Recurrent Neural Networks

Data Science Immersive

FLATIRON SCHOOL

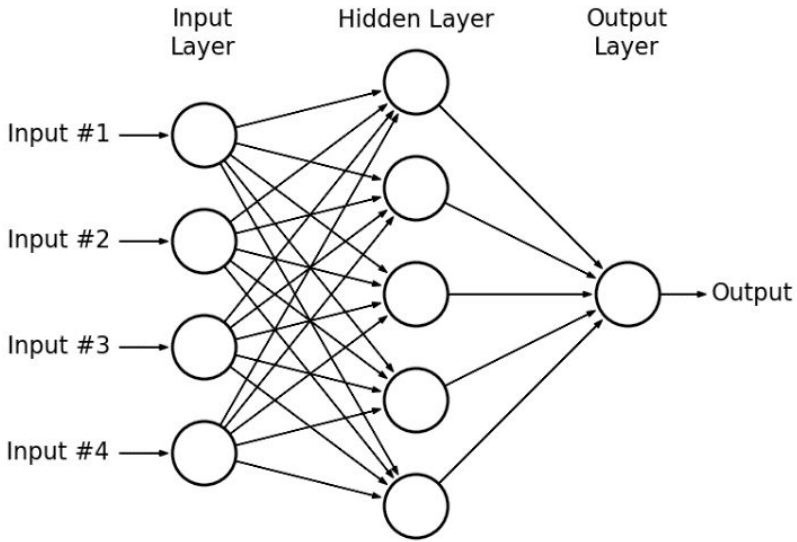
Agenda & Objectives

- Review MLP and its limitations
- Introduction to RNNs
- Introduction to LSTMs

- Understand the basic intuition of RNNs & LSTM
- Explain the use cases and applications for RNNs
- Understand how LSTM works to improve and address the shortcomings of RNNs

Why isn’t MLP enough?

- Let’s take a simple example - univariate time series forecasting. Why is a multilayer perceptron model not enough?



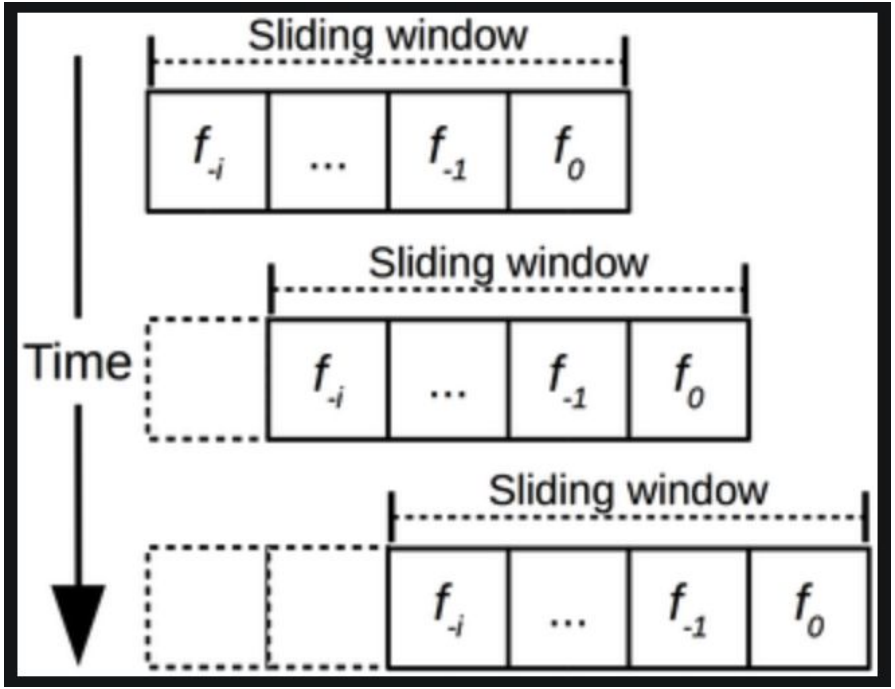
Sequence Learning

- Items in a collection has orders, or sequence
  - A,B,C produces a different outcome from A,C,B

- Several considerations
  - How can each element be represented, either numerically or as a vector?
  - Sequence length - is it variable or constant?

