CS 4824/ECE 4424: Recurrent Neural Networks

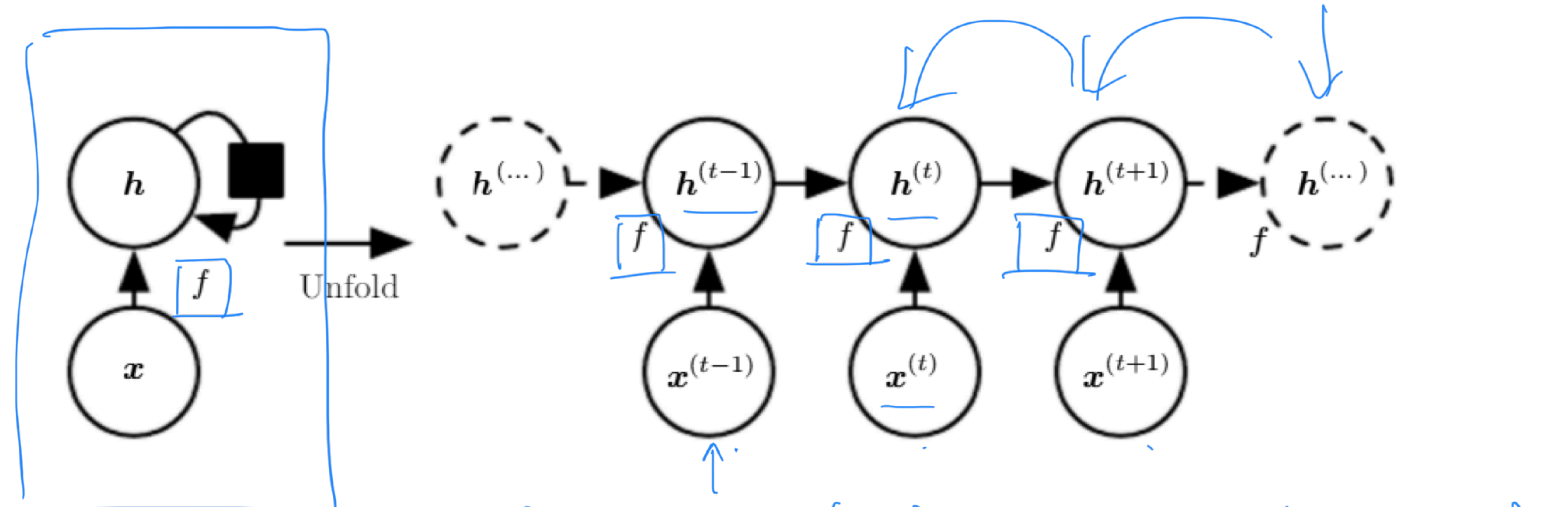
Acknowledgement:
Many of these slides are derived from Tom Mitchell, Pascal Poupart, Pieter Abbeel, Eric Eaton, Carlos Guestrin, William Cohen, and Andrew Moore.

Variable length data

- Traditional feed forward neural networks can only handle fixed length data
- Variable length data (e.g., sequences, time- series, spatial data) leads to a variable # of parameters
- Solutions:
  - Recurrent neural networks
  - Recursive neural networks

Recurrent Neural Network (RNN)

- In RNNs, outputs can be fed back to the network as inputs, creating a recurrent structure that can be unrolled to handle varying length data



$$h^{(t)} = f(h^{(t-1)}, x_t) = \sigma(W^{(h)} h^{(t-1)} + W^{(x)} x_t)$$

Training

- Recurrent neural networks are trained by backpropagation on the unrolled network
  - **backpropagation through time**

