AI Society

LSTMs

Agenda

- What is sequence data

- Problems with convolutional and feed forward networks

- Why are recurrent networks better
    - Long short-term memory (LSTM) networks
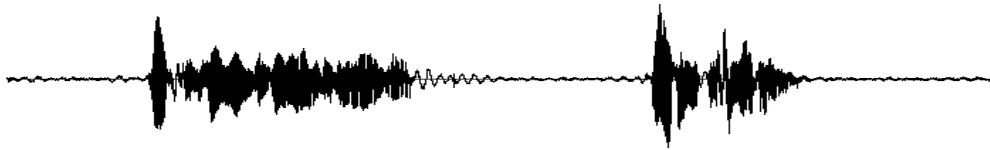
- Training LSTMs

What Is Sequence Data

A sentence

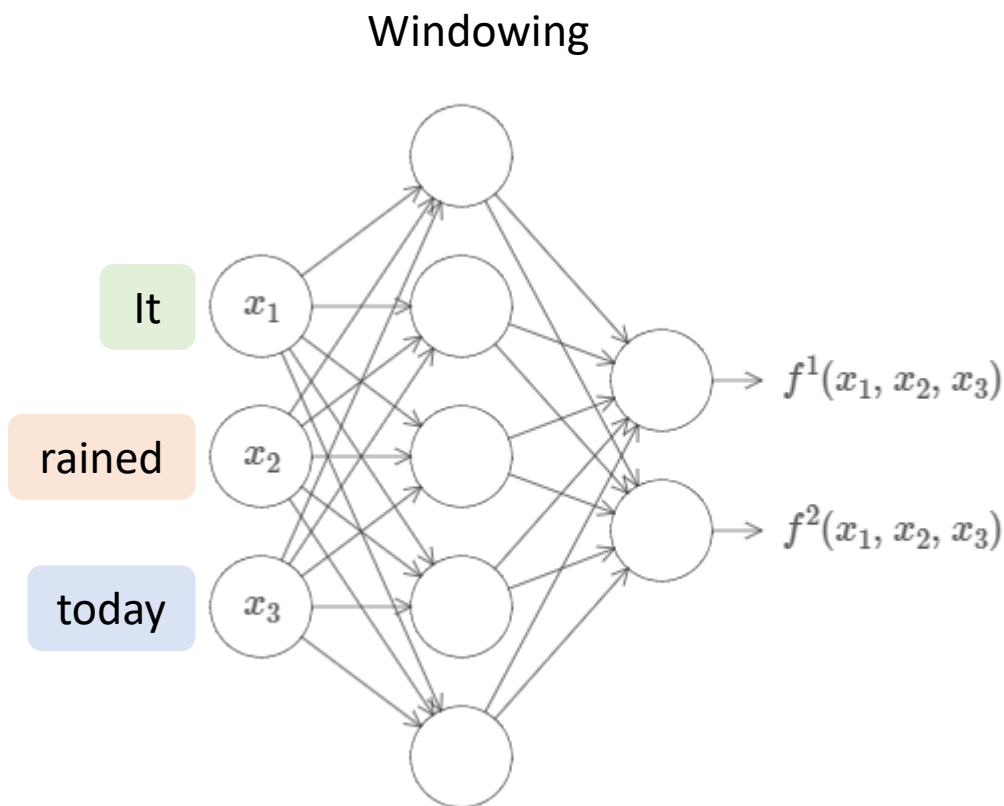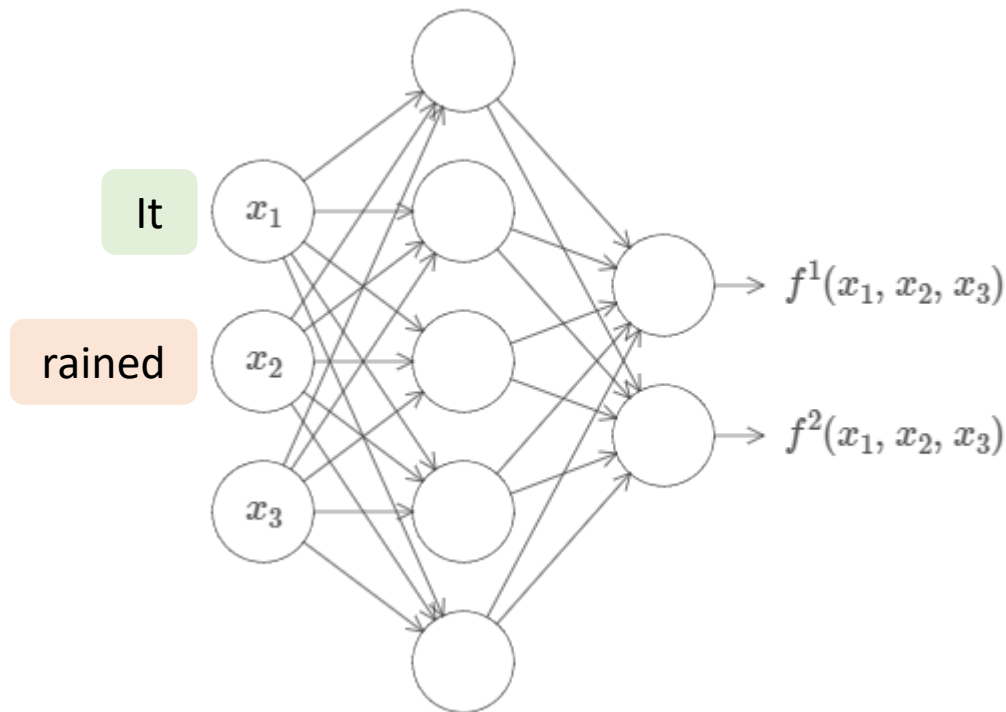| It | is | going | to | rain | today |