Sequence Learning: Examples

- ARIMA (and variants)
- Markov Models
- Sliding Window Models



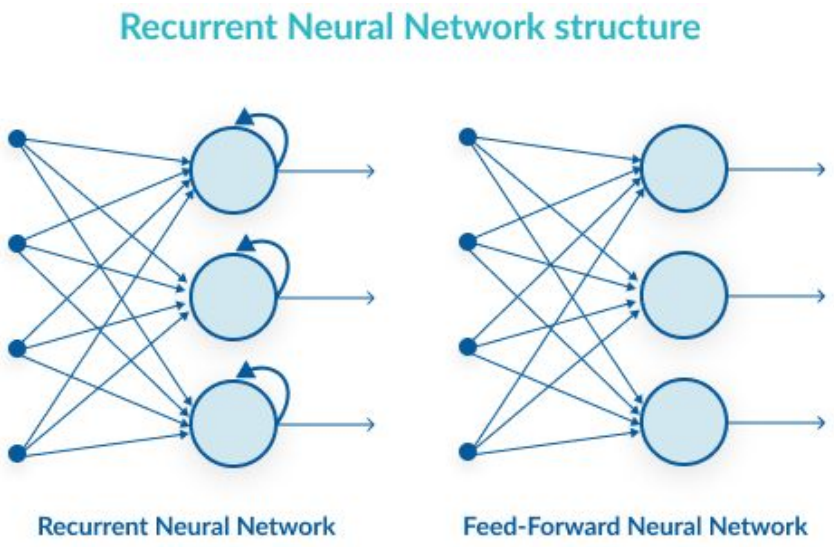
RNN Applications - Examples

- NLP
  - Predictive text
  - Sentiment analysis
- Recommendation systems
  - Collaborative filtering - (forget matrix factorization!)
  - Content filtering
- Multivariate time series forecasting
- Audio interpretation
  - Spectral analysis -> RNN -> Output
- Rudimentary image recognition
  - (CNNs are better)
- Recurrent convolutional neural networks

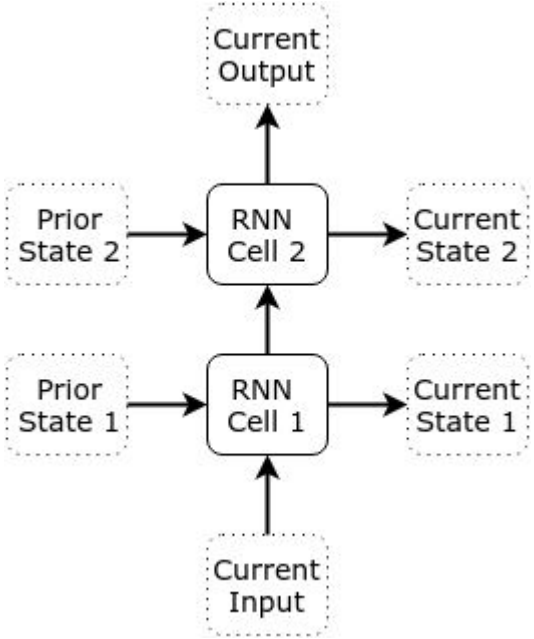
Recurrent Neural Networks

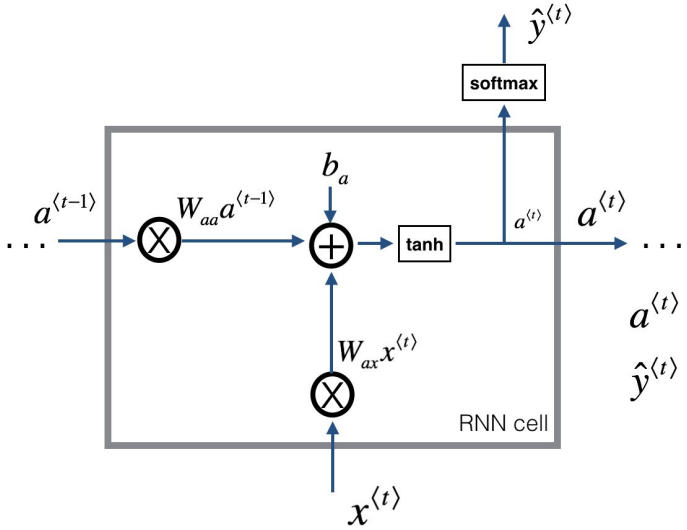
- Regular “forward feed” neural networks have an activation function that’s fed all the inputs, and then goes to the output.

- RNNs use a recurrent activation that’s also forward-propagated, and is fed from the output of the previous step in the sequence.



