Recurrent Neural Networks

Hanna Bugler
Biomedical Engineering PhD Candidate
AI^2 Lab and Harris Lab

Winter 2025

UNIVERSITY OF CALGARY

Outline

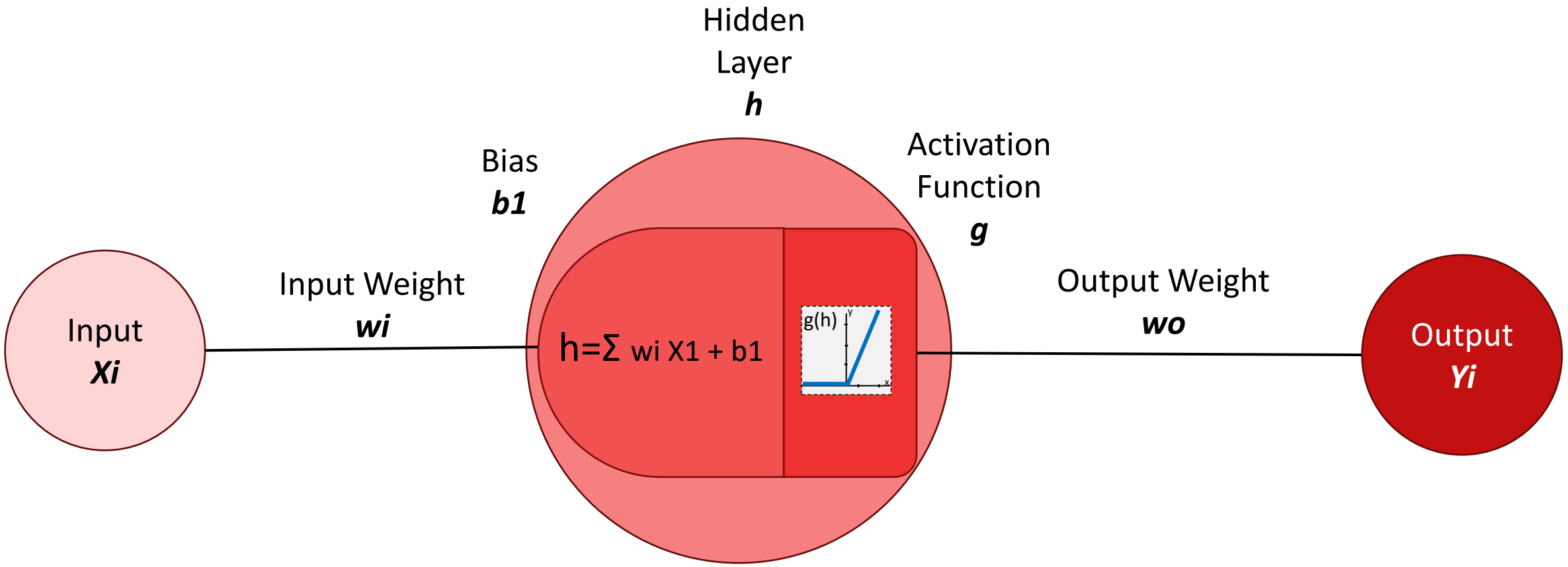
- Introduction to recurrent neural networks (RNNs)
- Challenges associated with RNNs
- Architectures derived from RNNs

Learning Goals

- Understand how recurrent neural networks (RNN) differ from traditional neural networks to solve data series tasks

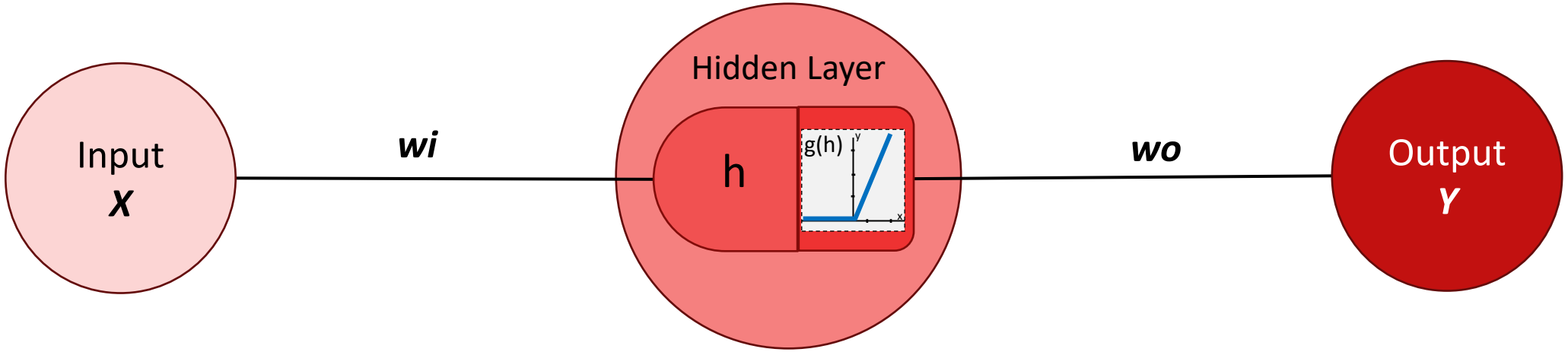
- Understand the challenges associated with simple RNNs and which architectural changes can improve performance

Traditional Neural Networks



UNIVERSITY OF CALGARY

Traditional Neural Networks



Vs.

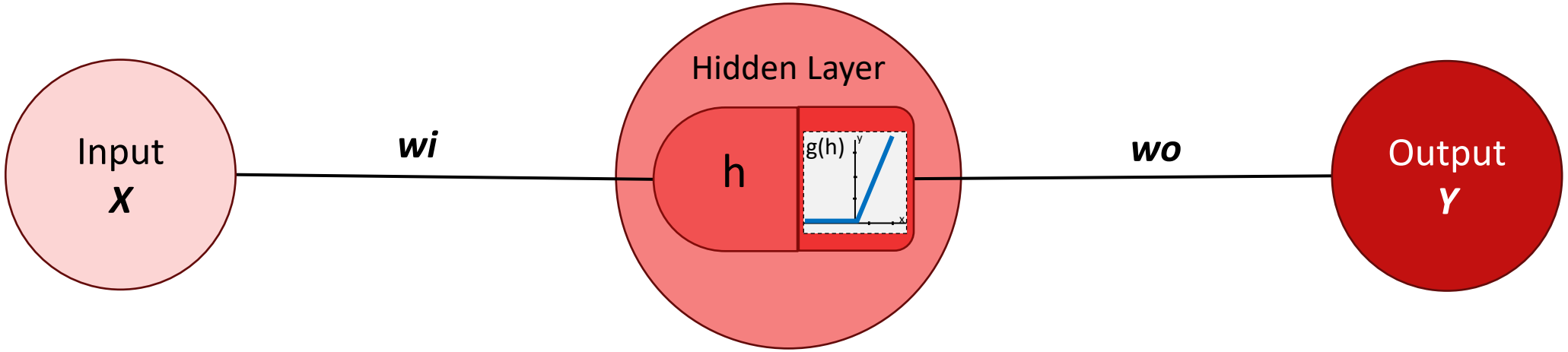
Last Game's Score: 4-3

Predicting which team will win next game

Calgary Wins

UNIVERSITY OF CALGARY

Traditional Neural Networks



Vs.

Previous Game's Score: 1-2

Vs.

Last Game's Score: 4-3

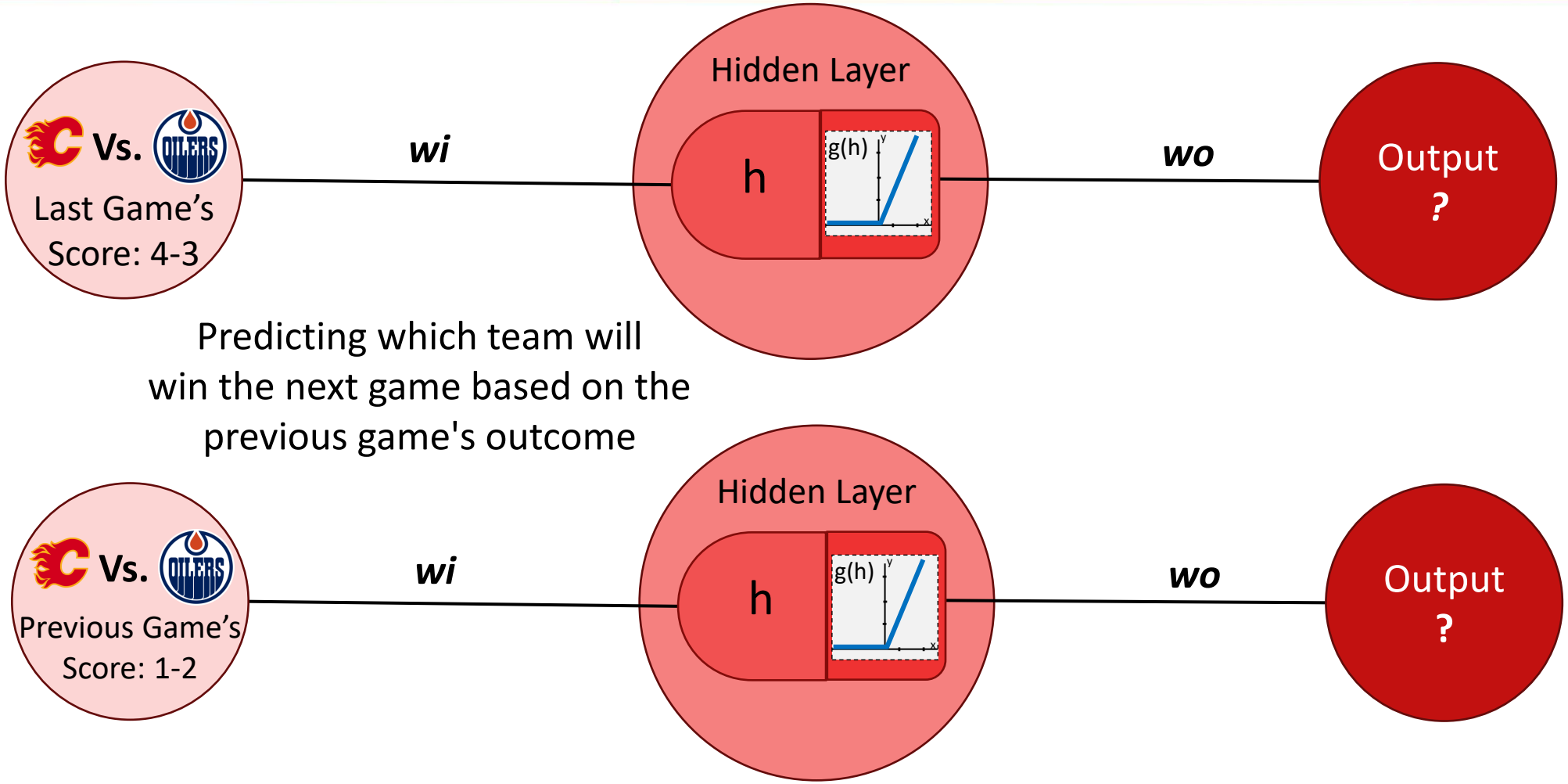
Predicting which team will win next game

?

What happens when the number of inputs changes?

UNIVERSITY OF CALGARY

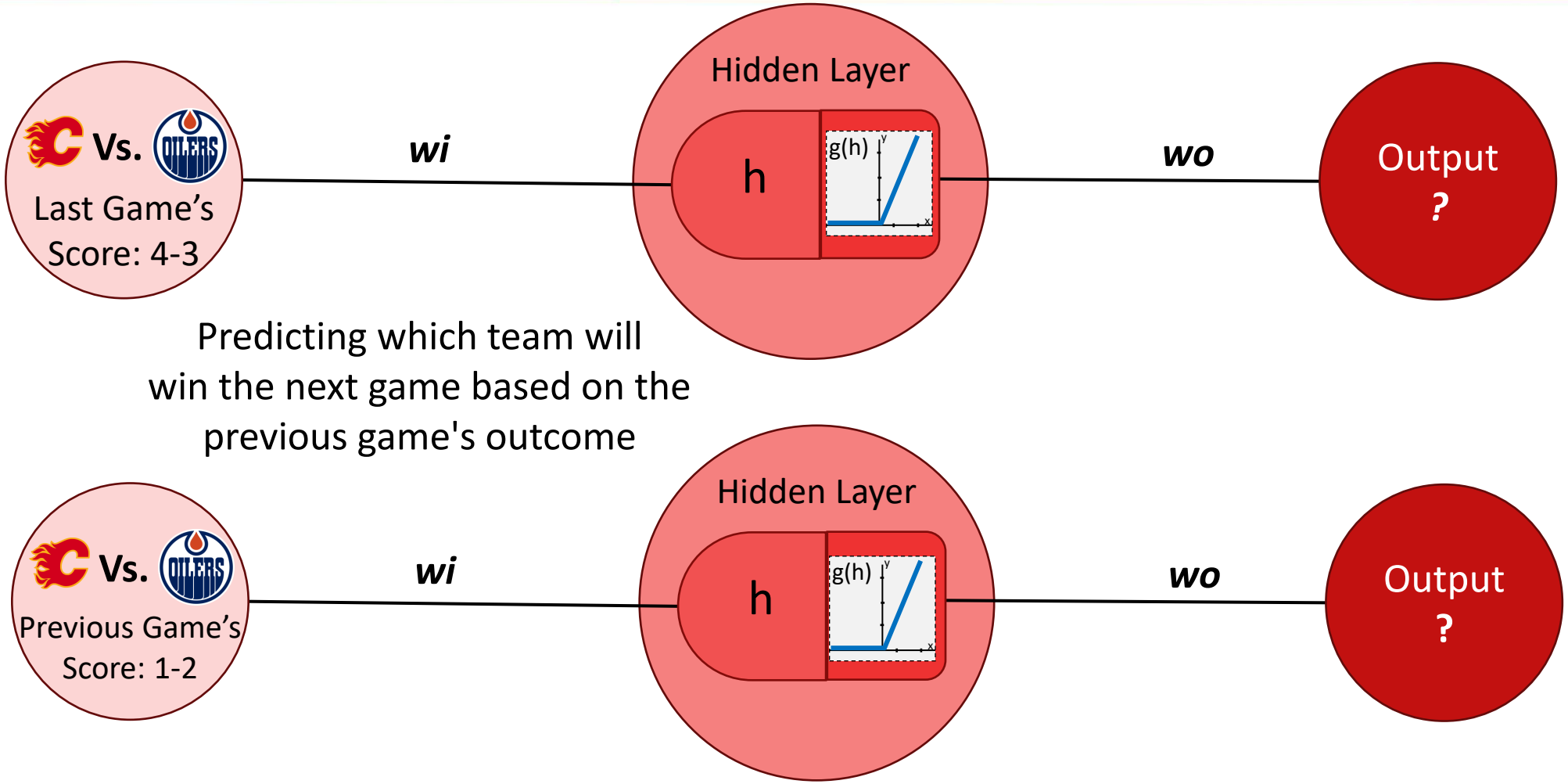
Traditional Neural Networks



What happens when we feed inputs individually through the network?

