

# Deep Learning

## MSDS 631

Sequence Models

Michael Ruddy

# Overview

- Sequences
- Amped up RNNs (LSTMs + GRUs)
- Encoder - Decoder (Seq2Seq)

# Deep Learning and Sequences

- Sequences
  - Variable length
  - Relationships between elements of sequence
- Examples
  - Text
  - Time Series
- Models
  - Continuous Bag of Words (CBOW)
  - 1D CNN
  - Recurrent Neural Network (RNN)

# Deep Learning and Sequences

- Sequences
  - Variable length
  - Relationships between elements of sequence
- Examples
  - Text
  - Time Series
- Models
  - Continuous Bag of Words (CBOW)
  - 1D CNN
  - Recurrent Neural Network (RNN)
  - Attention (Next!)

# Deep Learning and Sequences

- Models

- Continuous Bag of Words (CBOW)

- 1D CNN

- Recurrent Neural Network (RNN)

only 0 pad in front, not in the end

- Average feature vectors together to get fixed length input

- Lose a lot information about the sequence

# Deep Learning and Sequences

- Models
  - Continuous Bag of Words (CBOW)
  - 1D CNN
  - Recurrent Neural Network (RNN)
- Doesn't care about sequence length
- Uses filters to construct features from local interactions
- Difficult to capture long range dependencies

# Deep Learning and Sequences

- Models
  - Continuous Bag of Words (CBOW)
  - 1D CNN
  - Recurrent Neural Network (RNN)
- Updates a hidden state as the sequence is fed into the RNN
- Vanishing/Exploding gradient problem
- Doesn't have great long-term memory
- Slow (can't parallelize updates to a hidden state)

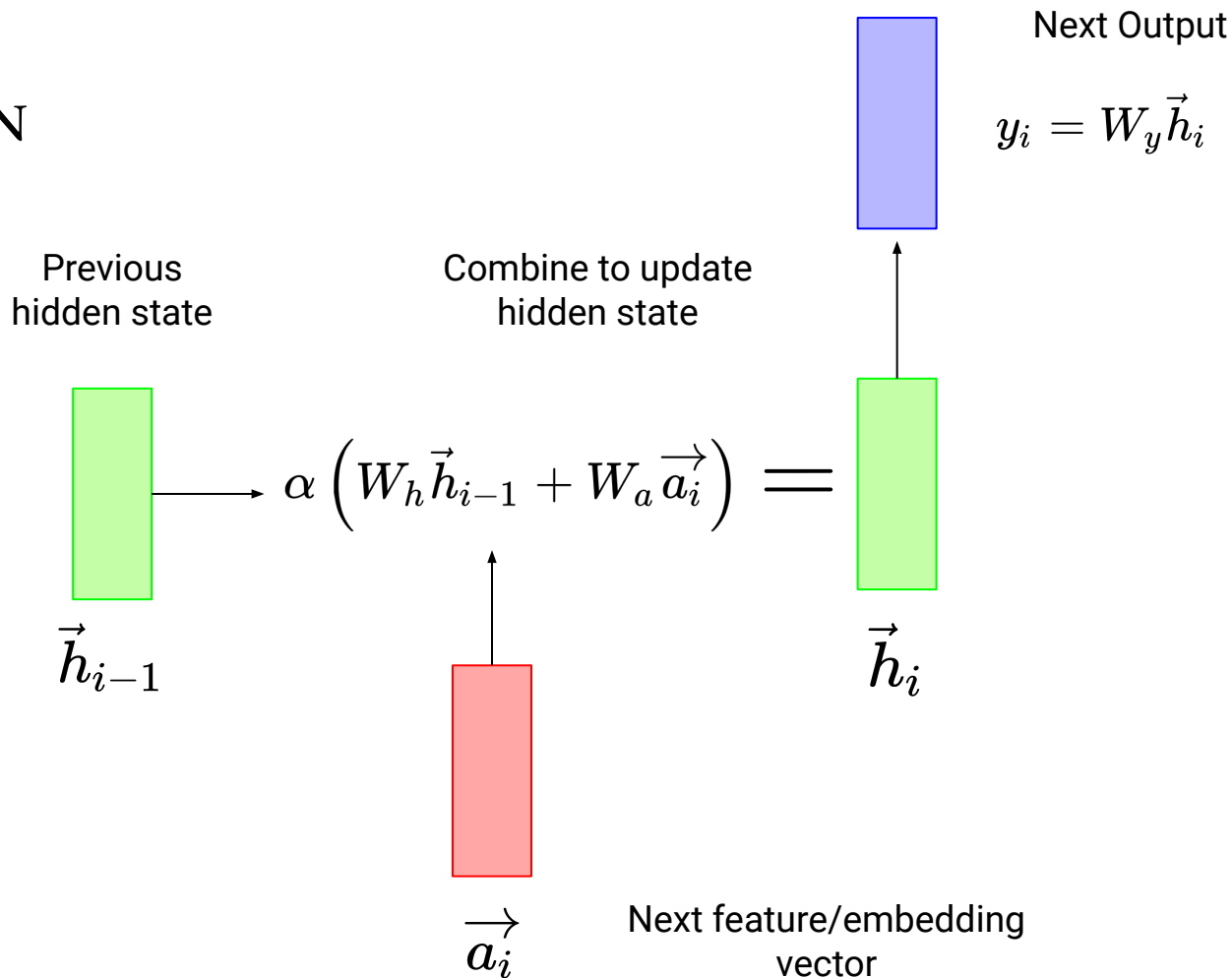
# Deep Learning and Sequences

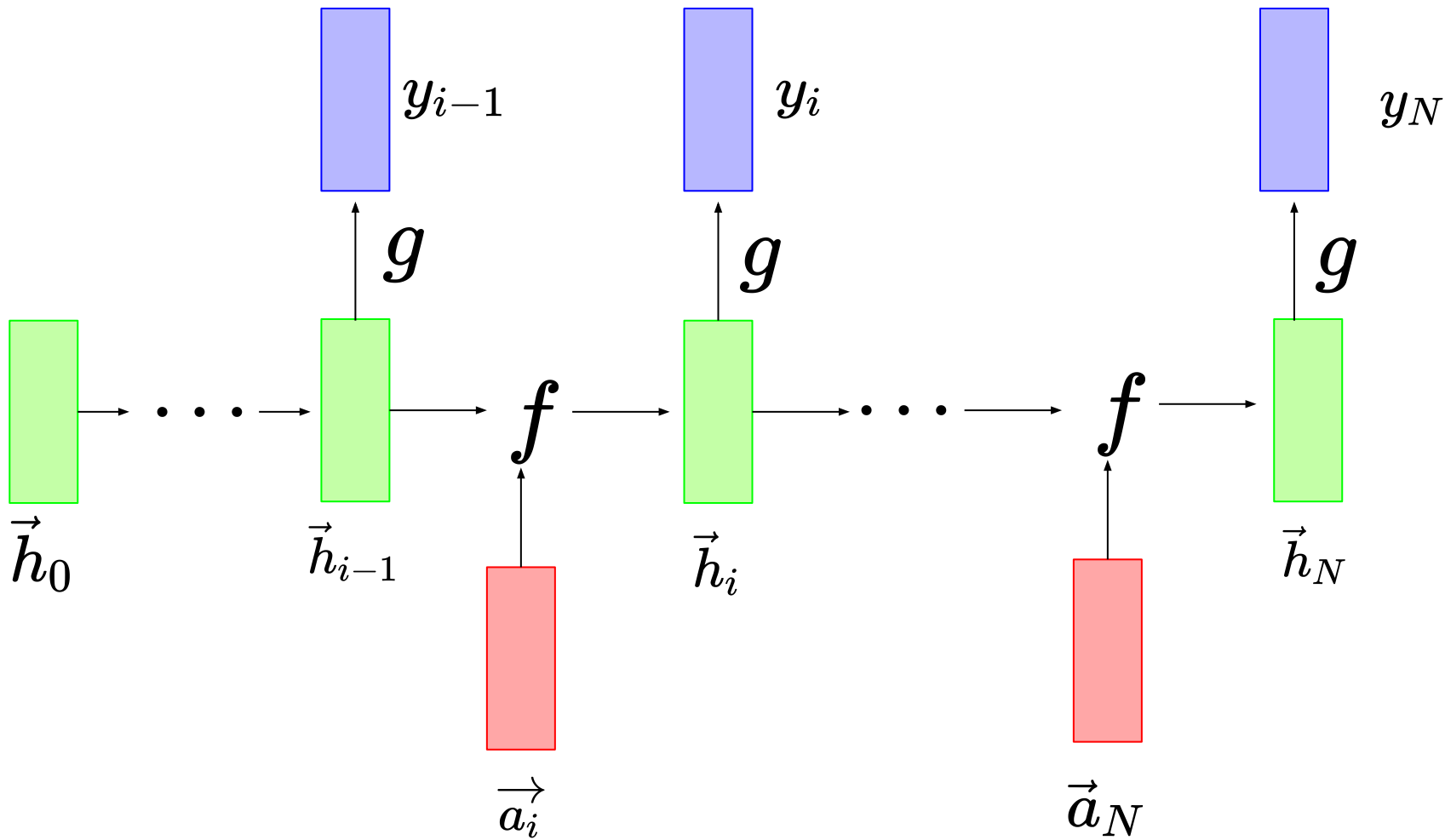
- Models
  - Continuous Bag of Words (CBOW)
  - 1D CNN
  - Recurrent Neural Network (RNN)
  - LSTMs, GRUs, and more!
- Fancier updates to a hidden state as the sequence is fed into the NN
- Helps with Vanishing/Exploding gradient problem
- Helps with long-term memory
- Still Slow (can't parallelize updates to a hidden state)



# RNNs

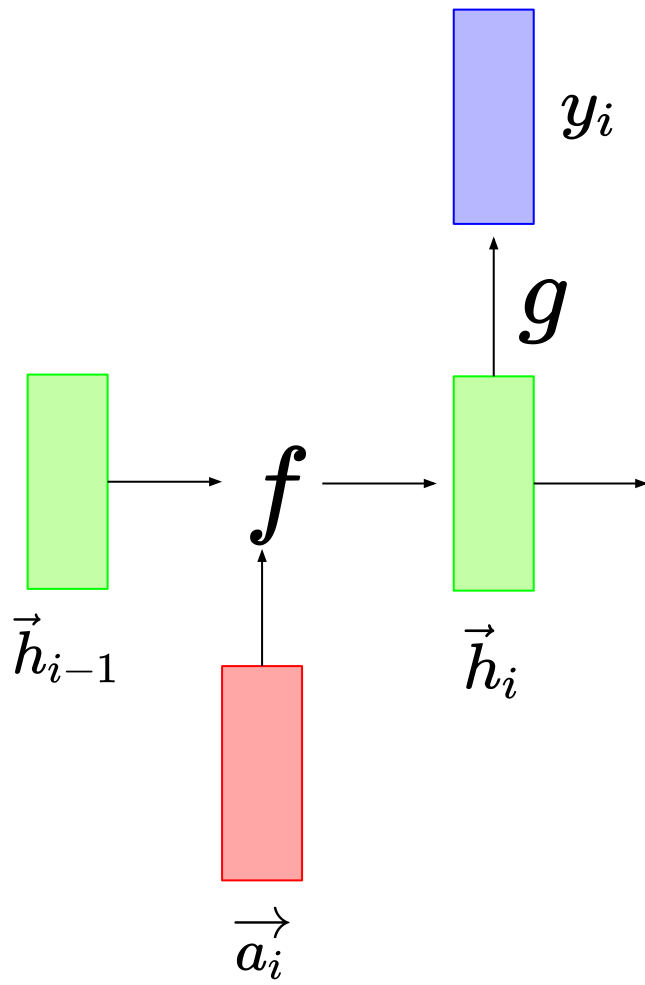
## - Vanilla RNN



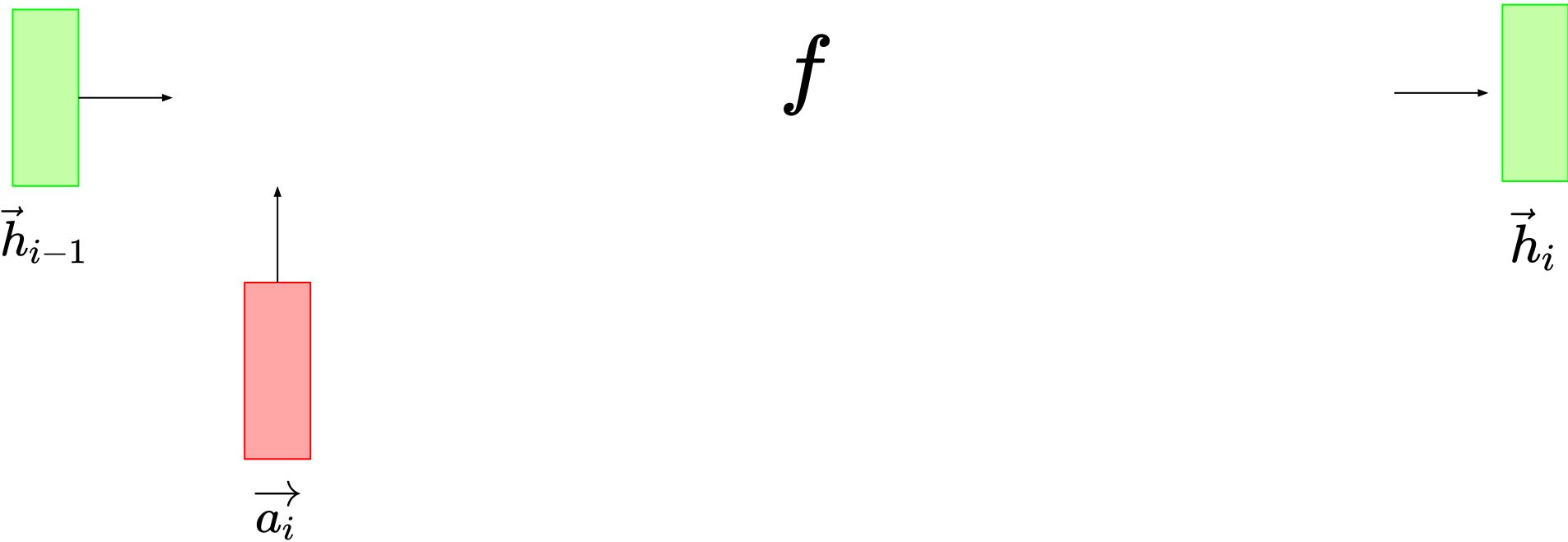


# The GRU

- Gated Recurrent Unit
- Idea: Change the function  $f$  to address common RNN problems