Forward Propagation In-Depth





$$a^{\langle t \rangle} = \tanh(W_{ax}x^{\langle t \rangle} + W_{aa}a^{\langle t-1 \rangle} + b_a)$$

$$\hat{y}^{\langle t \rangle} = soft\max(W_{ya}a^{\langle t \rangle} + b_y)$$

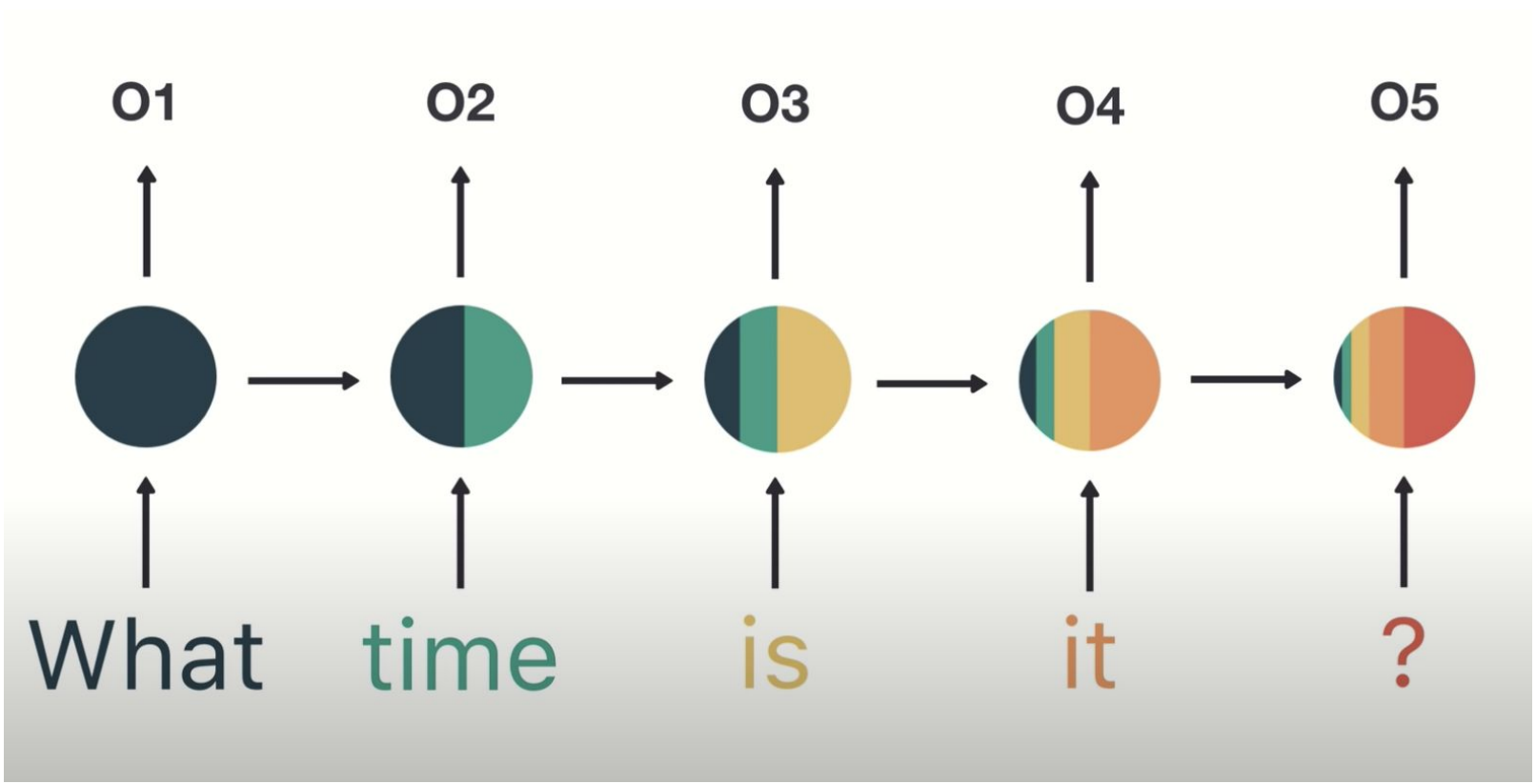
LSTM

- Diminishing gradient problem
- Short term-memory: the memory across a sequence
- Predictive text example:

There was a castle in the mountains on a river across from the valley where the king's army was marching.

