



# Zero-to-Hero Deep Learning in Python

## Intro to LSTMs

28/01/2020

Samuel White

[info@osgdigitallabs.com](mailto:info@osgdigitallabs.com) / [samwhiteoxford@gmail.com](mailto:samwhiteoxford@gmail.com)

Your top friends

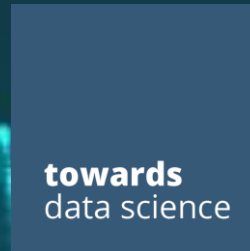


Also good places:

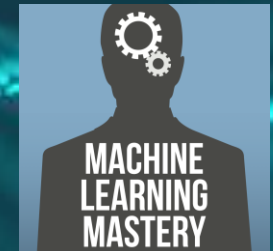


**Medium**

[medium.com](https://medium.com)



[towardsdatascience.com](https://towardsdatascience.com)



[machinelearningmastery.com/](https://machinelearningmastery.com/)



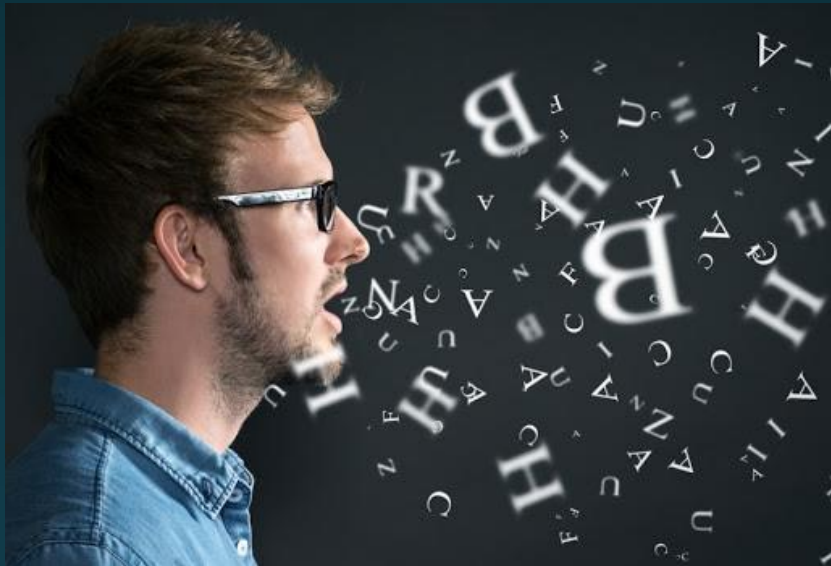
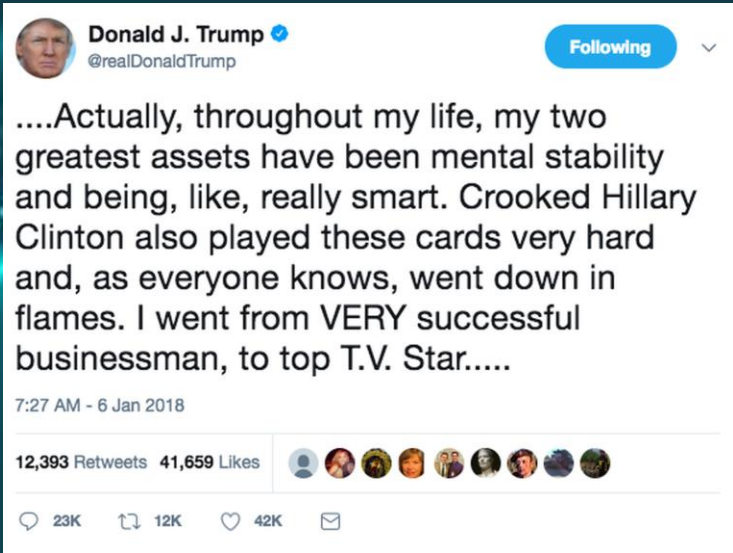
**Keras**

[keras.io/](https://keras.io/)



[stackoverflow.com/](https://stackoverflow.com/)

# Examples of time series





# Kind of stuff you can do

## TRUMP RALLY

INT. BIG ARBY'S IN SOUTH WYOMKLAHOMA

PRESIDENT TRUMP forces himself on a podium.