A genome

| C | A | C | G | A | C | T | T | ... |

Voice

Problems with Convolutional and Feed Forward Networks
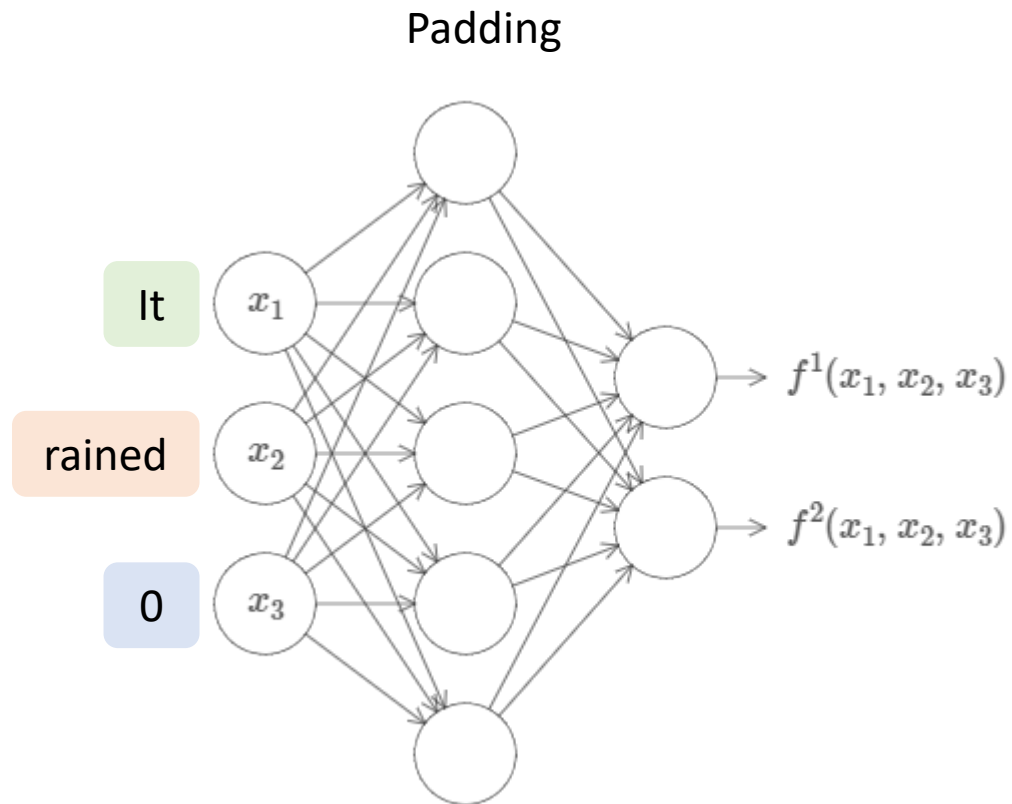
Problems with Convolutional and Feed Forward Networks