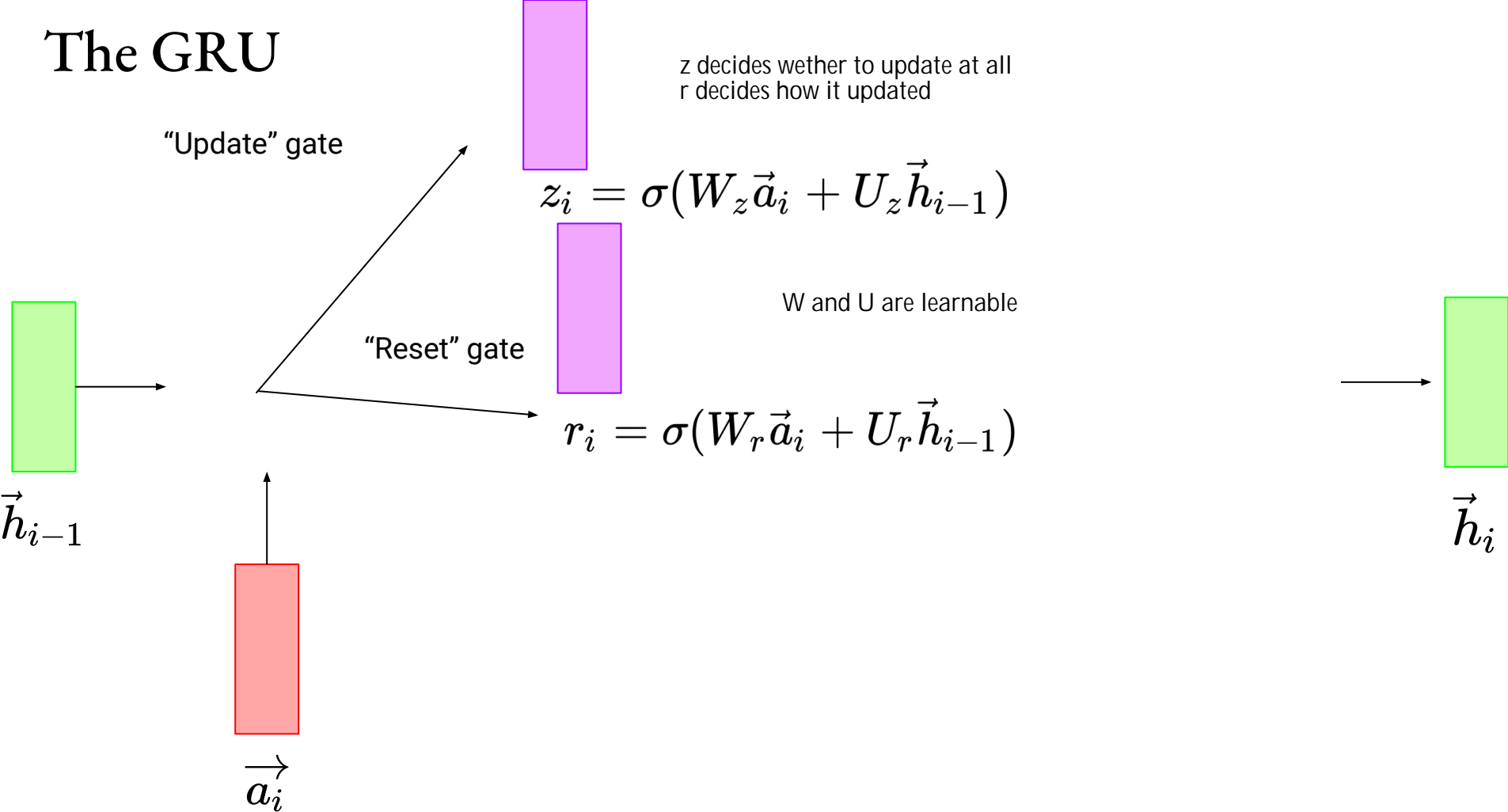
# The GRU



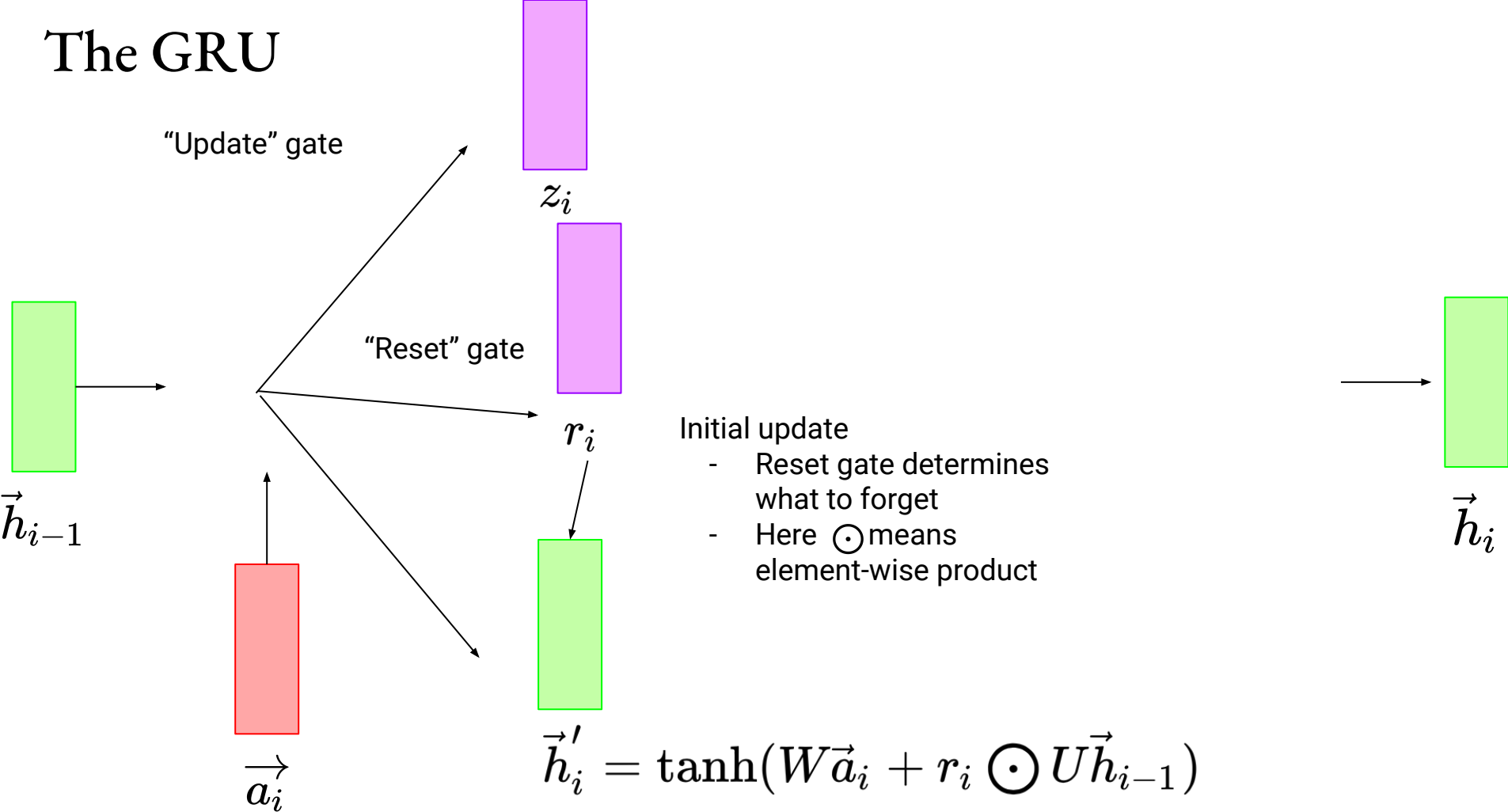
# The GRU



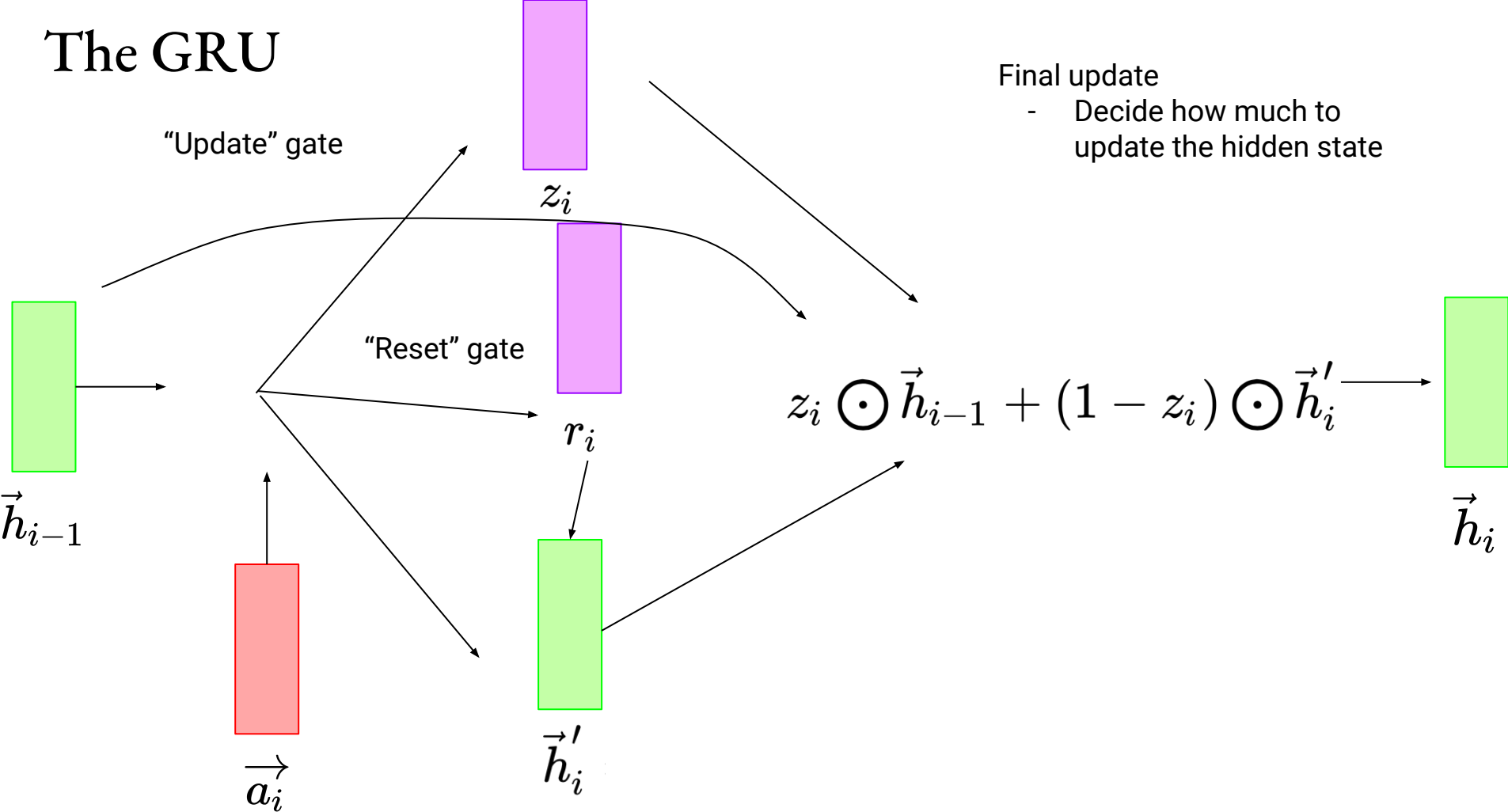
# The GRU



# The GRU



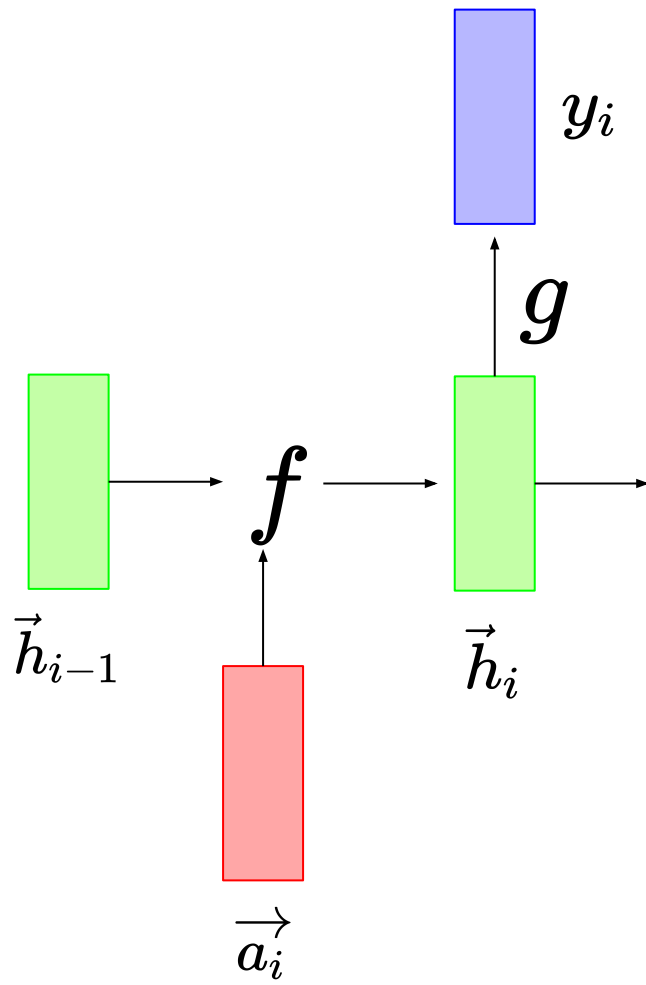
# The GRU





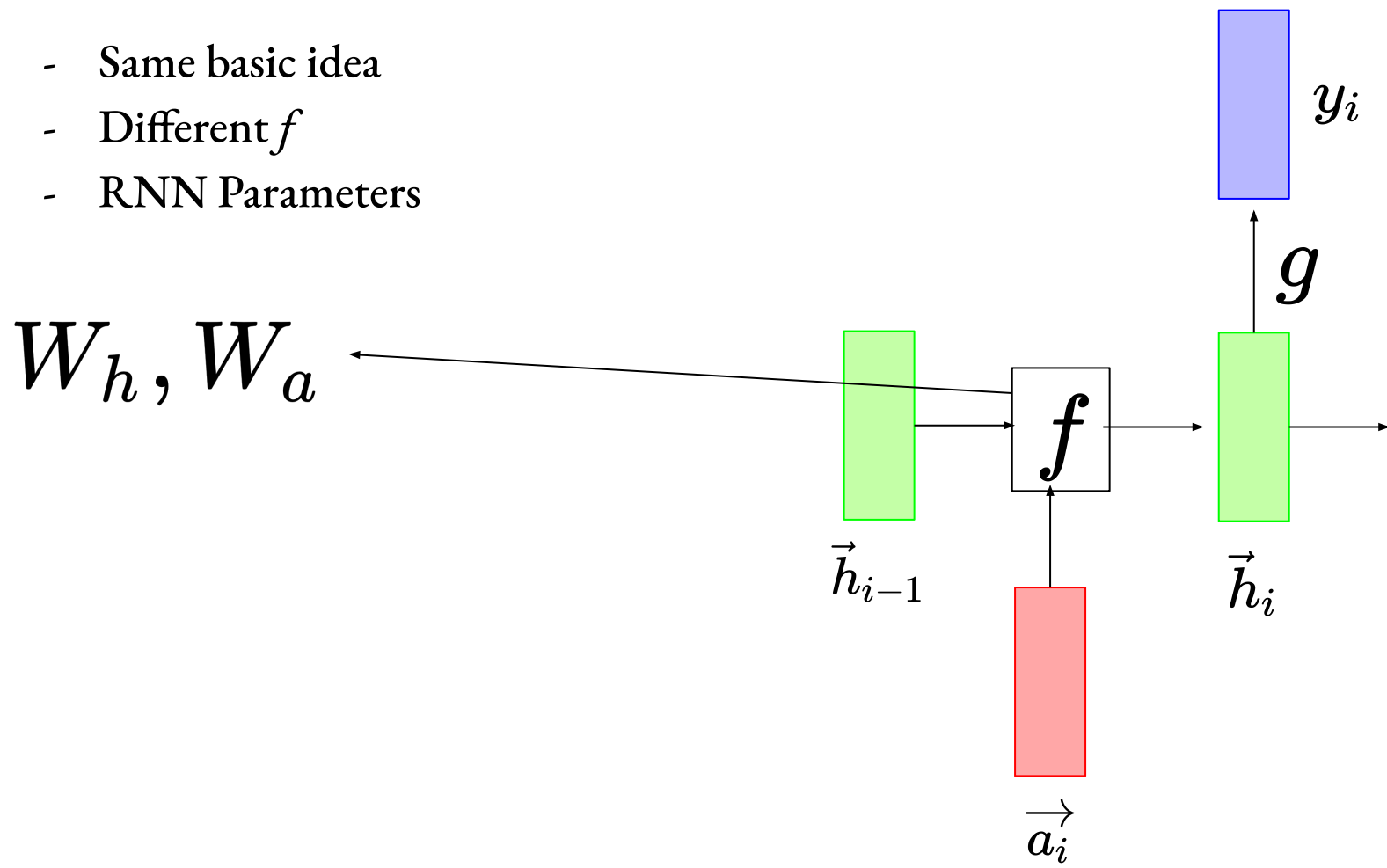
# Recurrent Unit

- Same basic idea
- Different  $f$



# Recurrent Unit

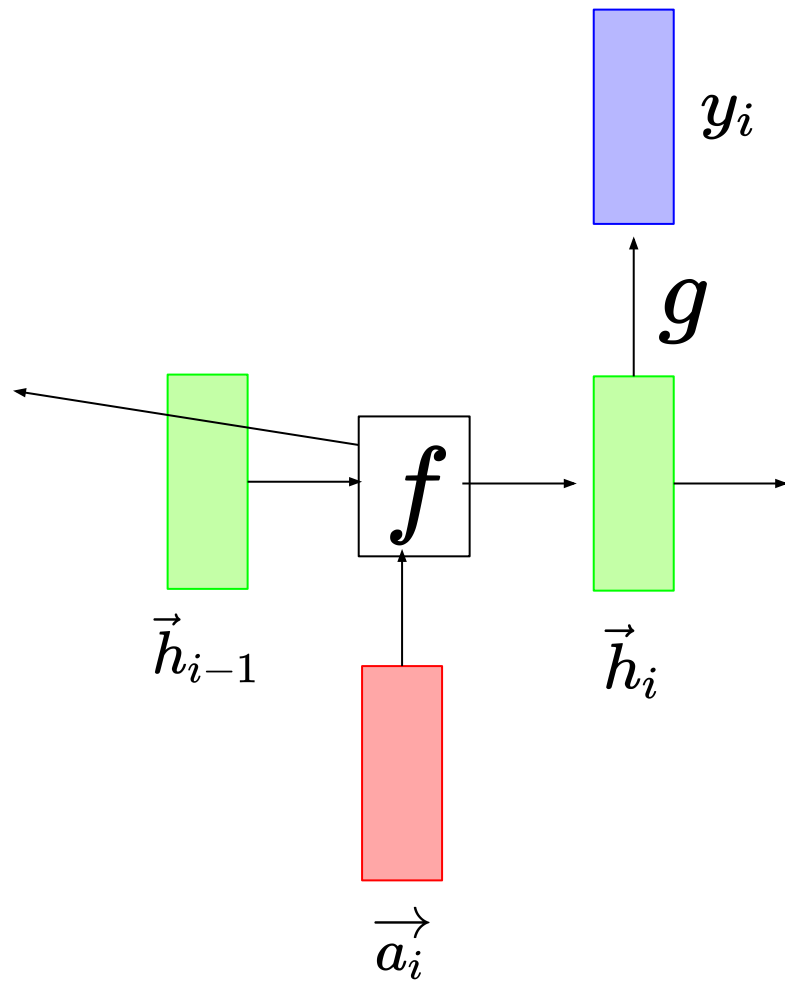
- Same basic idea
- Different  $f$
- RNN Parameters



