Intro to Neural Nets

Session 6: RNNs for Forecasting

Session Agenda

Sequence Data

- Recurrent Neural Networks (RNNs):
  - SimpleRNN
  - LSTM (Long Short-Term Memory)
  - GRU (Gated Recurrent Unit)
- Bidirectional RNN

A sequence modeling problem: predict the next word

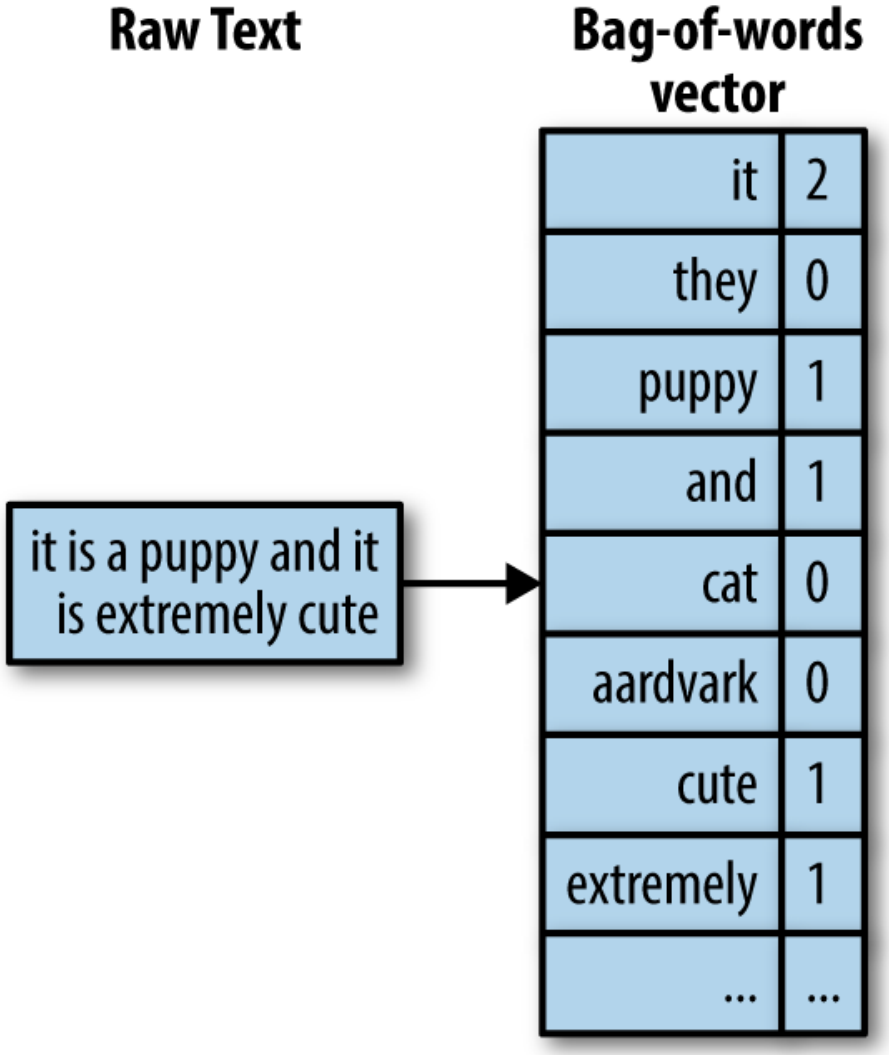
"This morning I took my cat for a walk."

given these words

predict the next word

Adapted from H. Suresh, 2018

Bag of Words Approach (Dense Layer)



Problem: Word Order Matters

The food was good, not bad at all.

vs.

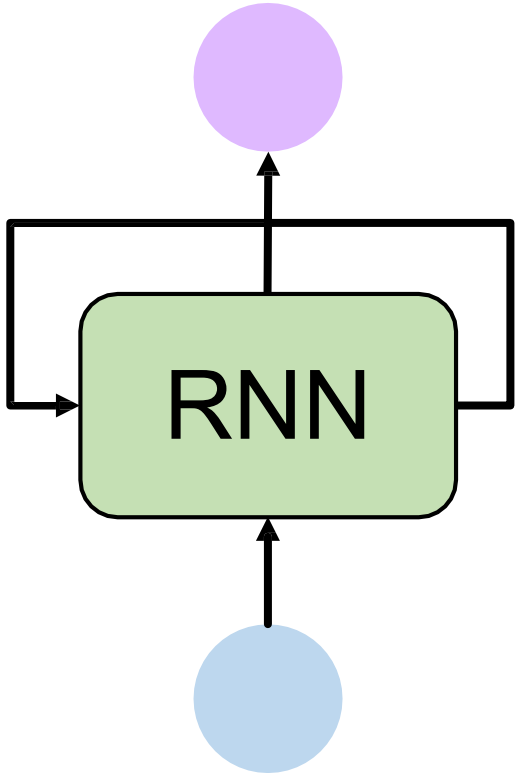
The food was bad, not good at all.

