$$h_t = \sigma(W_{ih}x_t + W_{hh}h_{t-1} + d_h)$$

$$y_t = W_{ho}h_t + b_o$$



unrolled in time



Recurrent Neural Network

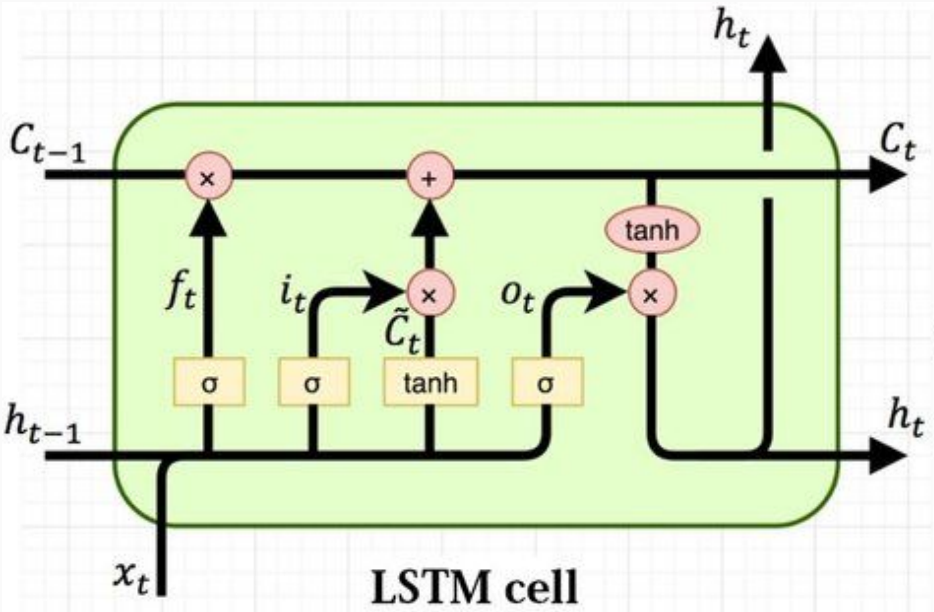
Drawbacks of vanilla RNNs:

- Can't retain long-term dependencies which are very important for many tasks. Consider the following example:

    1. Drake walked into the room. Jake walked in too. Drake said hi to —.
    2. Drake walked into the room. Jake walked in too. It was late in the day, and everyone was walking home after a long day at work. Drake said hi to —.

- It has been proven that RNNs can't answer the second question.

Long short term memory (LSTM) cell



$$i_t = \sigma\left(x_t U^i + h_{t-1} W^i\right)$$

$$f_t = \sigma\left(x_t U^f + h_{t-1} W^f\right)$$

$$o_t = \sigma\left(x_t U^o + h_{t-1} W^o\right)$$

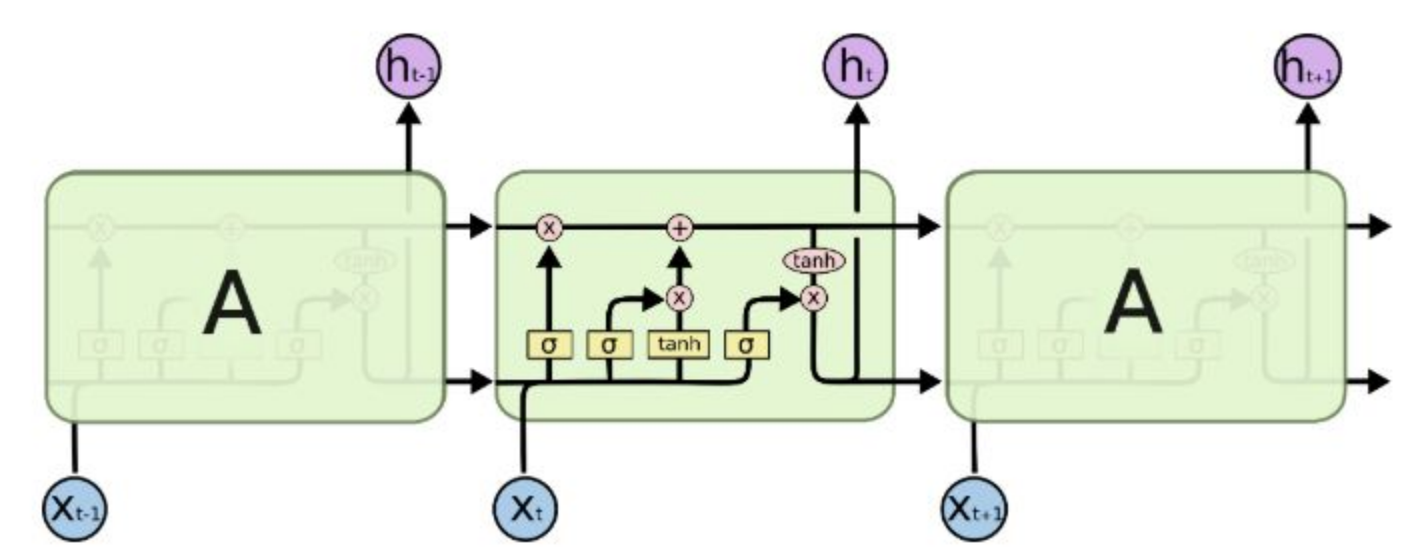
$$\tilde{C}_t = \tanh\left(x_t U^g + h_{t-1} W^g\right)$$