# Recurrent Unit

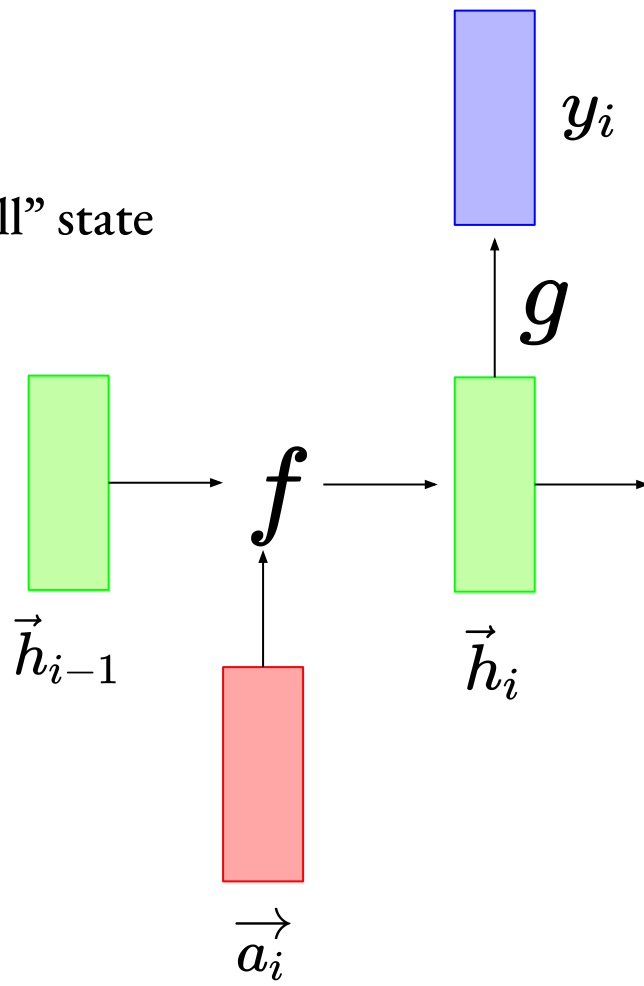
- Same basic idea
- Different  $f$
- GRU Parameters

$W, U, W_z, U_z, W_a, U_a$



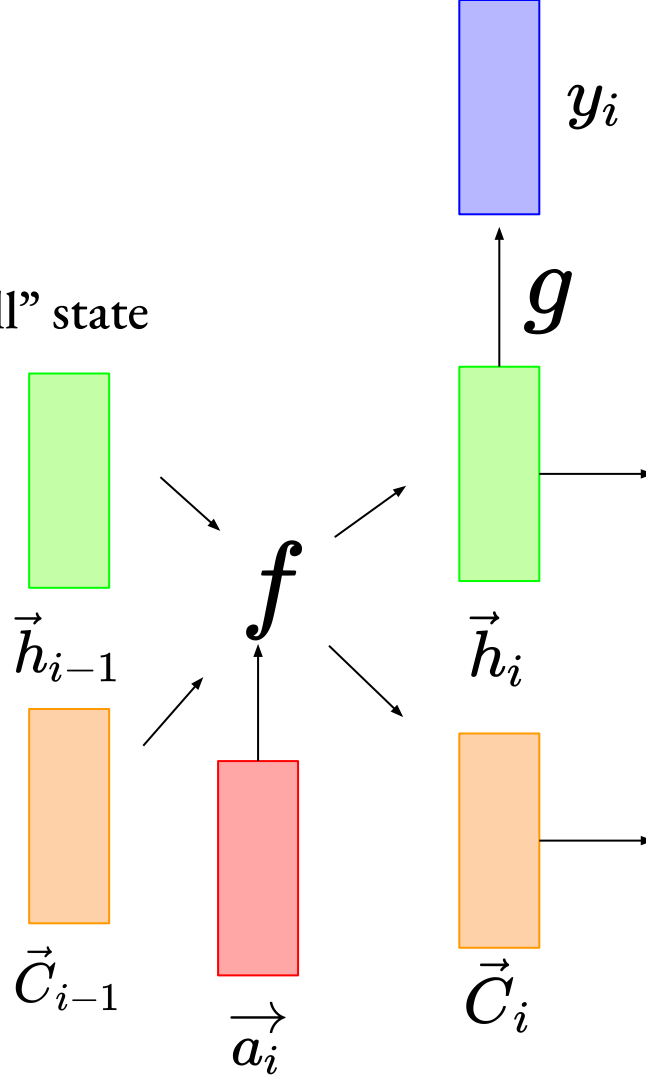
# The LSTM

- Long Short-Term Memory
- Idea: Change the function  $f$  to address common RNN problems and add a “cell” state

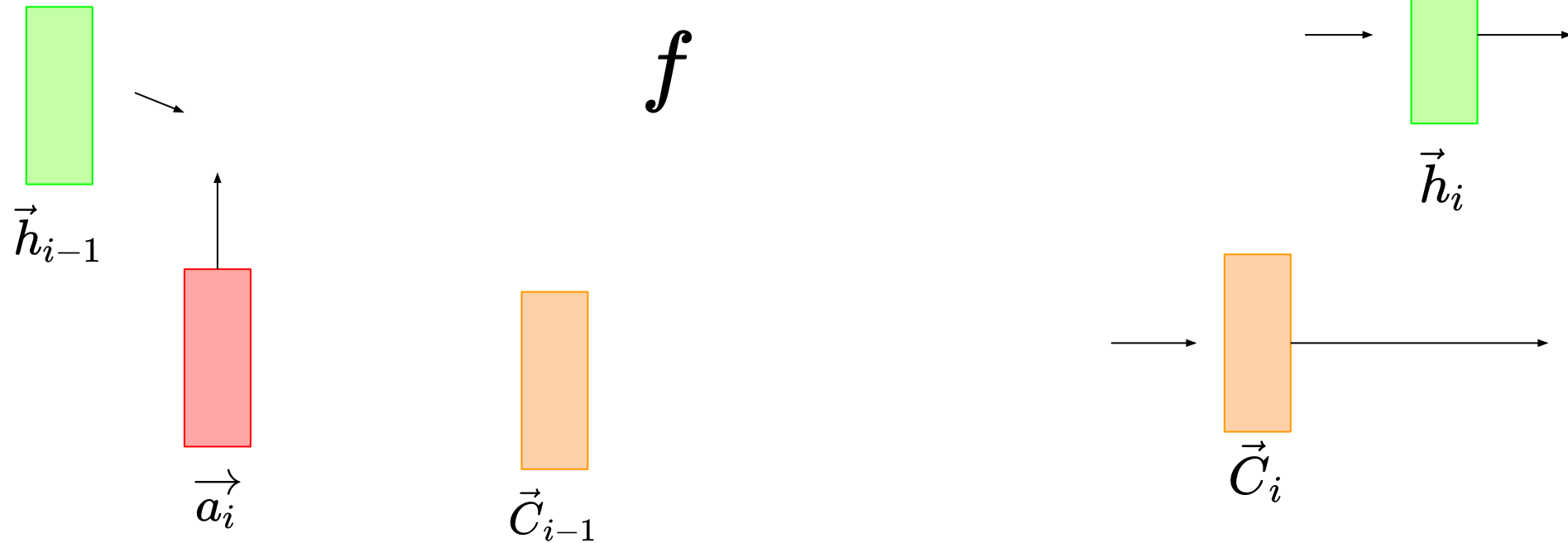


# The LSTM

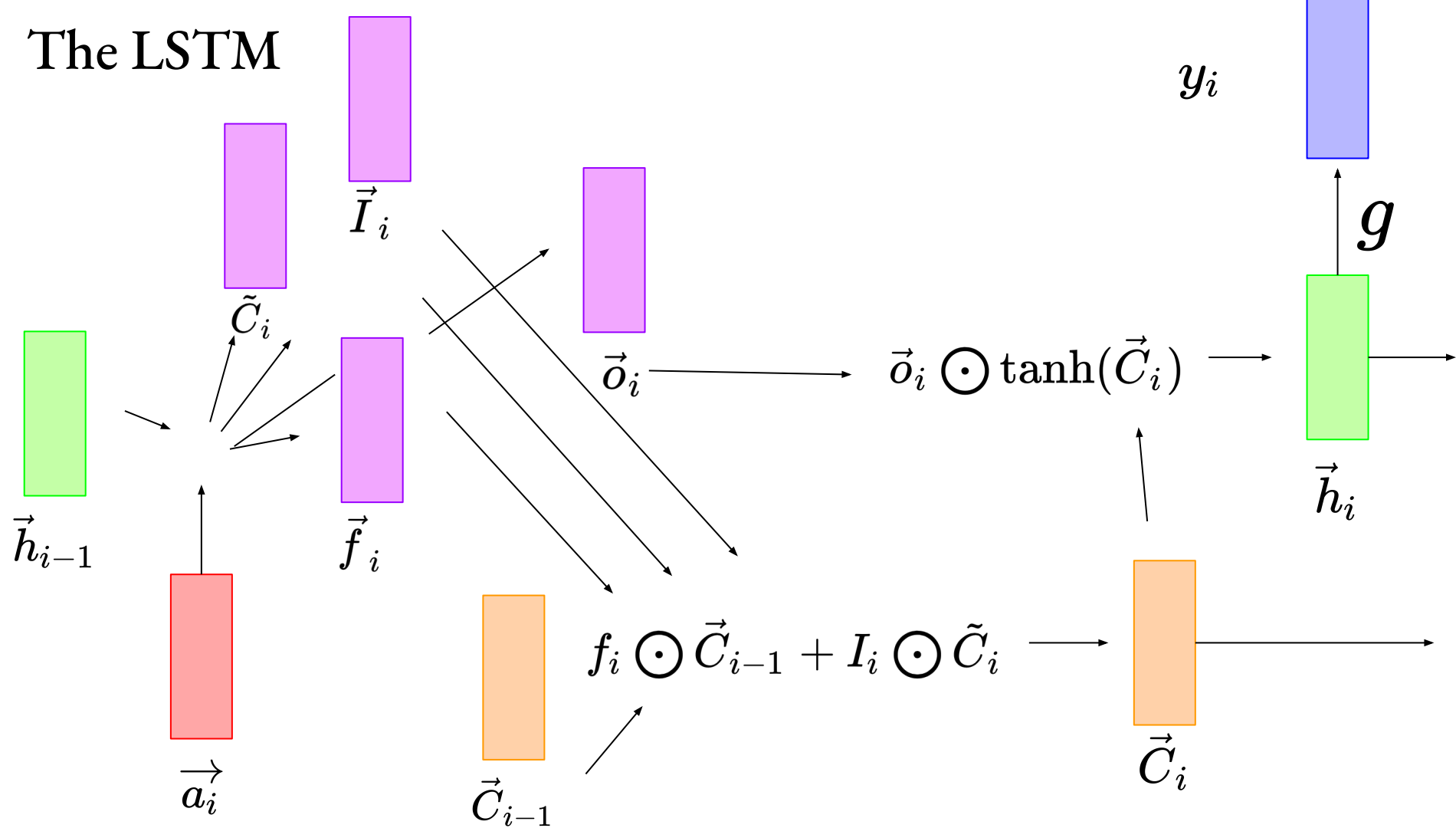
- Long Short-Term Memory
- Idea: Change the function  $f$  to address common RNN problems and add a “cell” state



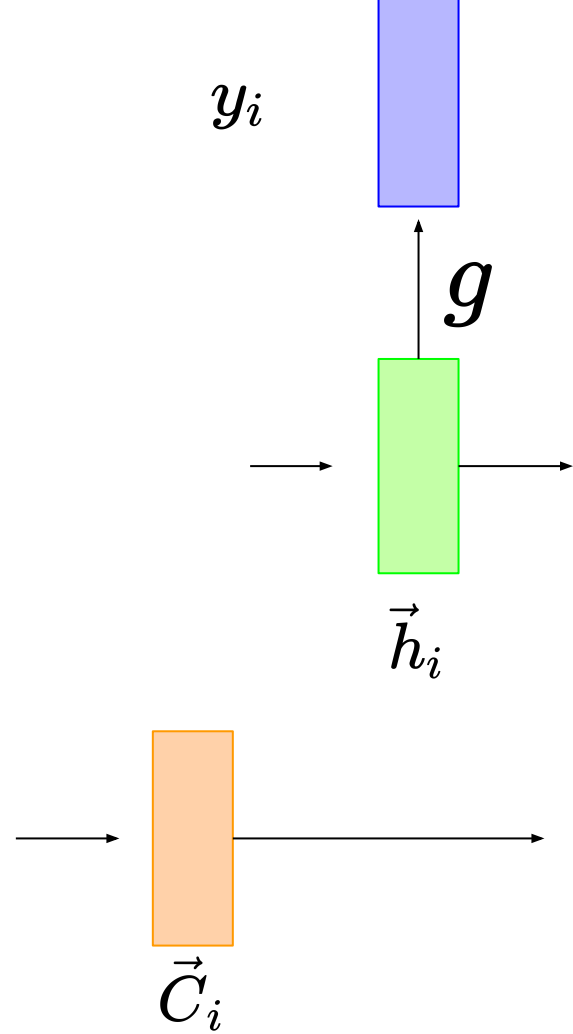
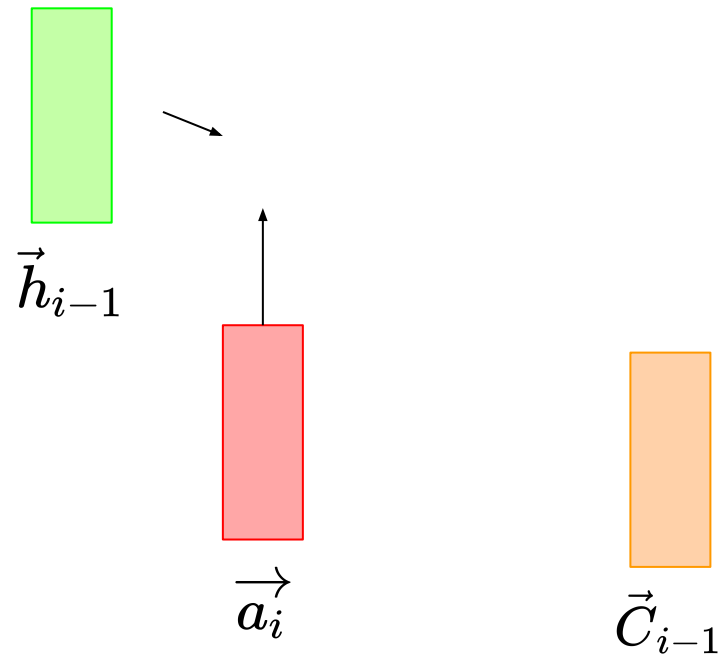
# The LSTM



# The LSTM

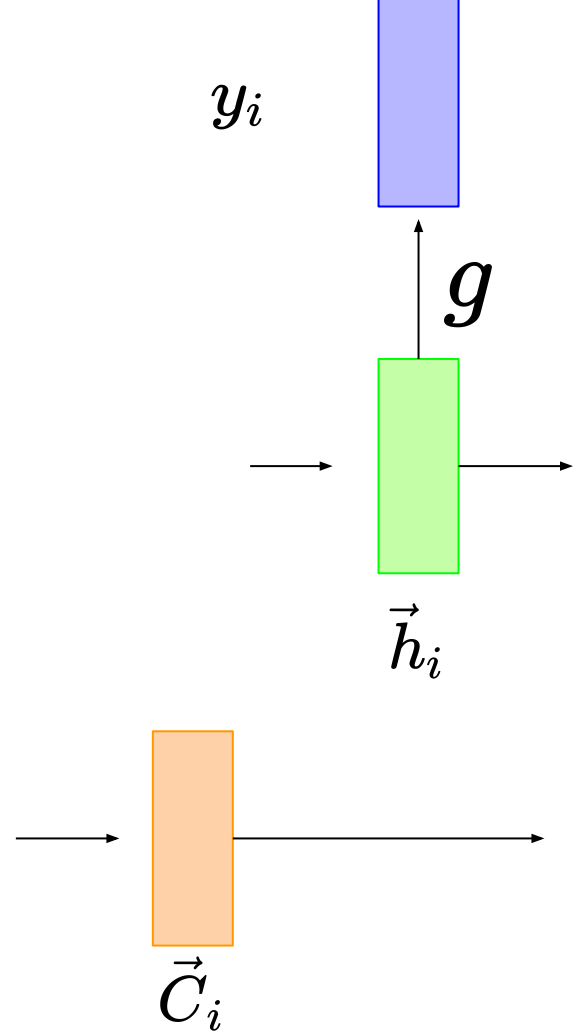
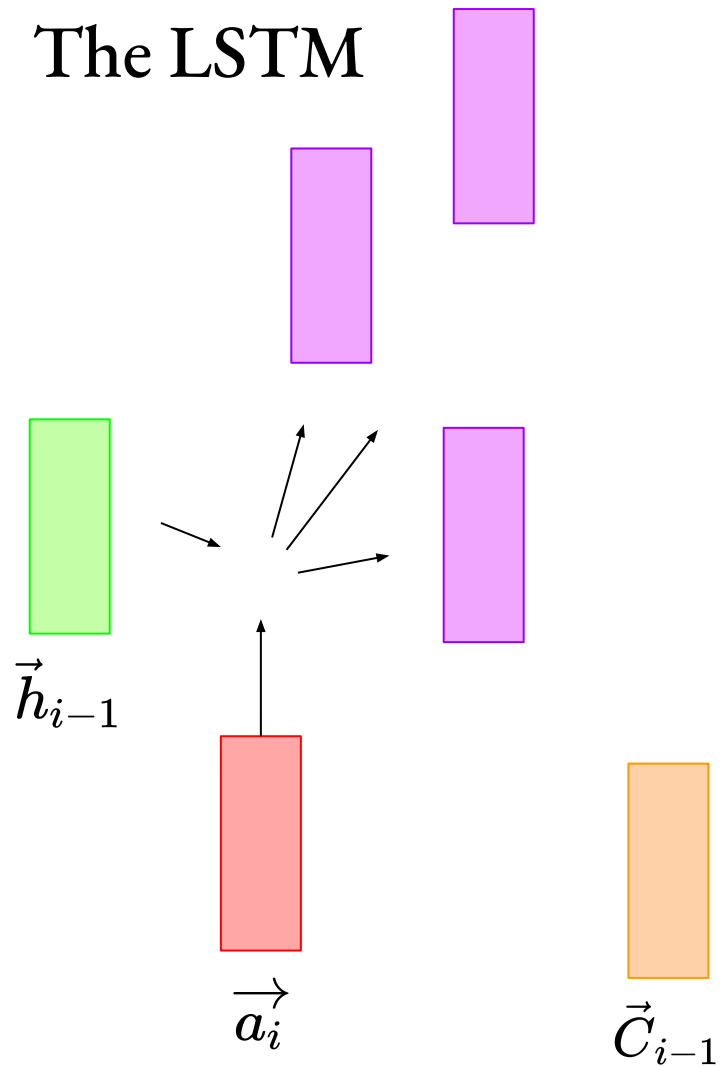


