Applied Deep Learning

Dr. Philippe Blaettchen
Bayes Business School (formerly Cass)

www.bayes.city.ac.uk

Learning objectives of today

Goals: Introduce recurrent neural networks (RNNs) as a means to work with sequence data

- Understand the importance of sequences and the difficulty of working with them using the neural network architectures we have learned about so far
- Develop the knowledge to use basic RNNs in practice, as well as critical extensions

How will we do this?

- We start by introducing sequence data and its relevance to machine learning tasks
- We then build up the concept of recurrence underlying RNNs
- We consider limitations of standard RNNs and introduce extensions that allow for "long-term memory"
- We take a brief detour to the TensorFlow Data API that allows us to manipulate datasets effectively, which is often needed when working with RNNs

Working with sequences

The importance of sequences

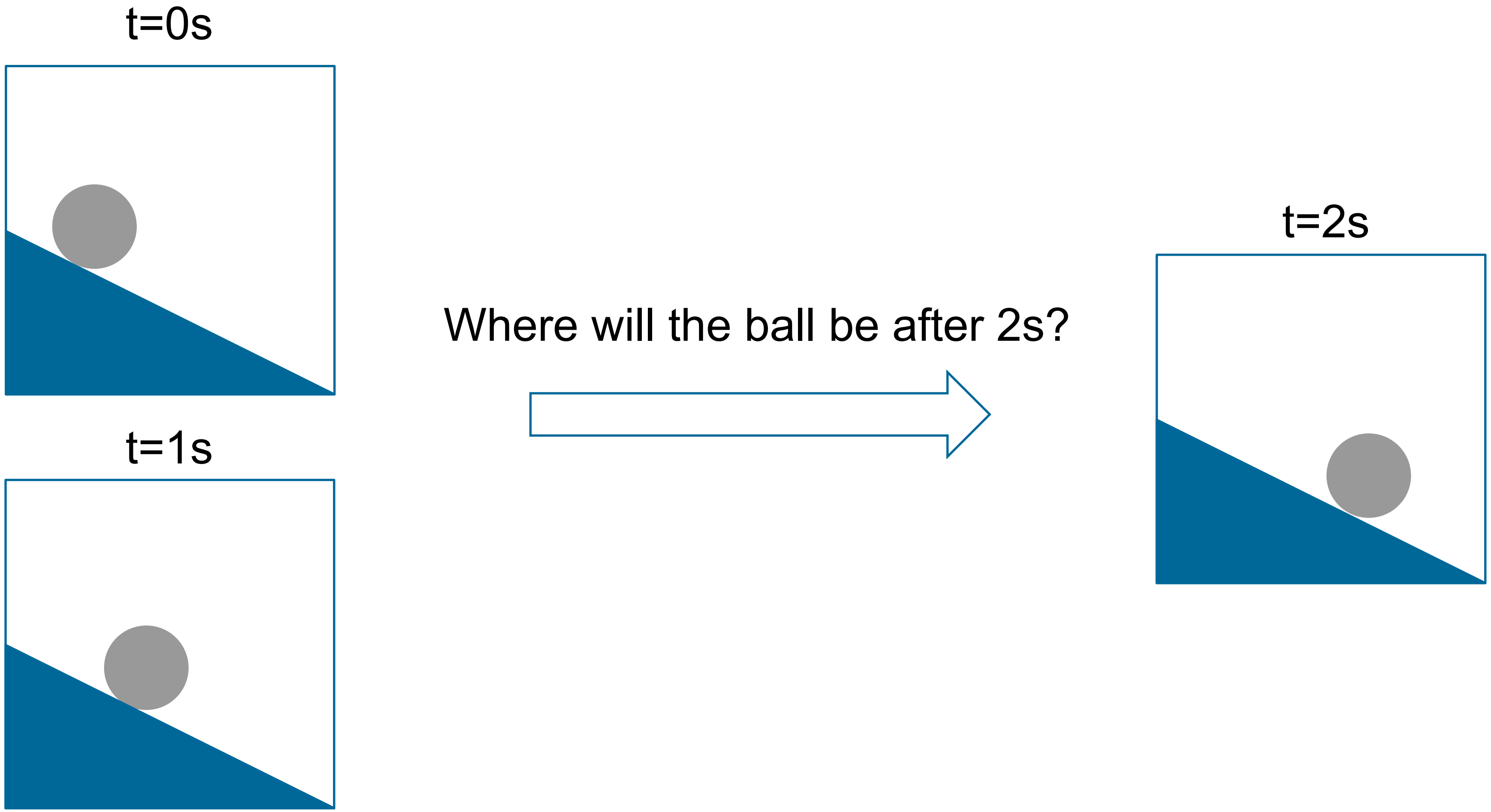
t=0s

Where will the ball be after 2s?

t=2s

?

The importance of sequences



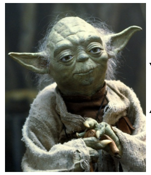
The importance of sequences

"Why do we care about sequences?"

"Why" "do" "we" "care" "about" "sequences" "?"

≠

"care" "about" "sequences" "Why" "do" "we" "?"

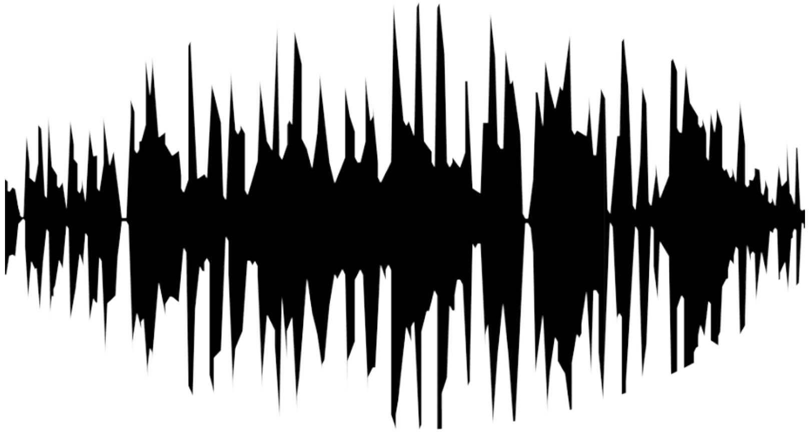
(unless you are )

Sequences

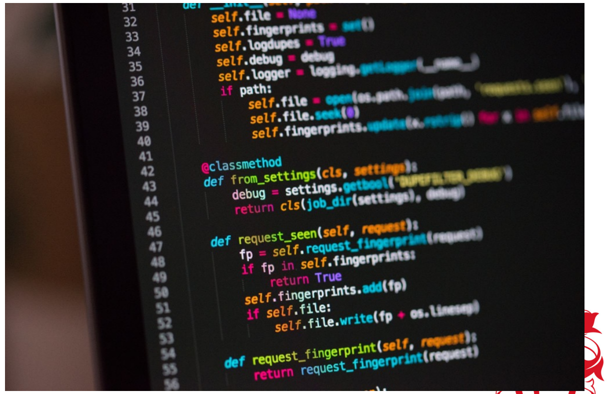
Sequences are collections of multiple elements (i.e., data points), where:

- The order matters
- Elements may be repeated
- The length is variable (and lengths of inputs and outputs don't have to match)