UNIVERSITY OF CALGARY

Traditional Neural Networks

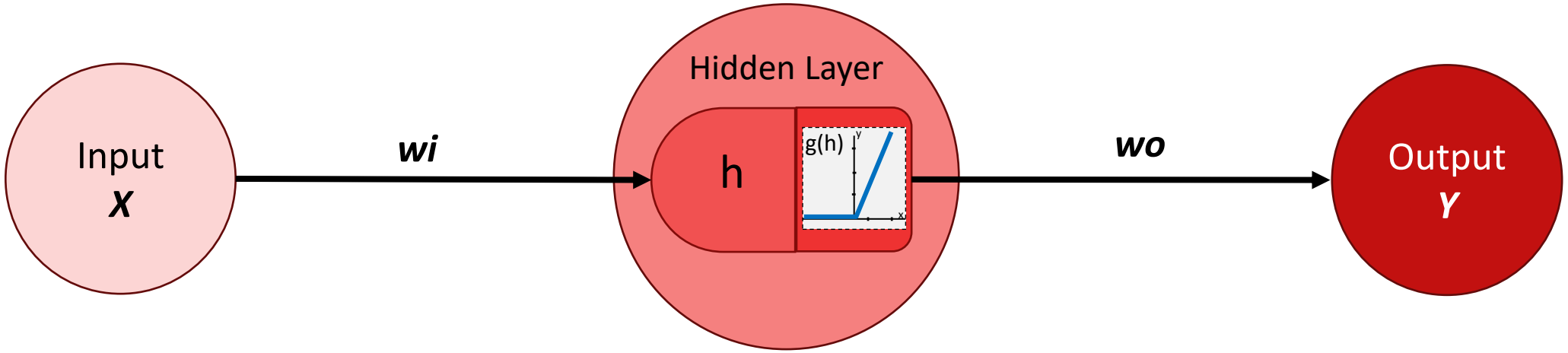


Traditional Neural Networks are not equipped to deal with variable input sizes

UNIVERSITY OF CALGARY

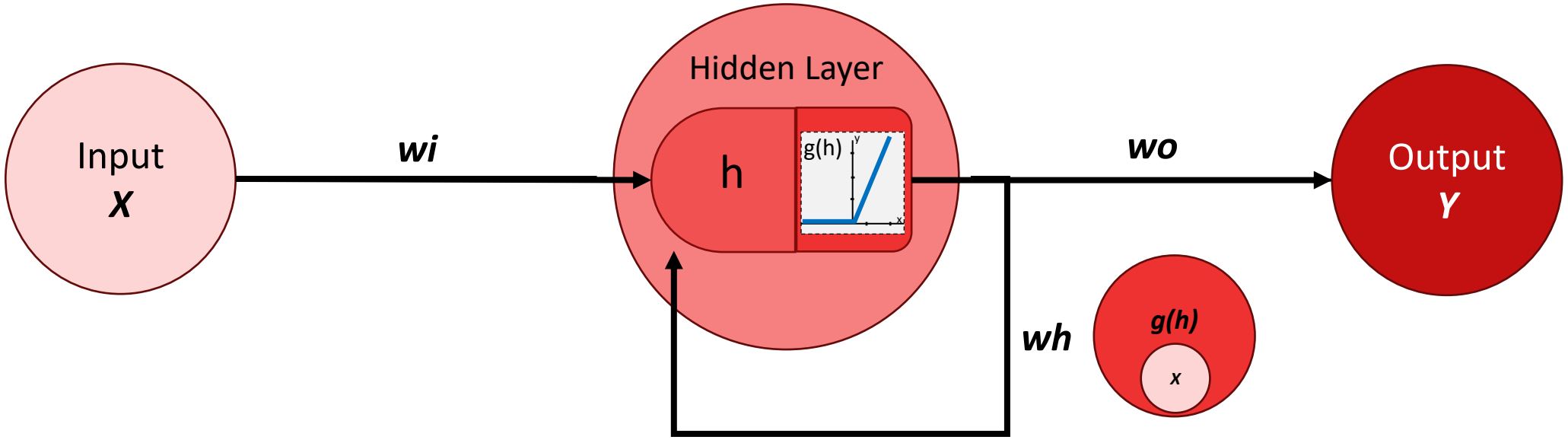
Introduction to Recurrent Neural Networks

