AI Society

LSTMs

USAIS

University of Southampton AI Society

Agenda

- What is sequence data
- Problems with convolutional and feed forward networks
- Why are recurrent networks better
    - Long short-term memory (LSTM) networks
- Training LSTMs

What Is Sequence Data

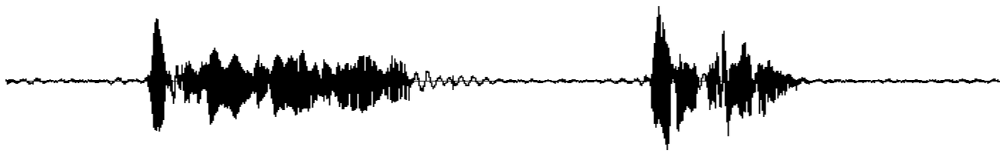
A sentence

It is going to rain today

A genome

C A C G A C T T …

Voice

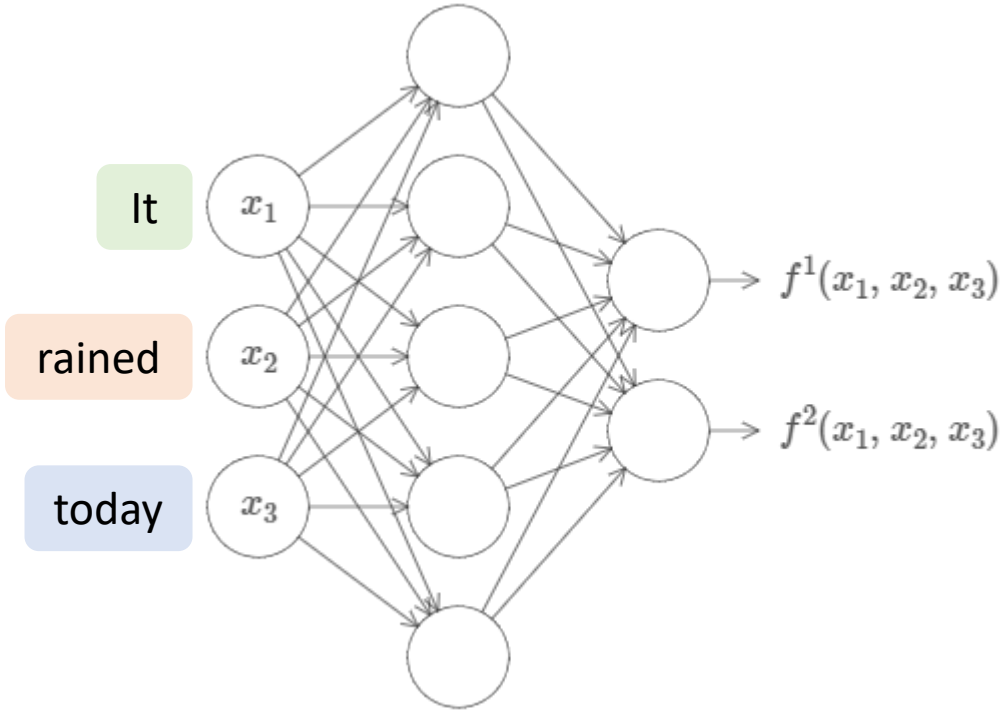


https://cmusphinx.github.io/wiki/tutorialconcepts/

Problems with Convolutional and Feed Forward Networks

Problems with Convolutional and Feed Forward Networks

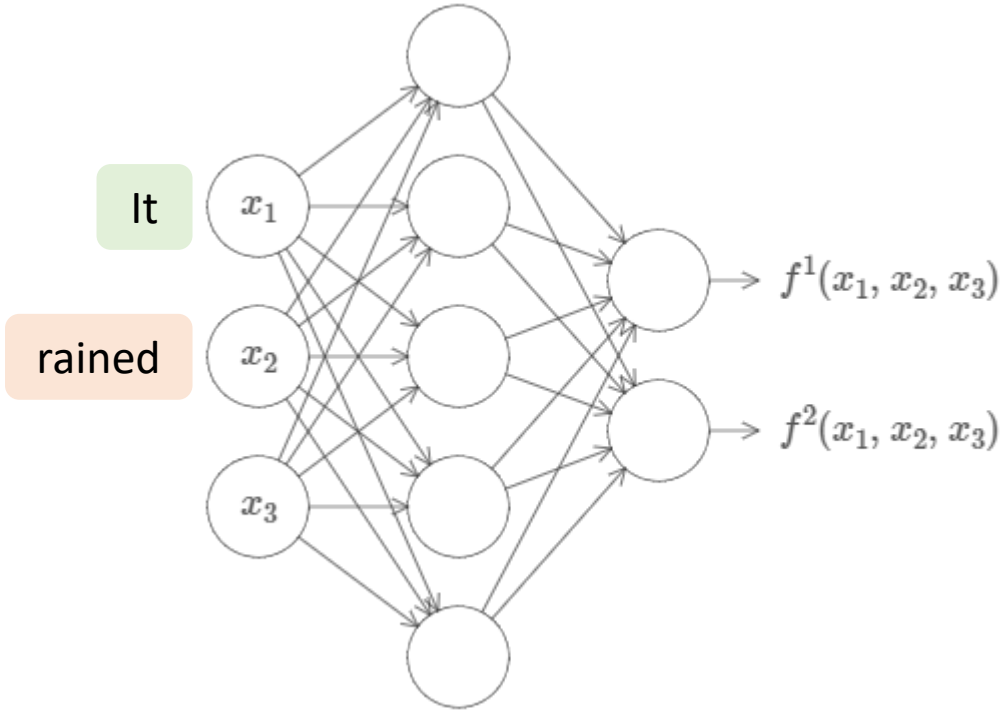
Windowing



http://neuralnetworksanddeeplearning.com/chap4.html

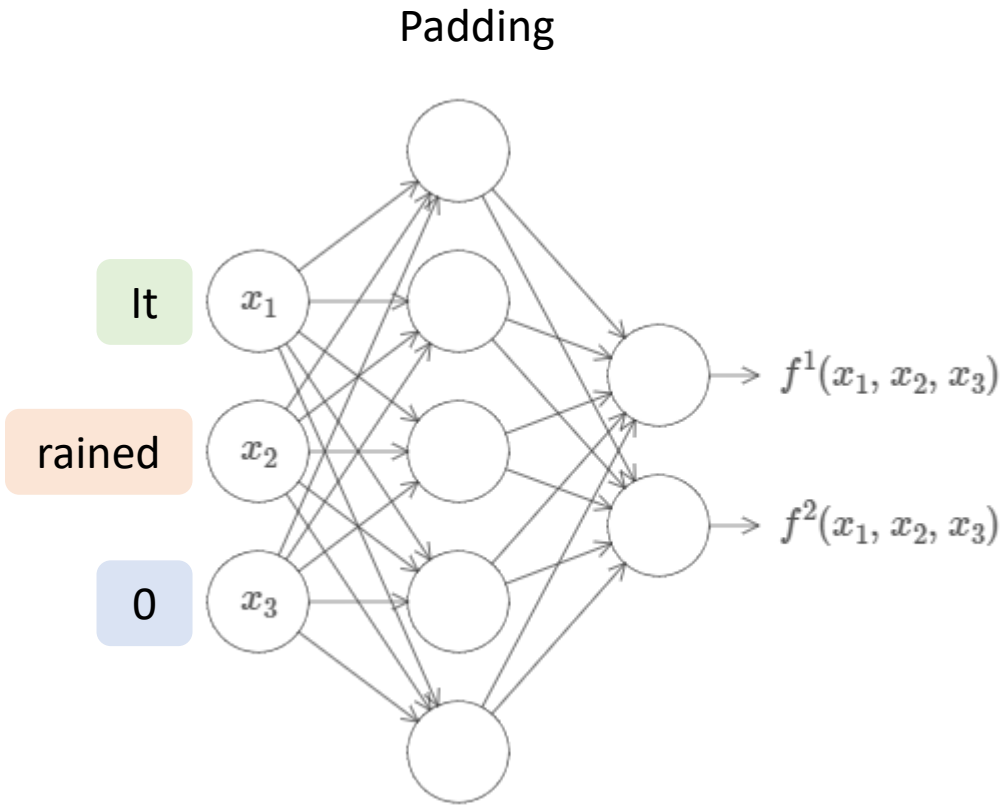
Problems with Convolutional and Feed Forward Networks