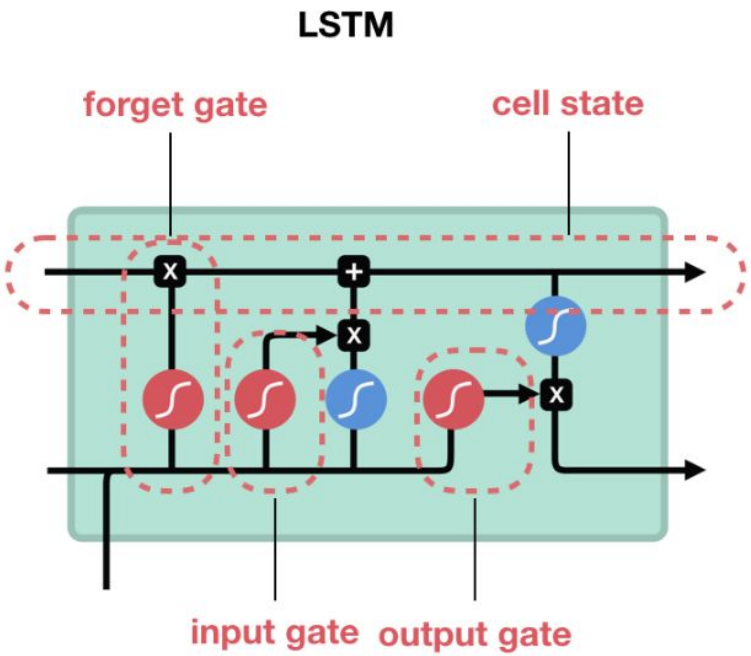
- King's -- Army
- Army -- Marching
- Castle -- King should be an obvious connection but is too separated by other words - the model has forgotten.

LSTM

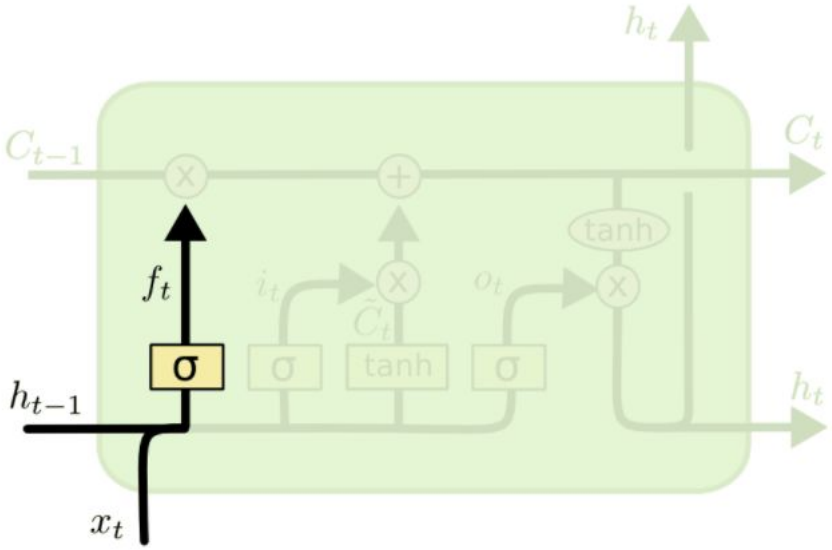


Logic Gates

- Forget gate
  - What information to keep or throw away
- Input gate
  - Updates cell state but not the output
- Output gate
  - What goes to cell state versus output
- Vectors are going into all these gates and have weights and biases associated with each, which are learned in backpropagation