- Weight sharing:
  - Combine gradients of shared weights into a single gradient

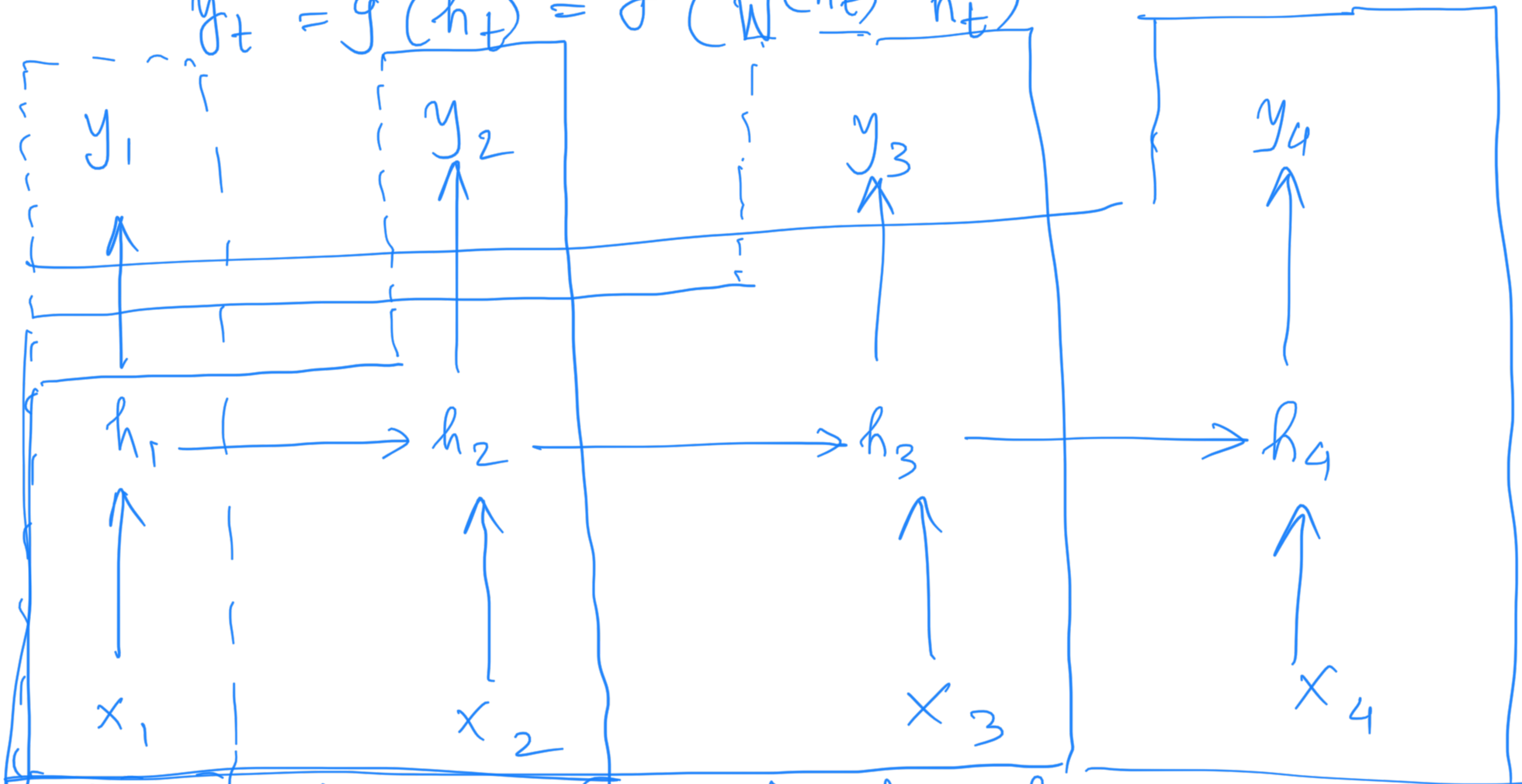
- Challenges
  - Gradient vanishing (and explosion)
  - Long range memory
  - Prediction drift