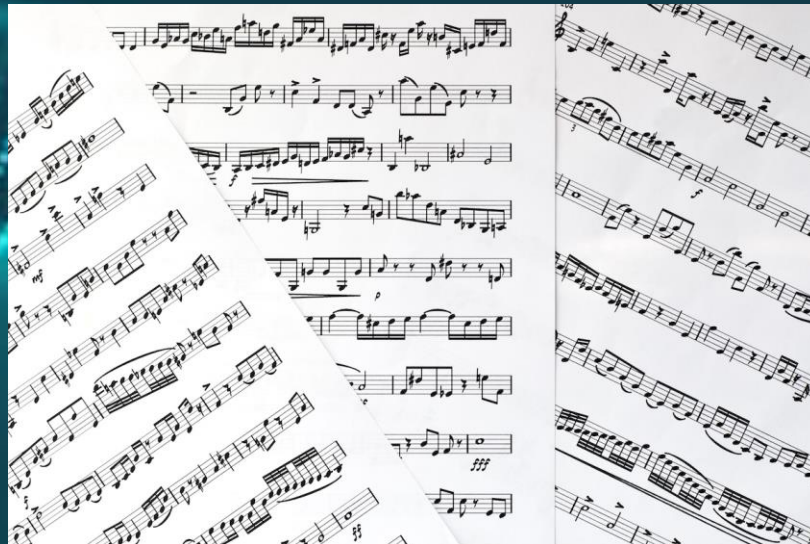
PRESIDENT TRUMP  
I just had a phone call with the  
economy. Jobs poured out of the  
phone. Great jobs. Tall jobs. Steve  
Jobs. All at Kinko's.

The crowd cheers. It is full of real Americans (man with hard  
hat, man with harder hat, gun that is alive).

PRESIDENT TRUMP (CONT'D)  
The United Snakes is doing so good.  
Other countries are on fire. All  
the people on fire. Hot fire too.  
Not us. Our flag is so beautiful.

President Trump salutes a flag that says: **ARBY'S FOOD IS FINE  
TO EAT.** The crowd howls. They love this flag of America.

PRESIDENT TRUMP (CONT'D)  
I signed a bill. No more swamp.  
Swamp gone. Swamp is in Mexico now.  
It's on fire. Great deal for us.