Sequences in real life

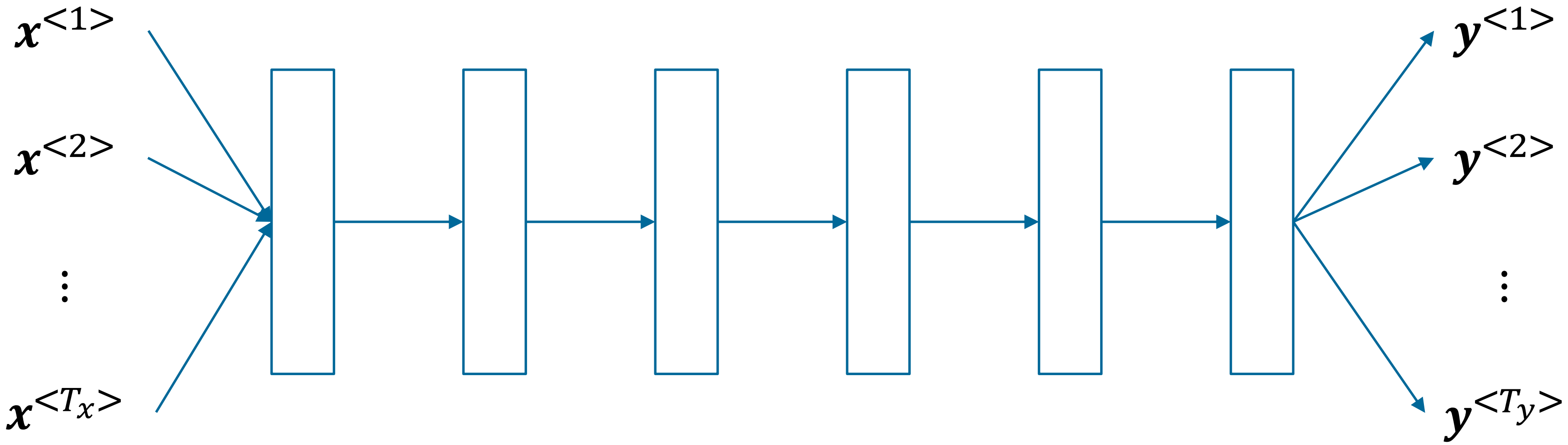


"And here I am, for all my lore, The wretched fool I was before"



Recurrence in neural networks

Why don't we do this?

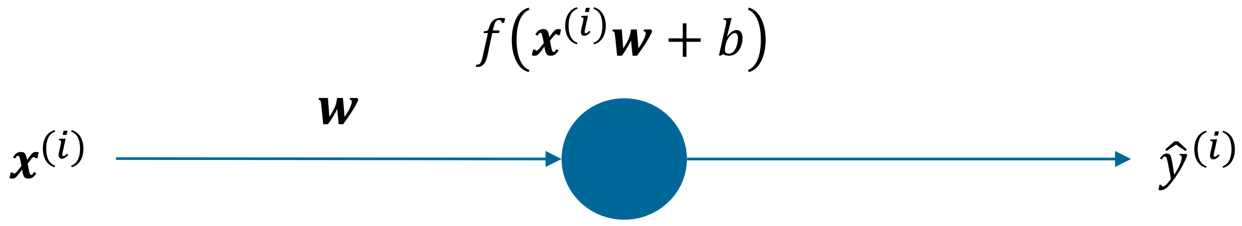


Issues:

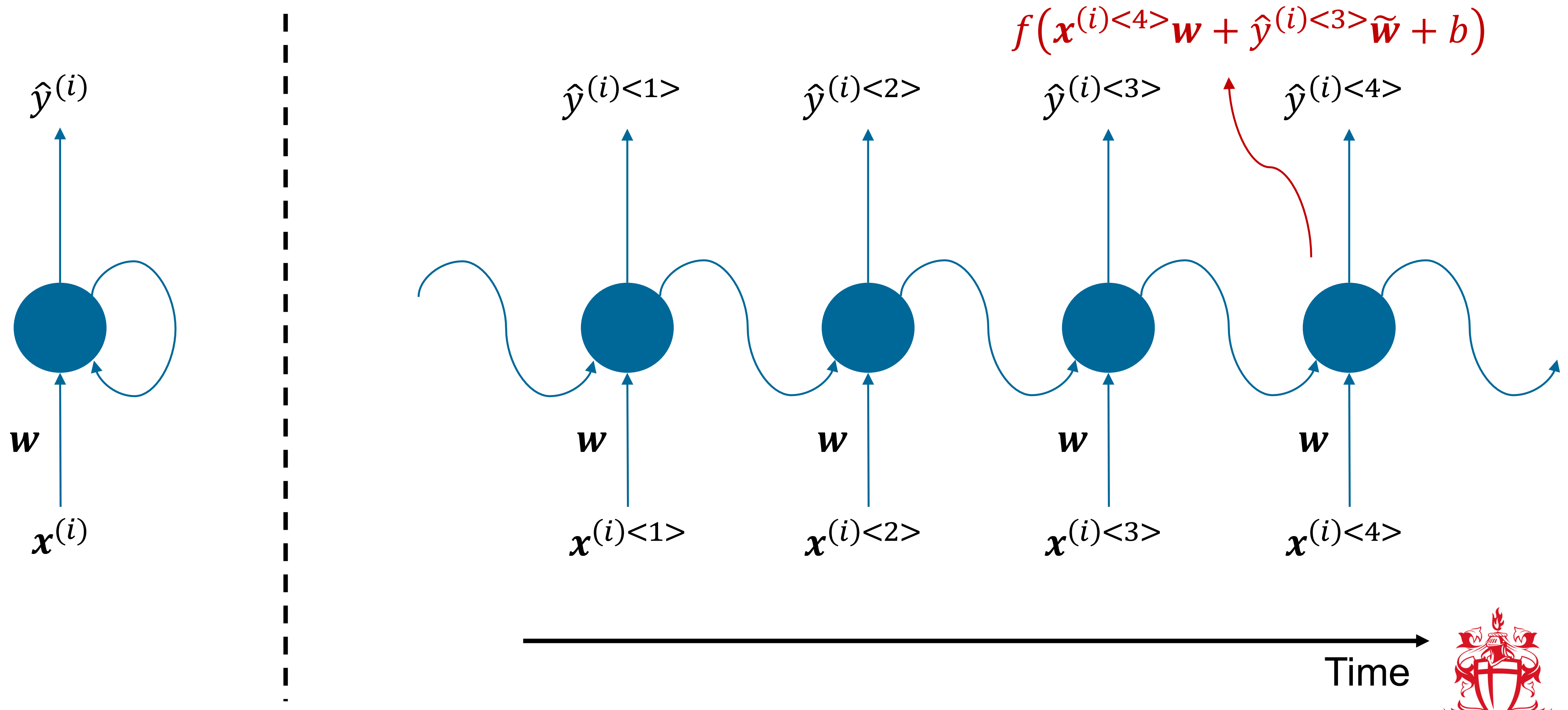
- Sequence lengths vary
- No definition of order
- Lack of parameter sharing: imagine a minute-long ECG