Windowing

http://neuralnetworksanddeeplearning.com/chap4.html

Problems with Convolutional and Feed Forward Networks

What if the sentence is shorter?

It x_1

rained x_2

x_3

$f^1(x_1, x_2, x_3)$

$f^2(x_1, x_2, x_3)$

Problems with Convolutional and Feed Forward Networks

Padding

It $\quad x_1$

rained $\quad x_2$

0 $\quad x_3$

$f^1(x_1, x_2, x_3)$

$f^2(x_1, x_2, x_3)$

Problems with Convolutional and Feed Forward Networks

What if the sentence is longer?

It x_1

is x_2

going x_3

to
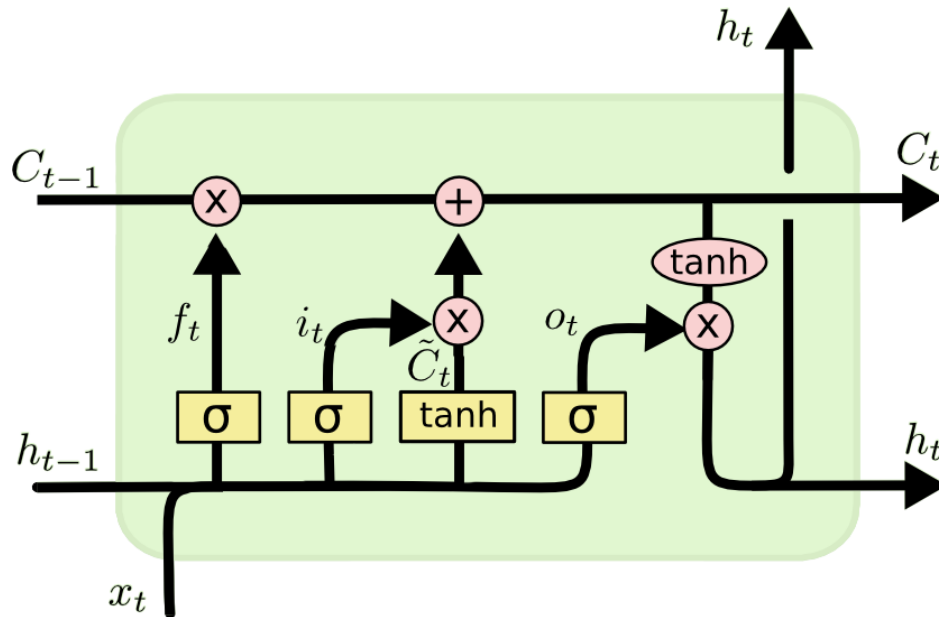
rain

$f^1(x_1, x_2, x_3)$

$f^2(x_1, x_2, x_3)$
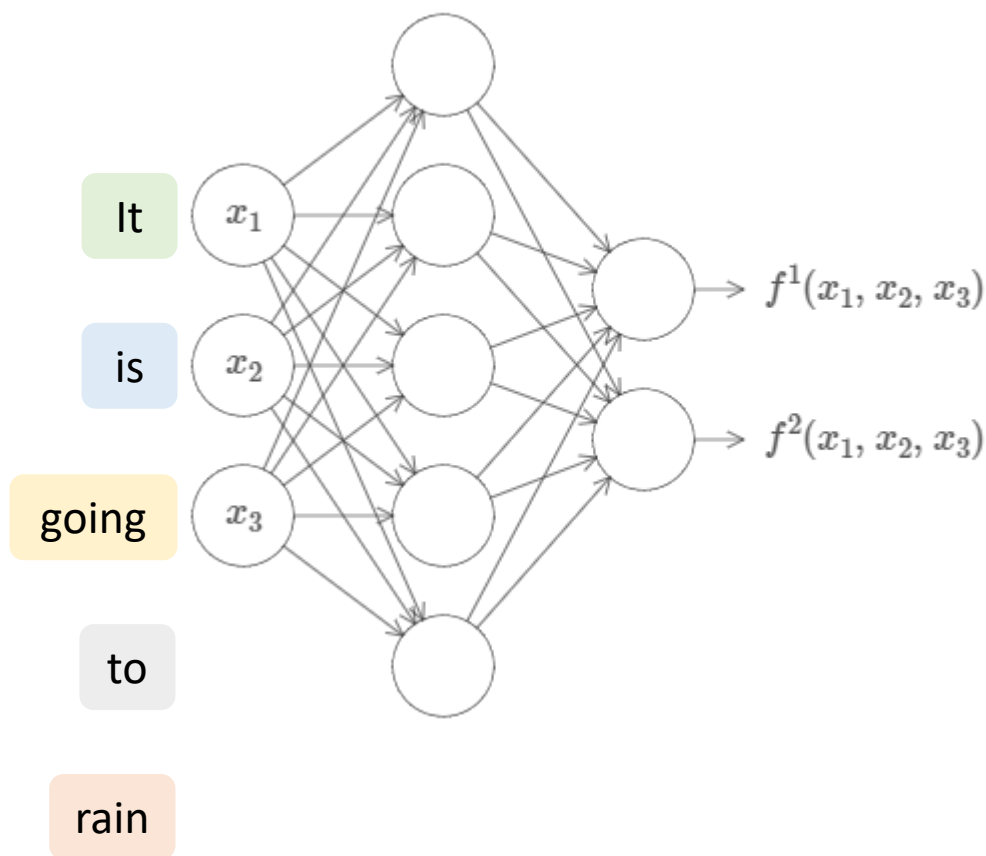
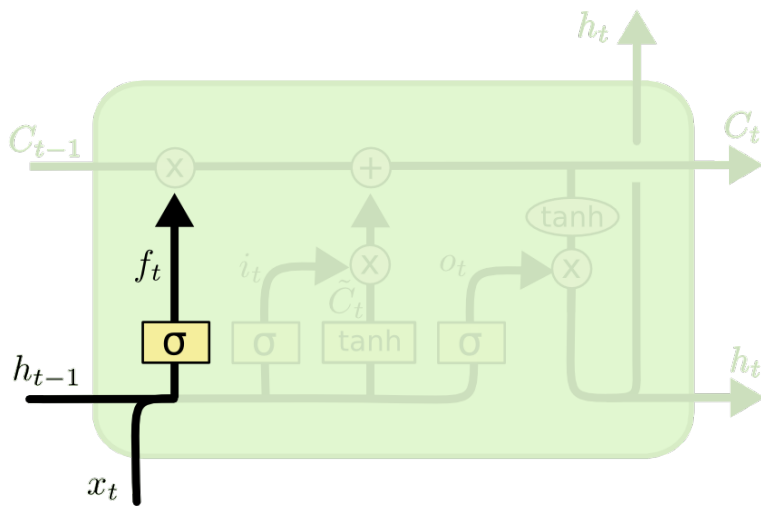# Why Are Recurrent Networks Better

