Section 9

- Homework 8 due last night
  - Can submit by 10/28 with penalties
  - What L1 scores did you get?
  - Things you did differently from HW7?
- Homework 9
  - Sequence Prediction
- Quiz 8 and HW7 grades posted
  - Please recheck your submissions before deadline
  - Confirm model checkpoint and models.py have the same model

Questions?

Agenda

- Recurrent Neural Networks
- Long Short Term Memory Networks
- Enhancements in LSTM
- PyTorch Code Walkthrough

Recurrent Neural Networks

- For Modeling sequential data
  - Text, Videos, Images
- Building “context” around the input

Apple is good for health

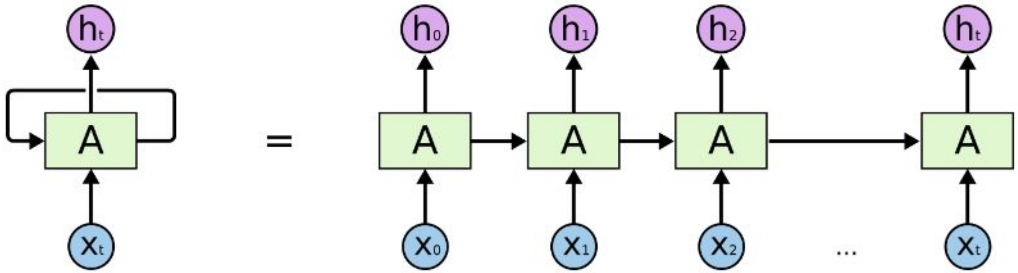
Fruit

Apple is a billion dollar company

Company

Recurrent Neural Networks

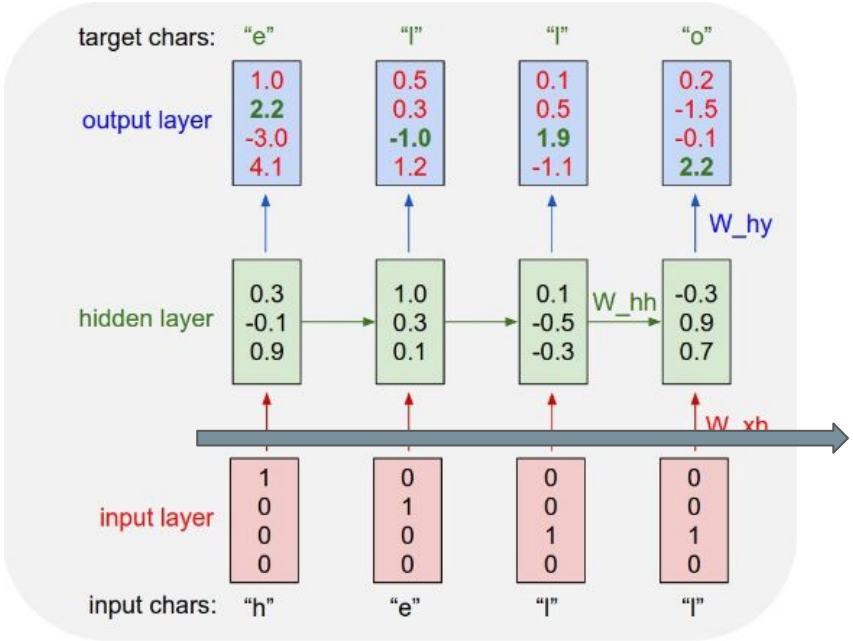
- How can we store this context?
- Handling variable sequences?



Recurrent passing of hidden state allows you to unroll the network dynamically

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

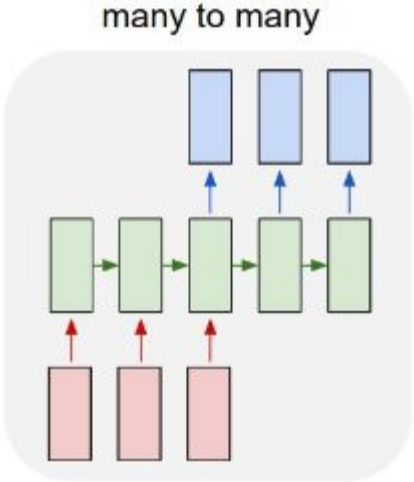
How does this happen?