What if the sentence is shorter?



http://neuralnetworksanddeeplearning.com/chap4.html

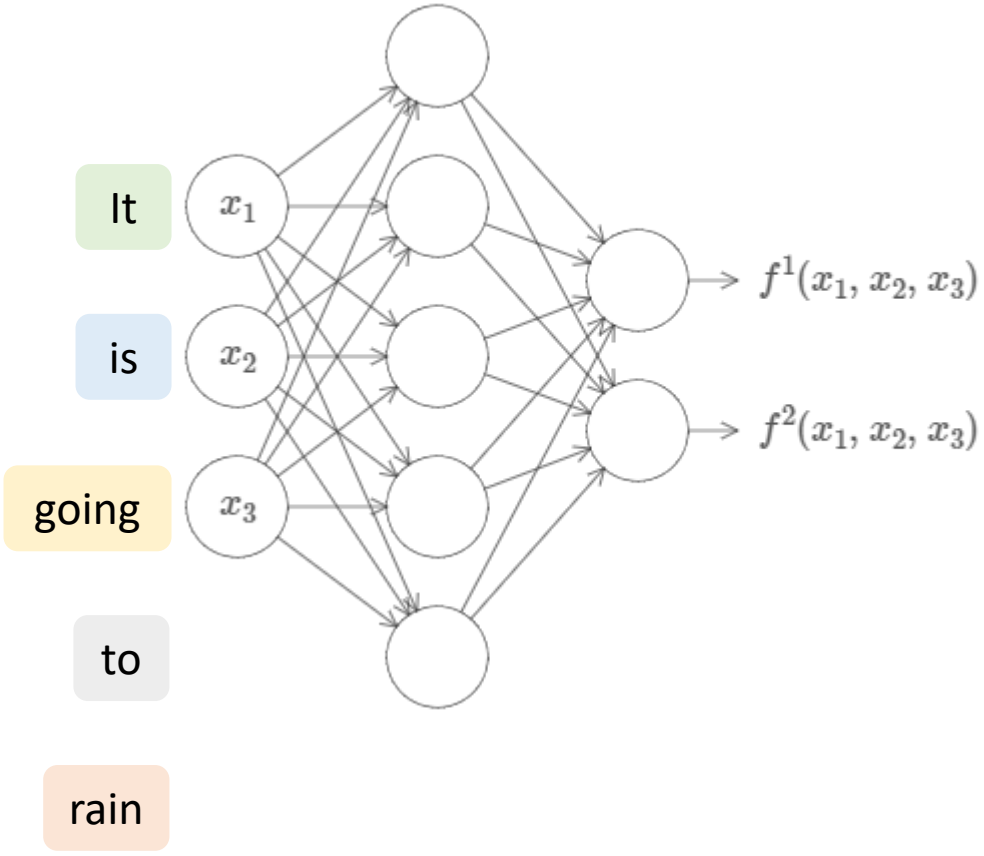
Problems with Convolutional and Feed Forward Networks



http://neuralnetworksanddeeplearning.com/chap4.html

Problems with Convolutional and Feed Forward Networks

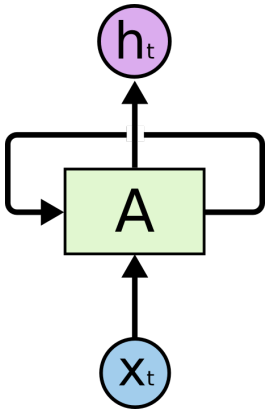
What if the sentence is longer?