Recurrent Neural Networks



UNIVERSITY OF CALGARY

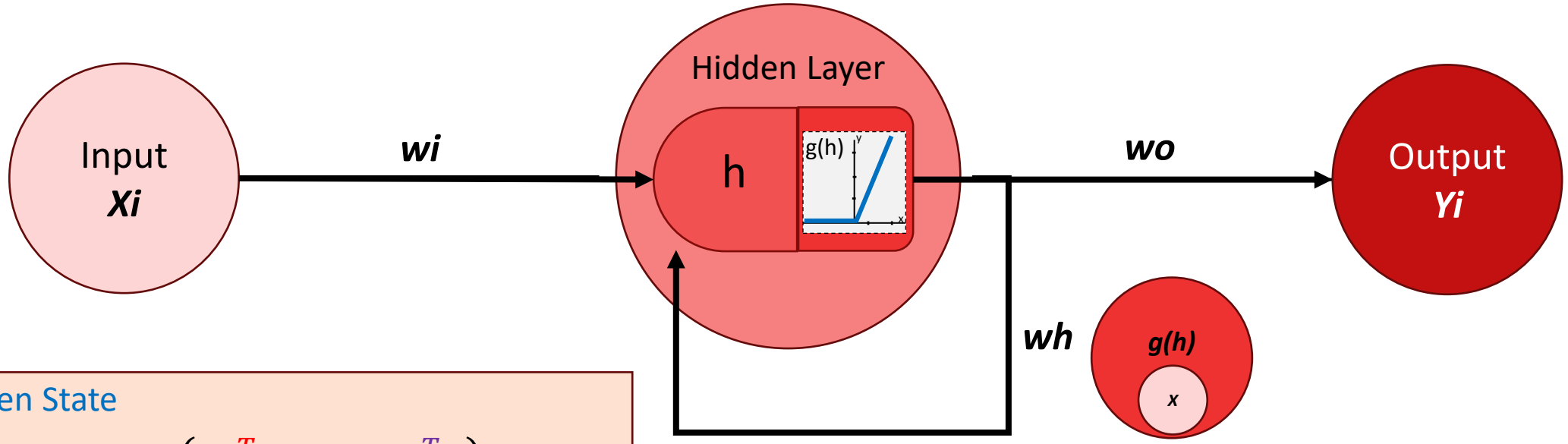
Recurrent Neural Networks - Feedback



Feedback

UNIVERSITY OF CALGARY

Recurrent Neural Networks - Feedback



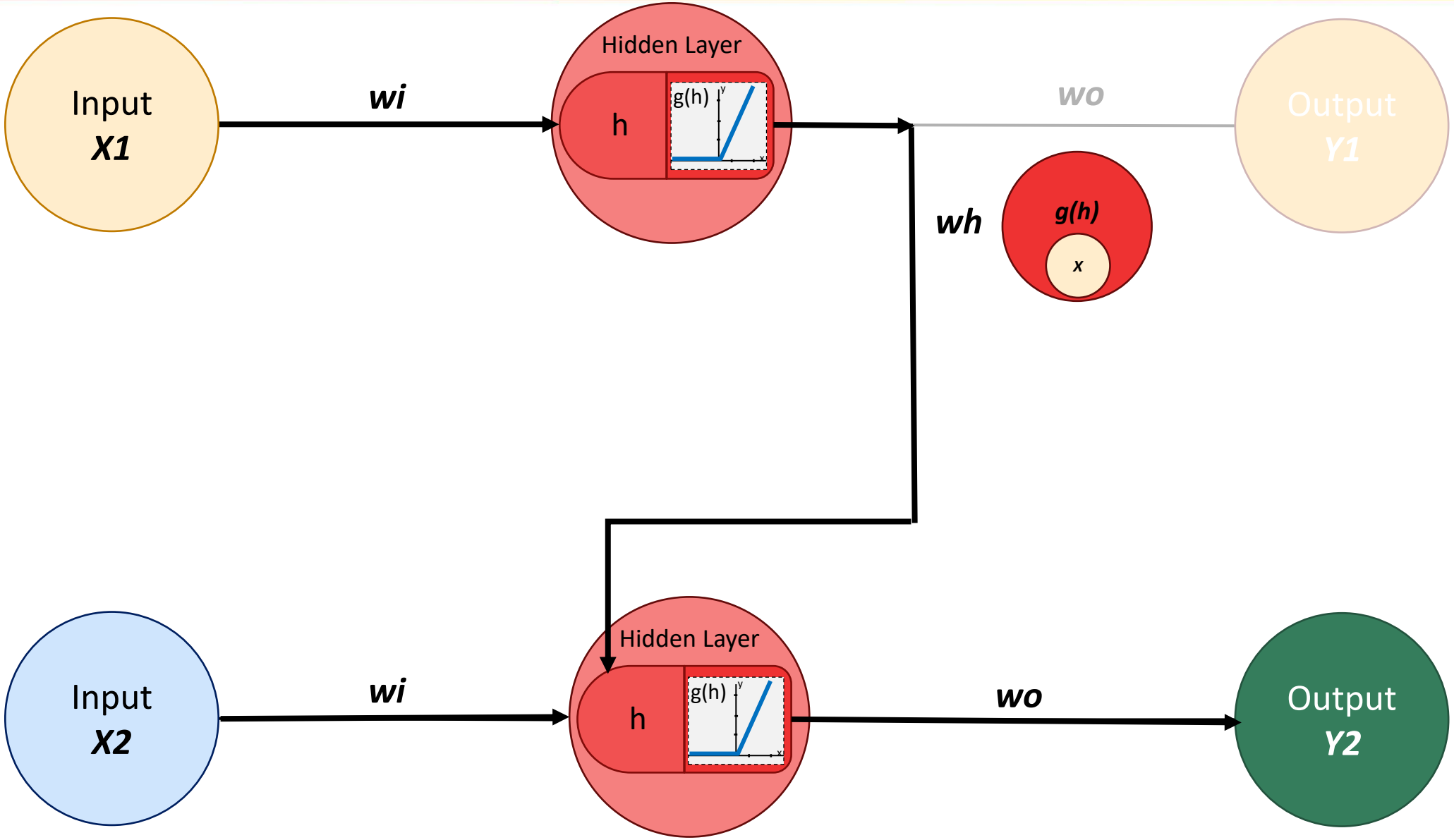
Hidden State

$$h_t = g\left(W_h^T h_{i-1} + W_i^T x\right)$$

Feedback Weight — Input Weight

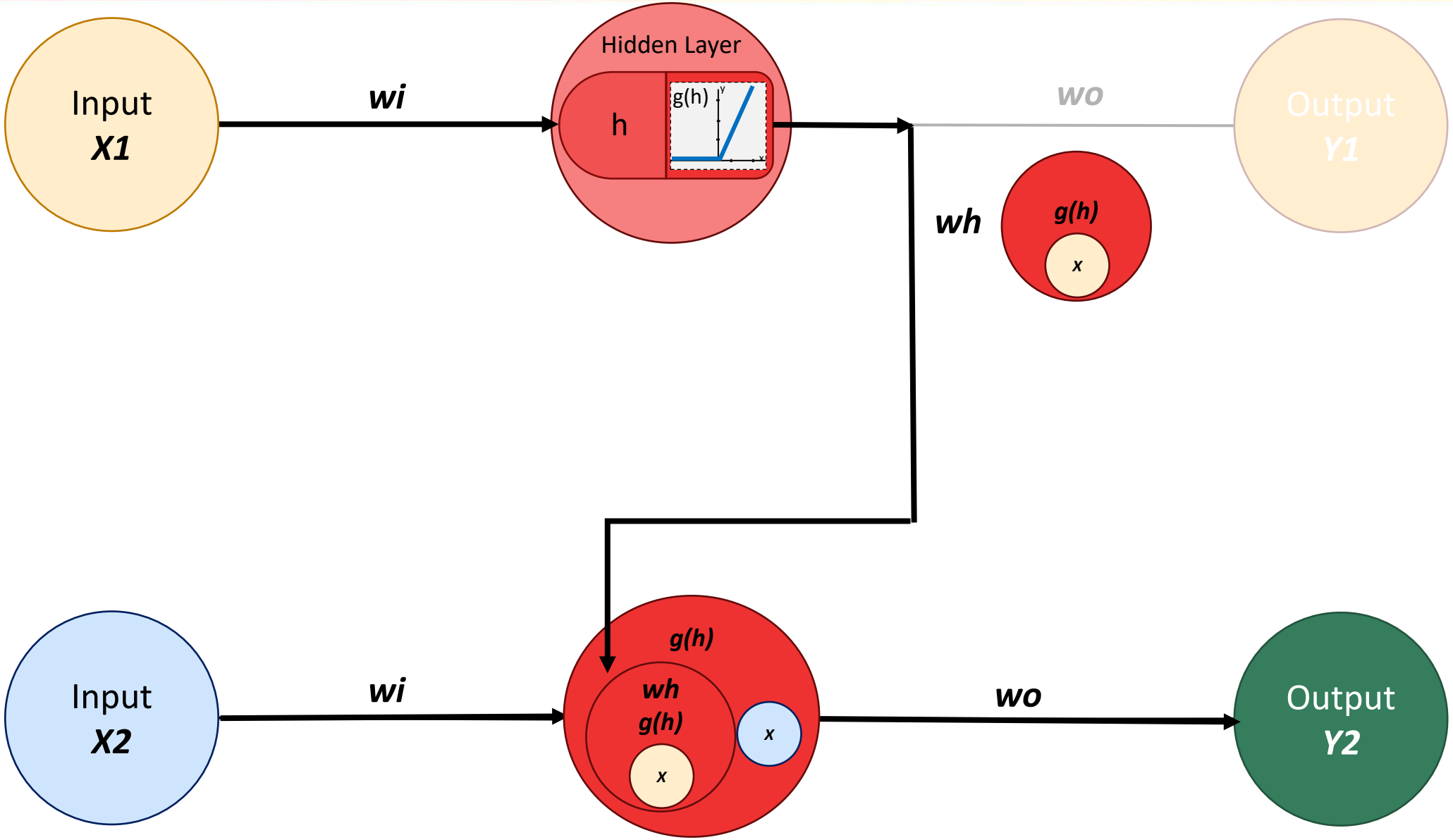
Previous Hidden State — Input Value

Recurrent Neural Networks - Unrolling



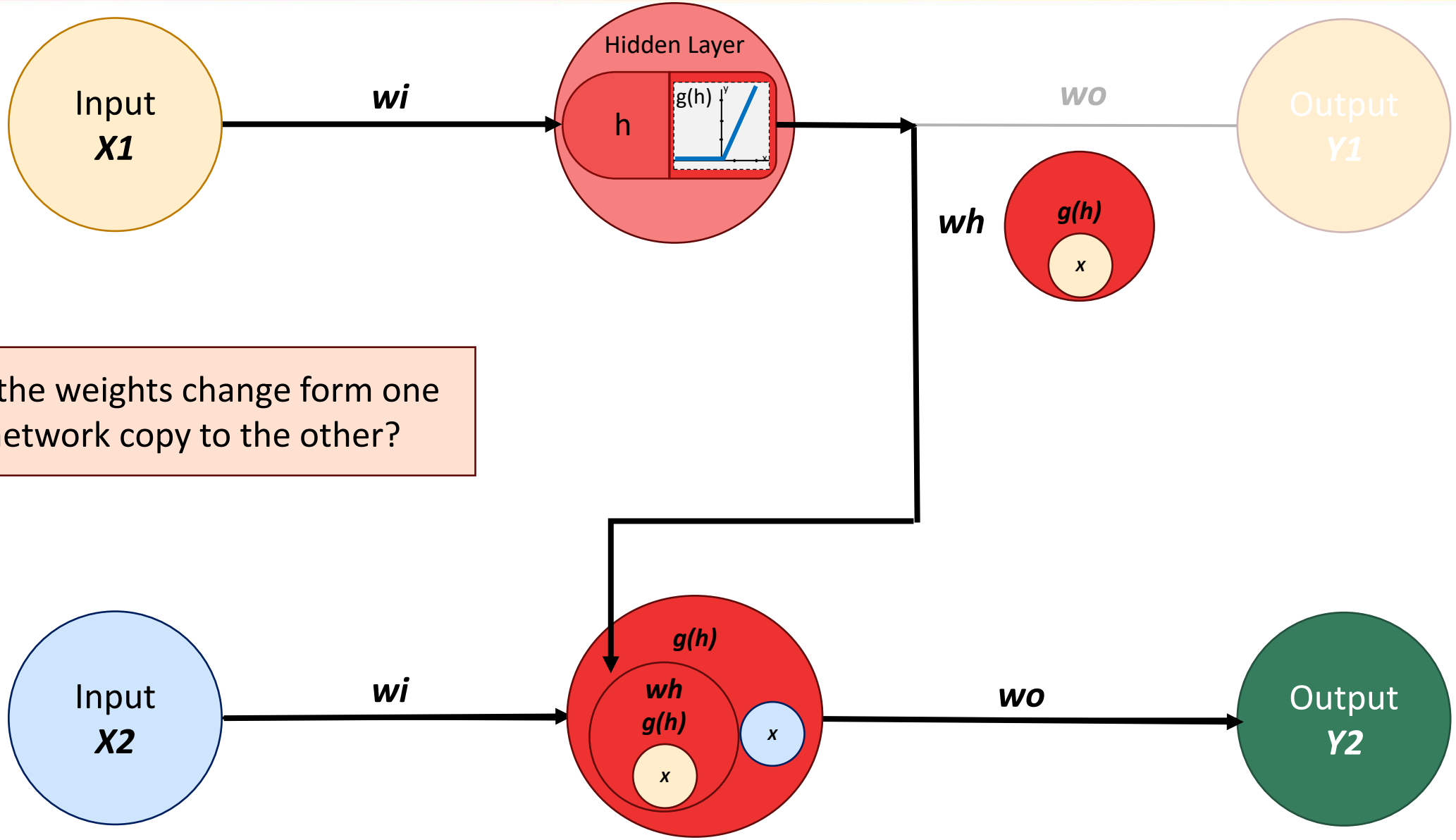
UNIVERSITY OF CALGARY

Recurrent Neural Networks - Unrolling



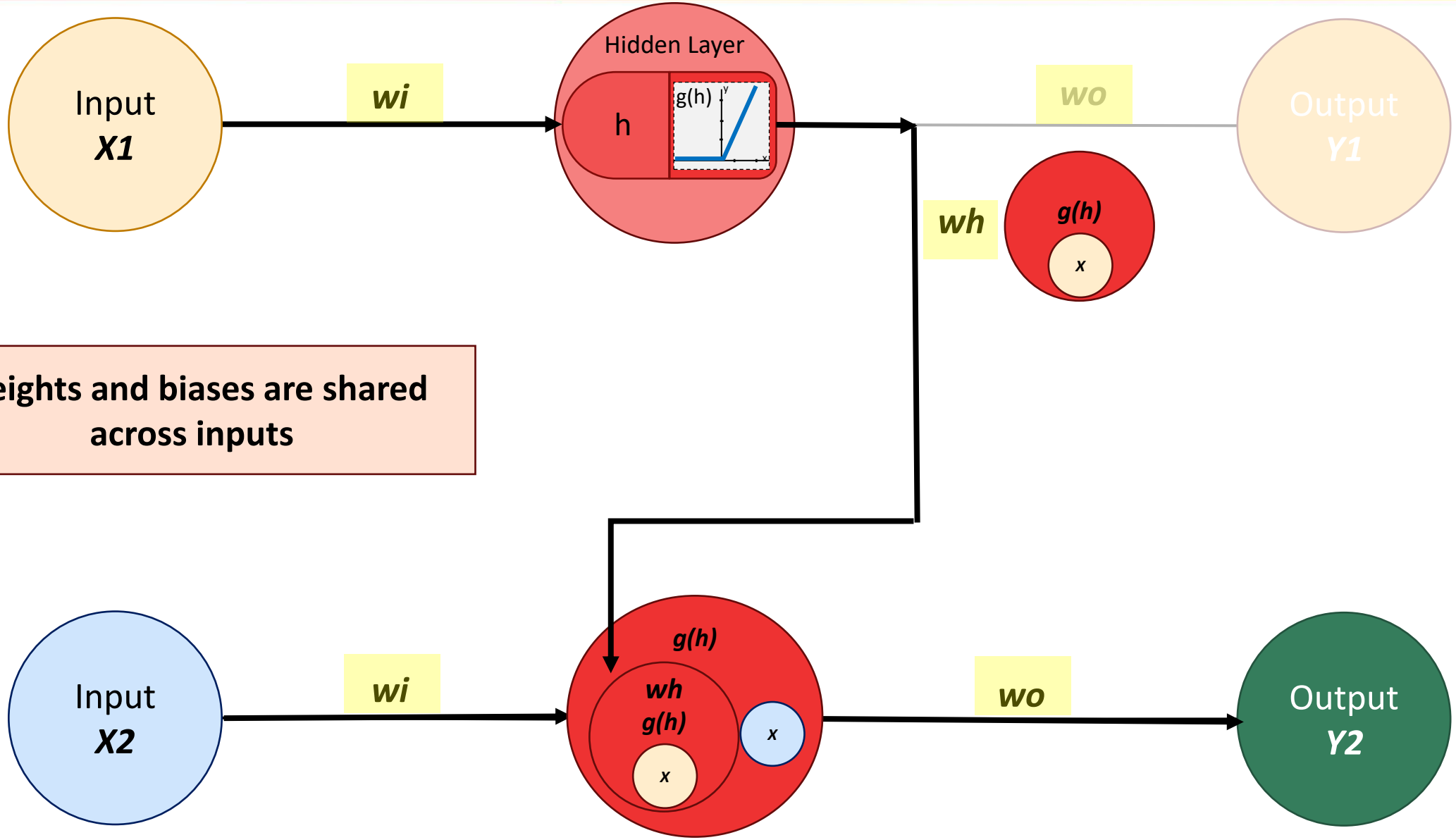
UNIVERSITY OF CALGARY

Recurrent Neural Networks - Unrolling



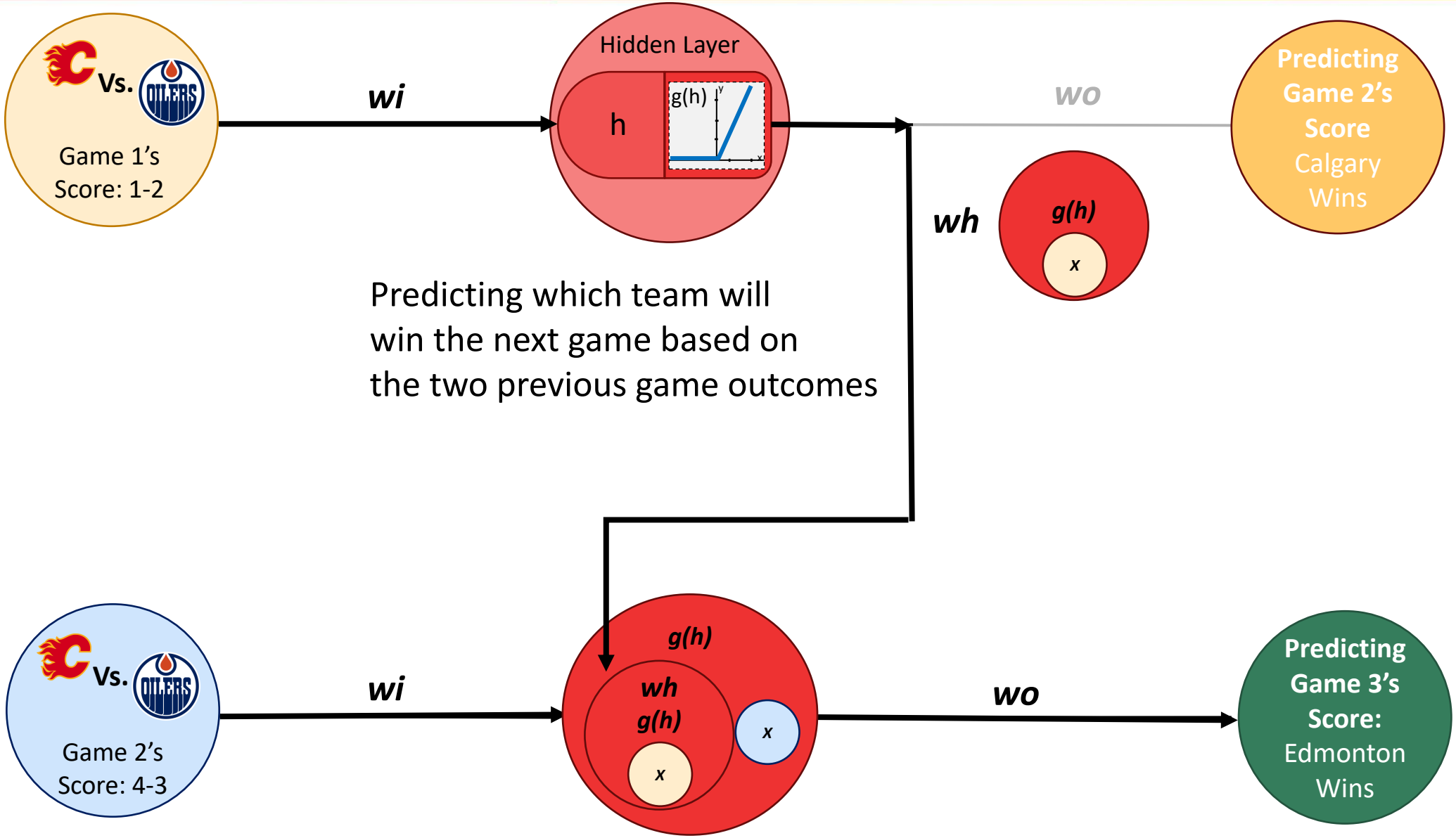
UNIVERSITY OF CALGARY

Recurrent Neural Networks - Unrolling



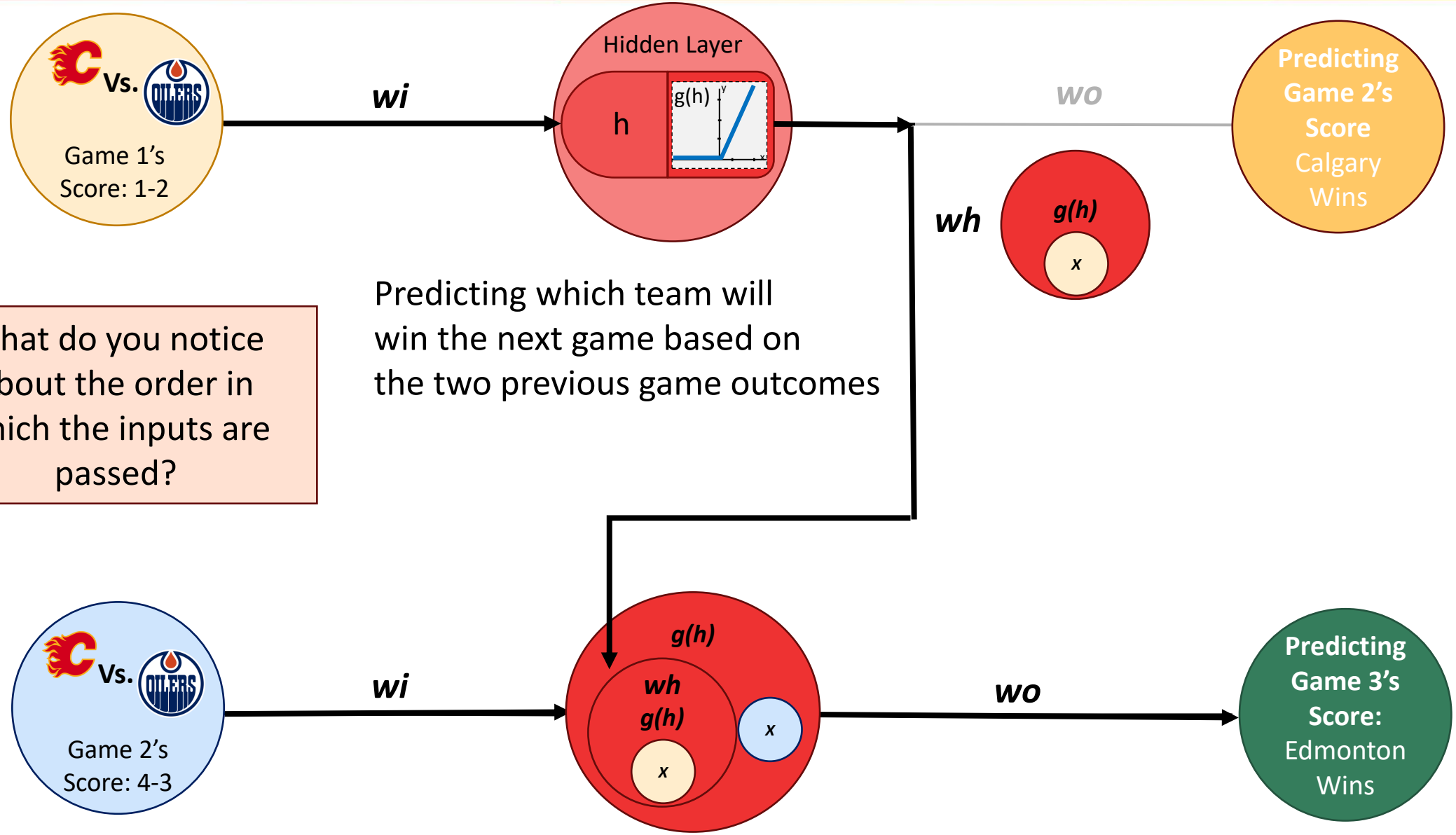
UNIVERSITY OF CALGARY

Recurrent Neural Networks - Unrolling



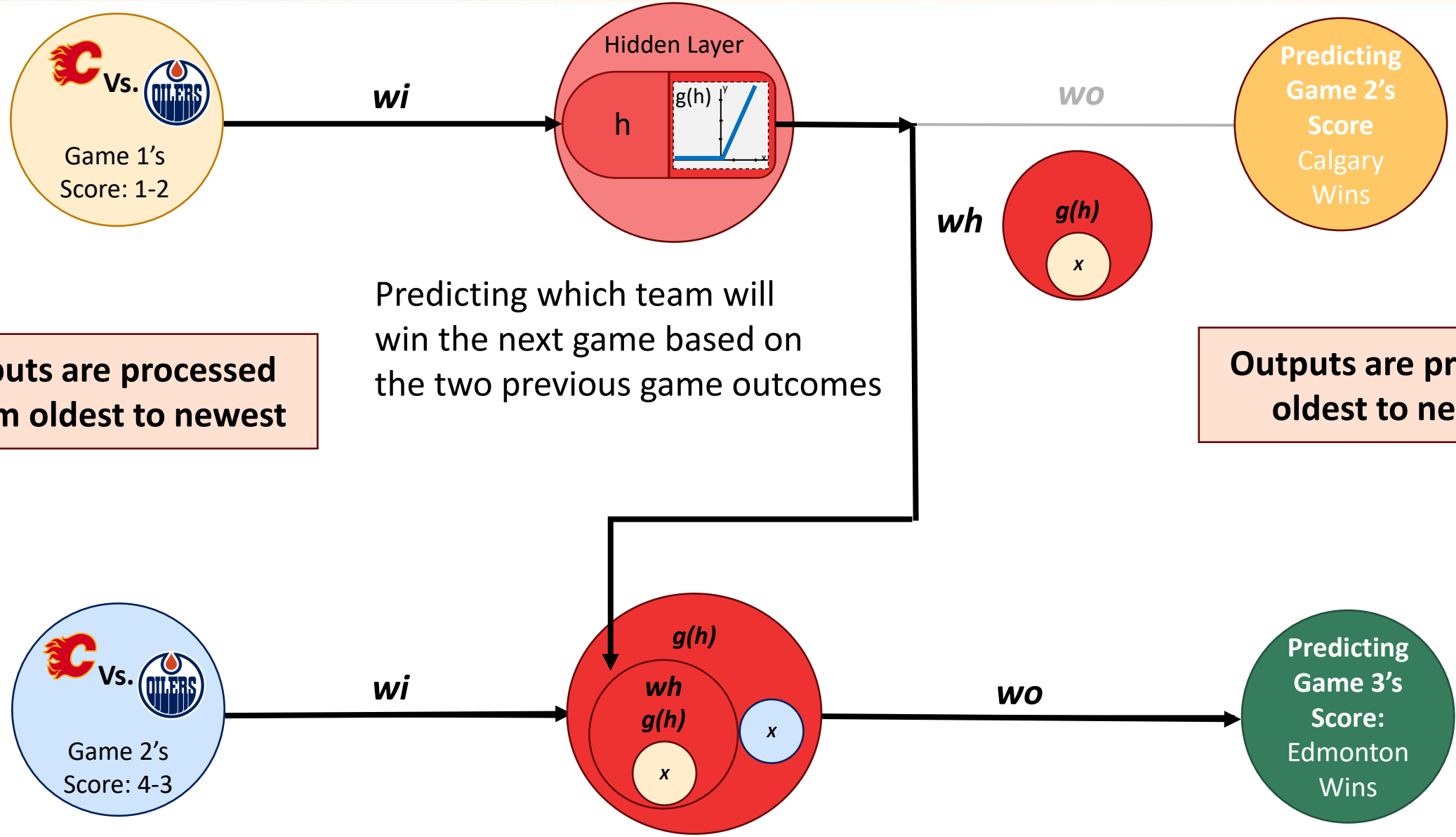
UNIVERSITY OF CALGARY

Recurrent Neural Networks - Unrolling



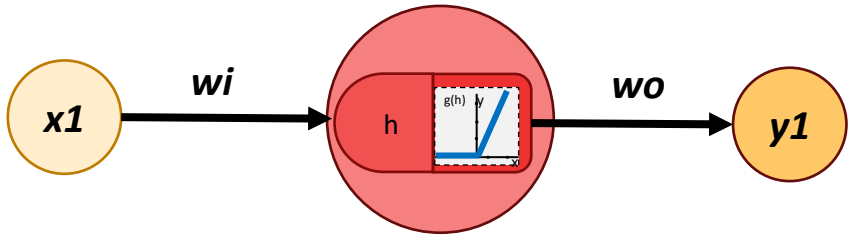
UNIVERSITY OF CALGARY