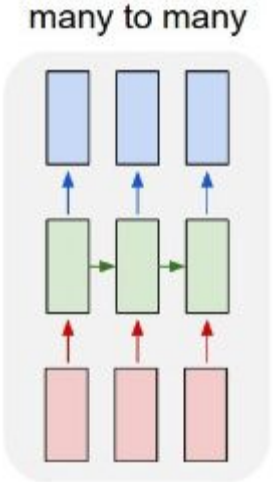


Can use an intermediate Embedding Layer

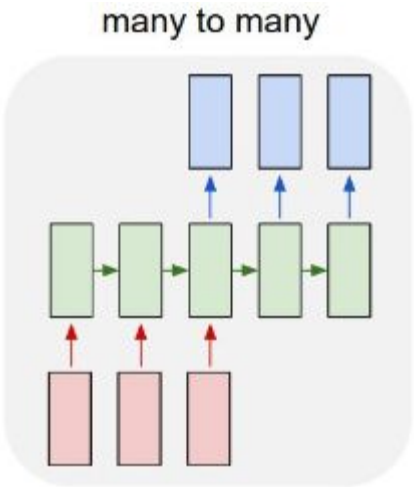
Image Credit: https://cs.stanford.edu/people/karpathy/

When to use these RNNs?

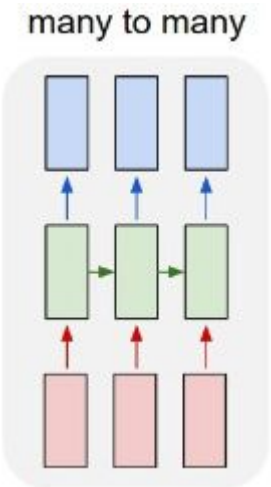




When to use these RNNs?

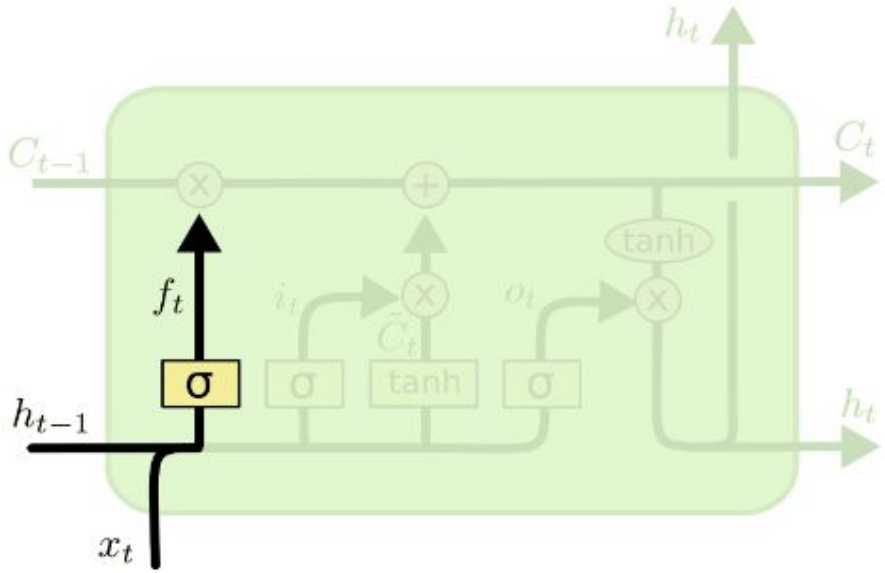


1. Input and output are not aligned
2. Input and output sequence length is different. E.g. Translation, Summarization
3. Problem: Remembering initial inputs