# The LSTM

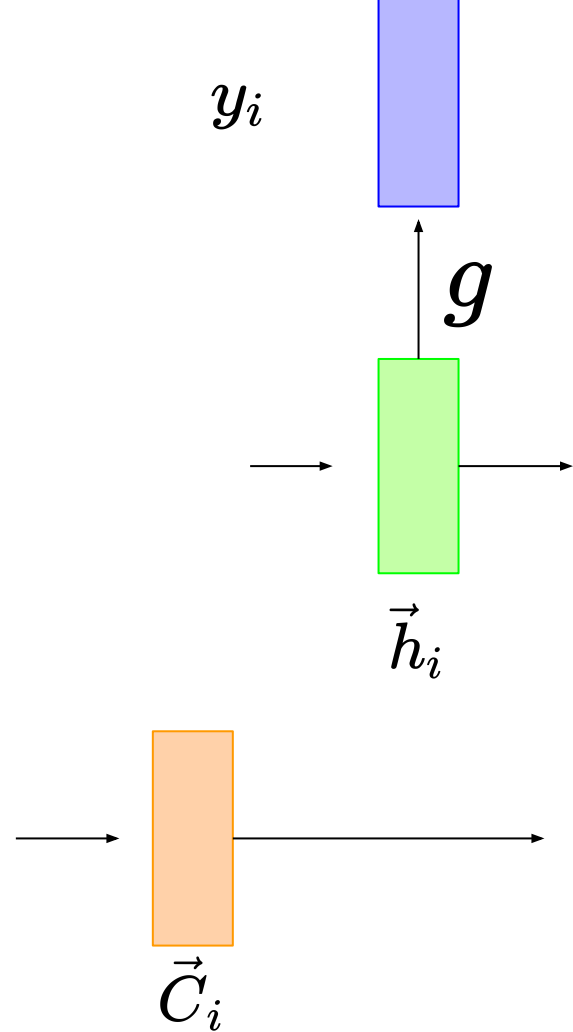
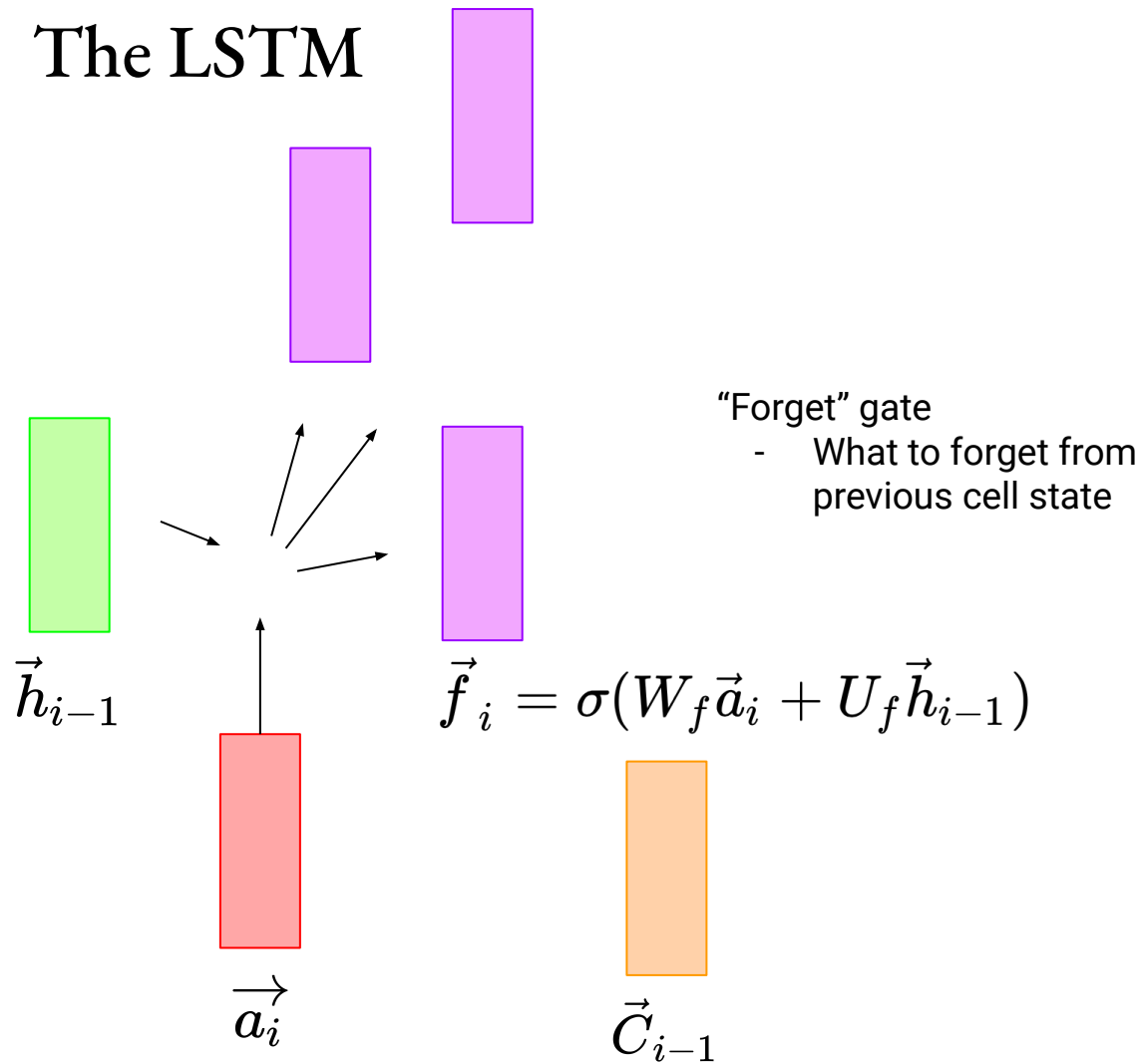




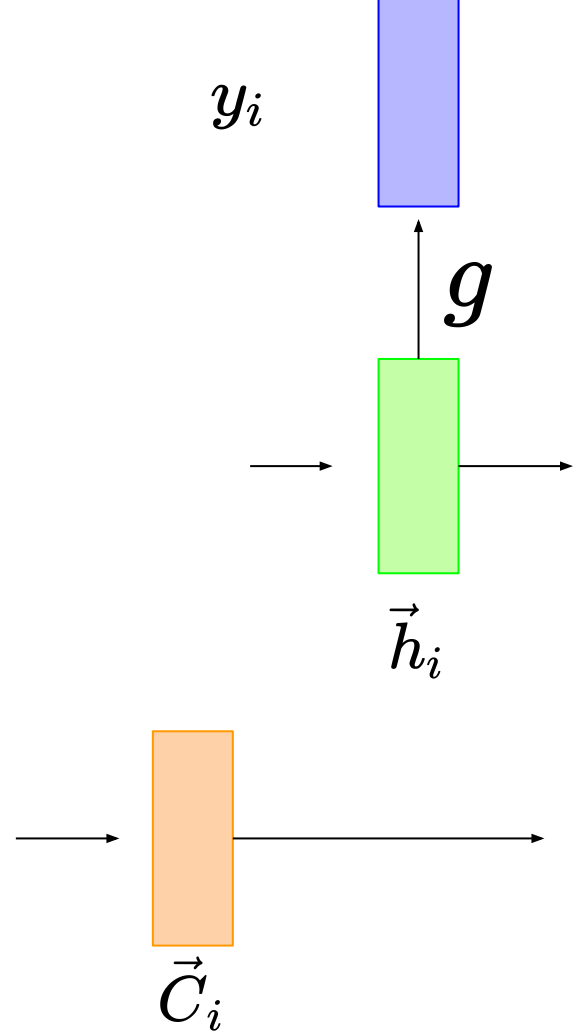
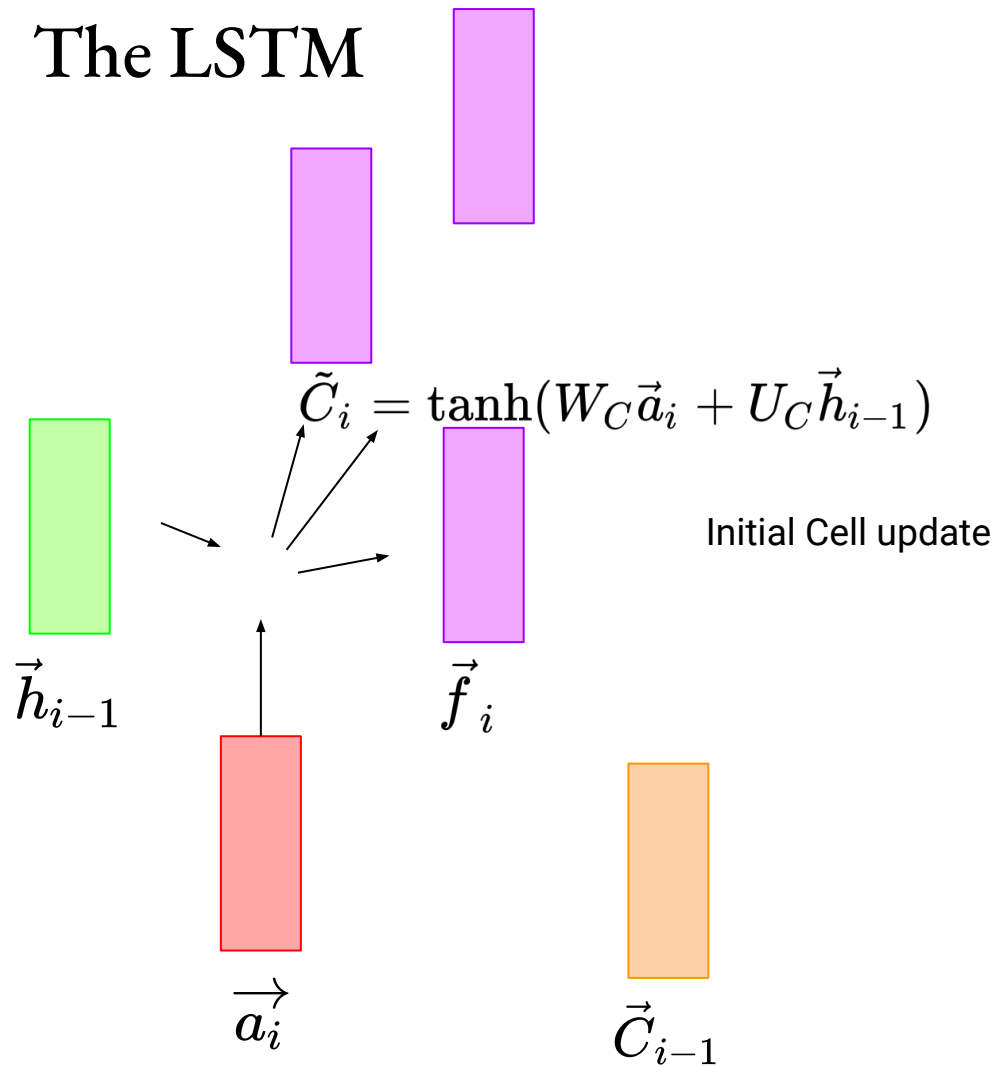
# The LSTM



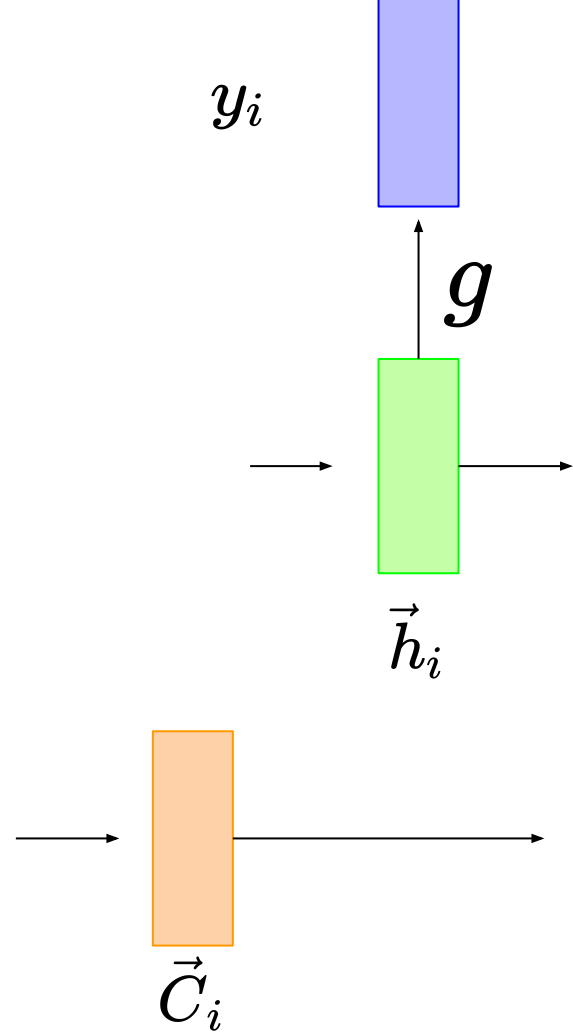
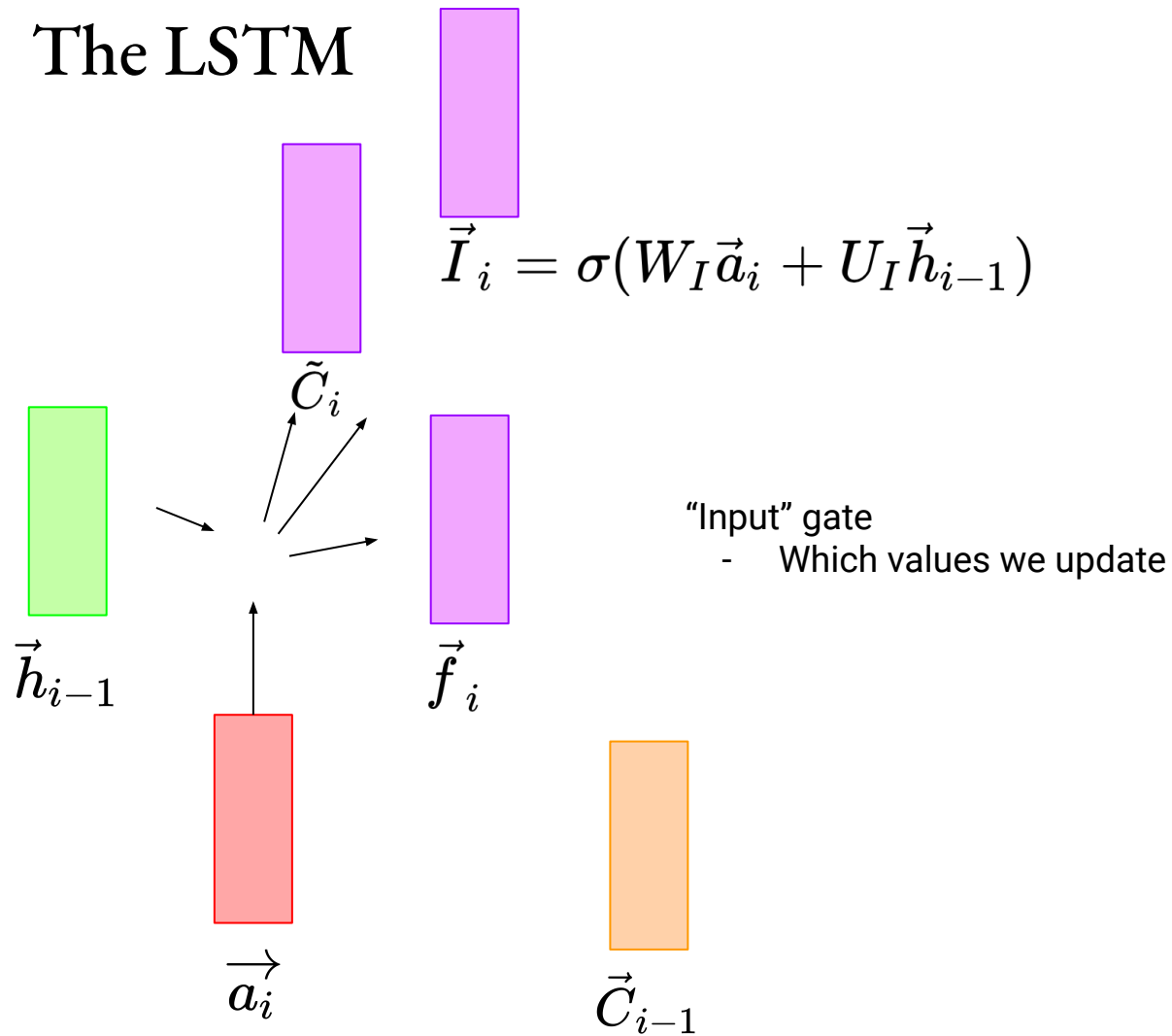
# The LSTM