http://neuralnetworksanddeeplearning.com/chap4.html

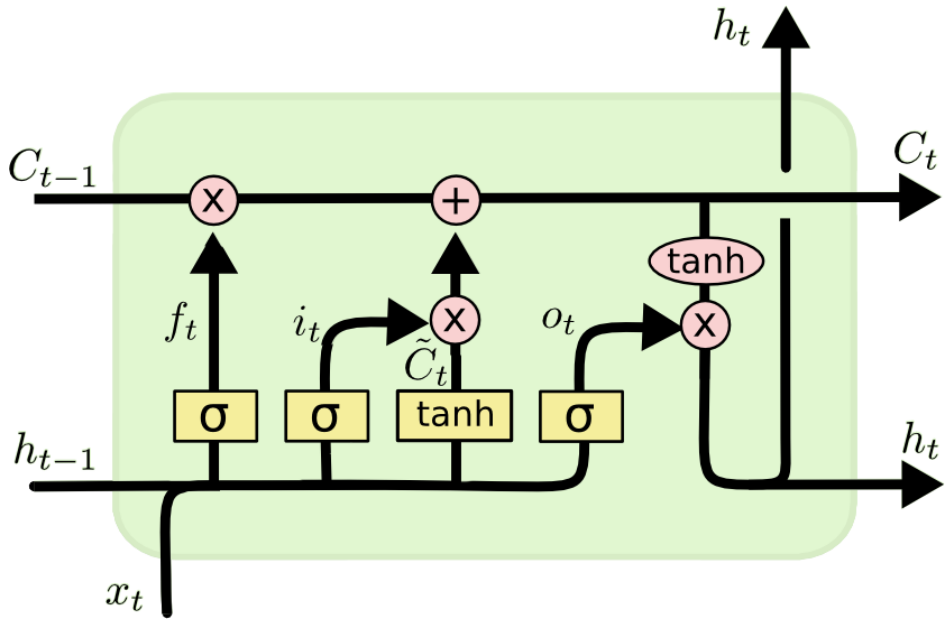
Why Are Recurrent Networks Better

Recurrent networks have loops



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

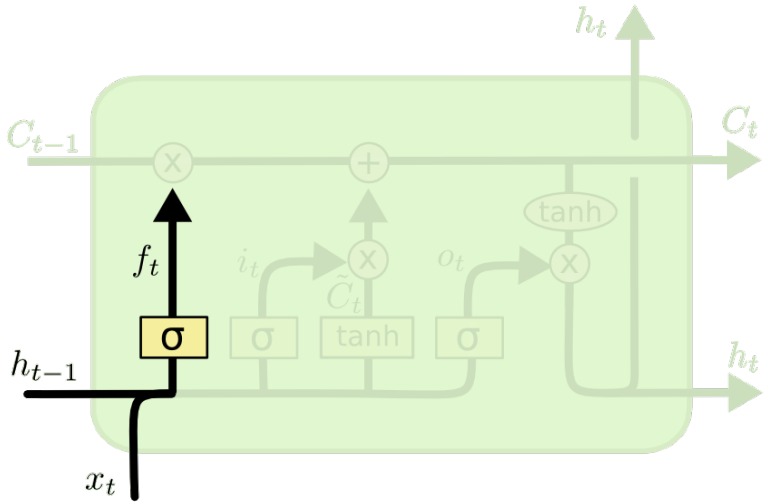
Long Short-Term Memory (LSTM) Networks



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