# Quick catch up on feed forward neural nets

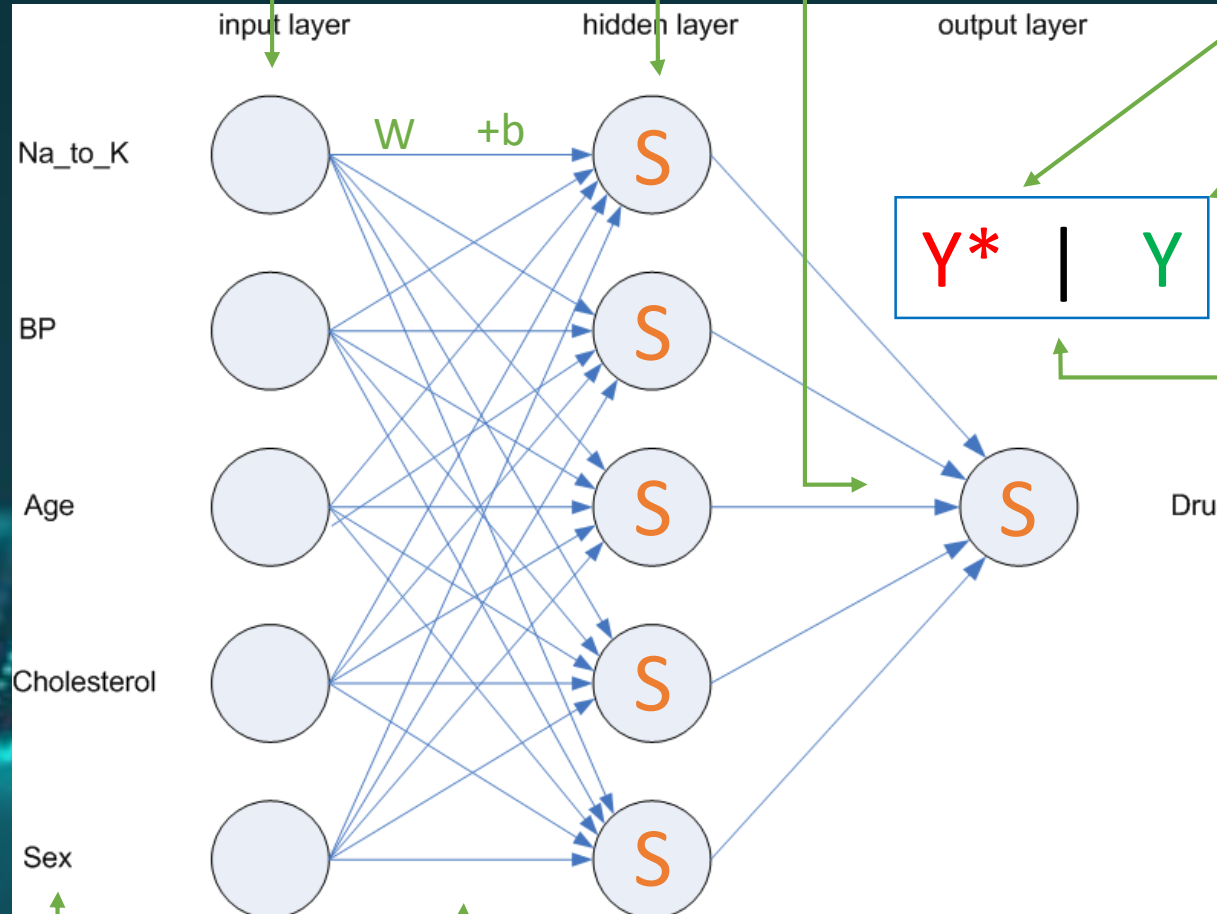
## Layers:

The entire structure of a neural net is based around multiplying an input (say "Age") by a weight ( $W$ ), adding a little bit to shift the data ( $+b$ ), and then doing some transform of the data (the orange "S"s) eg. say that if the result was less than zero, then just make it zero instead.

*We'll explain this more on the next slide.*

## Inputs:

The vector of data that you want your model to predict on



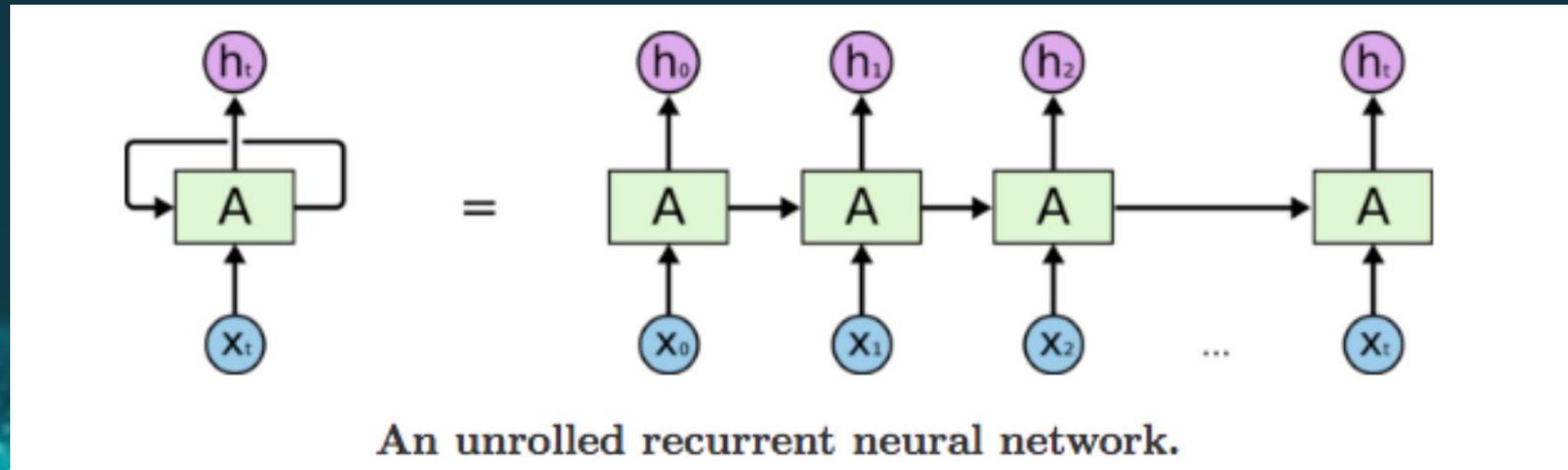
**Outputs:** what the model thinks the answer is