Adapted from H. Suresh, 2018

Sequence modeling: design criteria

To model sequences, we need to:

1. Track long-term dependencies
2. Maintain information about order
3. Share parameters across the sequence

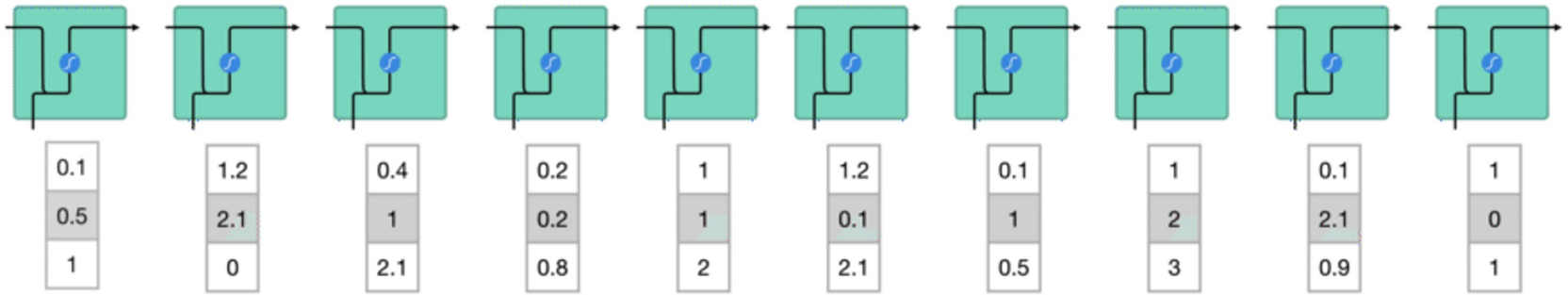


Recurrent Neural Networks (RNNs) as a core approach to sequence modeling problems

RNN Layers

SimpleRNN

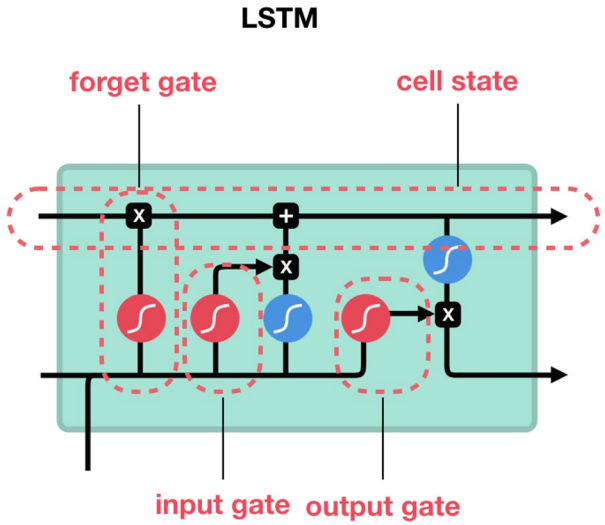
- We can, in practice, unroll an RNN layer into a series as follows.
- This is just a large Dense network with many inputs and many outputs. The inputs are arranged to interact with (feed back upon) each other based on their ordering in the data.



