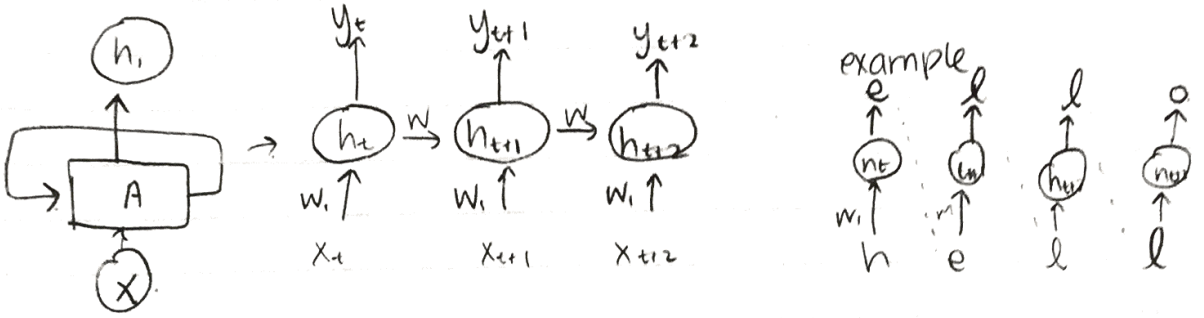
Recurrent Neural Network

- neurons are fed information not just from previous layer but also from themselves
- order that you feed input matters
- vanishing gradient problem because information gets lost over time

- you make multiple copies of the same network



→ Recurrent Neural Networks can work for short term dependencies

→ they have a "memory" which captures information about what has been calculated so far but in reality this memory is short

→ you take output of layer, put that output as input to the same neural network. This is why the order of inputs would matter.

youtube tutorial notes

- you can work with a sequence of inputs
  - → helpful for image captioning

- RNN with 100 steps = 100 layer MLP so you have a vanishing gradient descent. Each step runs basically a MLP because you repeat it over and over again

Long Short Term Memory

- you have gates and memory cell to address the vanishing gradient descent problem in RNN

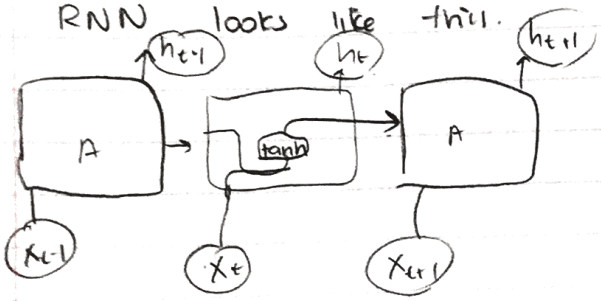
memory cell

input: how much info from previous layers get stored in the cell

output: how much of next layer gets to know about state of this cell

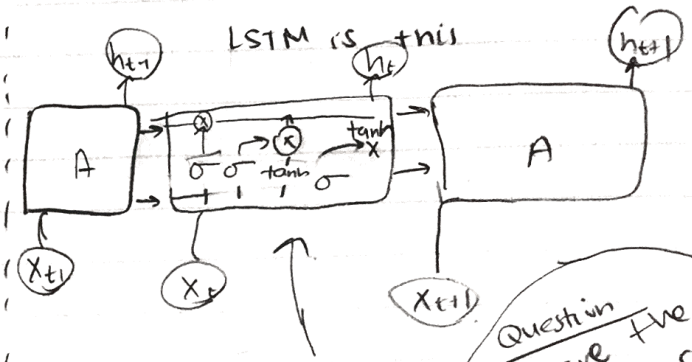
forget good to forget (forget gender of old subject of sentence if new subject is introduced.)

RNN looks like this.



each repeated neuron is a normal, standard single layer neural network

LSTM is this



blown up in pic below

Question: what are the $+$ or cell state? $\times$