Recurrent Neural Networks - Unrolling

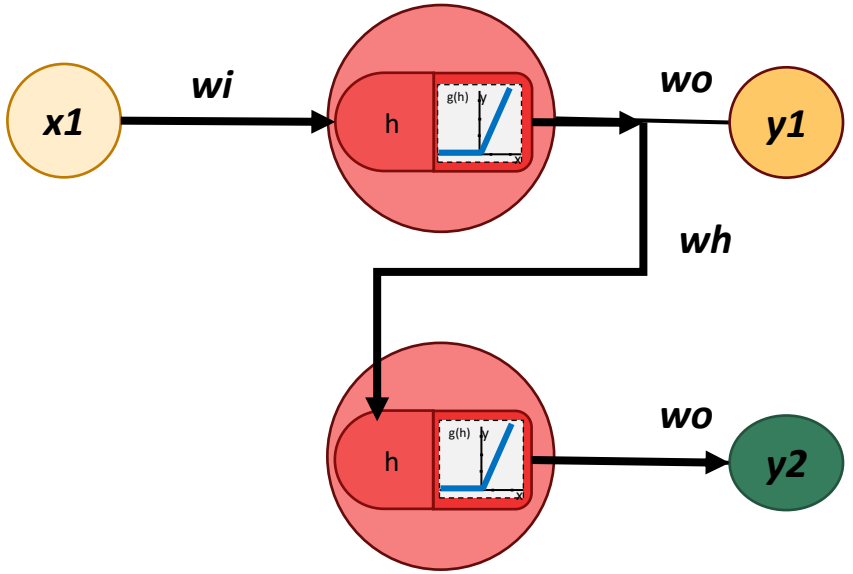


UNIVERSITY OF CALGARY

Recurrent Neural Networks – I/O Properties



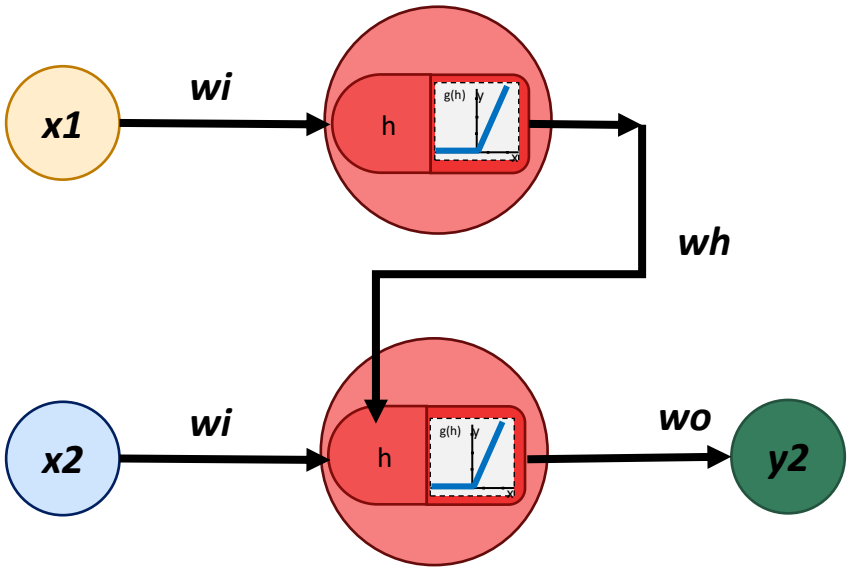
One [input] **to One** [output]



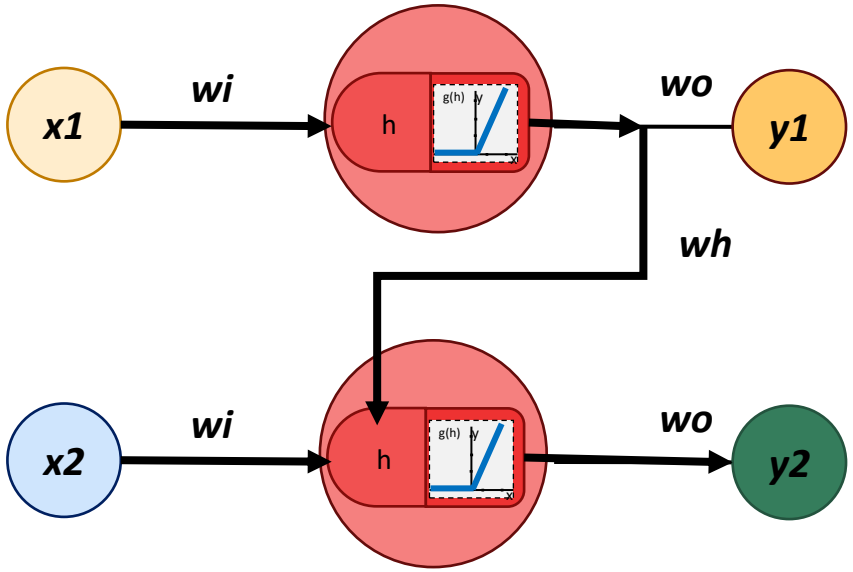
One [input] **to Many** [outputs]

UNIVERSITY OF CALGARY

Recurrent Neural Networks – I/O Properties



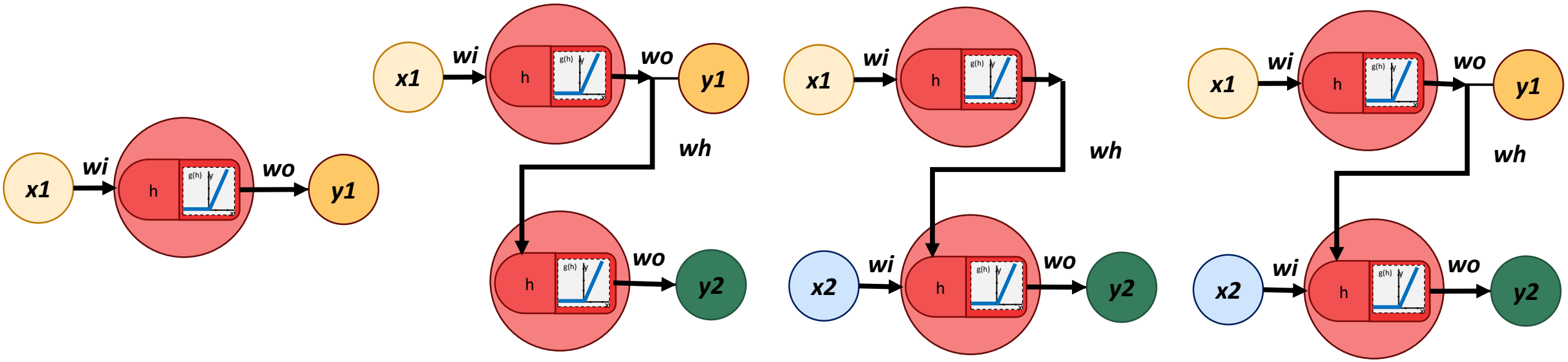
Many [inputs] **to One** [output]



Many [inputs] **to Many** [outputs]

UNIVERSITY OF CALGARY

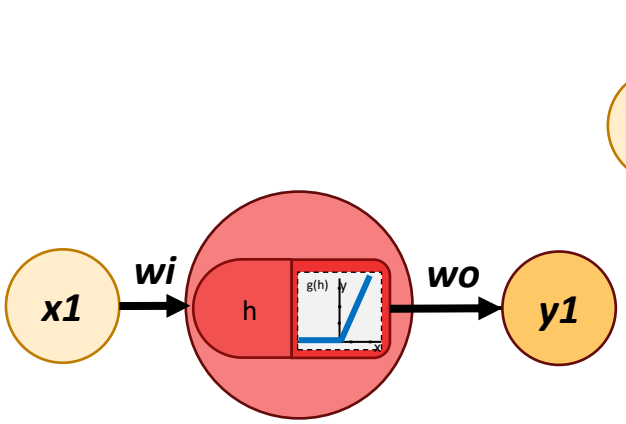
Recurrent Neural Networks – I/O Properties

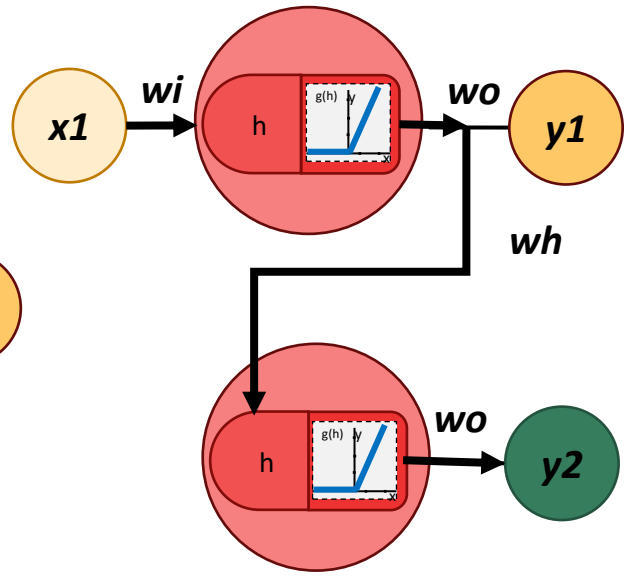


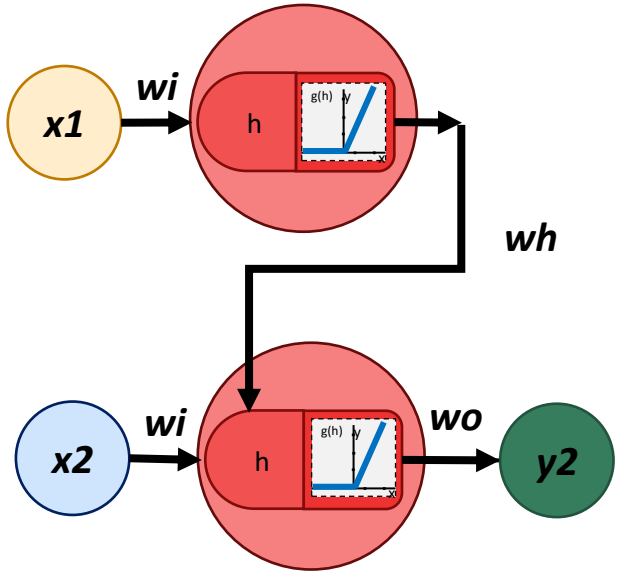
What format is our hockey game predictor that uses several past games to predict the next game's outcome?

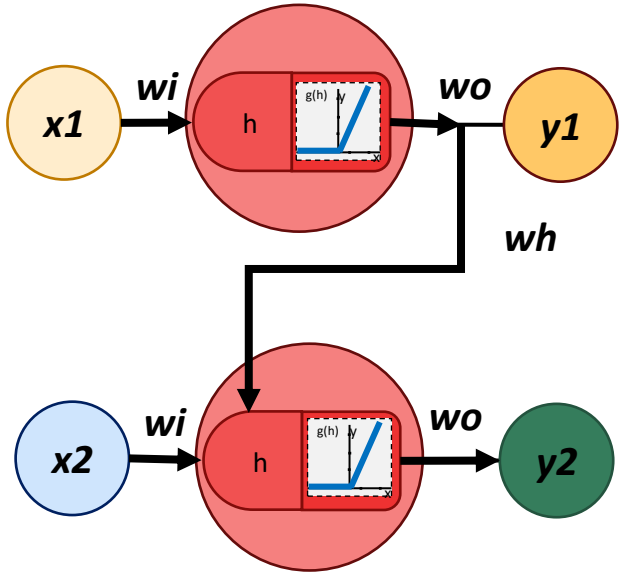
UNIVERSITY OF CALGARY

Recurrent Neural Networks – I/O Properties









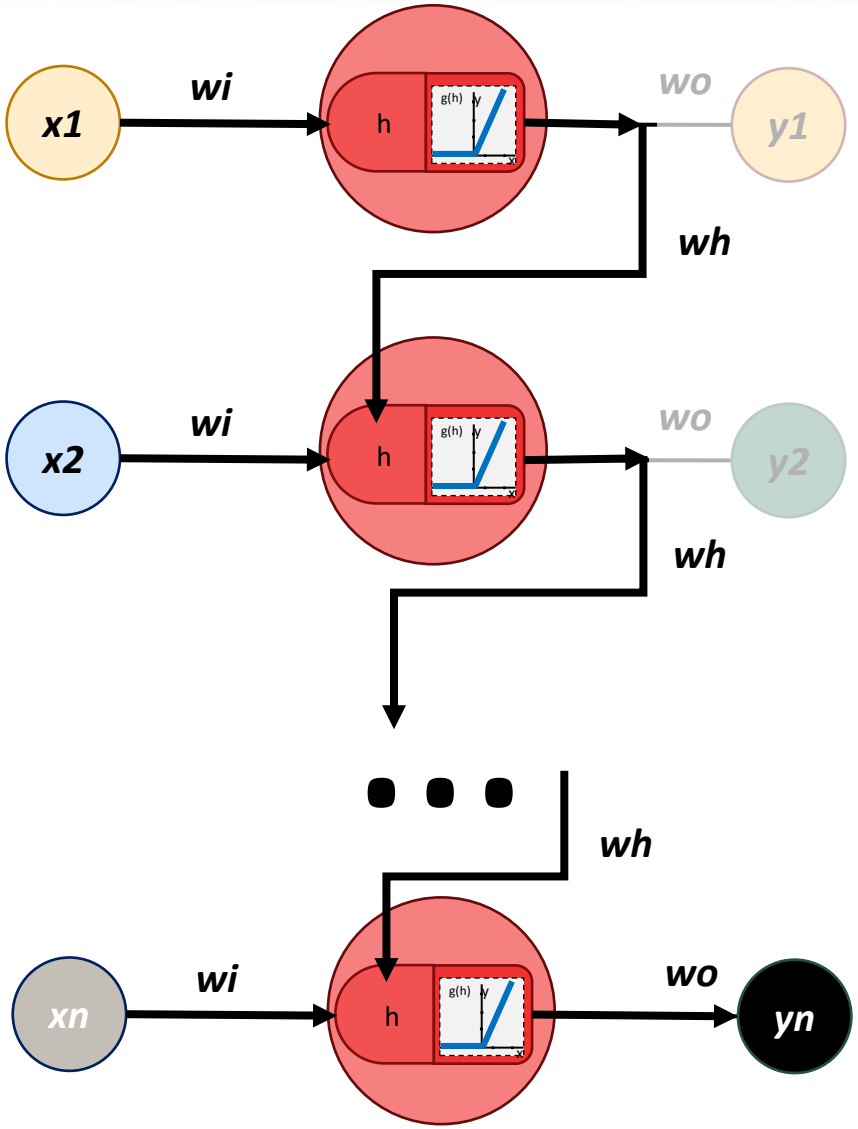
What format is our hockey game predictor that uses several past games to predict the next game's outcome?