RNN for forward propagation

$$\boxed{h_t} = f(h_{t-1}, x_t) = \sigma(W^{h_{t-1}} h_{t-1} + W^{x_t} x_t)$$

- The inputs enter and move forward at each time step

$$y_t = g(h_t) = \sigma(W^{(h_t)'} h_t)$$



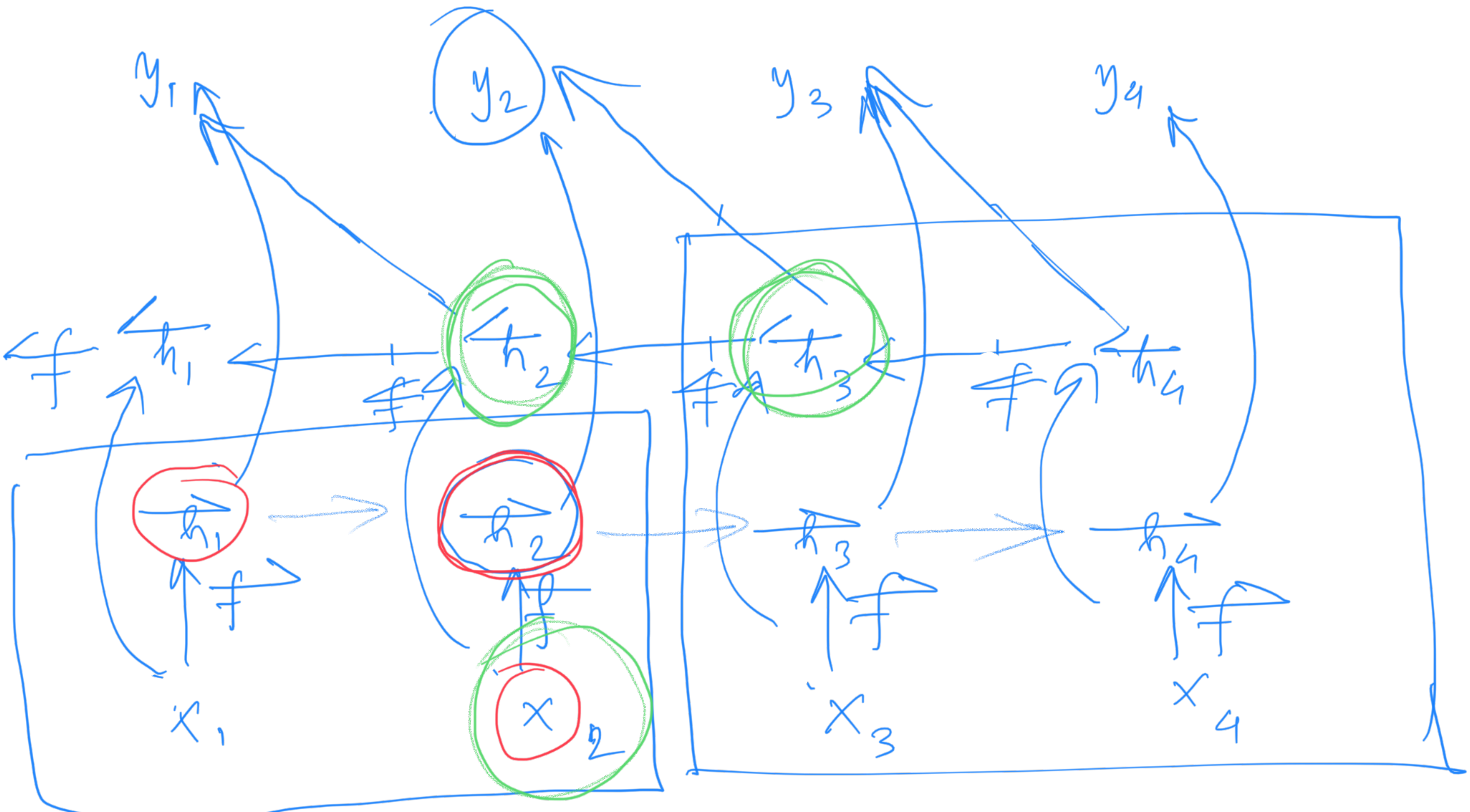
edges indicate functional dependencies

Limitation of RNN

- The inputs enter and **ONLY move forward** at each time step
- In some application, we would like to combine past and future evidence, i.e. perform backward propagation.