Recurrent networks have loops
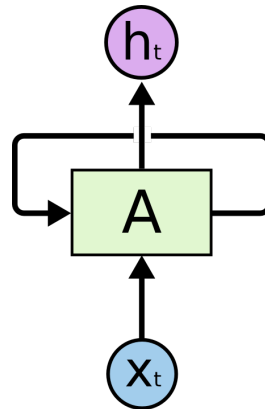
Long Short-Term Memory (LSTM) Networks
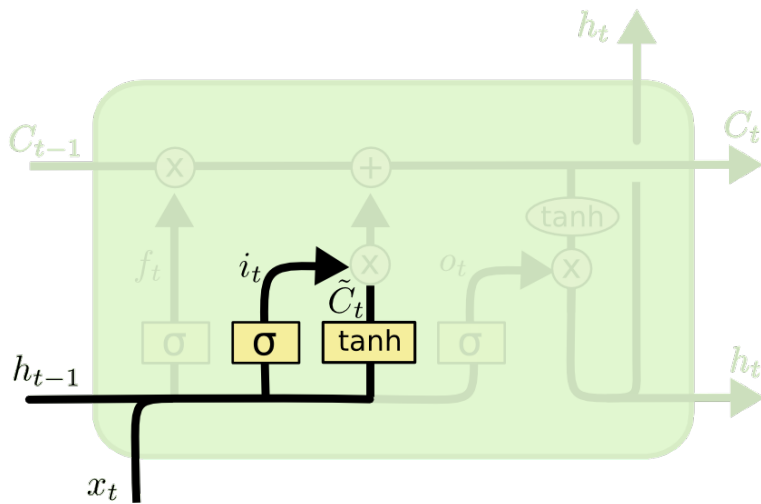
Long Short-Term Memory (LSTM) Networks

Forget gate layer

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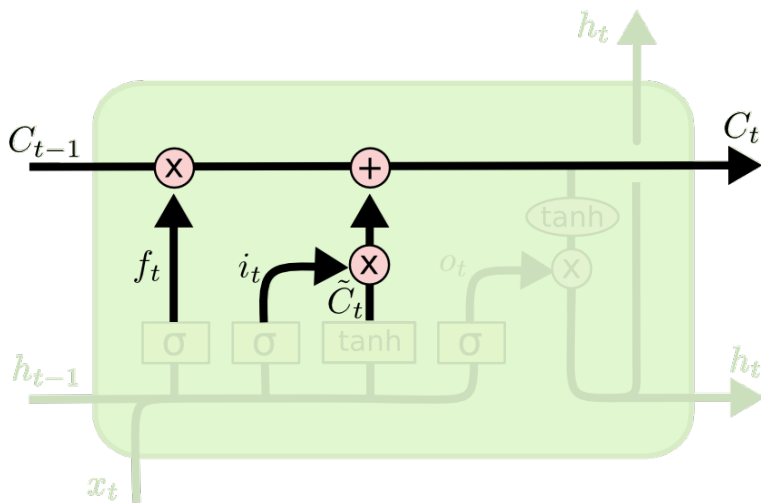
Long Short-Term Memory (LSTM) Networks

Input gate layer

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

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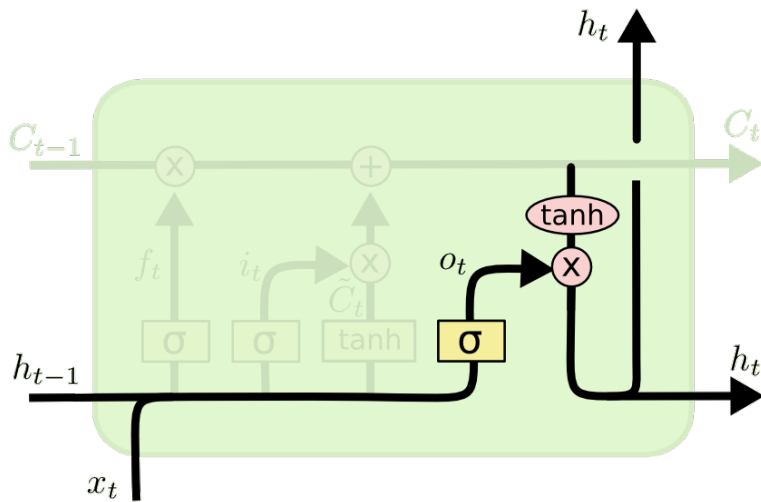
Long Short-Term Memory (LSTM) Networks