Many [past games scores]
to One [next game's score]

UNIVERSITY OF CALGARY

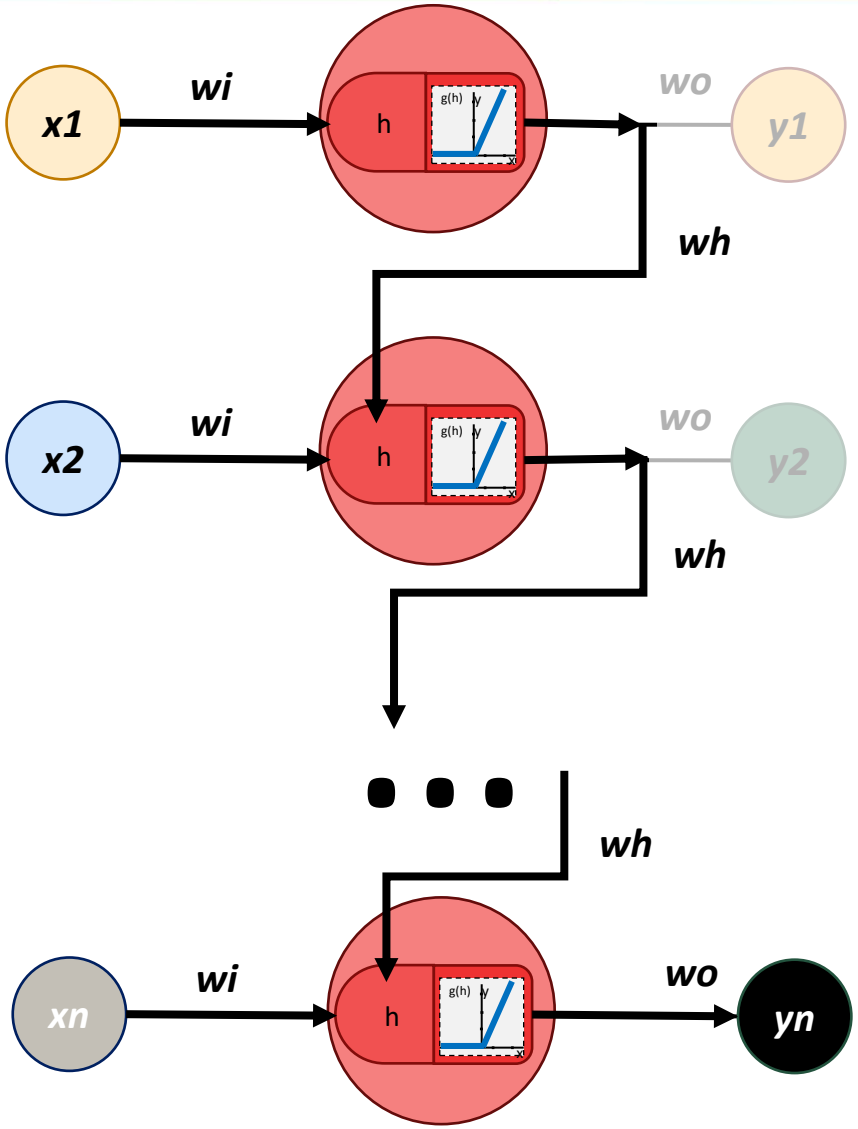
Challenges Associated with RNNs

RNN Backpropagation



What challenge do we encounter as the number of inputs increases?

RNN Backpropagation



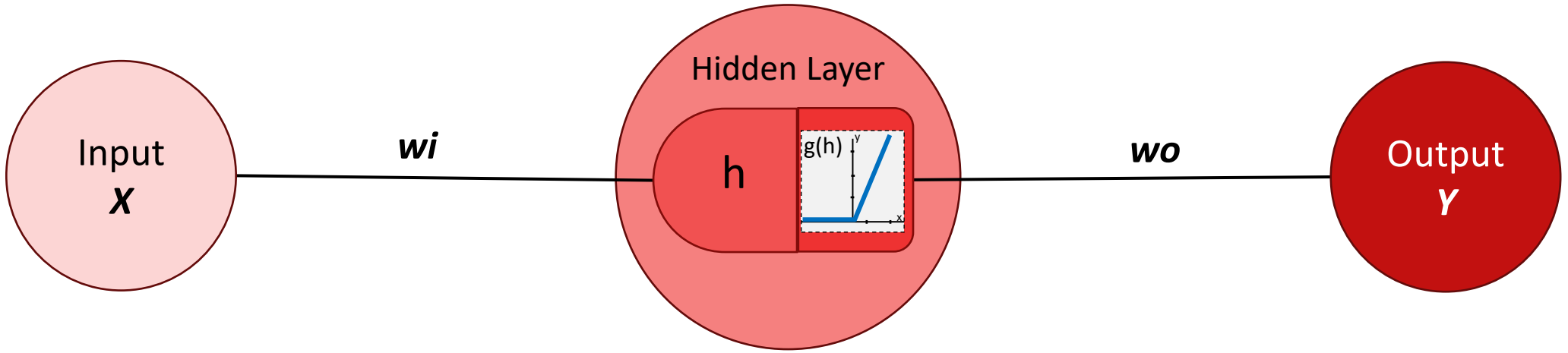
What challenge do we encounter as the number of inputs increases?

Exploding and Vanishing Gradient

Traditional Neural Network Backpropagation

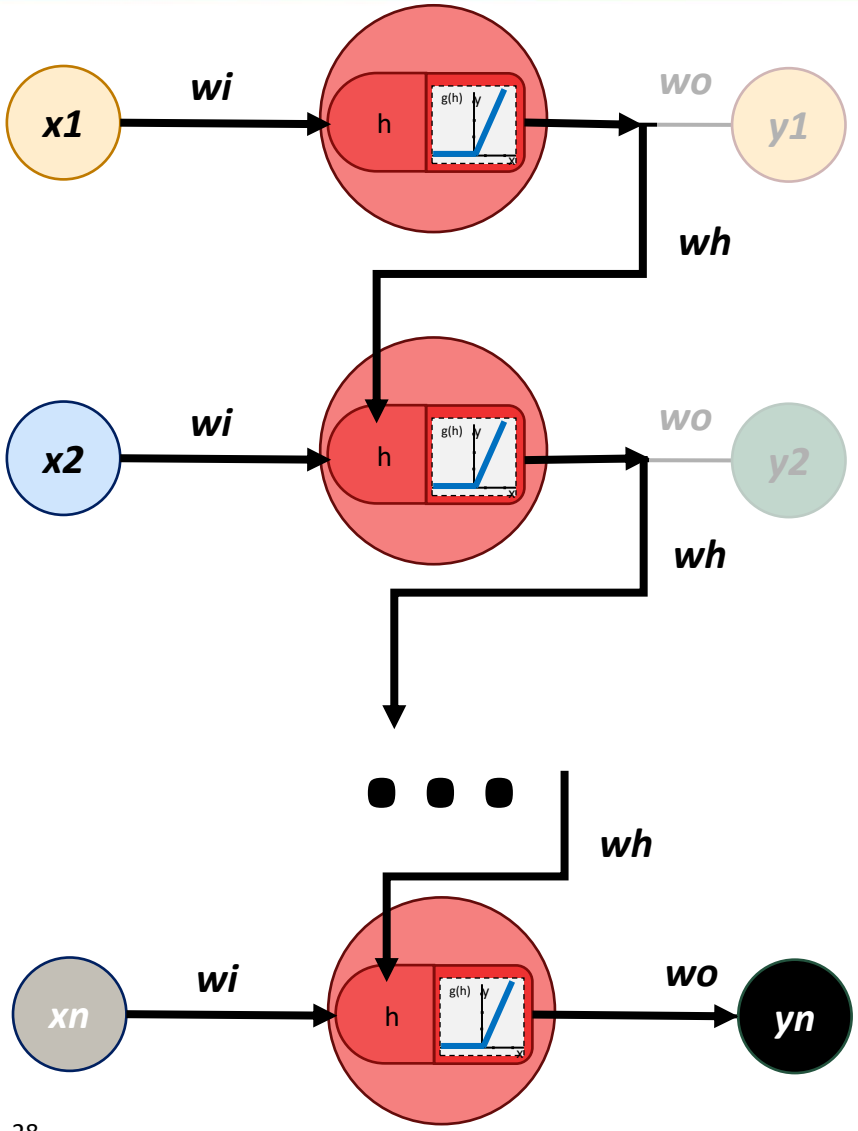
Backpropagation of Weight wi

$$\frac{\partial J(W)}{\partial W_i} = \frac{\partial J(W)}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial W_o} * \frac{\partial W_o}{\partial h} * \frac{\partial h}{\partial W_i}$$



UNIVERSITY OF CALGARY

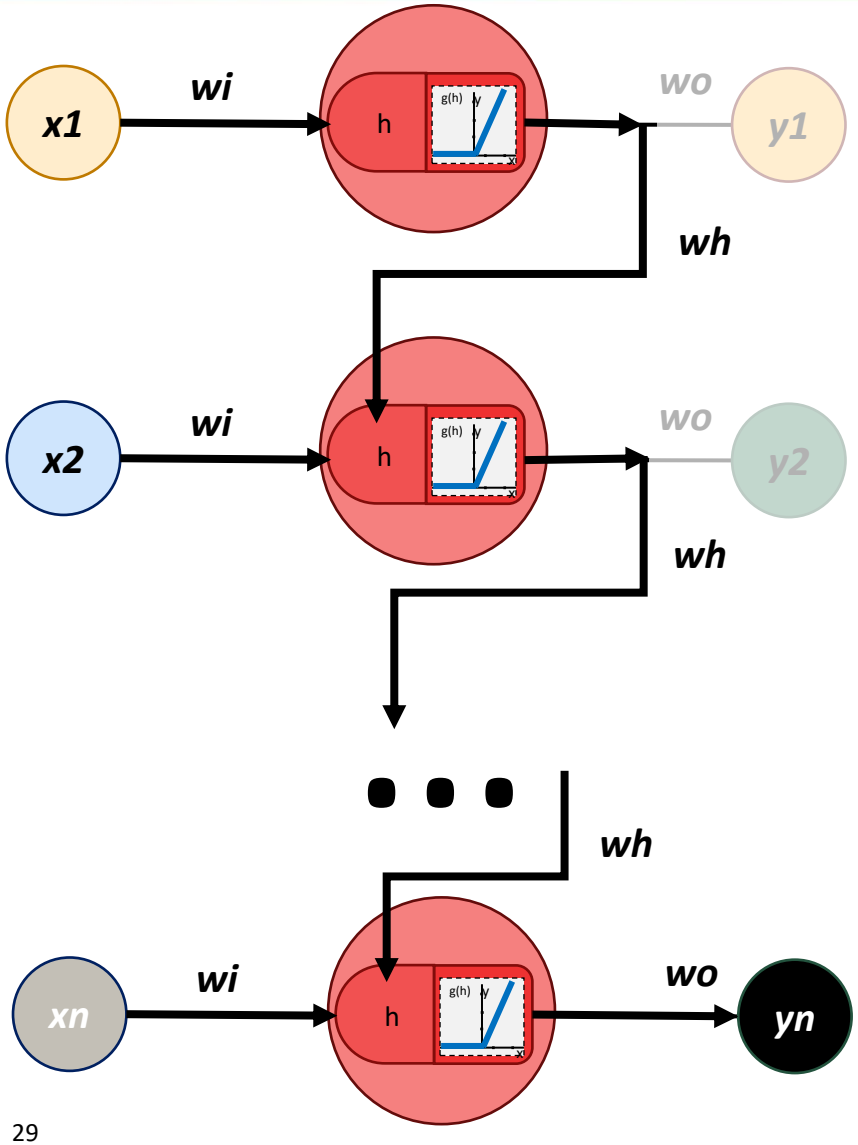
RNN Backpropagation



Backpropagation of Weight wi

$$\frac{\partial J(W)}{\partial W_i} = \frac{\partial J(W)}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial W_o} * \frac{\partial W_o}{\partial h} * \frac{\partial h}{\partial W_i} ?$$

RNN Backpropagation

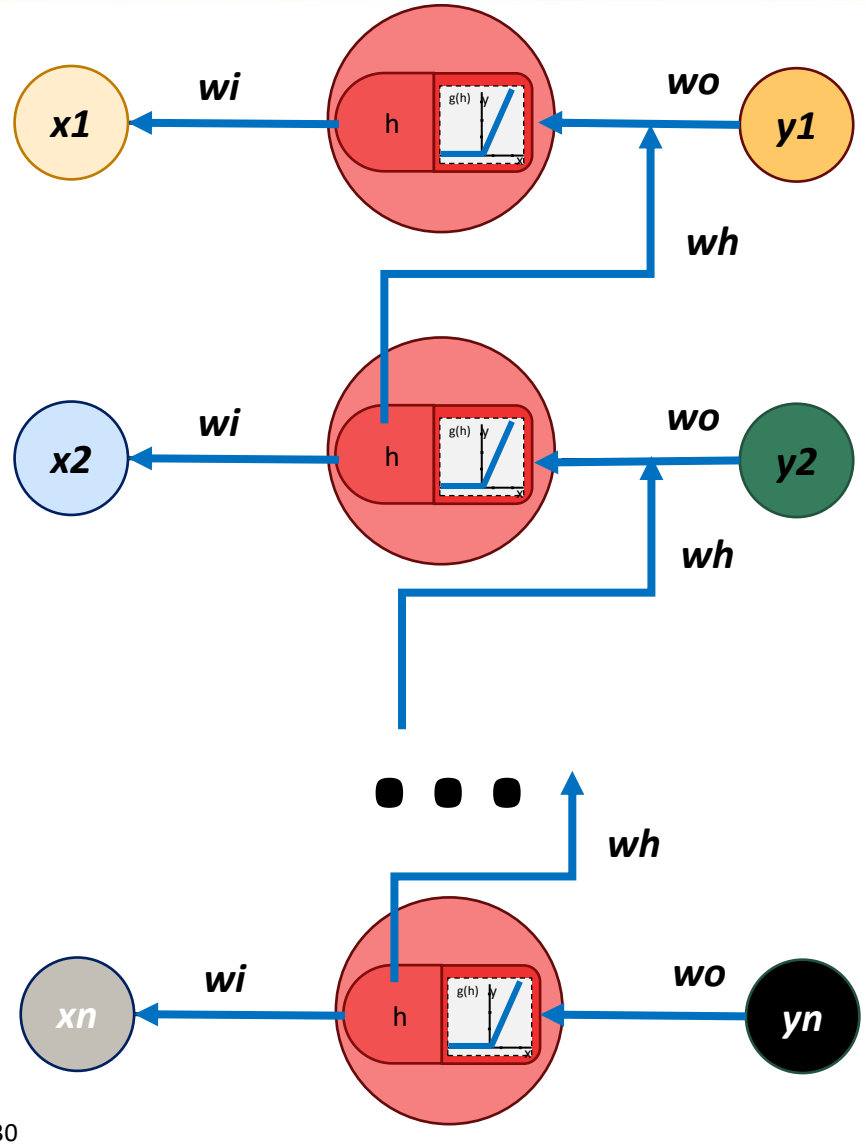


Backpropagation of Weight wi

$$\frac{\partial J(W)}{\partial W_i} = \frac{\partial J(W)}{\partial} \cdot \frac{\partial}{\partial W_o} \cdot \frac{\partial W_o}{} * \frac{\partial h}{\partial W_i} ?$$

UNIVERSITY OF CALGARY

RNN Backpropagation



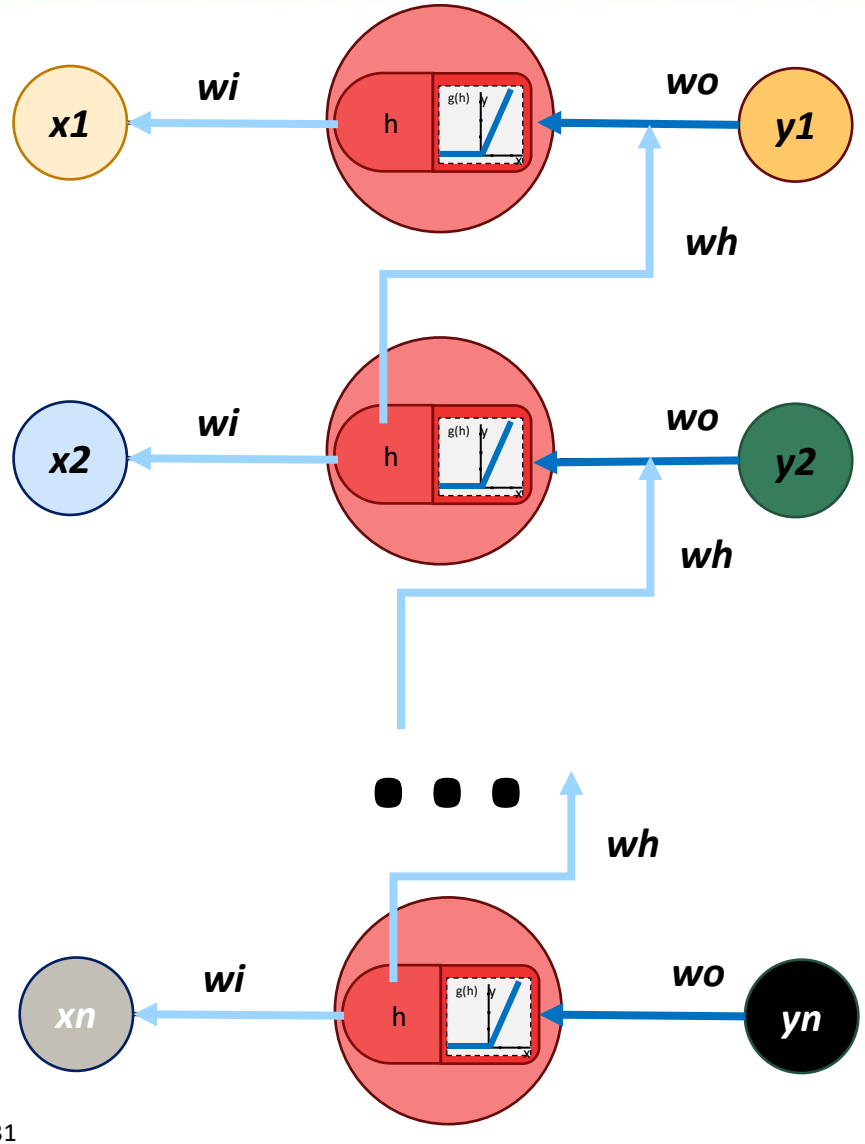
Backpropagation of Weight wi

$$\frac{\partial J(W)}{\partial W_i} = \frac{\partial J(W)}{\partial} \cdots \frac{\partial W_o}{\partial W_o} * \frac{\partial h}{\partial W_i} ?$$