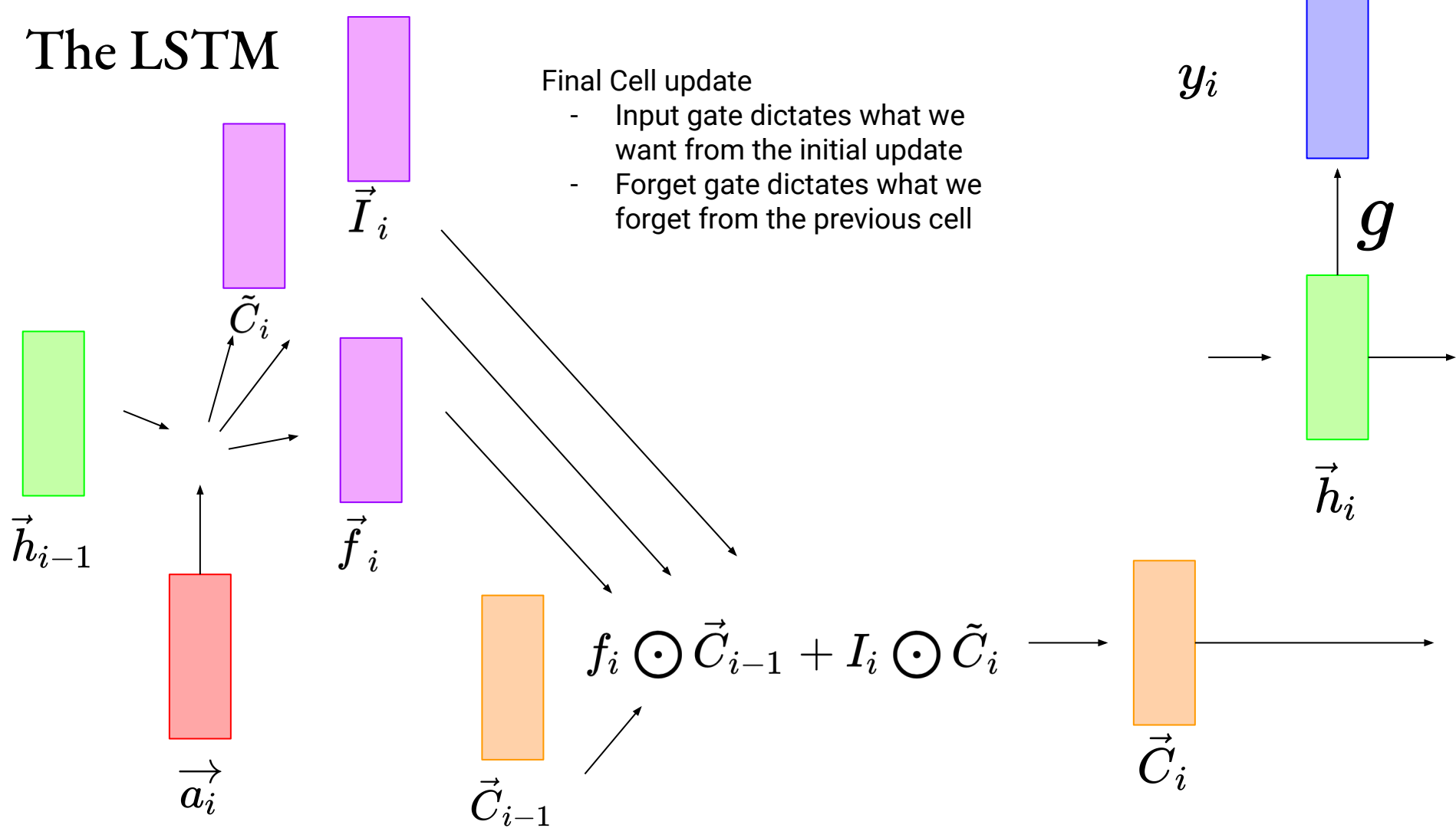
# The LSTM



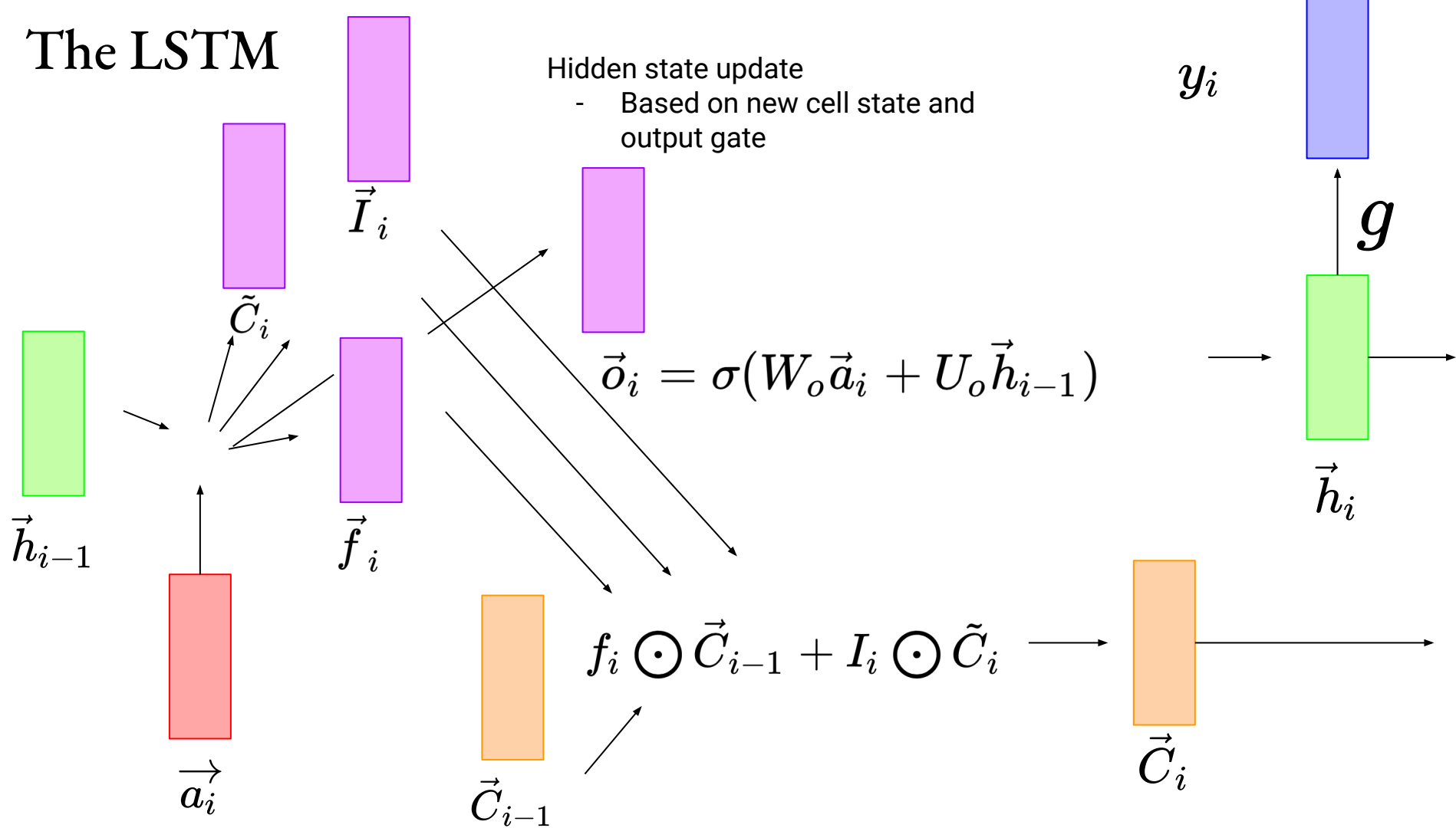
# The LSTM



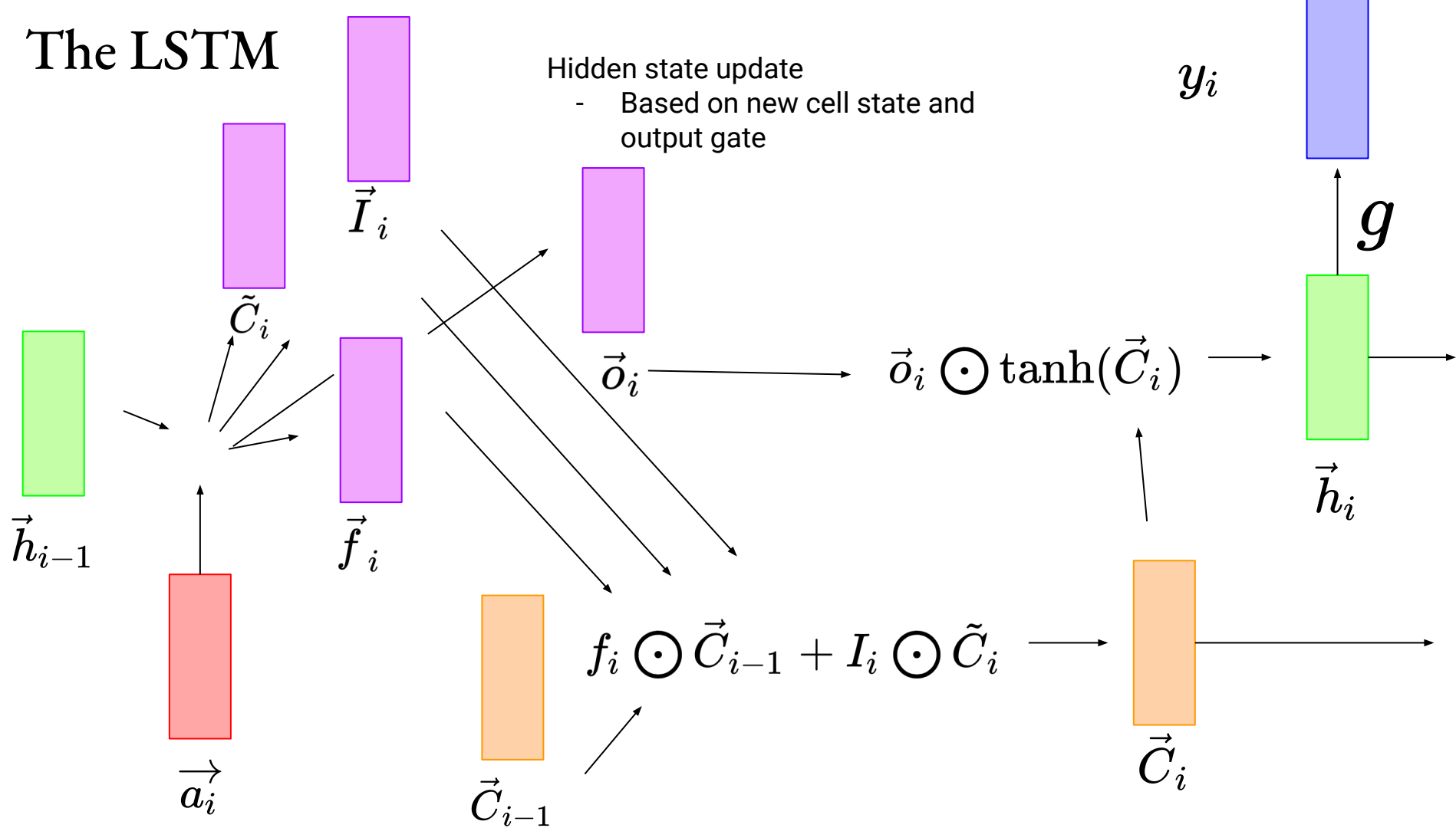
# The LSTM

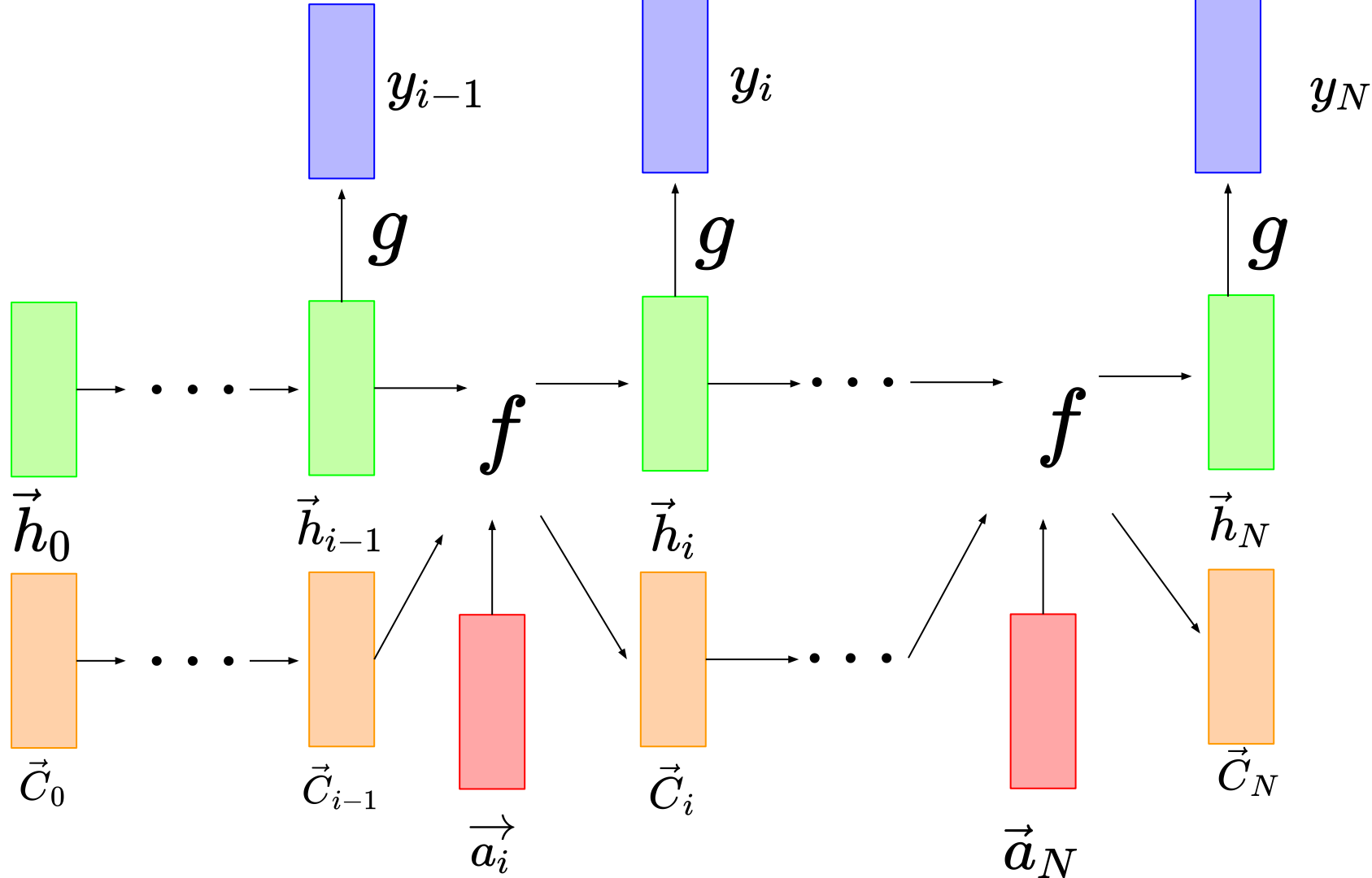


# The LSTM



# The LSTM



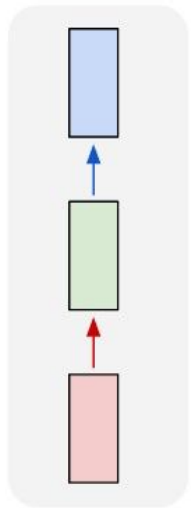




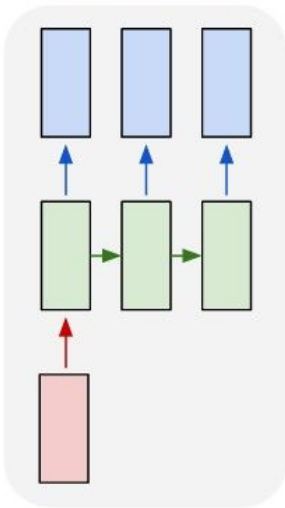
# RNNs

- Different styles based on desired inputs/outputs

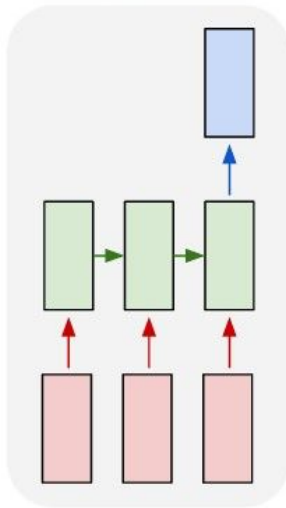
one to one



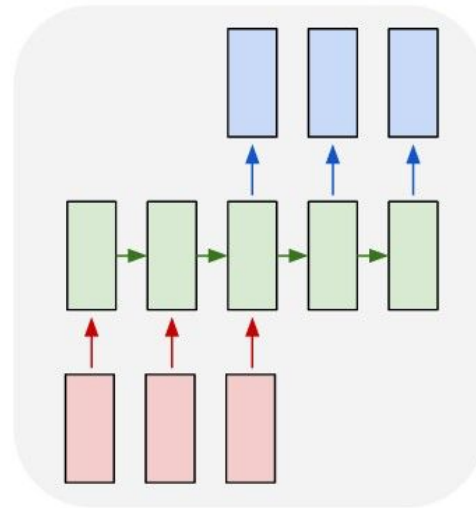
one to many



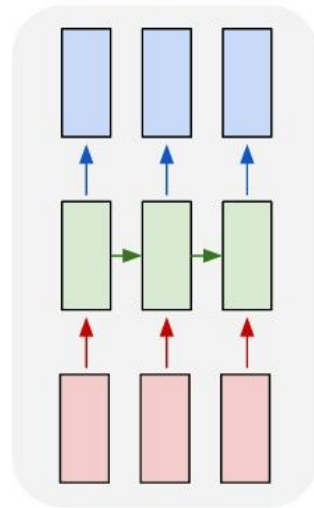
many to one



many to many



many to many



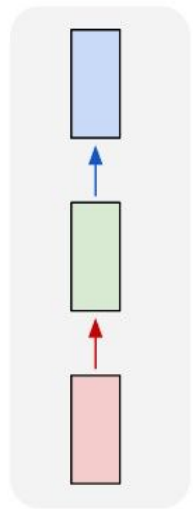
# Questions?

- From last lecture?
- From the lab assignment?