State update

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

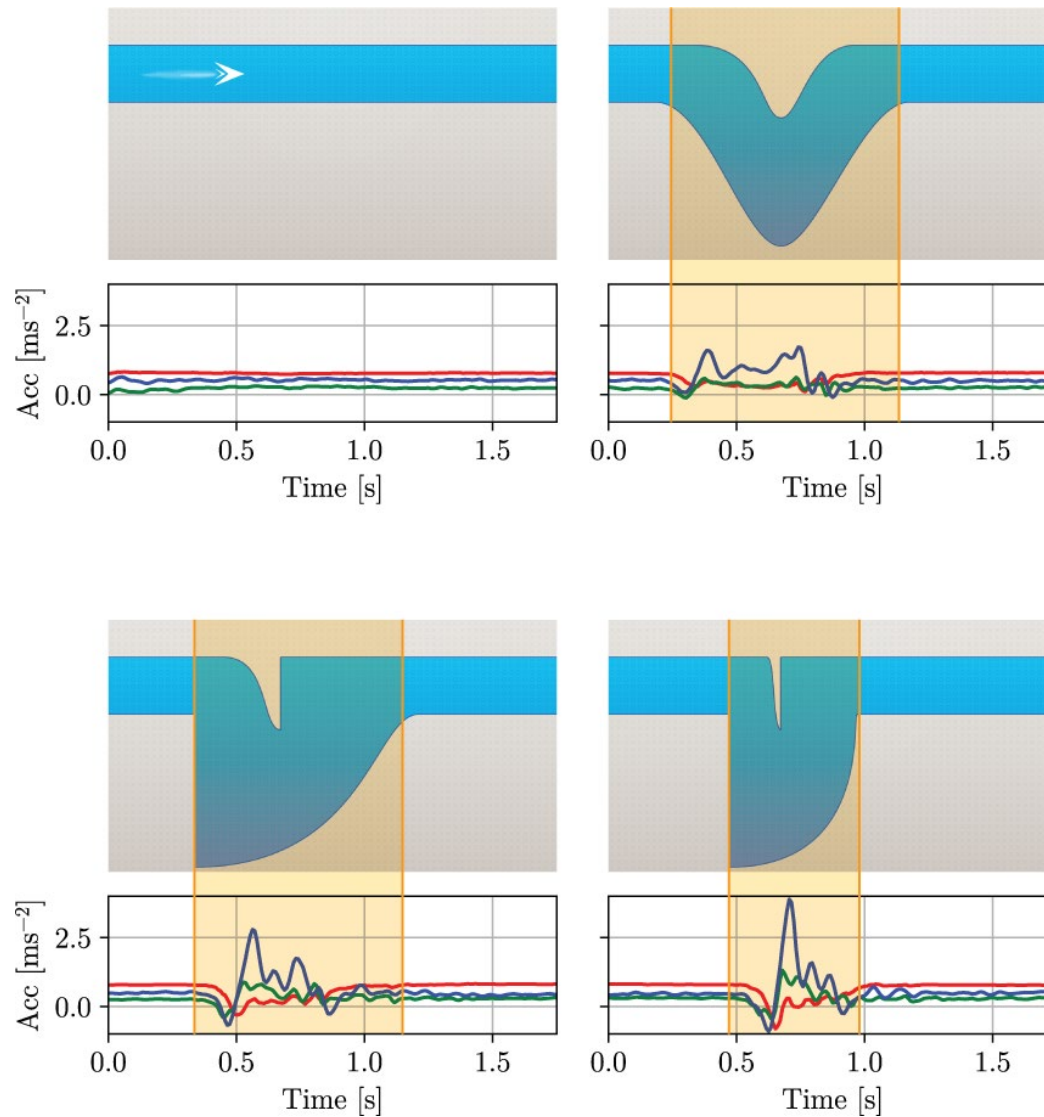# Long Short-Term Memory (LSTM) Networks

Decide what the output is

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

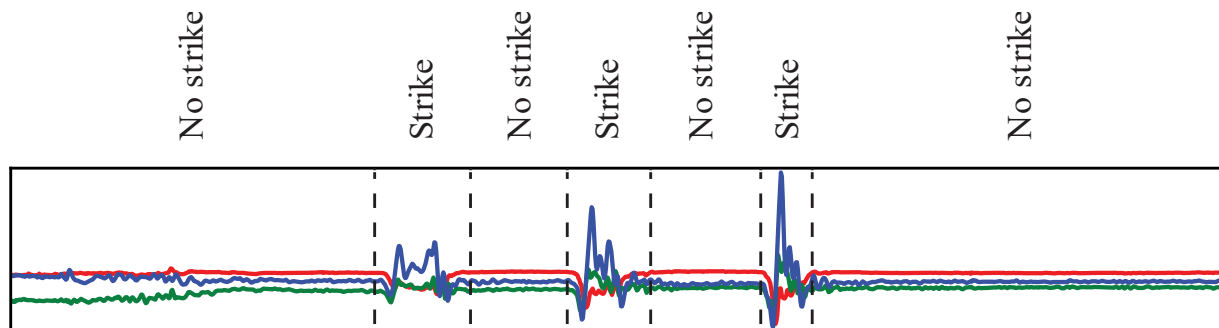$$h_t = o_t * \tanh\left(C_t\right)$$

Training LSTMs

Demo

Training LSTMs – labelling the data