Backpropagation must follow the sequential unrolling of the network

UNIVERSITY OF CALGARY

RNN Backpropagation

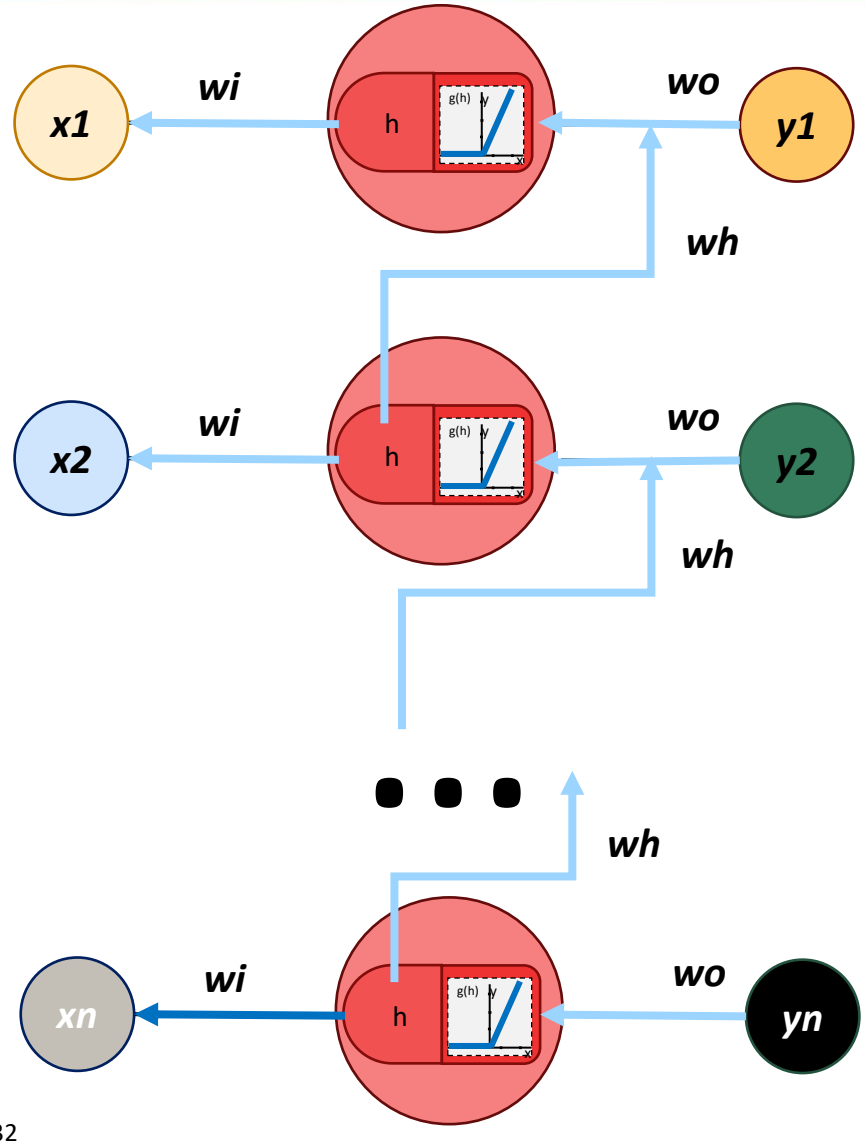


Backpropagation must follow the sequential unrolling of the network

1.) Gradient calculation from all outputs to last layer

UNIVERSITY OF CALGARY

RNN Backpropagation



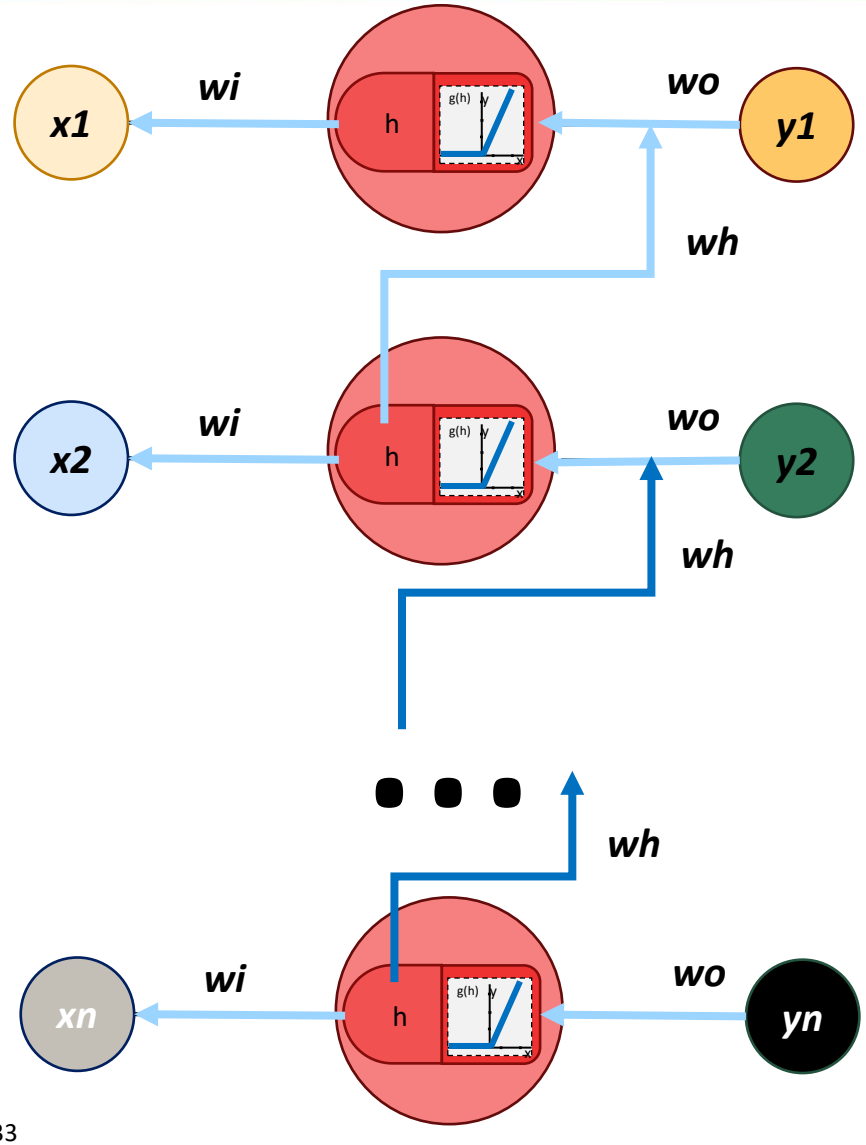
Backpropagation must follow the sequential unrolling of the network

1.) Gradient calculation from all outputs to last layer

2.) Gradient calculation from most recent input

UNIVERSITY OF CALGARY

RNN Backpropagation



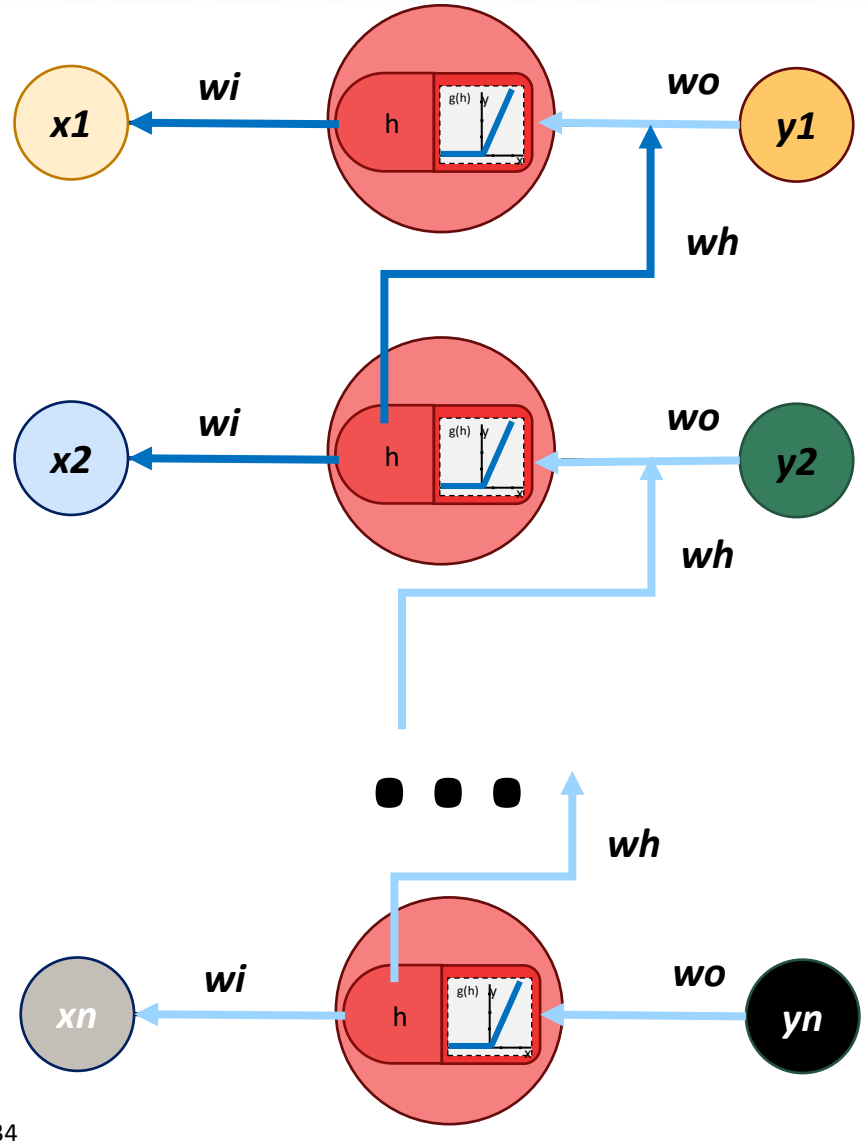
Backpropagation must follow the sequential unrolling of the network

1.) Gradient calculation from all outputs to last layer

2.) Gradient calculation from most recent input

3.) Gradient calculation from last feedback path

RNN Backpropagation



Backpropagation must follow the sequential unrolling of the network

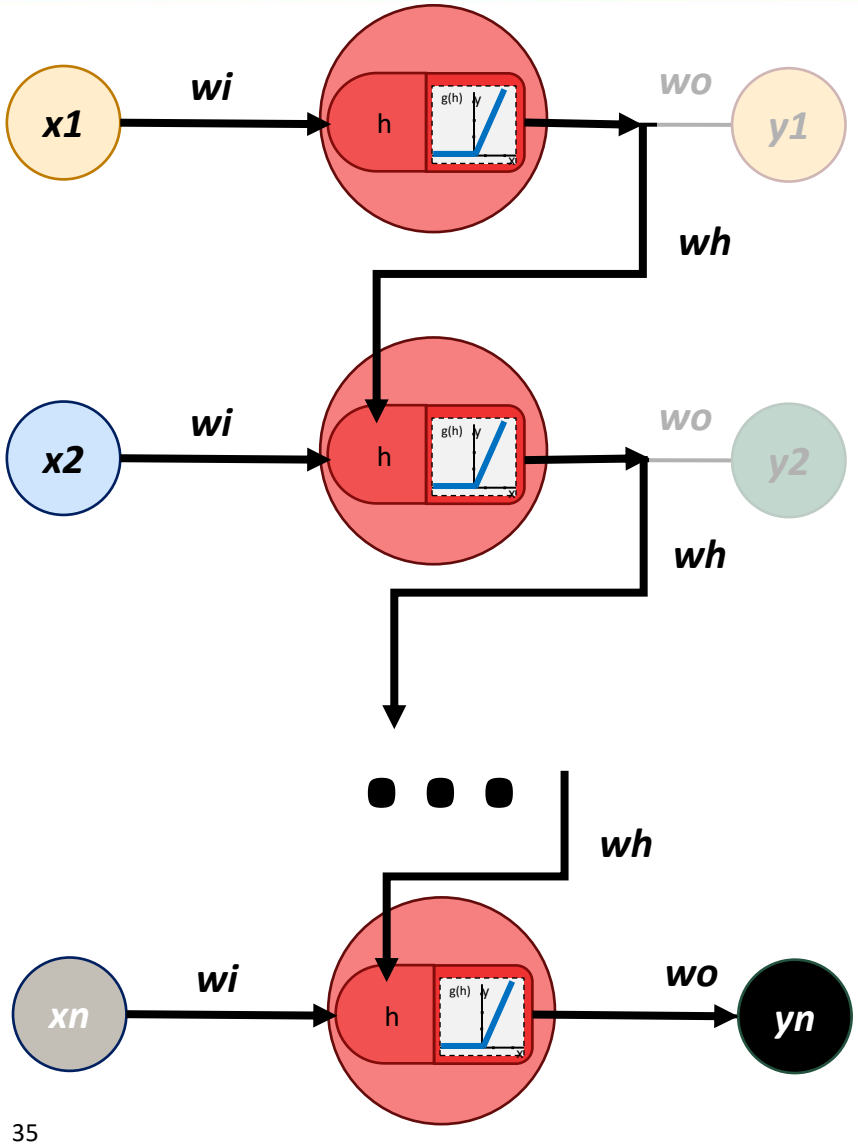
1.) Gradient calculation from all outputs to last layer