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

What we do instead – let's start with a single neuron

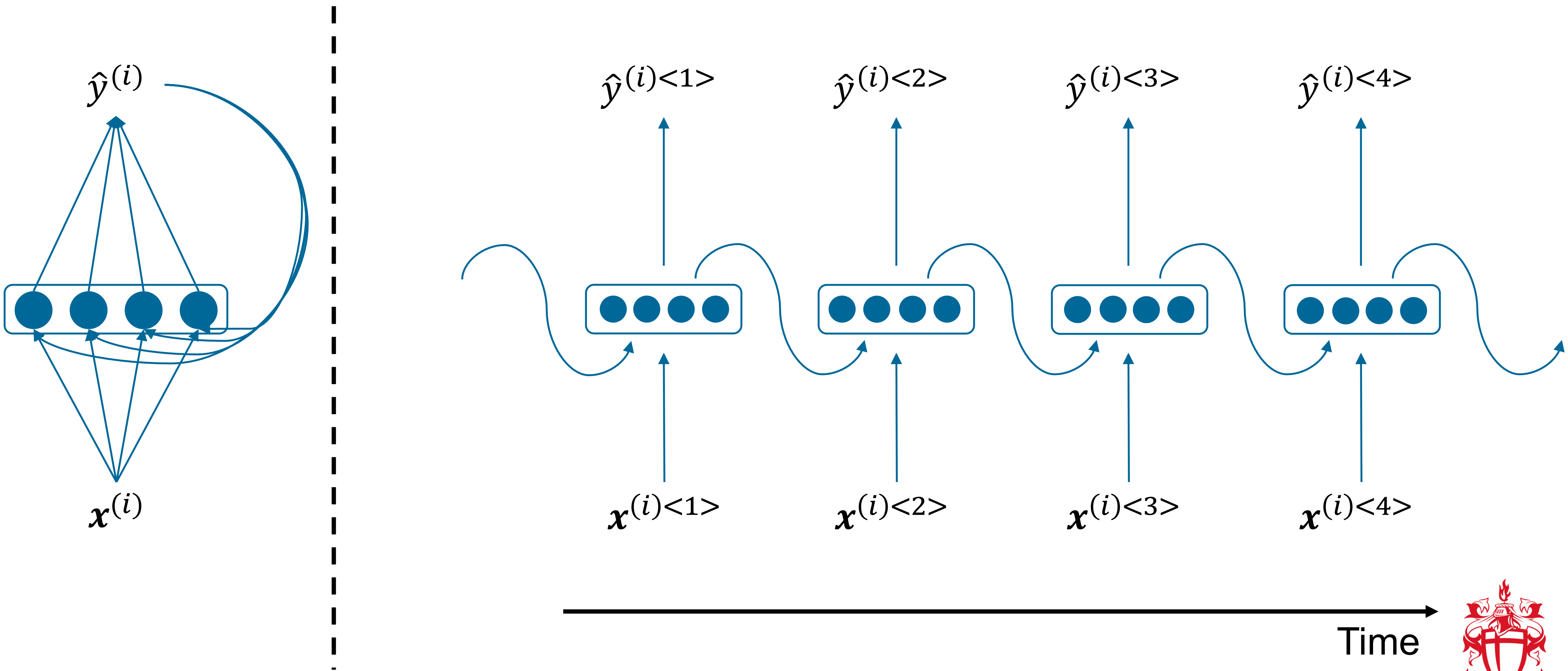


A recurrent neuron