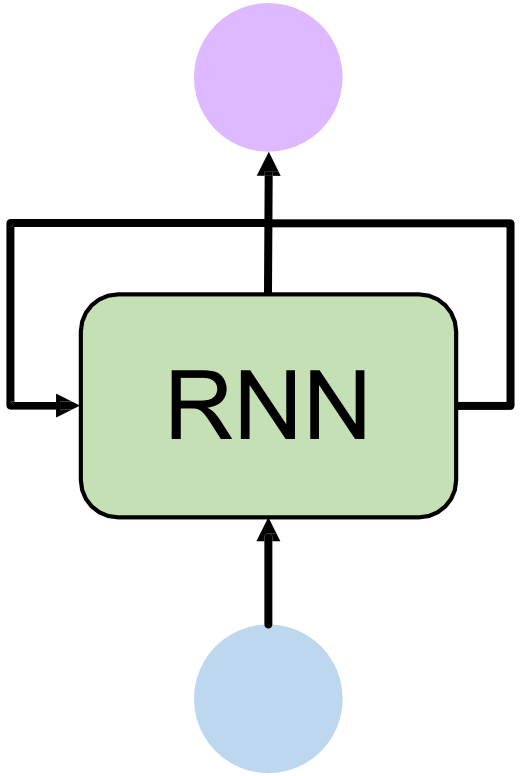
Keras RNN Layers

Long Short-Term Memory (LSTM)

- In practice, we cannot use SimpleRNNs to achieve meaningful memory (vanishing gradients arise quickly). We add what's called a "carry track" – this is an additional connection that combines output at step t, inputs at step t, and the last carry track's output. The topology means your network can 'learn' to use these connections as passthroughs for old info, or it can learn to 'block' that information in favor of more recent information (whatever is useful for accurate prediction).
- Carry tracks are basically just another degree of freedom for learning how to use lagged information.
- These ideas inspired the design of LSTM units, but nothing guarantees that these gates serve these functions..



Extensions of RNN



RNN (Recurrent Neural Network)

LSTM (Long short-term memory)