2.) Gradient calculation from most recent input

3.) Gradient calculation from last feedback path

4.) (2.) and 3.)) Continue as the network unrolls

RNN Backpropagation – Gradient Challenges



With long input sequences,
as wo **increases…**

$$X1 \times wo \times wo \times wo$$

If weight wo is 3

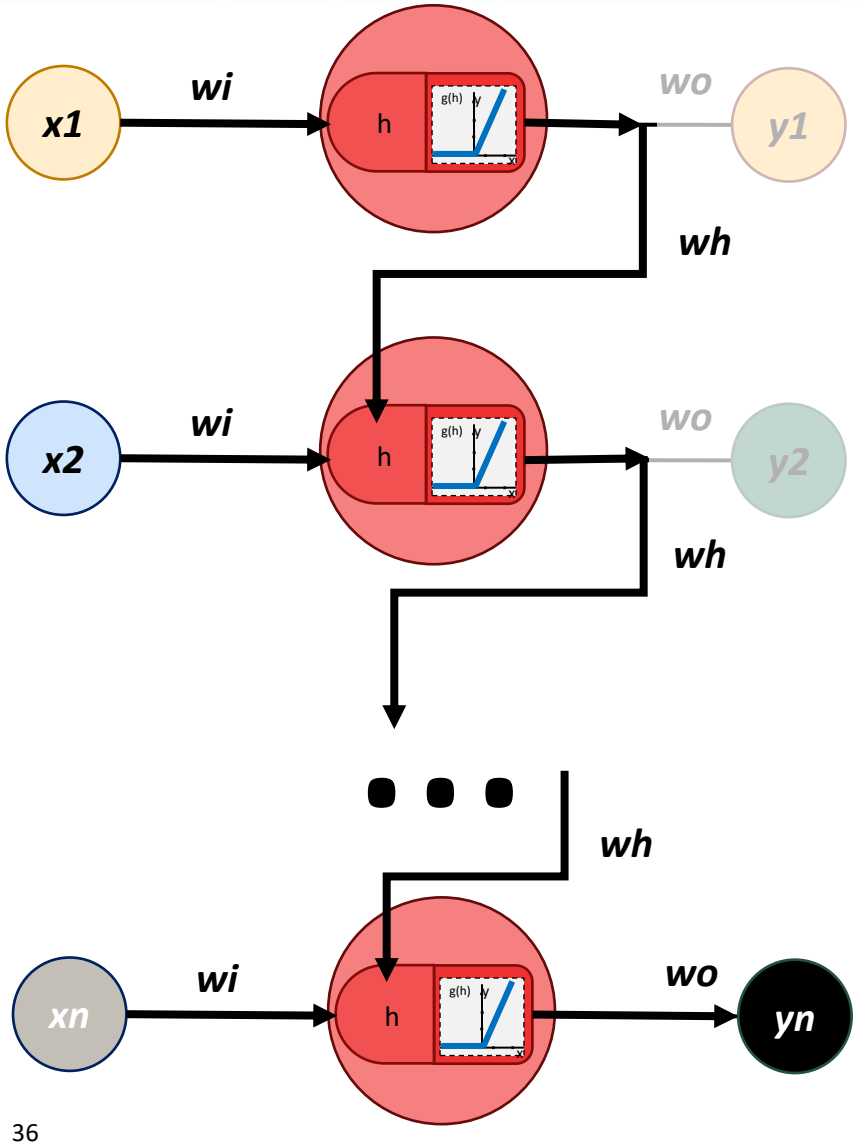
$$X1 \times 3^3$$

$$X1 \times 27$$

The calculated gradient **increases…**

Exploding Gradient

RNN Backpropagation – Gradient Challenges



With long input sequences,
as wo **decreases…**

$$X1 \times wo \times wo \times wo$$

If weight wo is 0.33

$$X1 \times 3^3$$

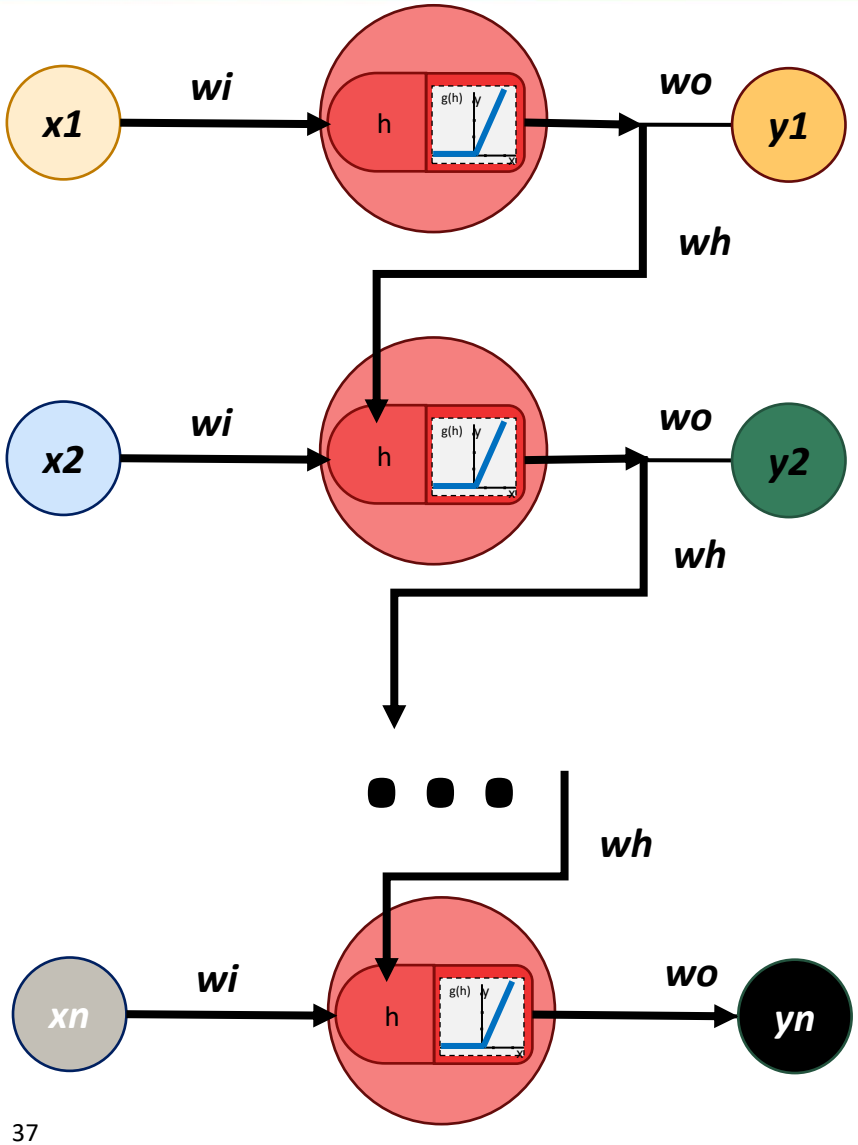
$$X1 \times 0.035$$

The calculated gradient **decreases …**

Vanishing Gradient

UNIVERSITY OF CALGARY

RNN Backpropagation – Challenges

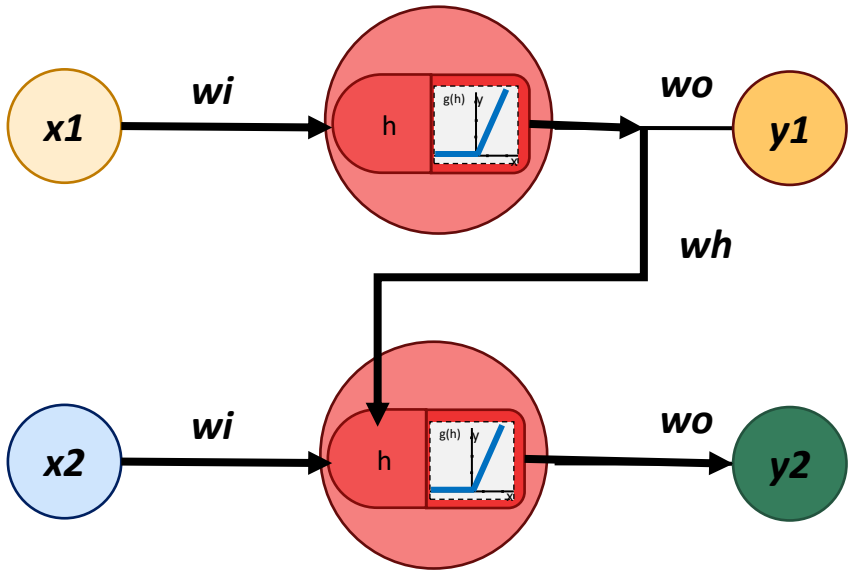


Modelling long-term dependencies (memory) with simple RNNs is challenging

Cannot parallelize operations leading to inefficient training

Architectures Derived from Simple RNNS

Long-Short Term Memory (LSTM) Network

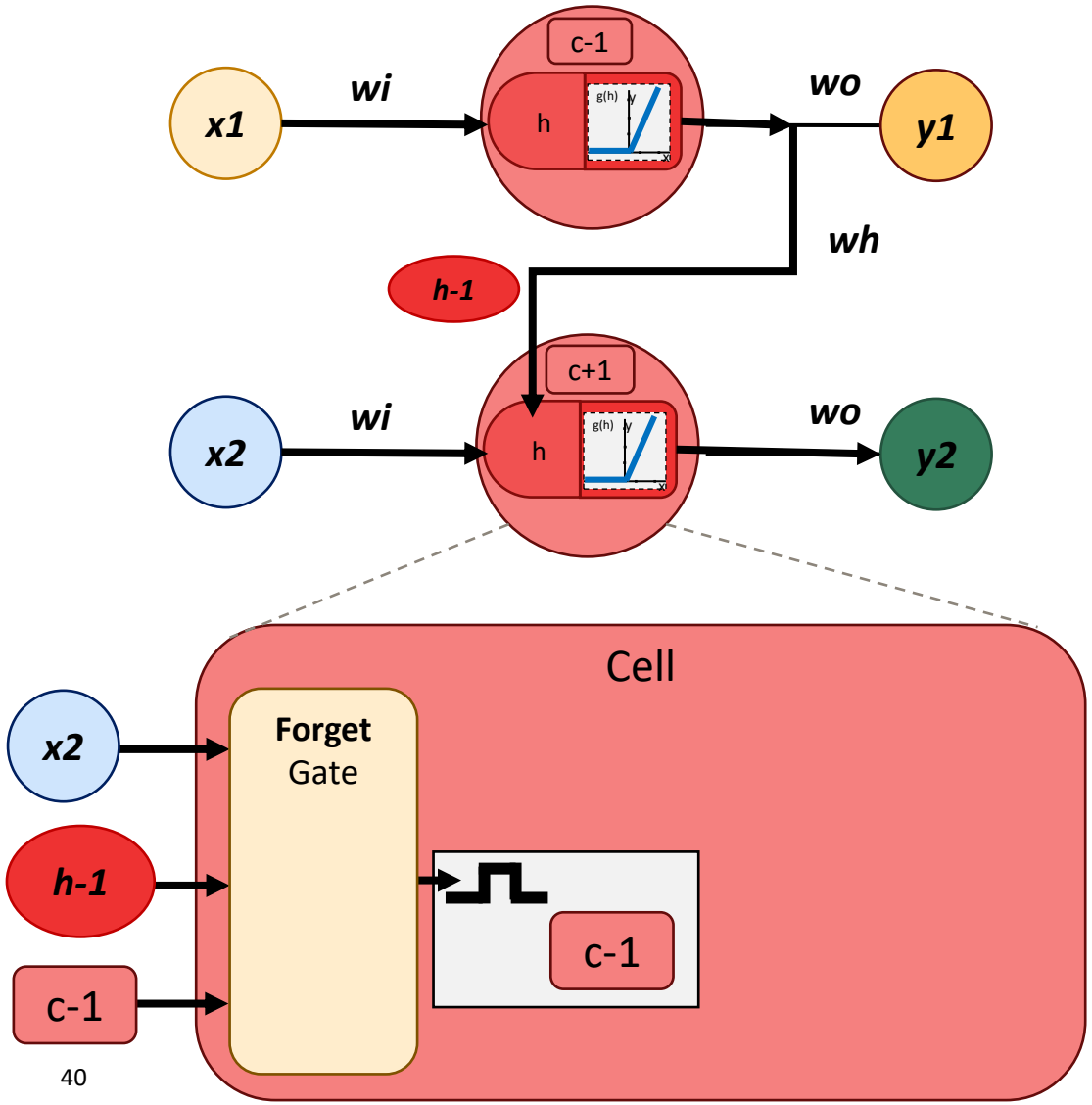


What shortcoming of the RNN does the LSTM try to overcome?

Vanishing Gradient
Using
Three Step Approach

UNIVERSITY OF CALGARY

Long-Short Term Memory (LSTM) Network Architecture

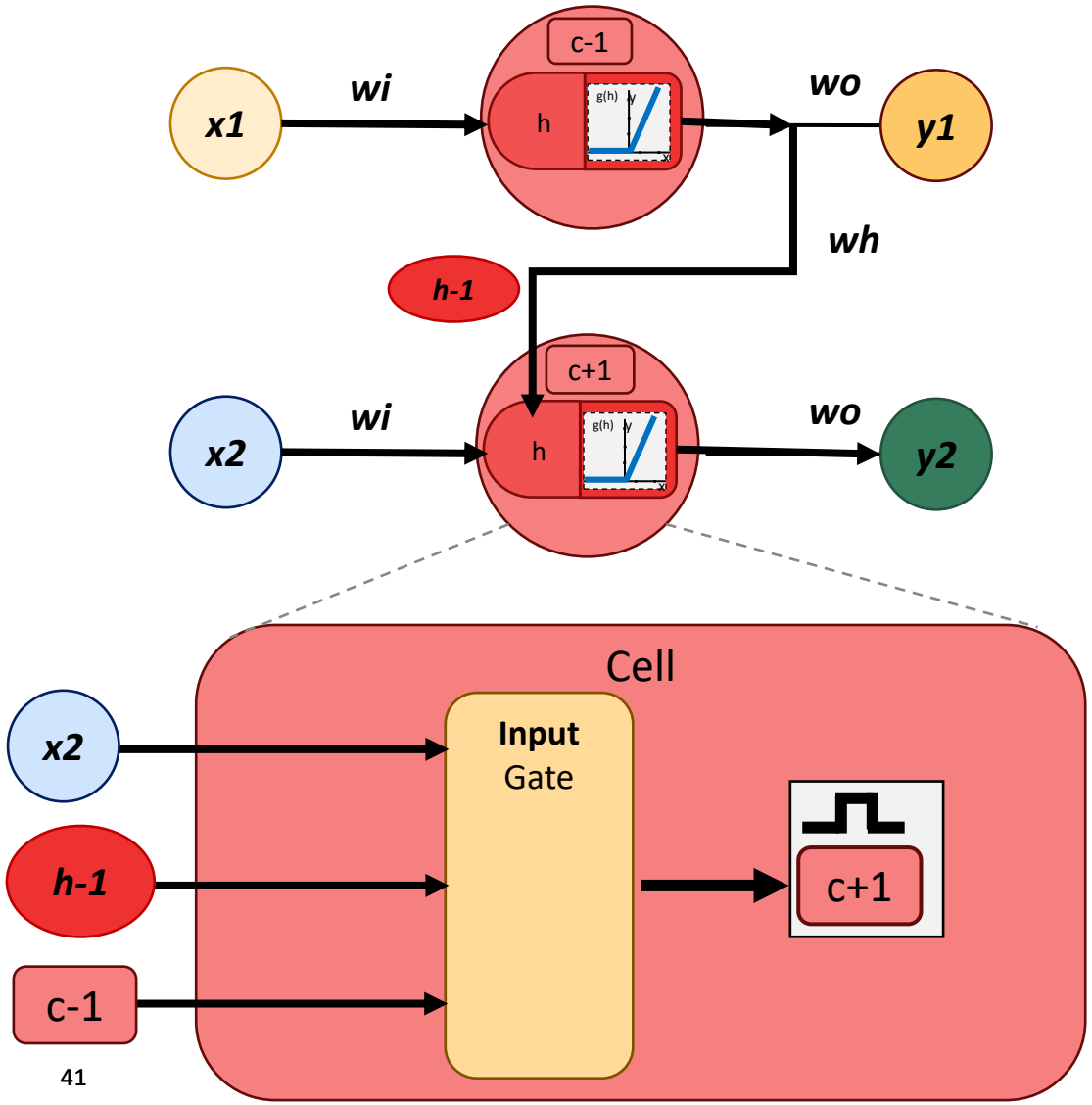


Step 1: What information will we keep from the previous cell state?