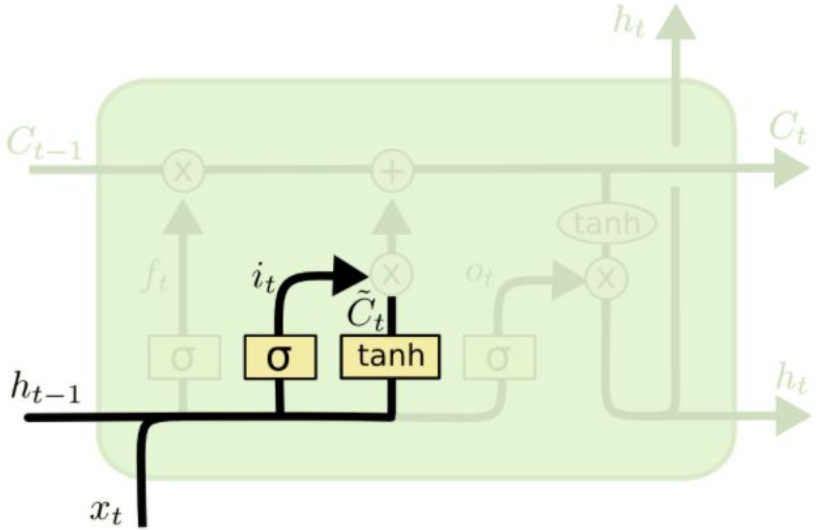


Logic Gates



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

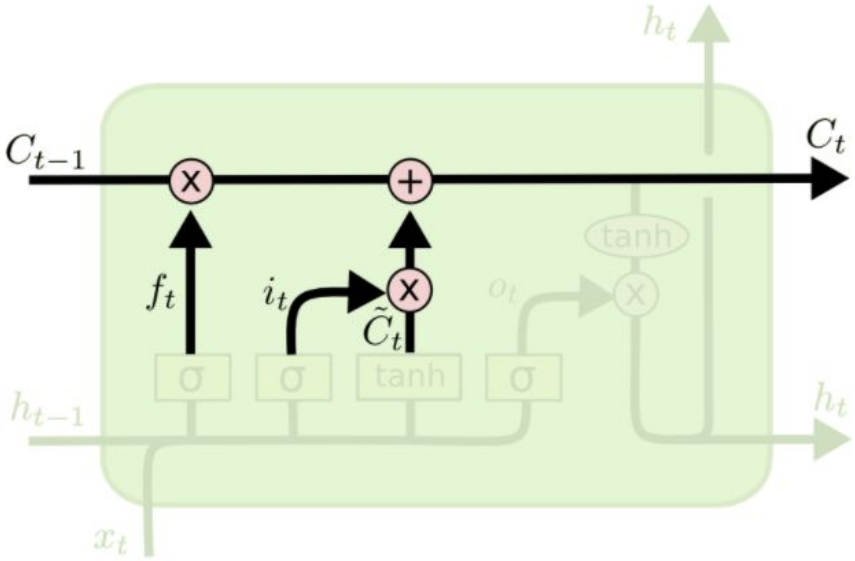
Logic Gates



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

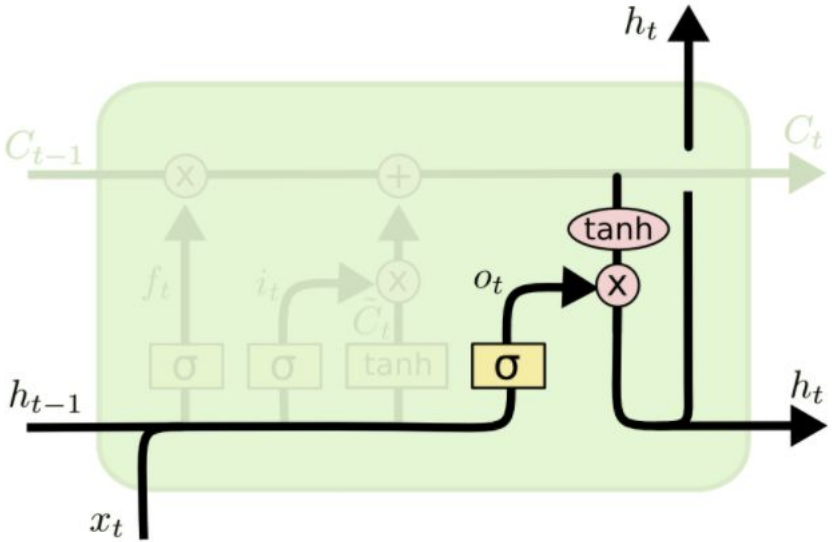
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Logic Gates



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

Logic Gates



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

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

Further Exploration

- Gated Recurrent Units
- Bidirectional RNNs

Resources

- Andrew Ng's coursera course
- Project example https://github.com/taylorhawks/RNN-music-recommender/blob/master/cloud/model.ipynb
- Live Coding RNNs: https://youtu.be/BSpXCRTOLJA
- Funny Russian guys: https://youtu.be/IycKqccytfU
- Illustrated guide to LSTM: https://youtu.be/8HyCNIVRbSU
- Famous RNN blog post: http://karpathy.github.io/2015/05/21/rnn-effectiveness/