Training LSTMs – creating the dataset

No strike   Strike   No strike   Strike   No strike   Strike   No strike
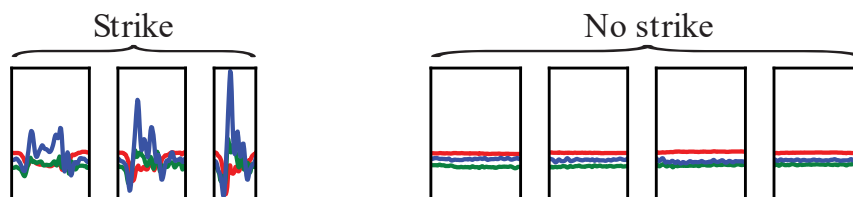
Step 1: split
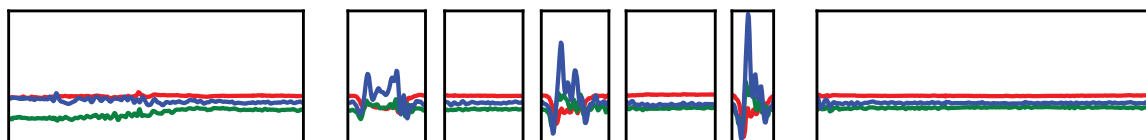
Step 2: trim

Step 3: group

Strike

No strike
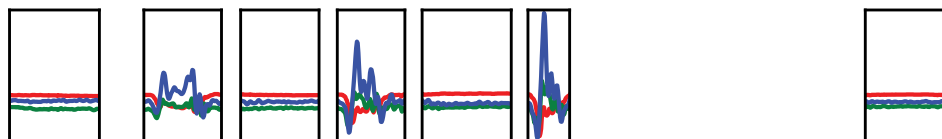
Training LSTMs

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)