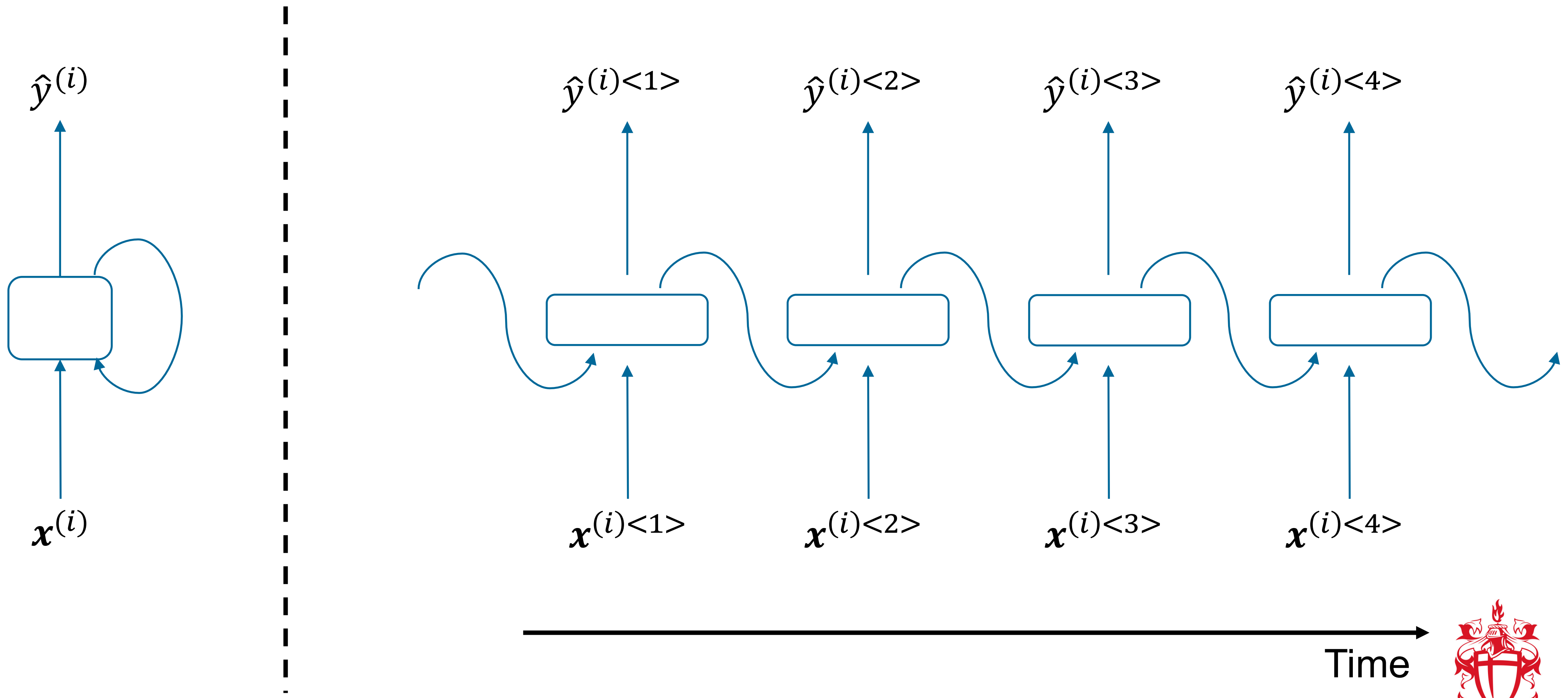


Let’s see a recurrent neuron in practice

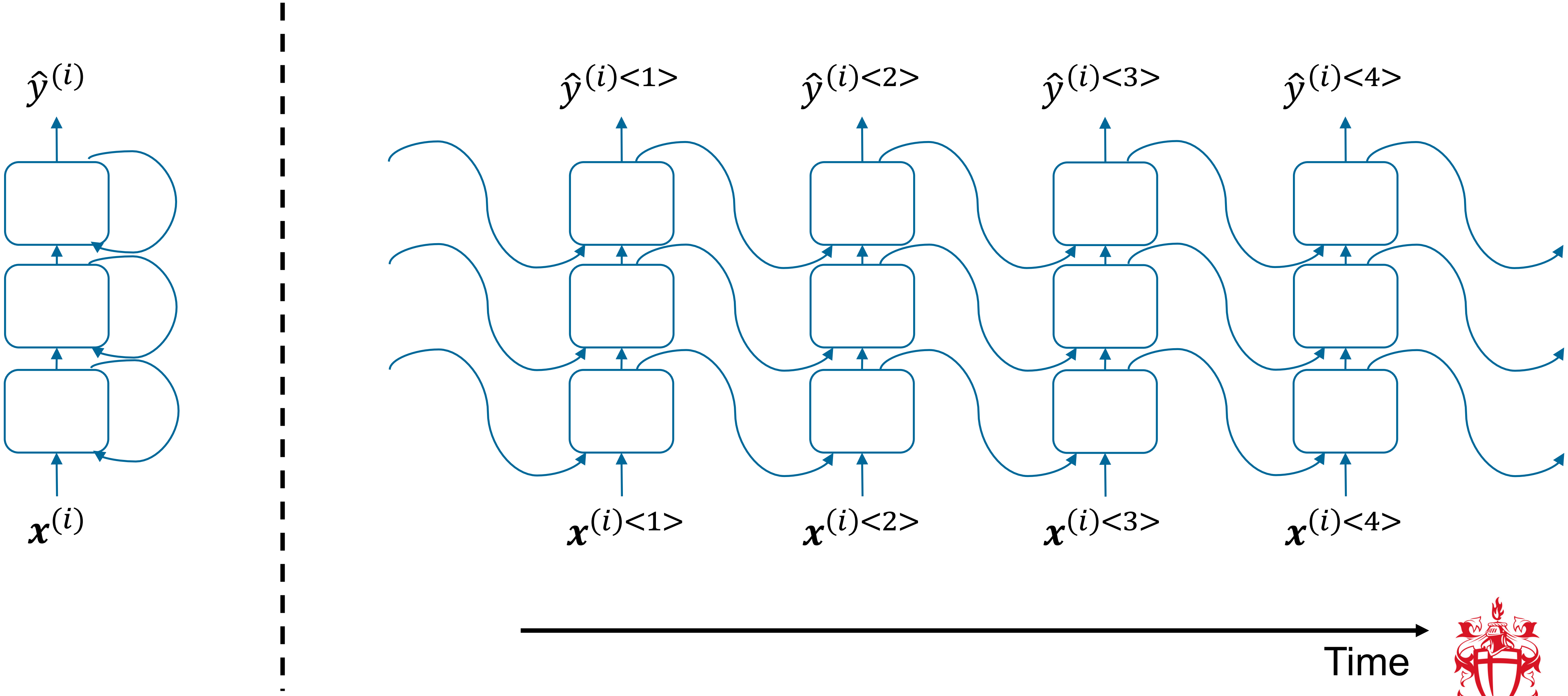
Layers of recurrent neurons – a recurrent neural network (RNN)