Long Short-Term Memory (LSTM) Networks

Forget gate layer

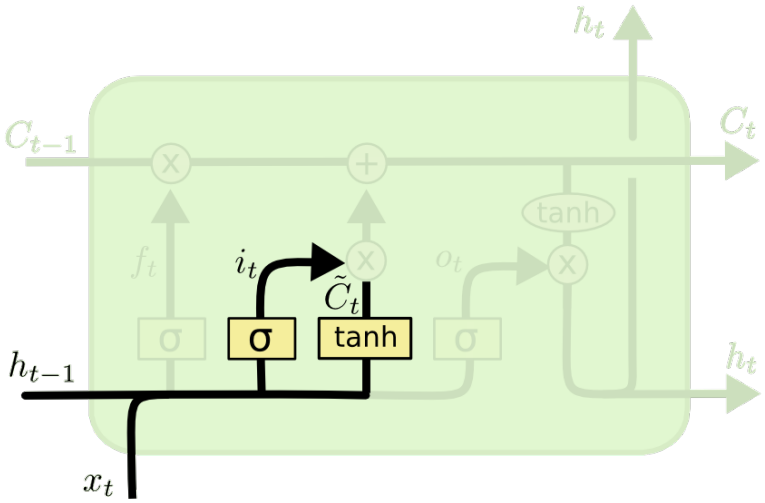


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

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

Long Short-Term Memory (LSTM) Networks

Input gate layer



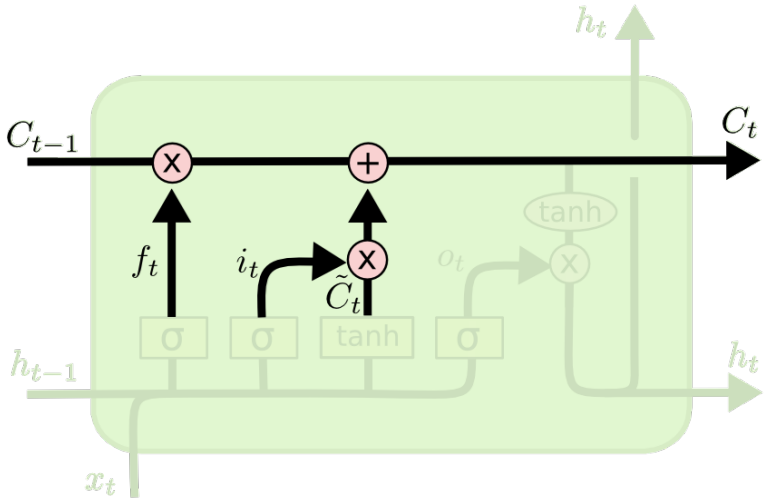
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