Transformer

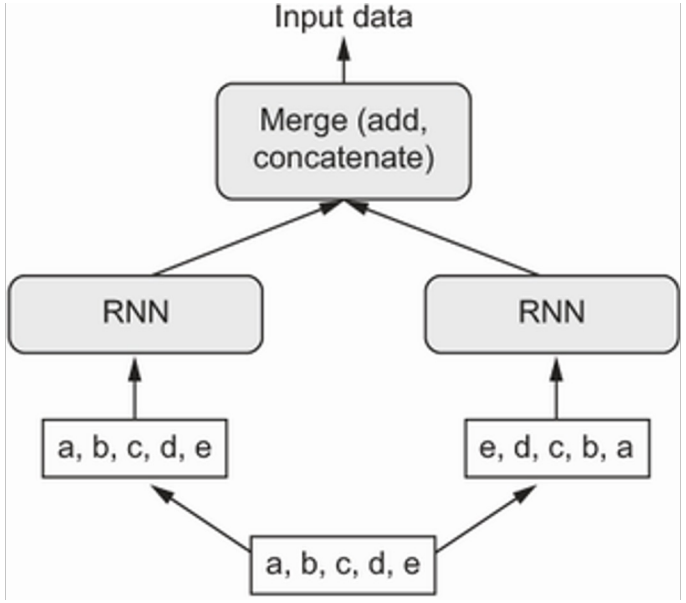
BERT

GPT

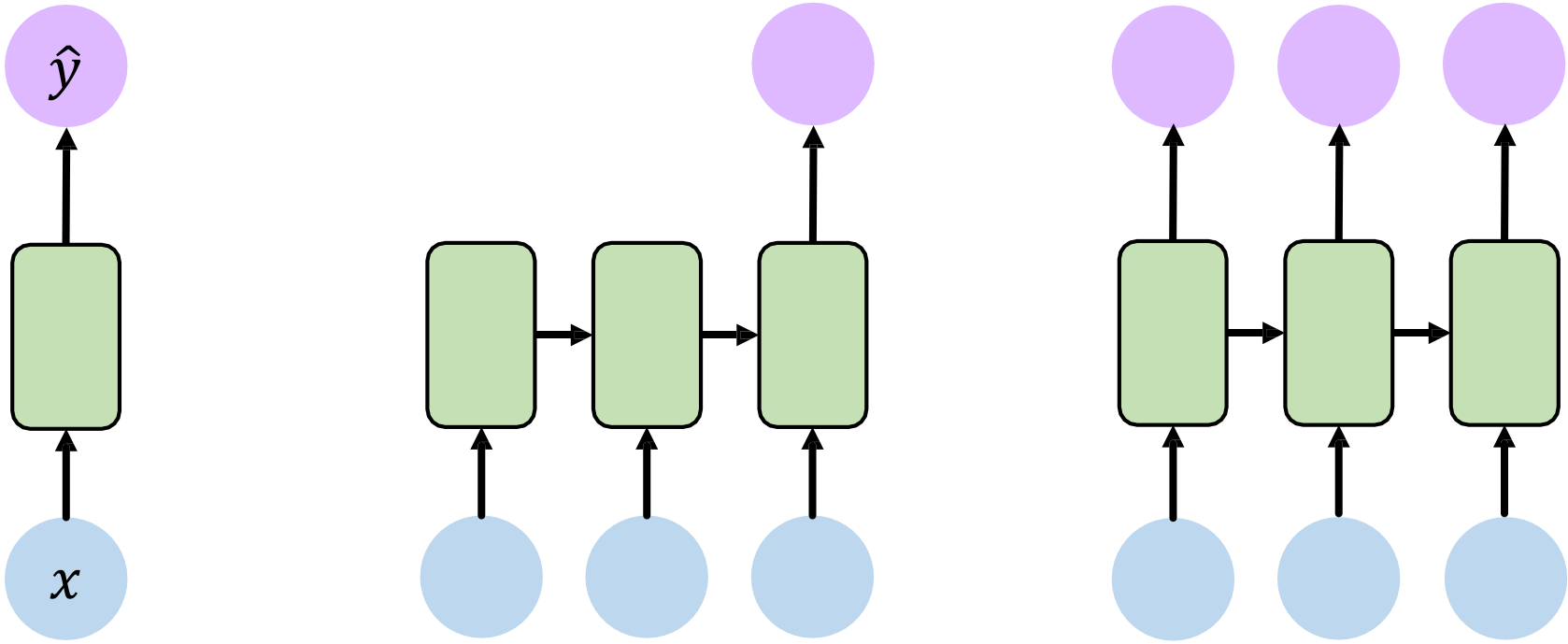
Bidirectional RNNs

Some Sequences Yield Information in Both Directions

- Consider that, in language, words that come later in a sentence can be predictive of what came before.
- Bidirectional RNNs implement a standard RNN, but they also incorporate a parallel layer implementation that takes the sequence ordered in reverse.