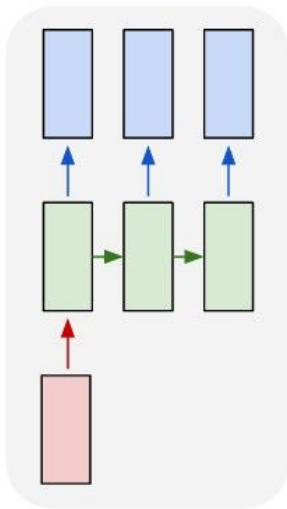
# RNNs

- Different styles based on desired inputs/outputs

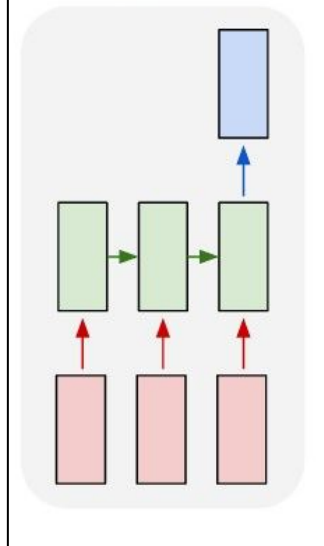
one to one



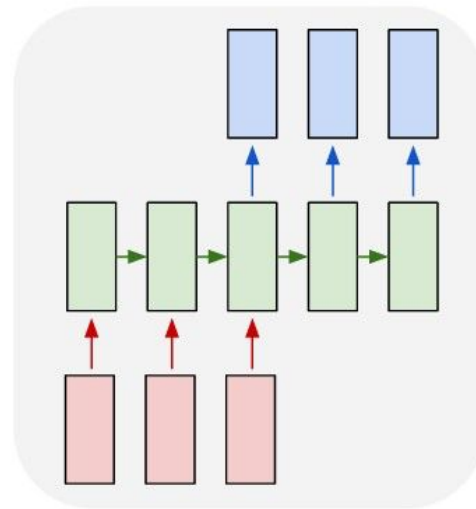
one to many



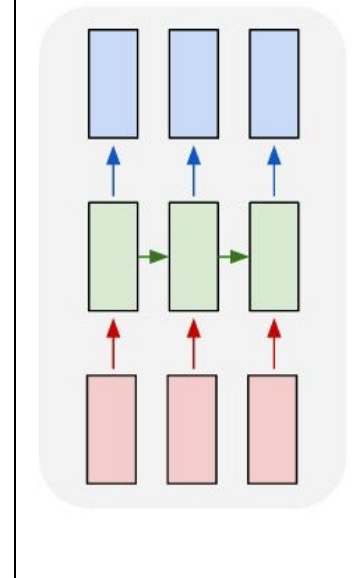
many to one



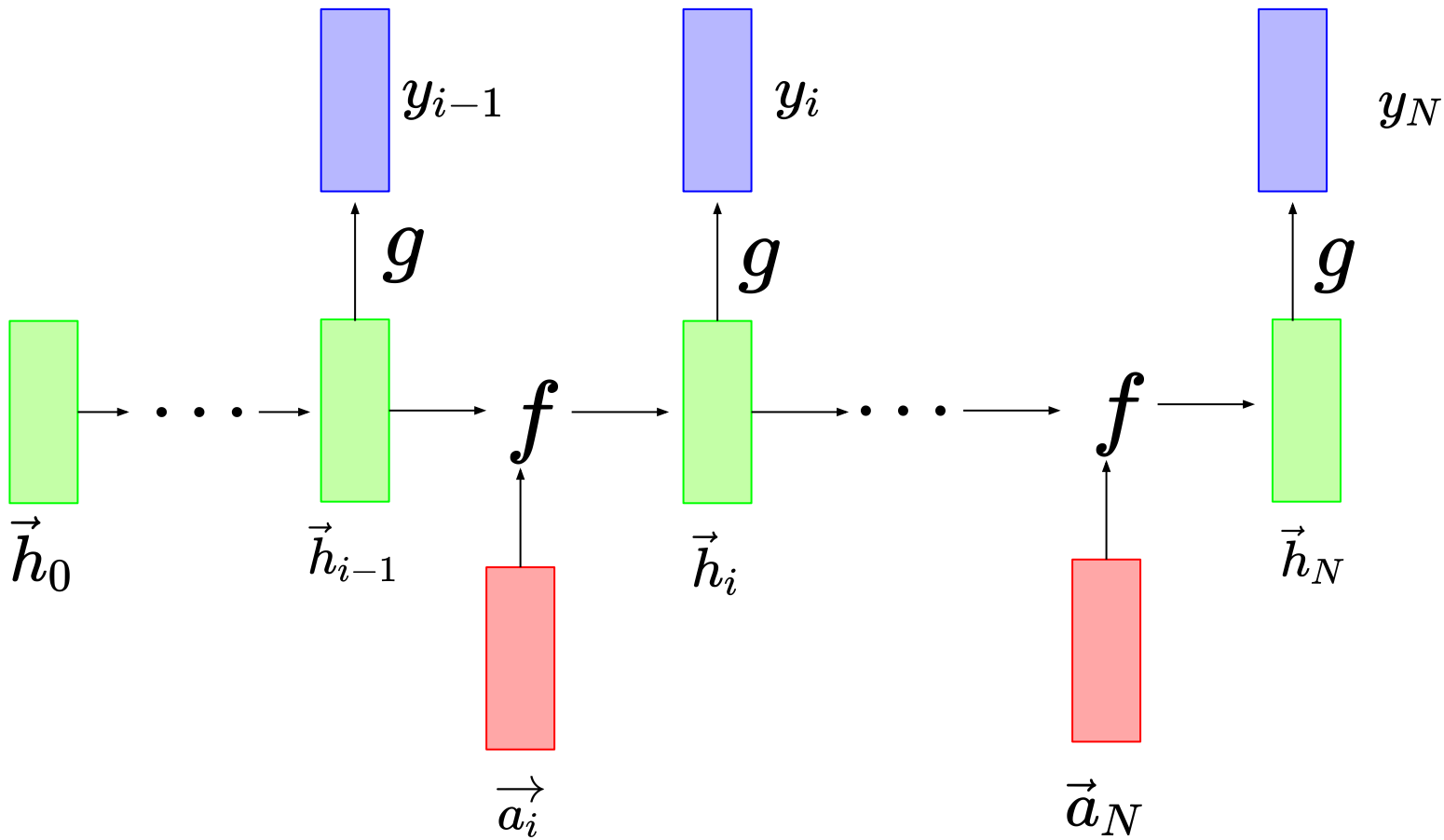
many to many



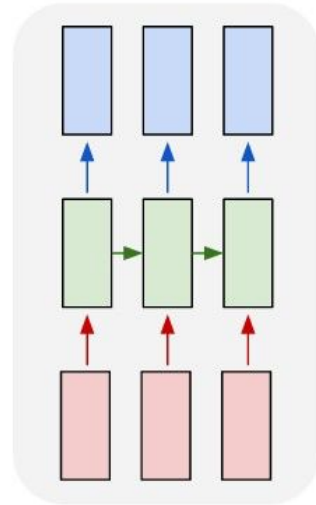
many to many



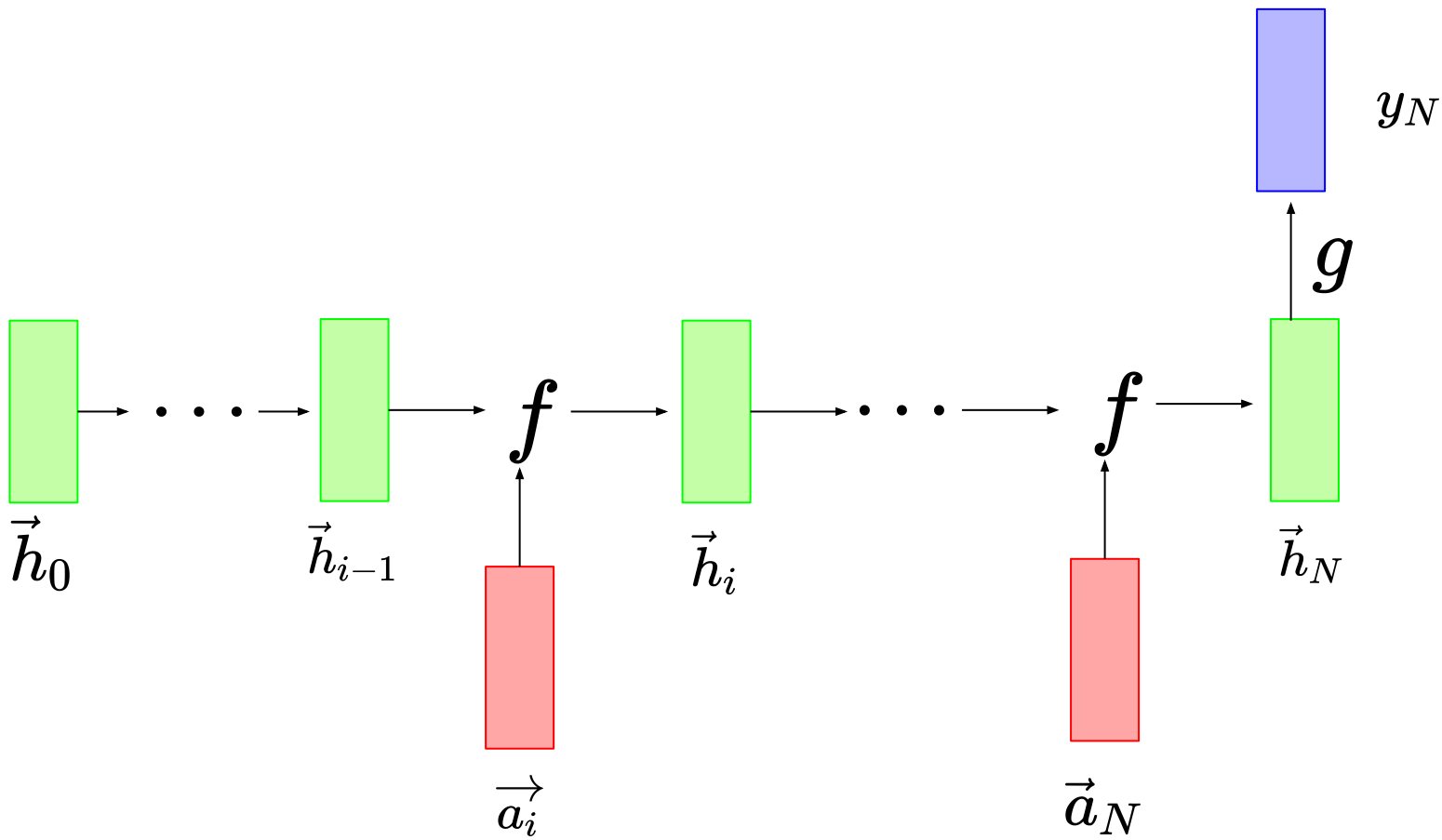
# RNNs



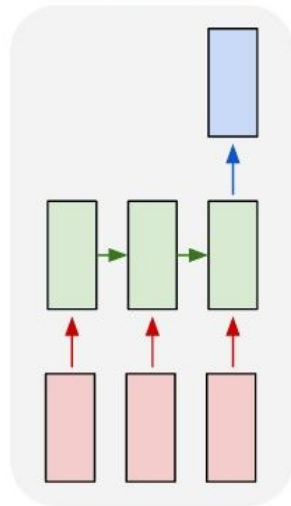
many to many



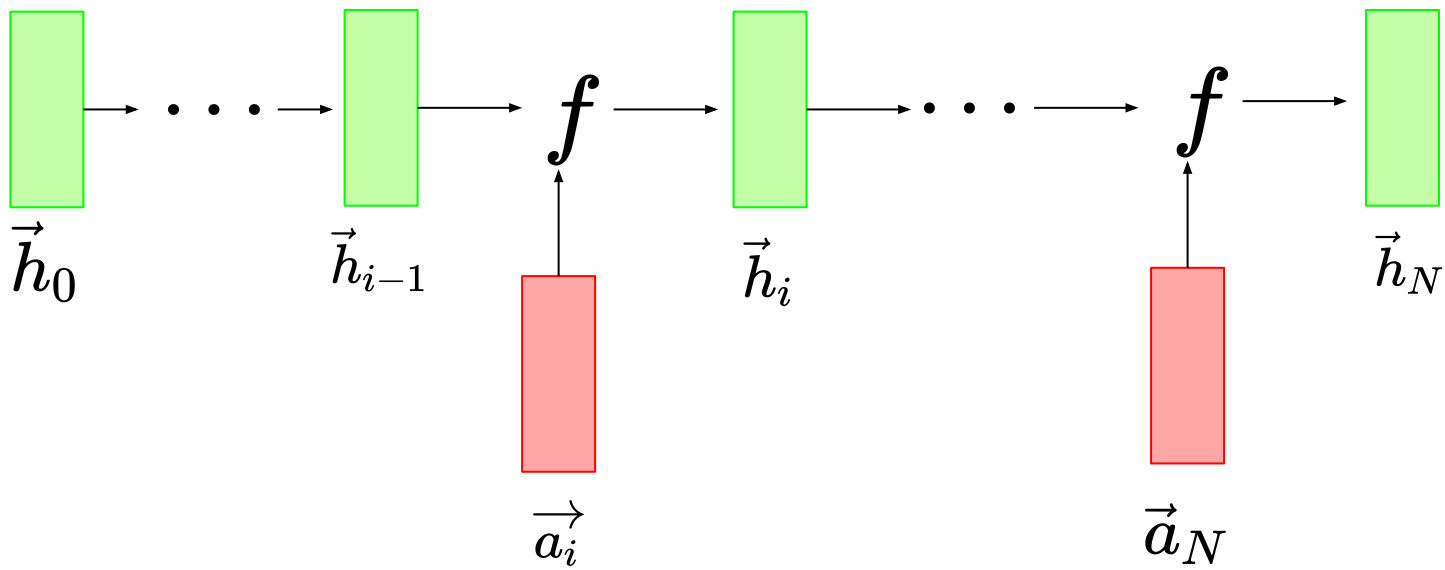
# RNNs



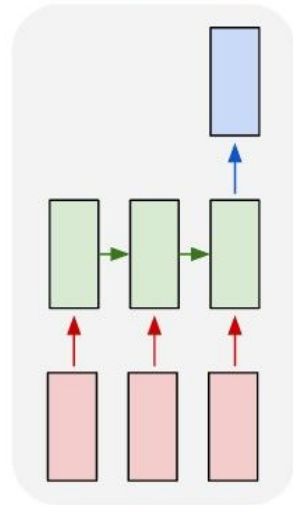
many to one



# RNNs

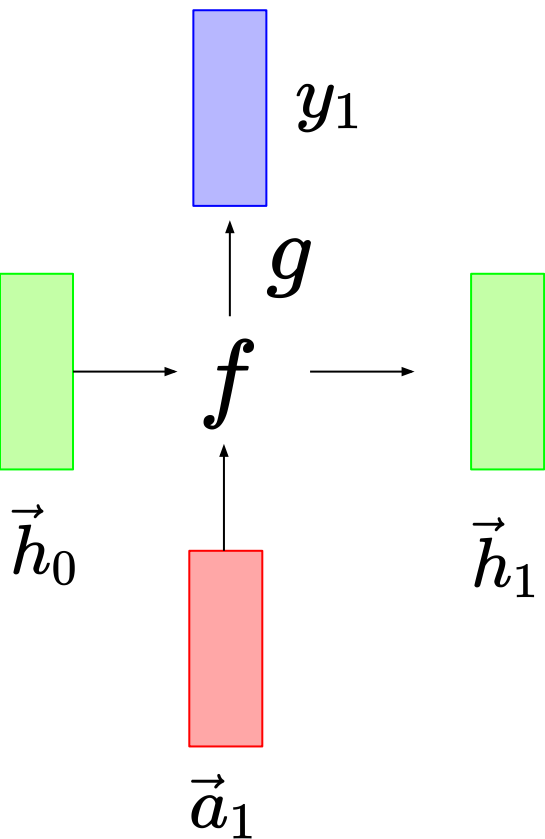


many to one

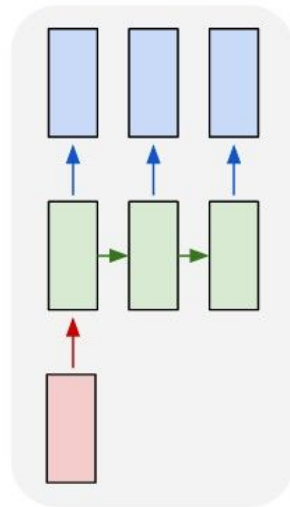


# RNNs

- Generate a sequence from a single input

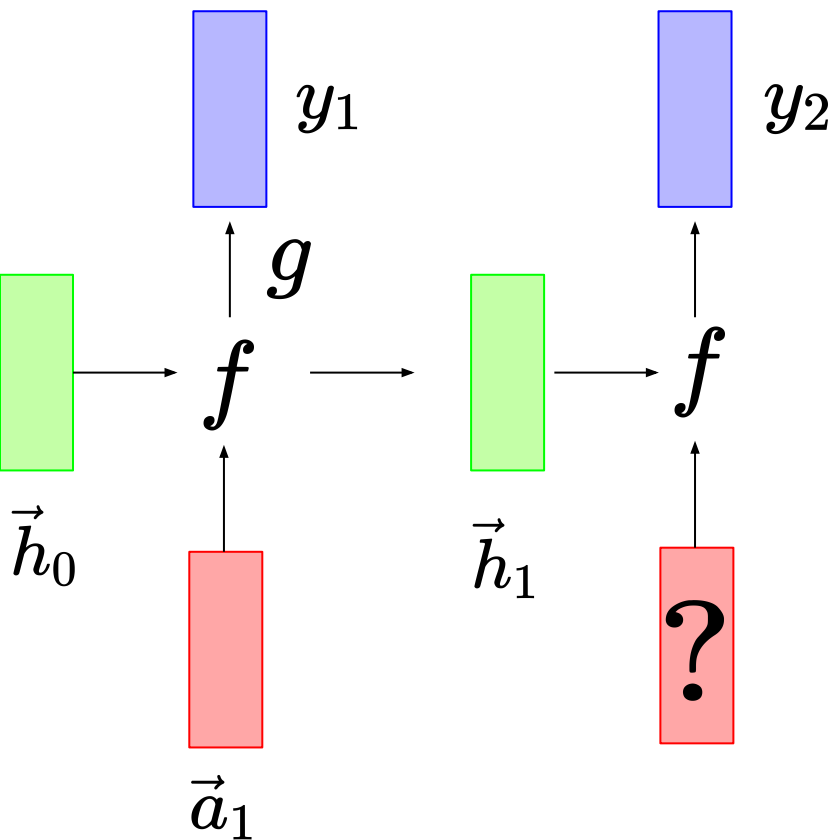


one to many

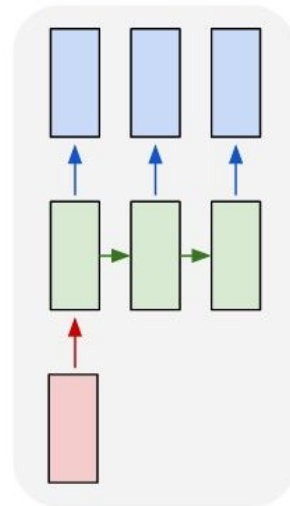


# RNNs

- Generate a sequence from a single input



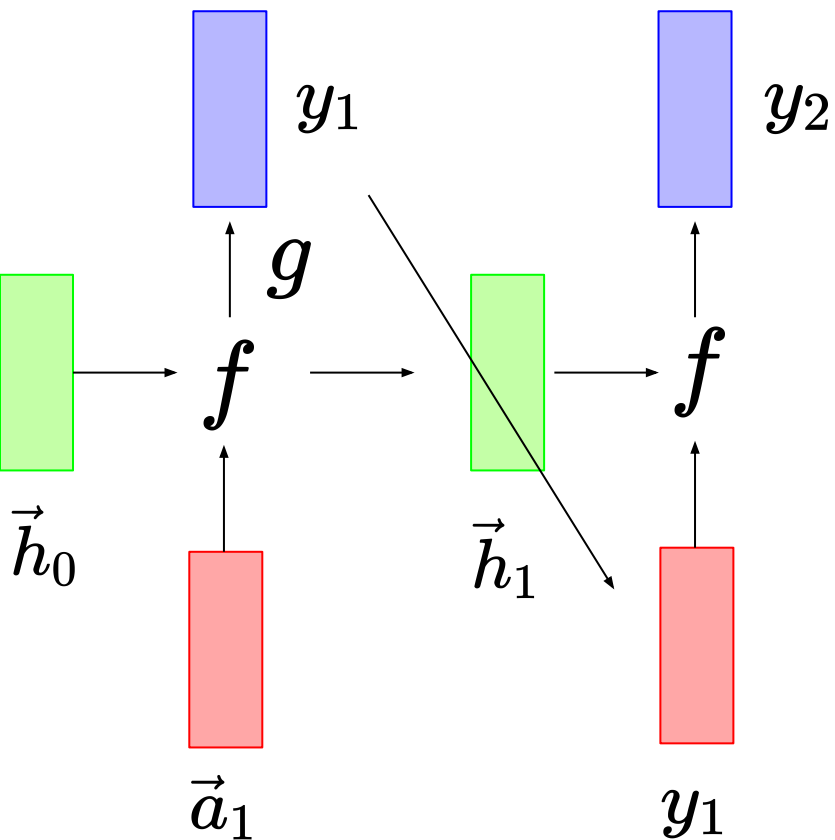
one to many



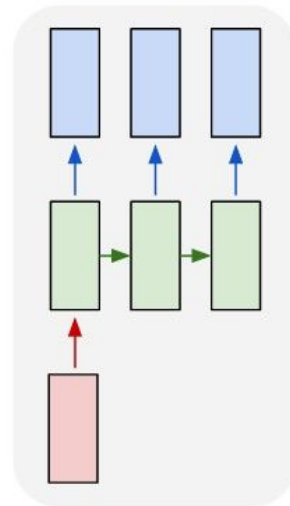


# RNNs

- Generate a sequence from a single input