Bi-Directional RNN (Bi-RNN)

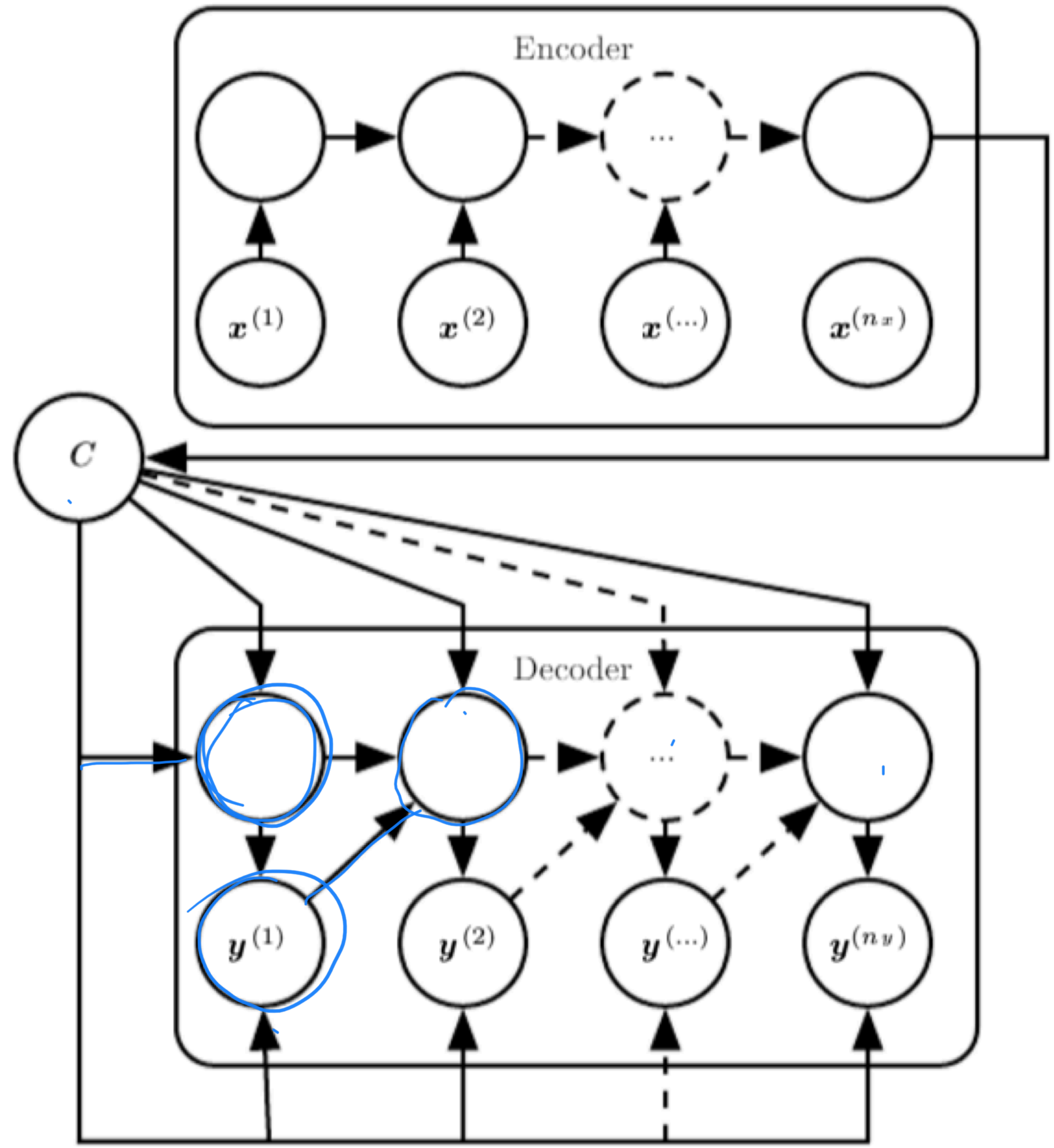
- We can combine past and future evidence in separate chains