Applications of RNNs



One to One

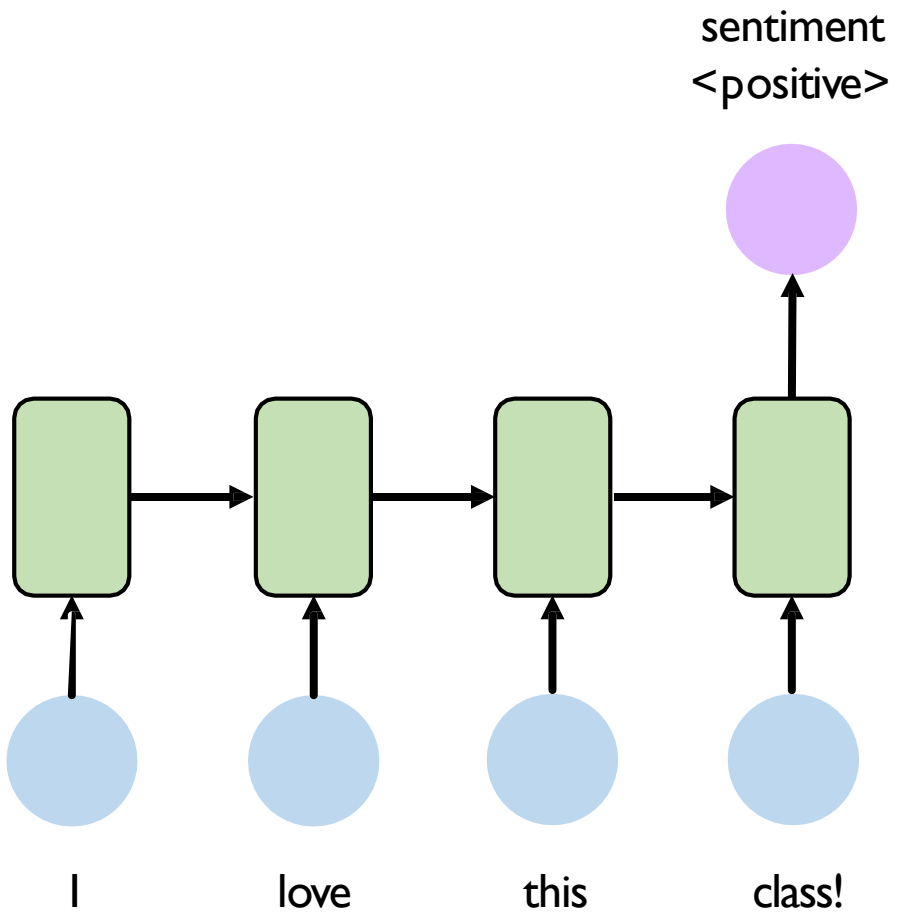
Multi-to-one

Sentiment Classification

Multi-to-Multi

Music Generation

Example task: sentiment classification

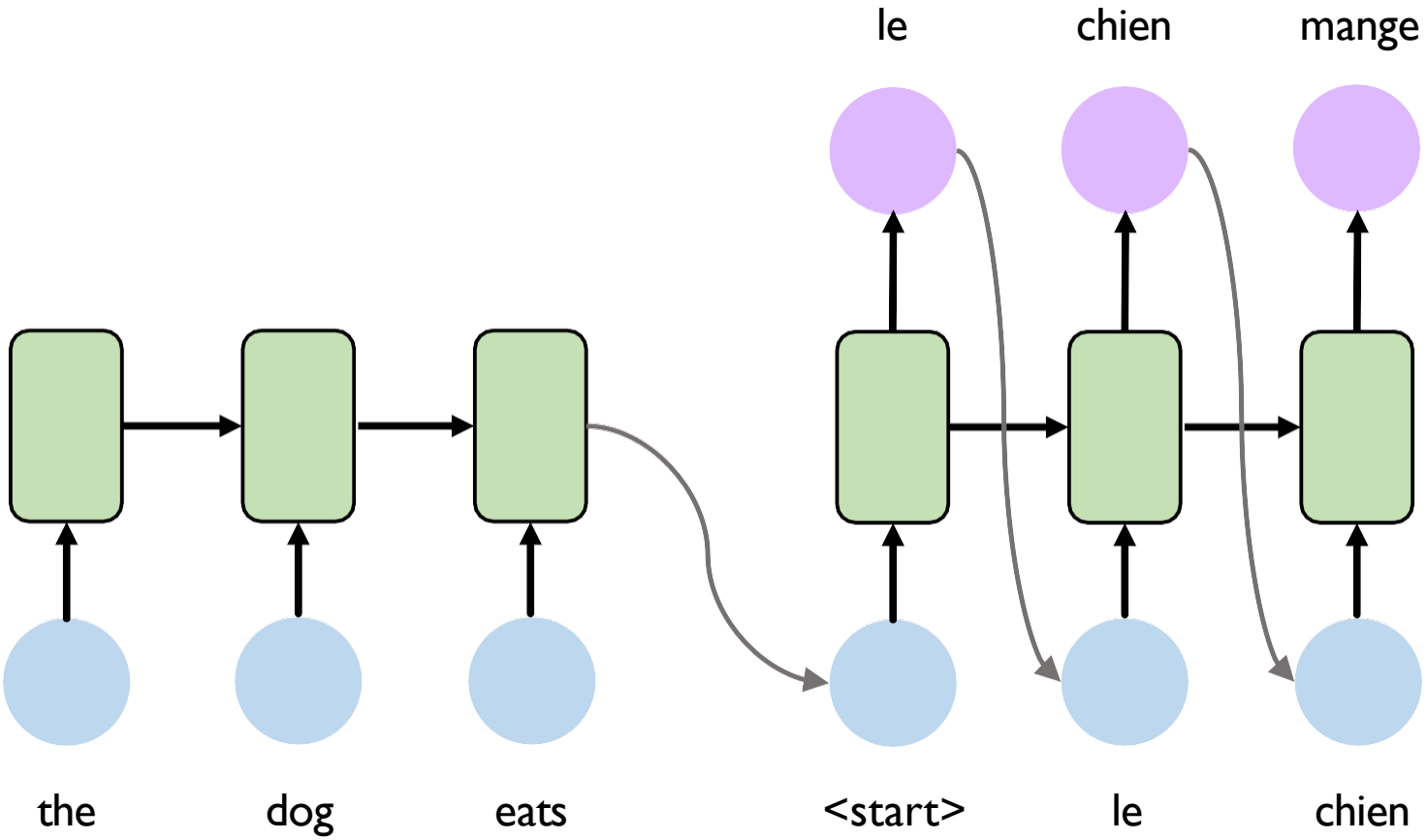


Input: sequence of words

Output: probability of having positive sentiment

Adapted from H. Suresh, 2018

Example task: machine translation



Adapted from H. Suresh, 2018

Questions?