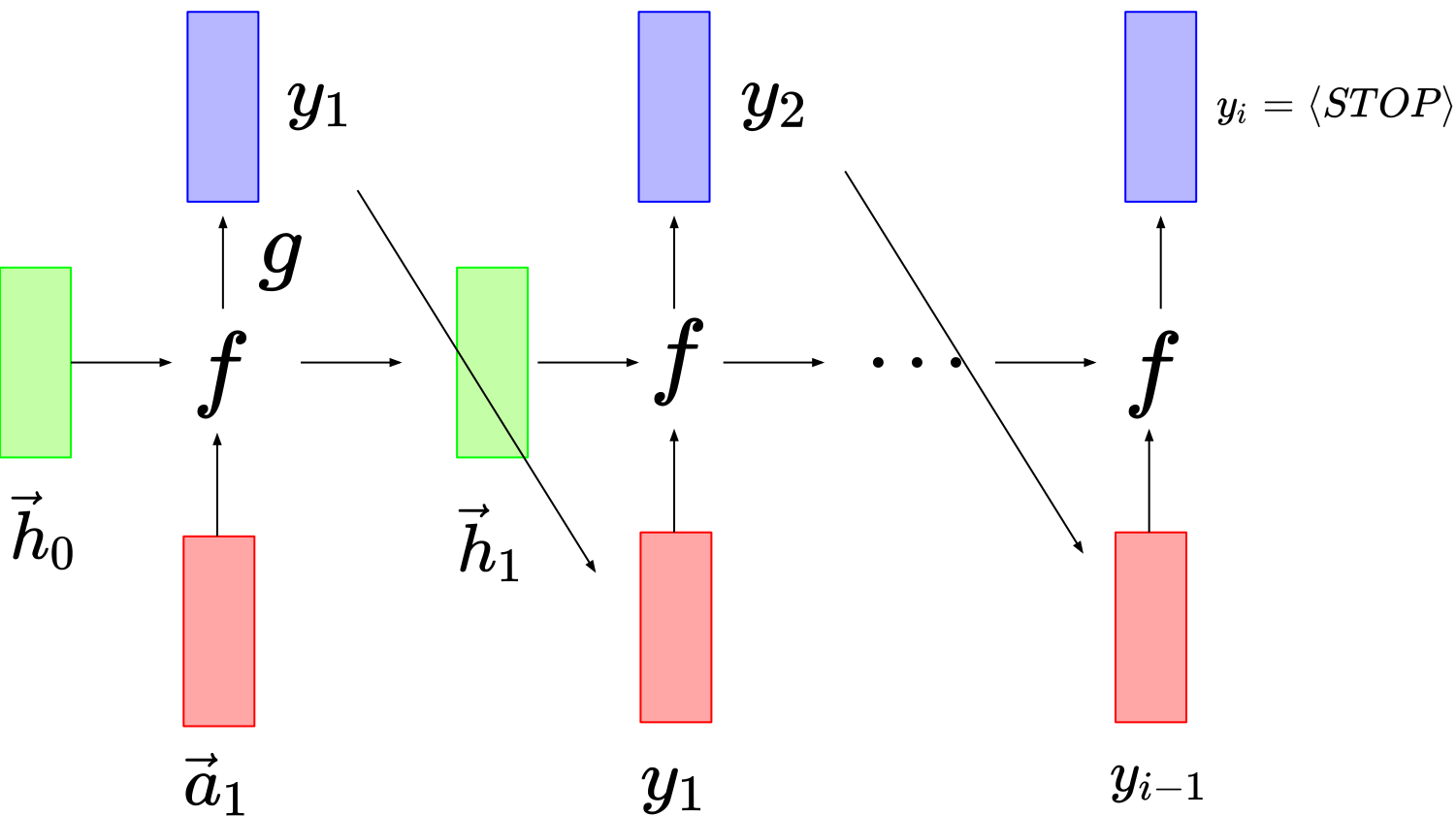


one to many

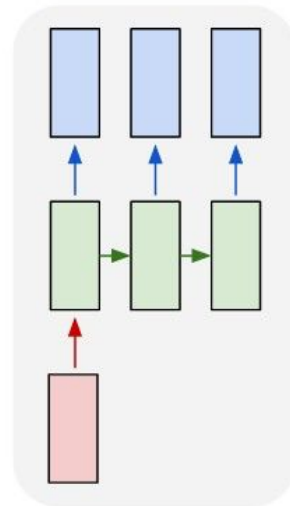


# RNNs

- Generate a sequence from a single input

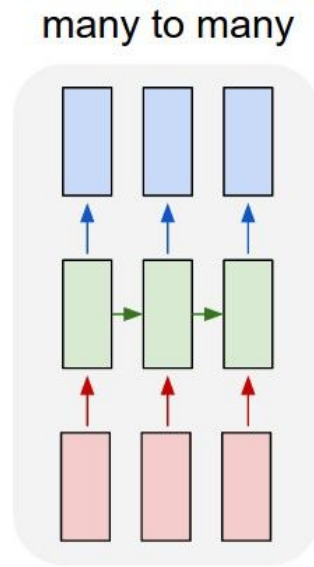


one to many



# Problems with this set-up

- Late parts of input sequence don't inform early predictions

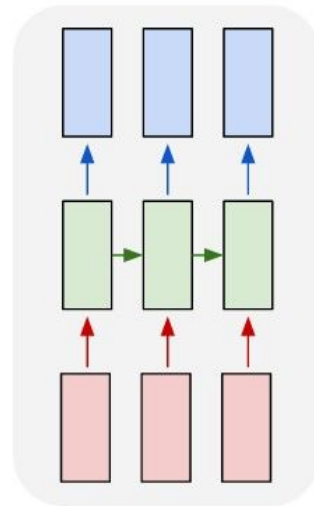


# Problems with this set-up

- Late parts of input sequence don't inform early predictions
- Problem in translation

Ich muss auf den Markt gehen. → I must go to the Market

many to many

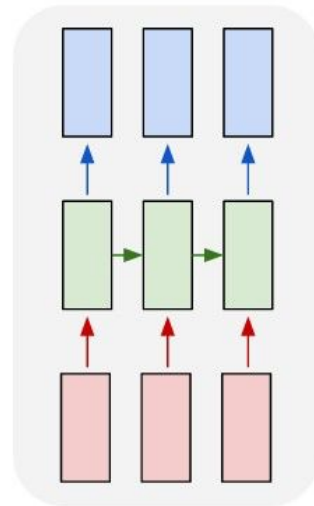


# Problems with this set-up

- Late parts of input sequence don't inform early predictions
- Problem in translation

Ich muss auf den Markt gehen.  $\longrightarrow$  I must go to the Market

many to many

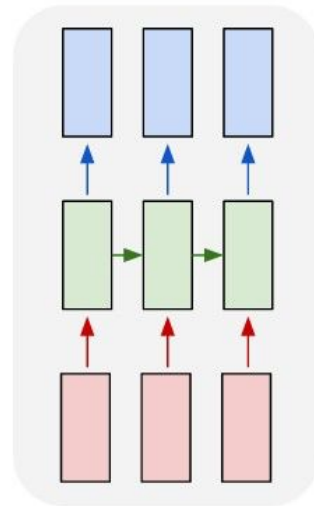


# Problems with this set-up

- Late parts of input sequence don't inform early predictions
- Problem in translation

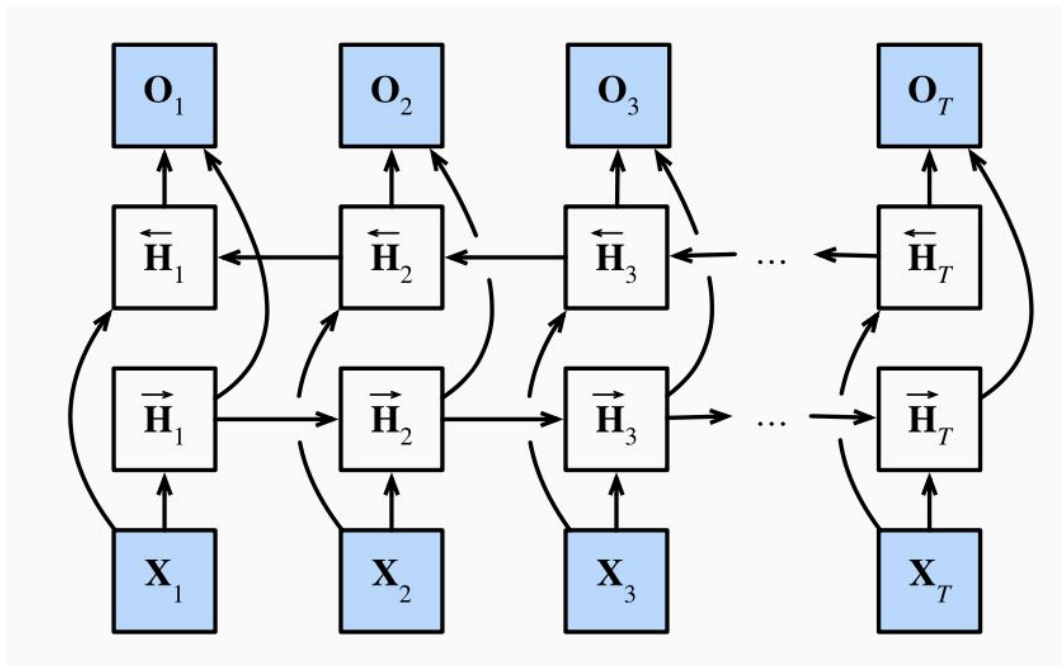
Ich muss auf den Markt gehen. → I must go to the Market  
?