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

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

Long Short-Term Memory (LSTM) Networks

State update

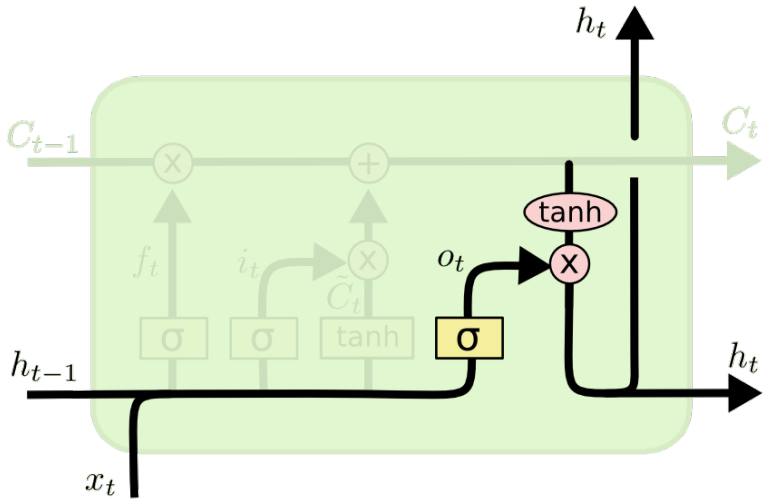


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

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

Long Short-Term Memory (LSTM) Networks

Decide what the output is



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

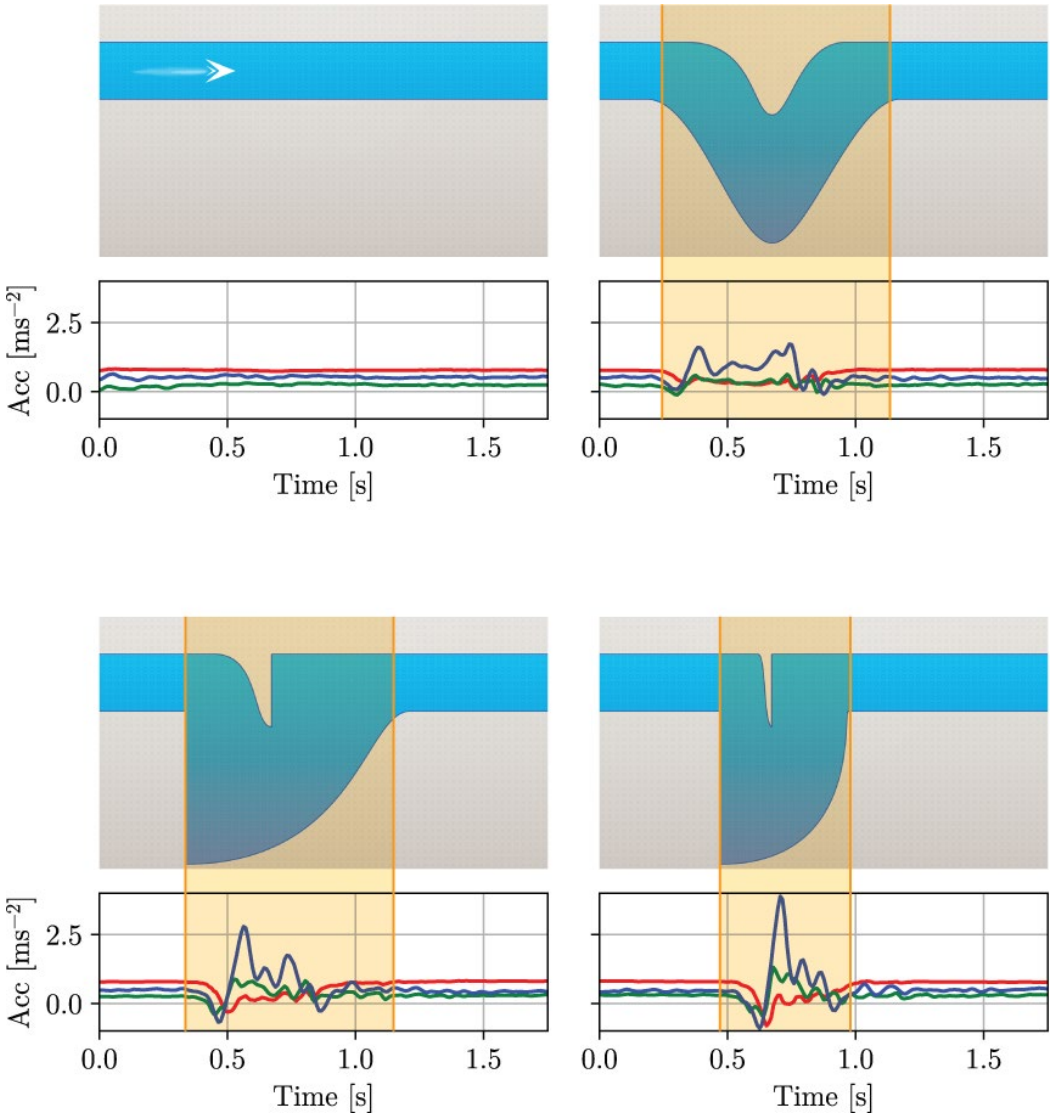
$$h_t = o_t * \tanh\left(C_t\right)$$

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

Training LSTMs

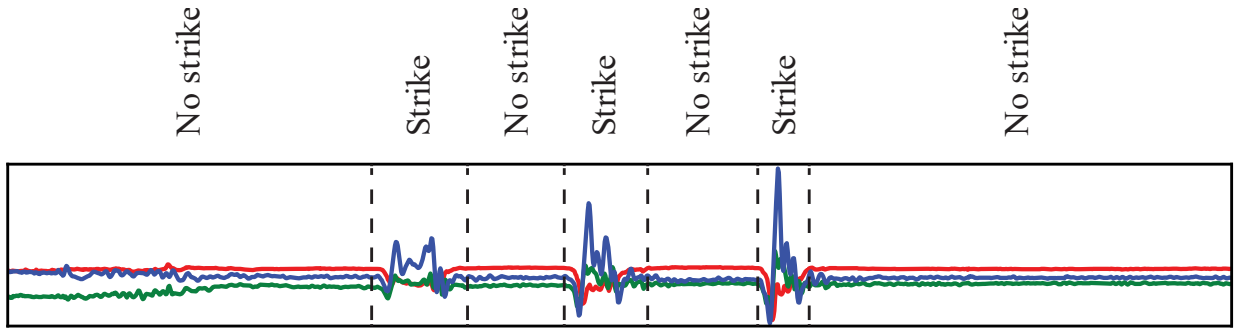
Demo