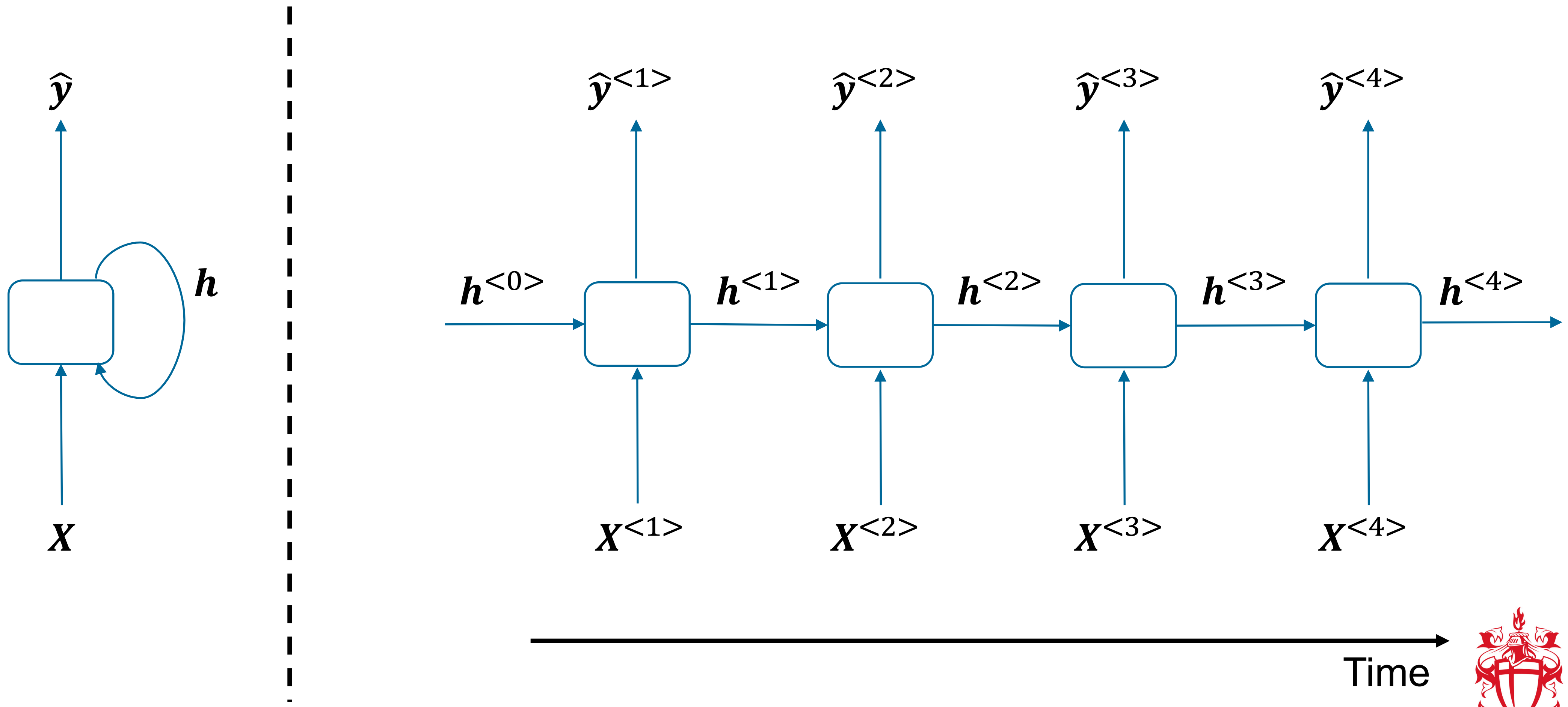
Layers of recurrent neurons – a recurrent neural network (RNN)