**Labels:** what the true answer is (in supervised learning)

**Loss function:** some measure of how different  $Y^*$  is from  $Y$ . It factors into how much the learning function adjusts the weights in the model.

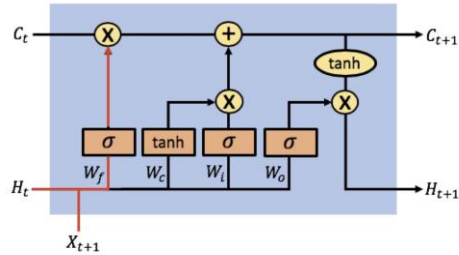
**Learning function:** uses the results from the loss to adjust the weights and biases that make up a model.

# How are they structured? (RNNs)



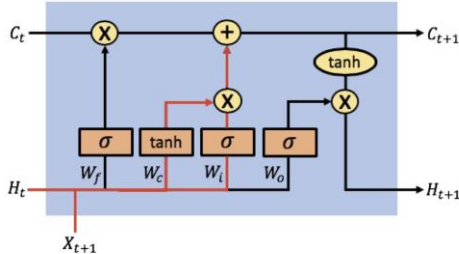


# How are they structured? (LSTMs)



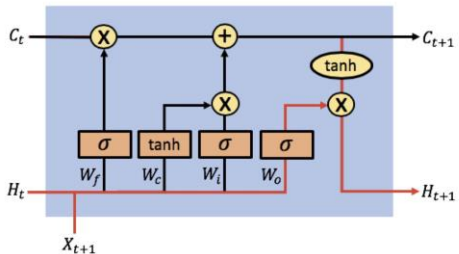
## Forget Gate:

The forget gate controls what information in the cell state to forget, given new information than entered the network.