1. Input and output are aligned
2. Common Problems: POS Tagging, NER, Slot Filling
3. Lesser flexibility

LSTM

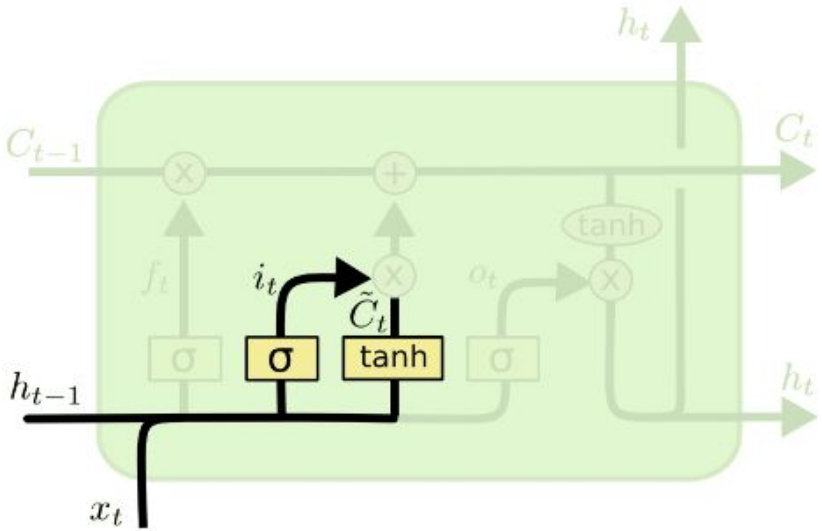


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

1. Forget Irrelevant Information

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

LSTM



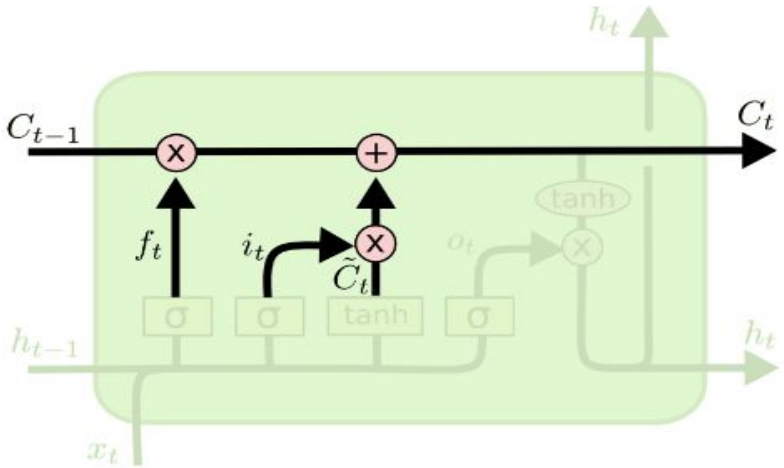
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

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

2. Select Information to Update

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

LSTM

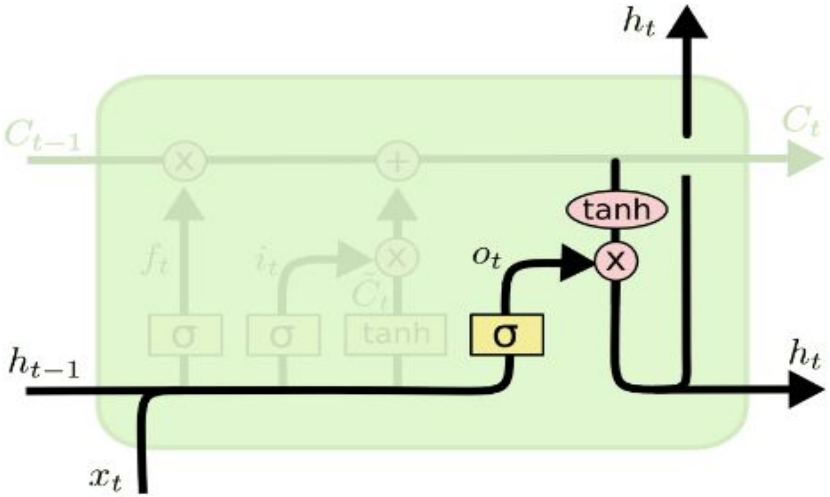


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

3. Update Cell Information

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

LSTM



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

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

4. Compute Output

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

Optimizing Backpropagation

- Set Hidden state to zero (easier)
- Detach the hidden state to prevent backpropagation (recommended)
  - https://r2rt.com/styles-of-truncated-backpropagation.html

Takeaways

- LSTMs/GRUs are better than generic RNNs
- Optimizing hacks like gradient clipping, detaching hidden state are helpful

PyTorch Code