many to many

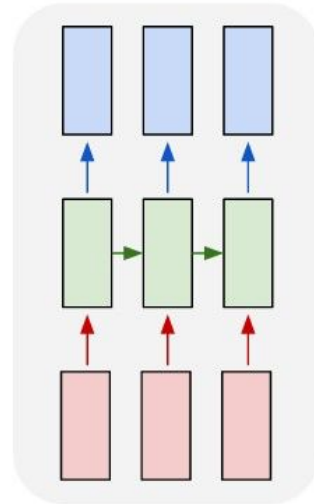


# Problems with this set-up

- Late parts of input sequence don't inform early predictions
- Problem in translation
- Bidirectional RNN



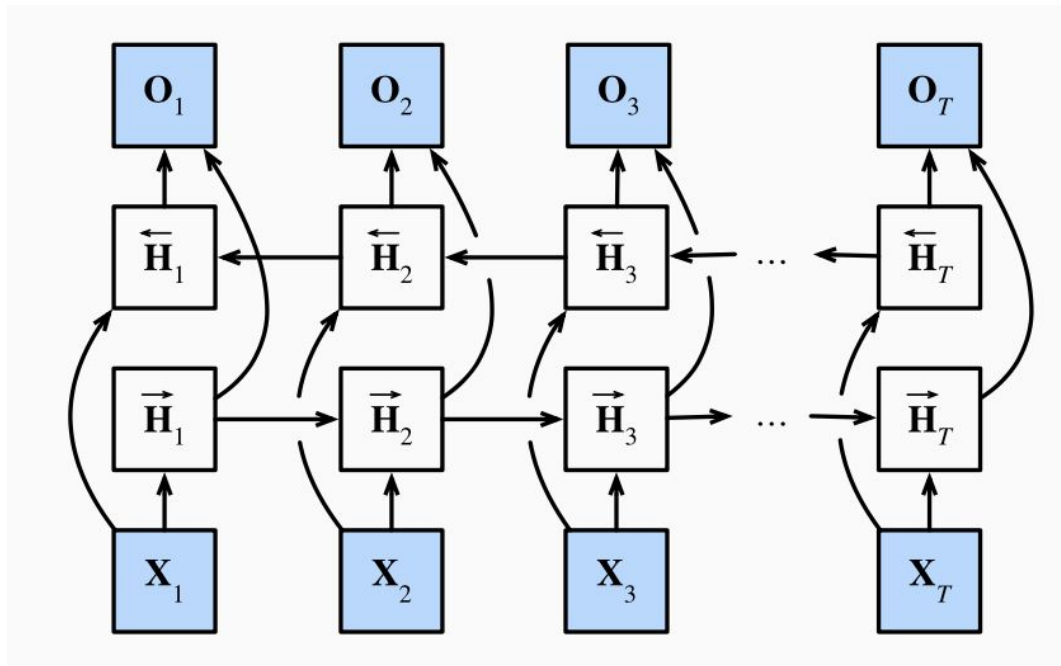
many to many



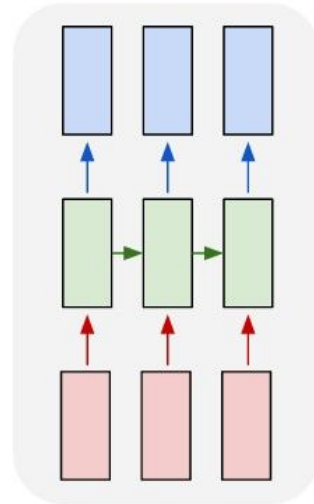
[Image source](#)

# Problems with this set-up

- Late parts of input sequence don't inform early predictions
- Problem in translation
- Bidirectional RNN
  - SLOW



many to many

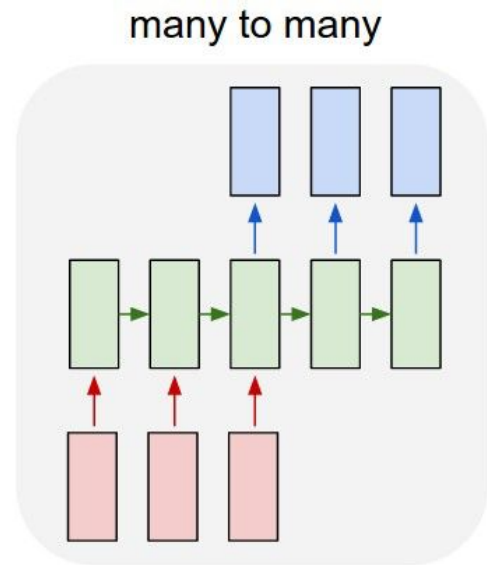


[Image source](#)



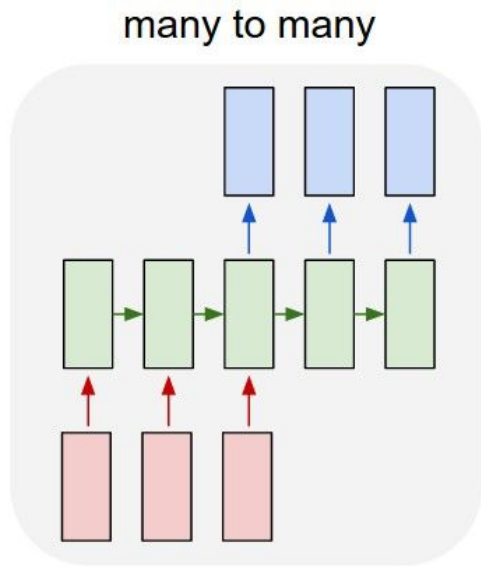
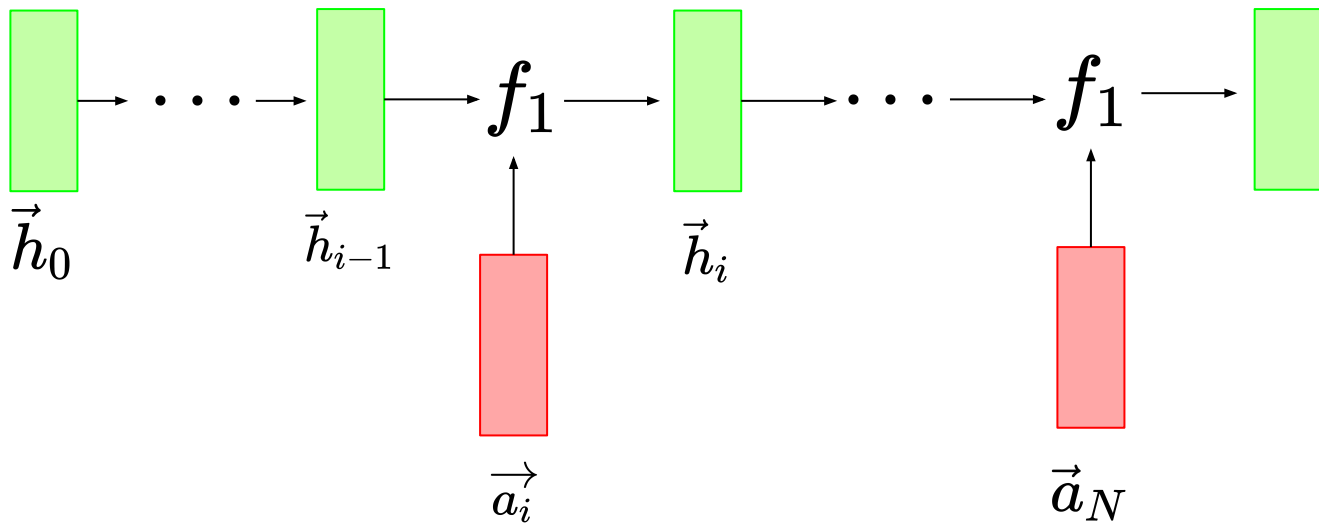
# Seq2Seq Models

- Generate a sequence using encoder/decoder framework
- Idea: Different RNNs for encoding vs. decoding



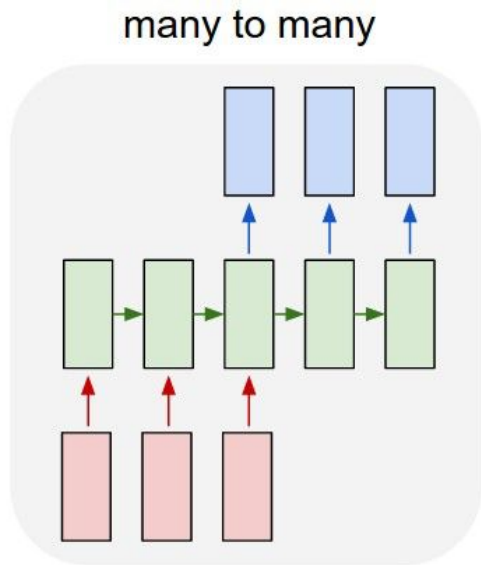
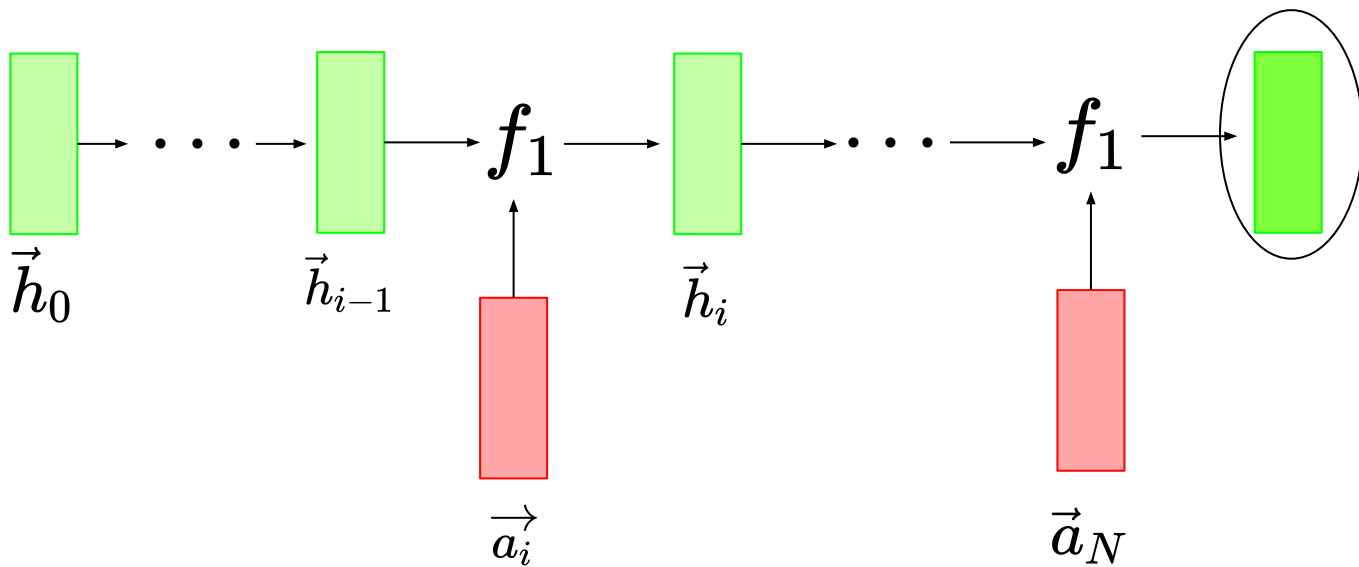
# Seq2Seq Models

- Generate a sequence using encoder/decoder framework
- Idea: Different RNNs for encoding vs. decoding



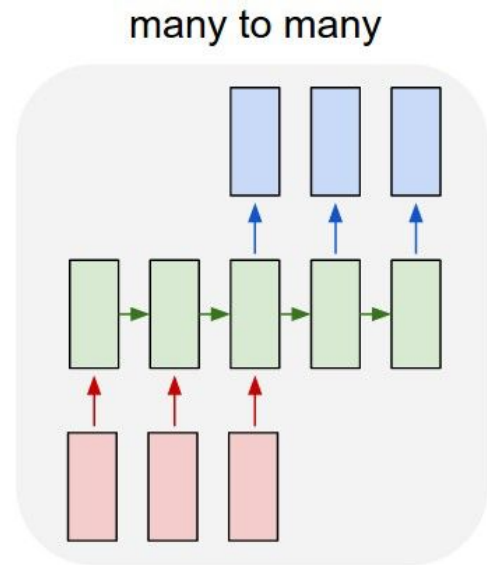
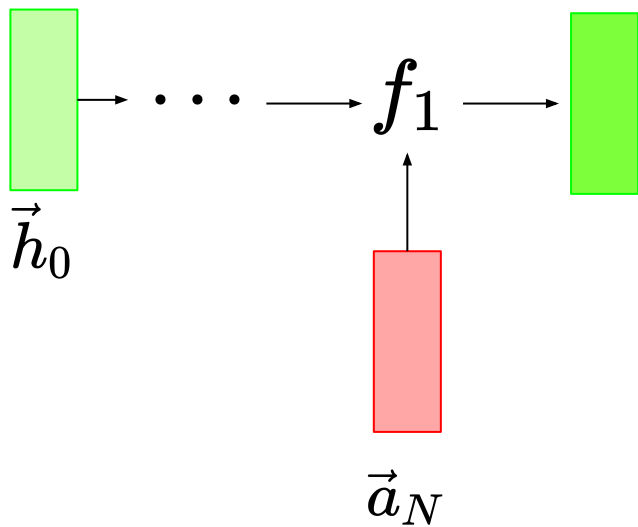
# Seq2Seq Models

- Generate a sequence using encoder/decoder framework
- Idea: Different RNNs for encoding vs. decoding



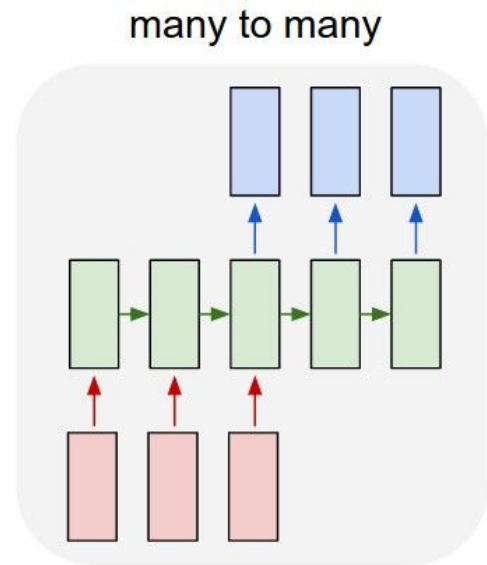
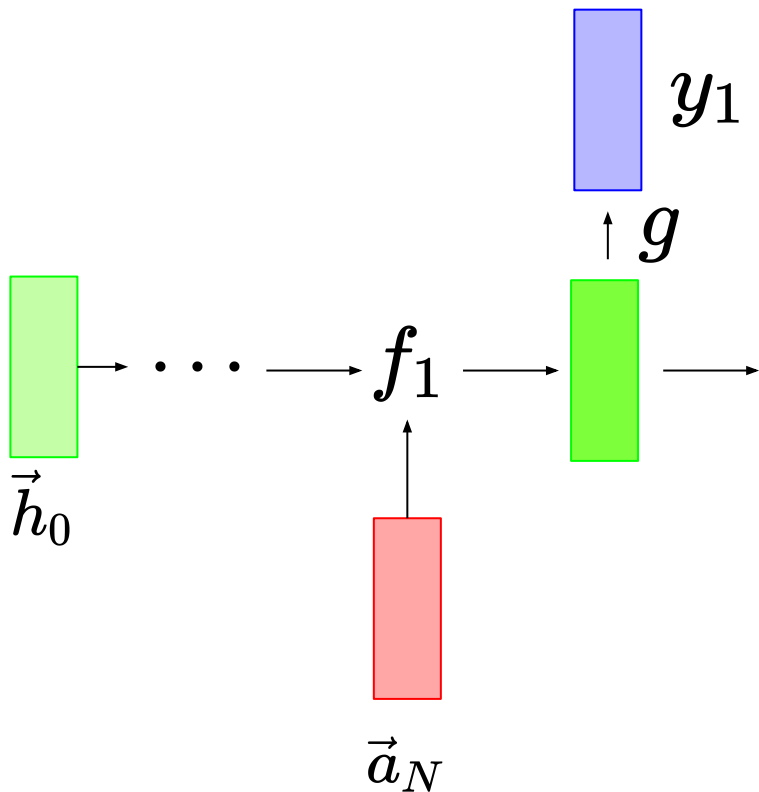
# Seq2Seq Models

- Generate a sequence using encoder/decoder framework
- Idea: Different RNNs for encoding vs. decoding



# Seq2Seq Models

- Generate a sequence using encoder/decoder framework
- Idea: Different RNNs for encoding vs. decoding



# Seq2Seq Models

- Generate a sequence using encoder/decoder framework
- Idea: Different RNNs for encoding vs. decoding