## Input Gate:

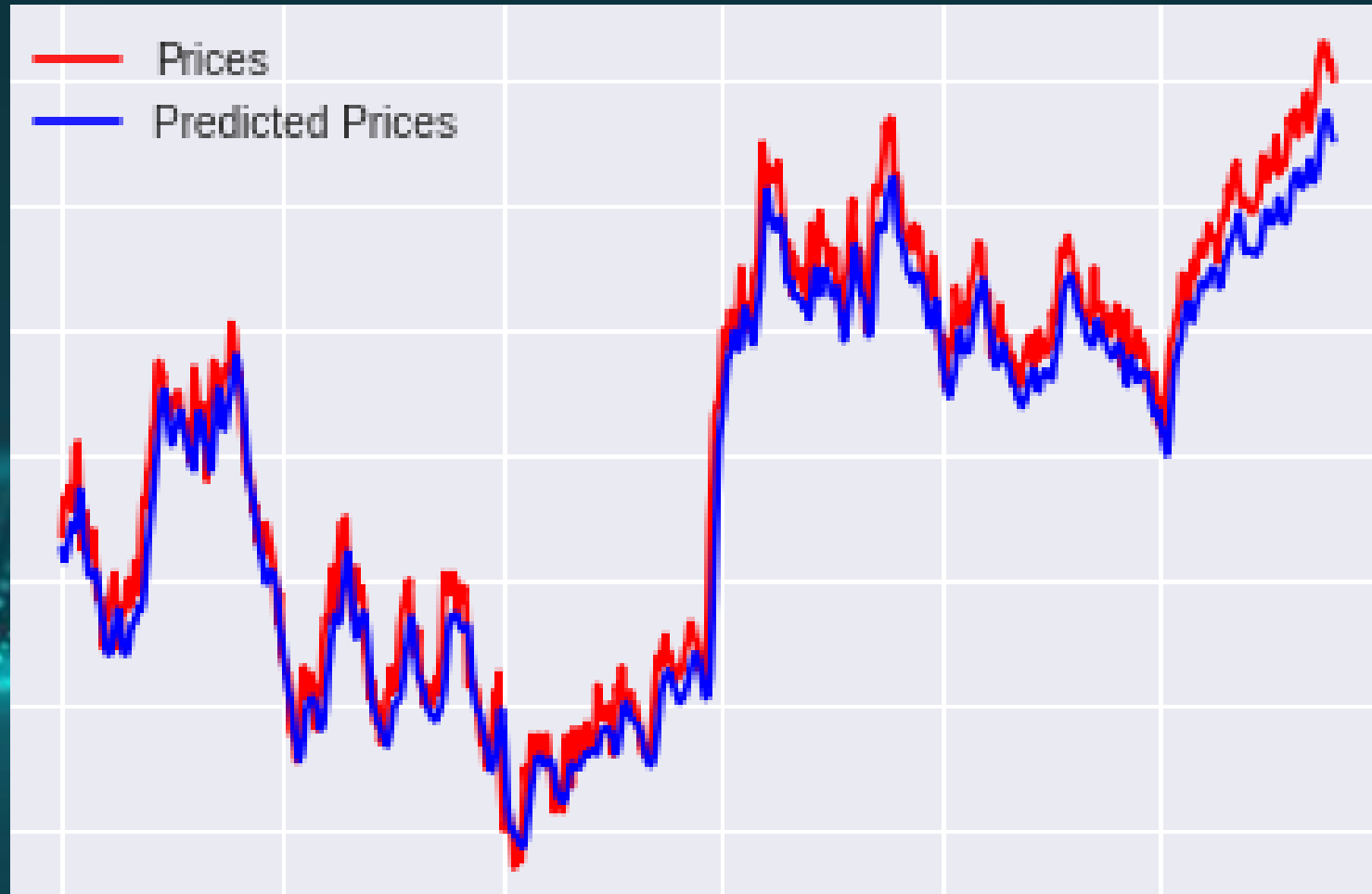
The input gate controls what new information will be encoded into the cell state, given the new input information.



## Output Gate:

The output gate controls what information encoded in the cell state is sent to the network as input in the following time step, this is done via the output vector  $H_t$ .

# Problems...





# Adaptions...



**Stacked:** LSTMs on LSTMs, the output sequence for one is the input to the next.

**Bidirectional:** The input sequence is fed in both forwards and backwards so that the model can attempt to interpret the result using both directions.

**CNN-LSTM:** The 1D input sequence is first read by a CNN to translate it to a different sequence, this is then the input for an LSTM.

**ConvLSTM:** Related to the CNN-LSTM, but the CNN module is built into the LSTM cell.

**Multi-Variate:** Instead of taking one input (stock price) take multiple (stock price, momentum, volume etc...) and try to predict off that.

**Encoder-Decoder:** One LSTM maps the input sequence to some kind of encoded sequence, this is then fed into a second LSTM that attempts to output the correct sequence from the encoded representation.

**Finance Prediction Workbook:**

*<https://tinyurl.com/osgd-lstm-MDT-MRK>*