Training LSTMs – labelling the data

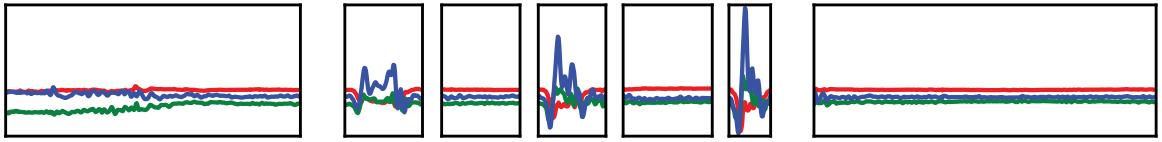


https://yannidd.github.io/posts/2019-10-10-drumless/index.html

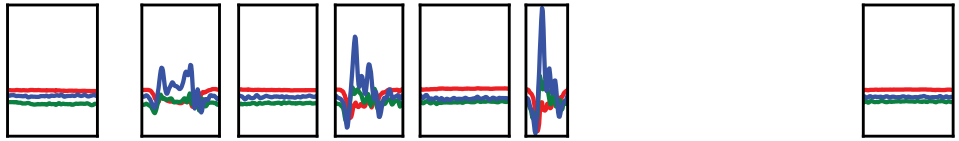
Training LSTMs – creating the dataset



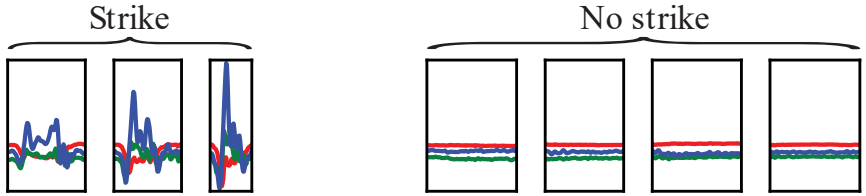
Step 1: split



Step 2: trim



Step 3: group



https://yannidd.github.io/posts/2019-10-10-drumless/index.html

Training LSTMs

Further Reading

[1] Understanding LSTM Networks. C. Olah. 2015. (online)

[2] Long Short-Term Memory. S. Hochreiter, et al. 1997 (online)

[3] Recurrent Neural Networks (RNN) with Keras. (online)

[4] Masking and Padding with Keras (online)