c-1

UNIVERSITY OF CALGARY

Long-Short Term Memory (LSTM) Network Architecture

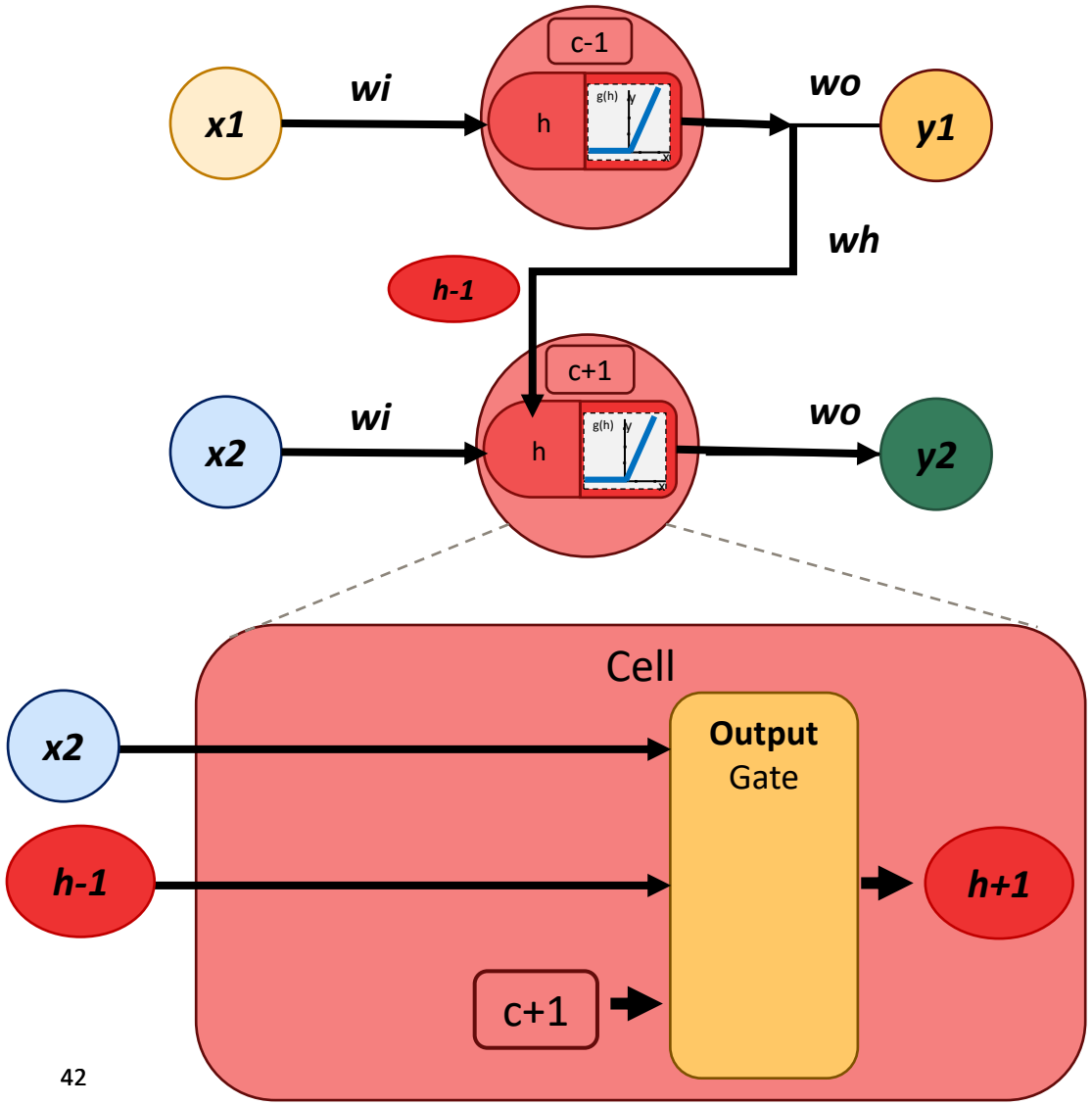


Step 2: What new information (from the input) do we want to store in the next cell state?

c+1

UNIVERSITY OF CALGARY

Long-Short Term Memory (LSTM) Network Architecture

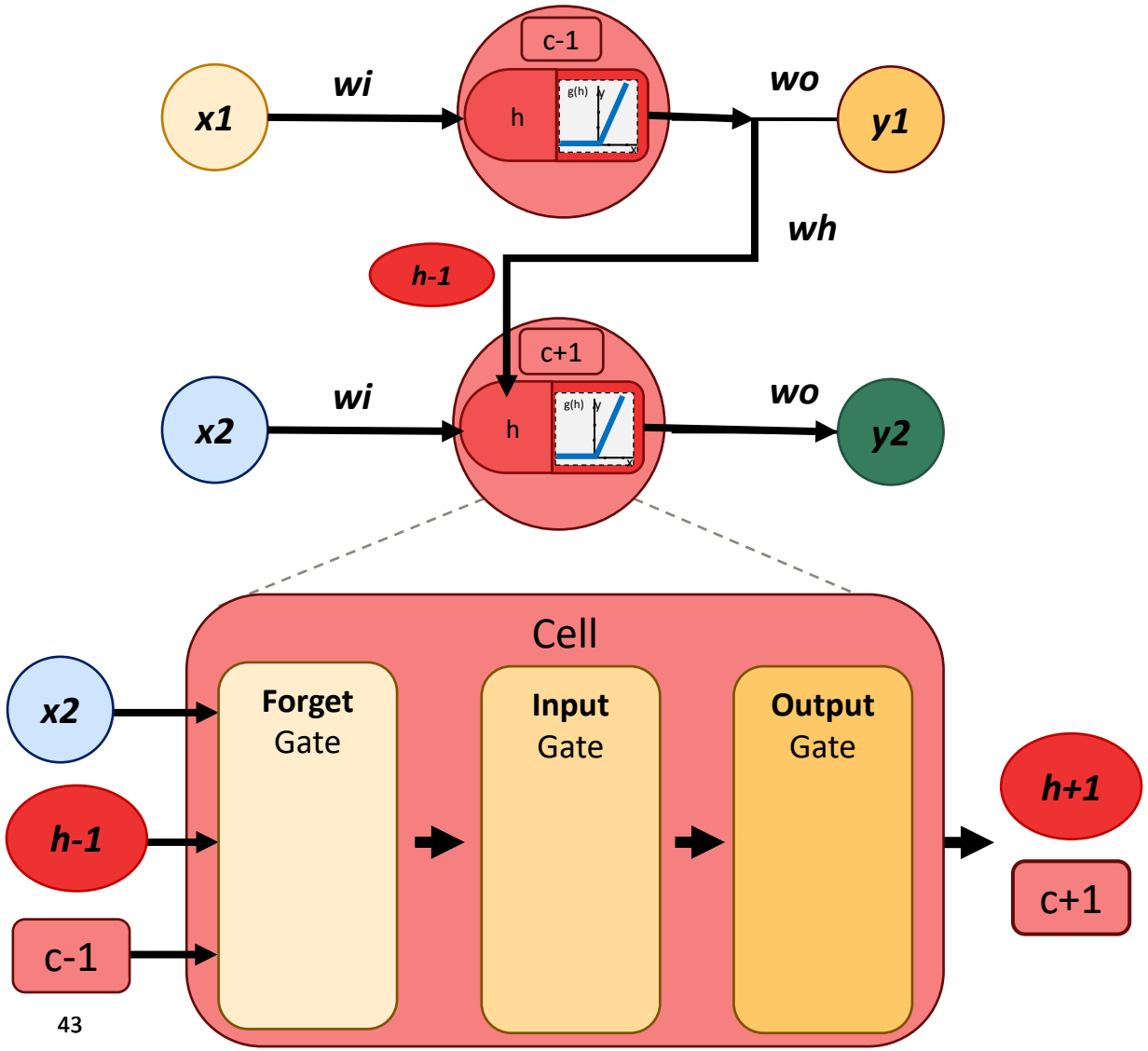


Step 3: What information from the cell state will we output?

h+1

UNIVERSITY OF CALGARY

Long-Short Term Memory (LSTM) Network

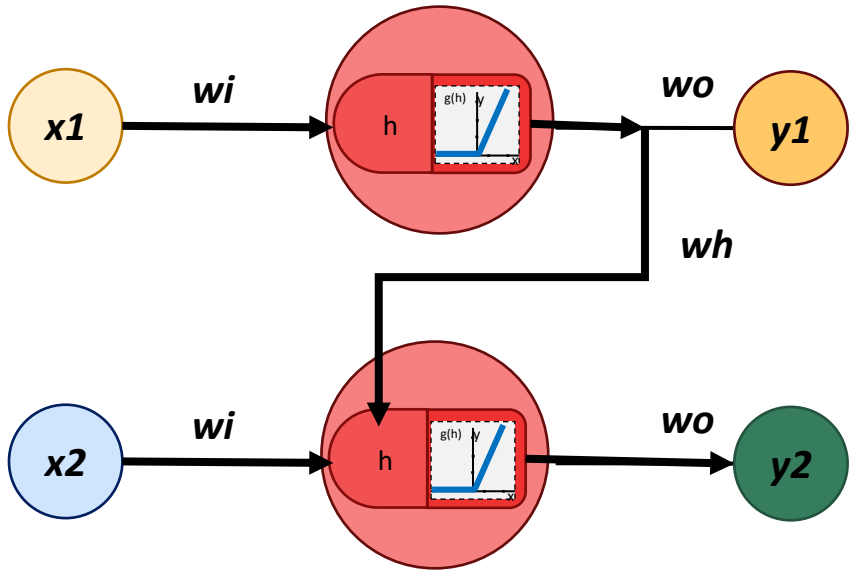


How does this architecture limit the vanishing gradient?

What are the limitations of the LSTM?

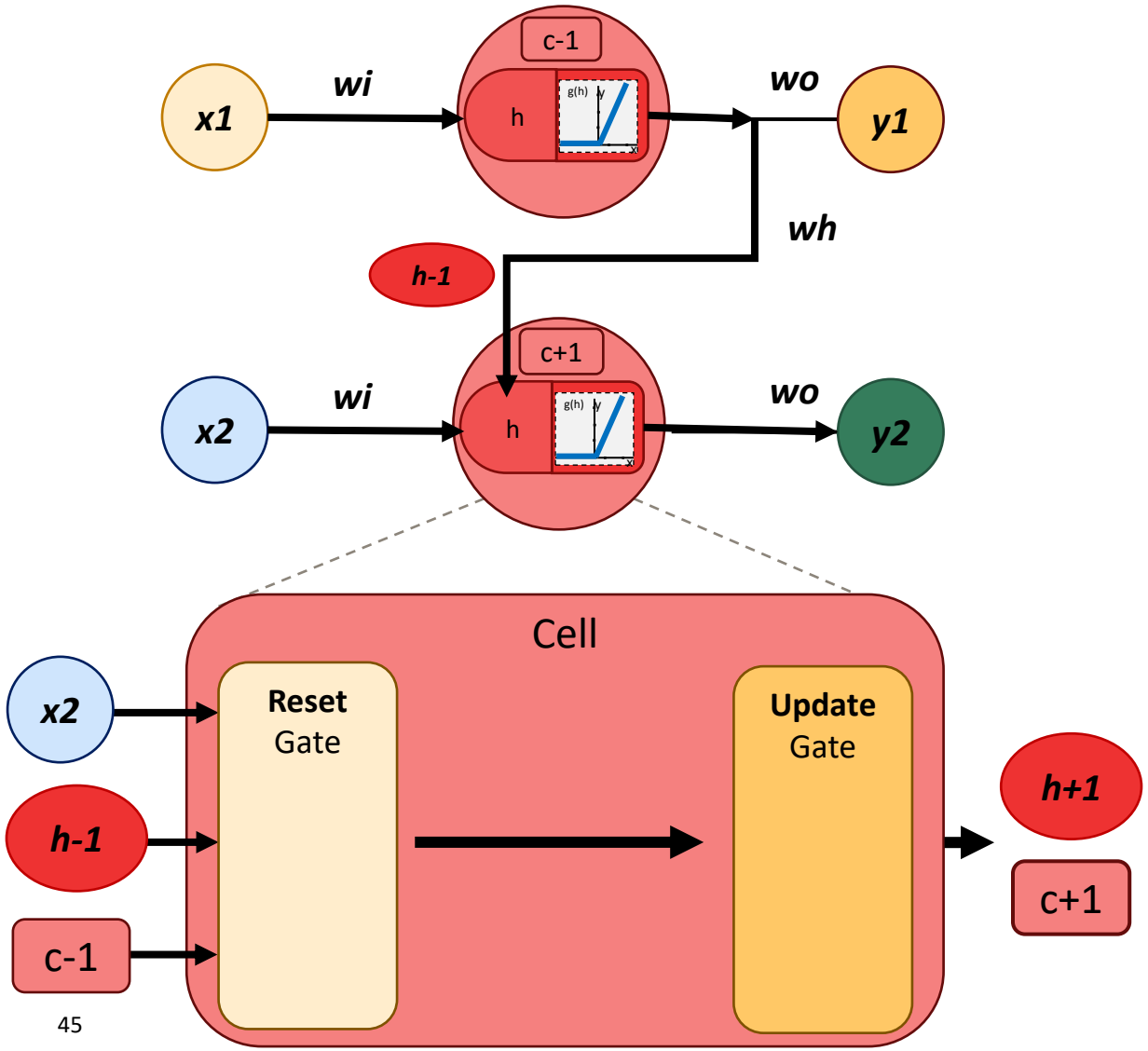
Exploding Gradient

Gated Recurrent Unit (GRU) Network



What is the difference between an LSTM and GRU?

Gated Recurrent Unit (GRU) Network



What is the difference between an LSTM and GRU?

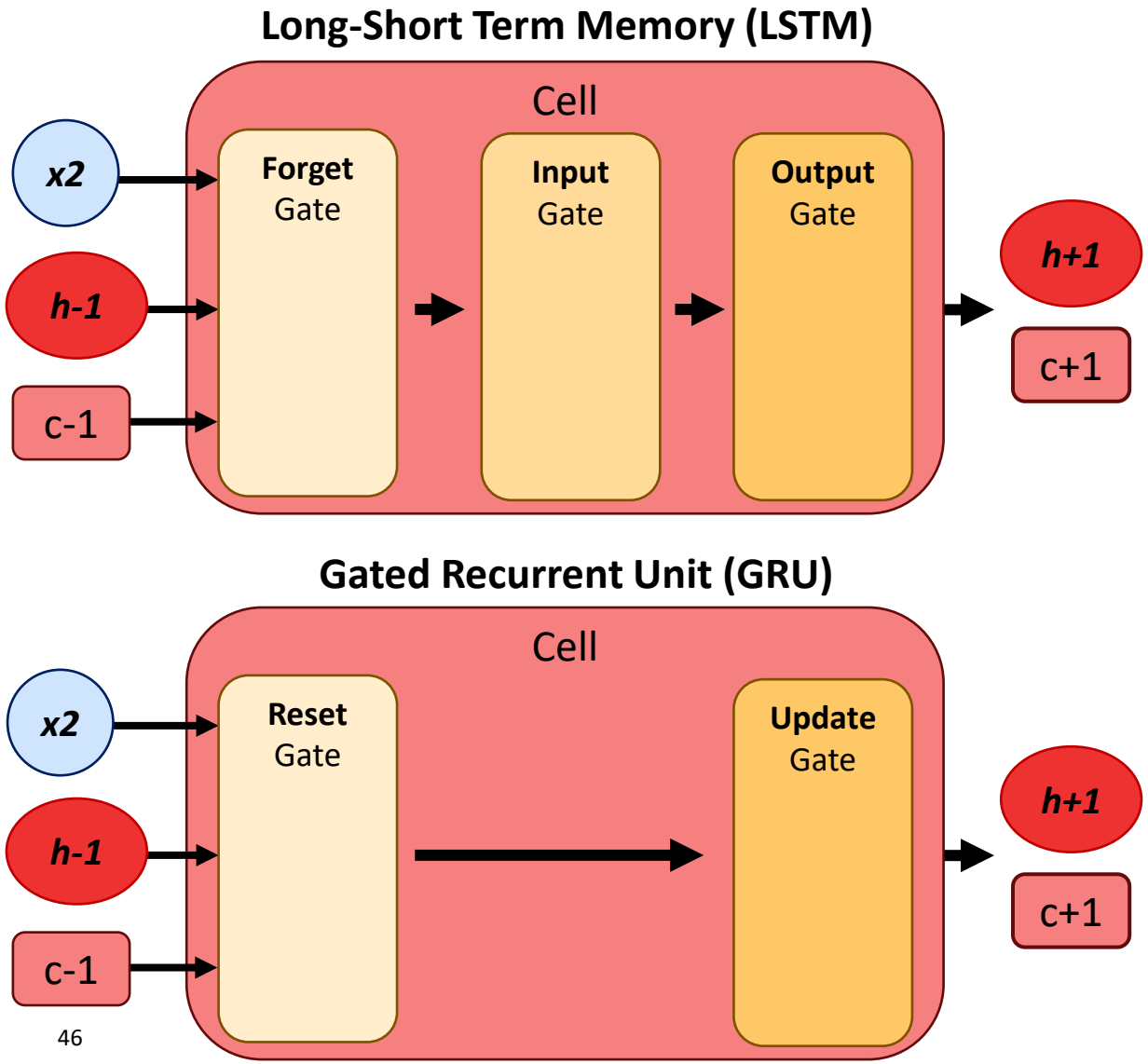
Capturing Dependencies using

Three Gates (LSTM)

vs.

Two Gates (GRU)

Gated Recurrent Unit (GRU) Network



When do we use an LSTM vs. GRU?

LSTM: Long term dependencies **>** Computational Efficiency

GRU: Long term dependencies **<** Computational Efficiency

UNIVERSITY OF CALGARY

Summary

- RNNs can provide better performance for the processing of series data
- Simple RNNs are at-risk of exploding and vanishing gradients
- LSTMs and GRUs provide an alternative to simple RNNs which minimize learning (gradient) pitfalls

References and Resources

Sherstinsky (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena. 404: 132306. doi: 10.1016/j.physd.2019.132306.

MIT 6.S191: Recurrent Neural Networks, Transformers, and Attention

https://weberna.github.io/blog/2017/11/15/LSTM-Vanishing-Gradients.html