$$C_t = \sigma\left(f_t * C_{t-1} + i_t * \tilde{C}_t\right)$$

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

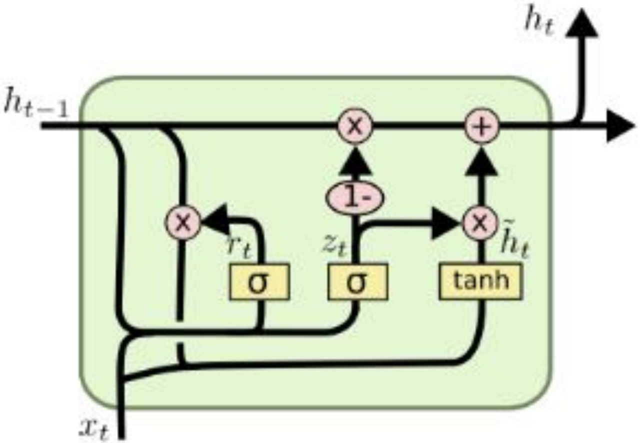
Source : http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Long short term memory (LSTM) cell



Source : http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Gated recurrent unit (GRU)



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

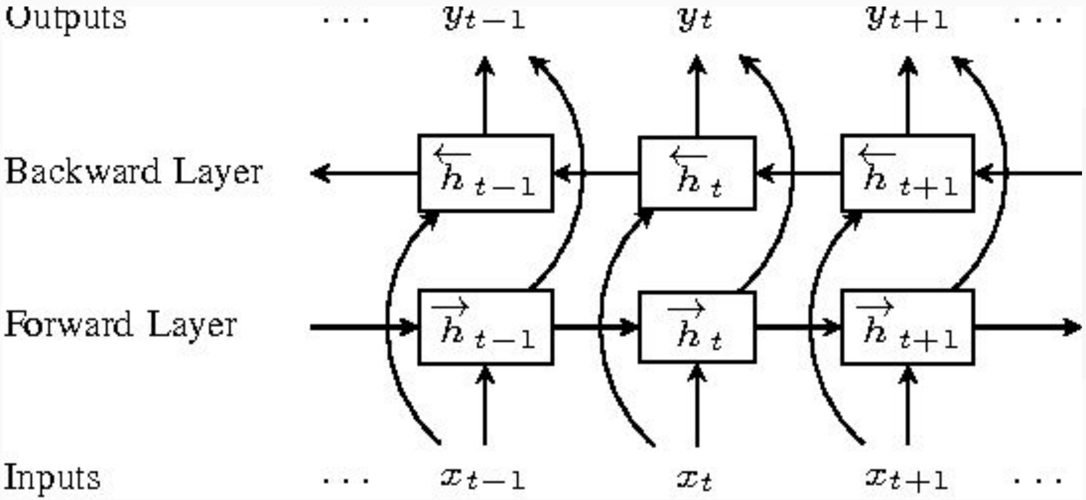
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Source : http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Bi-directional RNN



$$\overrightarrow{h}_t = f(\overrightarrow{W} x_t + \overrightarrow{V}\overrightarrow{h}_{t-1} + \overrightarrow{b}) \qquad (1)$$

$$\overleftarrow{h}_t = f(\overleftarrow{W} x_t + \overleftarrow{V}\overleftarrow{h}_{t+1} + \overleftarrow{b}) \qquad (2)$$

$$\hat{y}_t = g(Uh_t + c) = g(U[\overrightarrow{h}_t; \overleftarrow{h}_t] + c) \qquad (3)$$

Source : http://colah.github.io/posts/2015-08-Understanding-LSTMs/