Encoder-Decoder Model

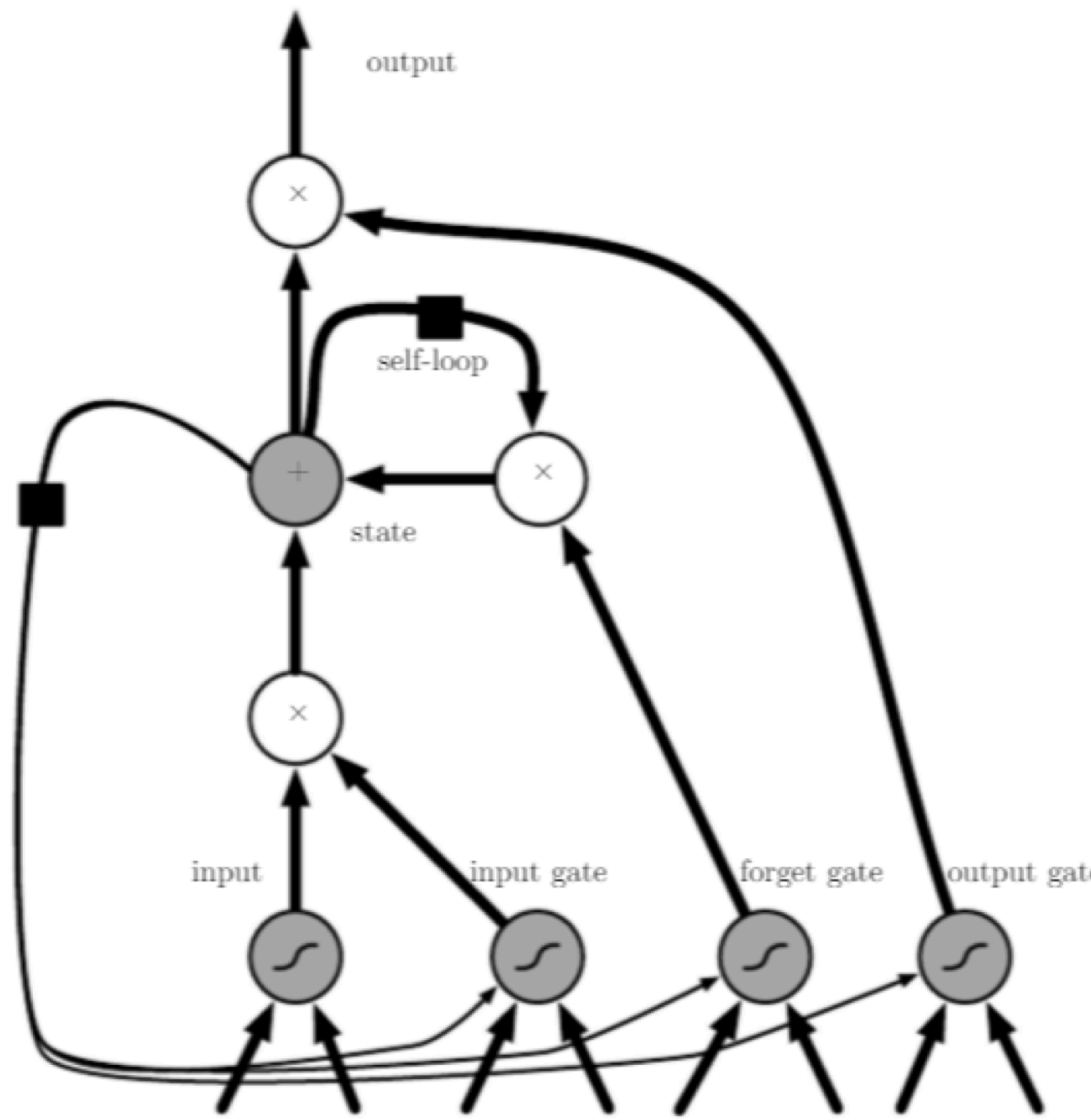
- Also known as sequence2sequene
  - $X^{(i)}$: i^{th} input
  - $y^{(i)}$: i^{th} output
  - c: context (embedding)

- Usage:
  - Machine translation
  - Question answering
  - Dialog



Long Short Term Memory (LSTM)

- Special gated structure to control memorization and forgetting in RNNs
- Facilitate long term memory



Unrolled LSTM

- Schematic