Deep RNNs



Representing RNNs and memory more generally



Let’s see more complex recurrent networks

The issues with training RNNs

Problem 1 – vanishing and exploding gradients

- In principle, same as with other networks

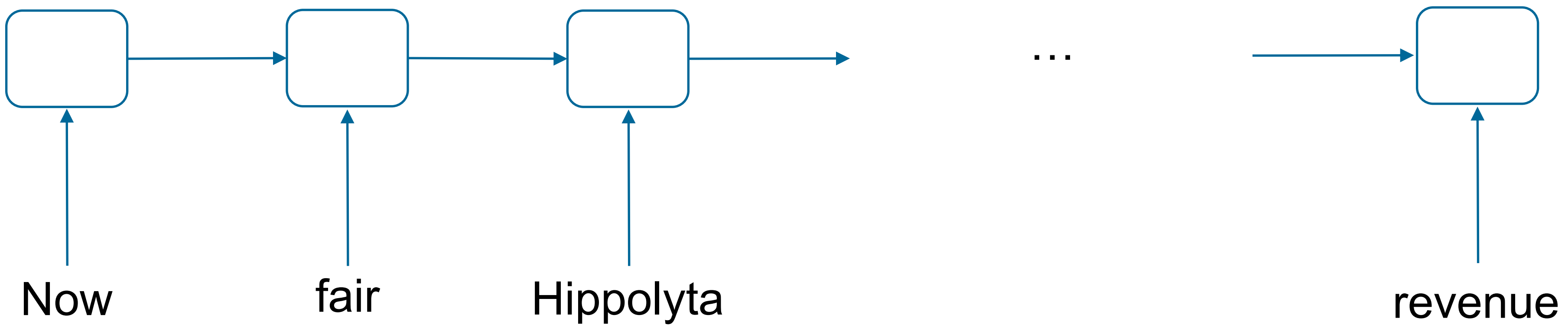
- Before, we mostly focused on vanishing gradients
  → use of non-saturating activation functions such as ReLU

- With RNNs, exploding gradients become more of a problem
  - Same weights used for different time steps can lead to self-reinforcing increases of gradients

  → We frequently use saturating activation functions, such as tanh, or other methods such as gradient clipping

Problem 2 – memory issues

- Vanishing gradients are still a problem (sometimes even more so than in other networks):



- This is essentially a very very deep neural network!
  - Some information is lost at each time step

- After just a few time steps, there is virtually no more information about the first input

When memory loss can be a problem

The BA students, which had been working for days on end, was finally done with their projects.

BAYES
BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

When memory loss can be a problem

The BA students, which had been working for days on end, was finally done with their projects.

BAYES
BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Adding long-term memory

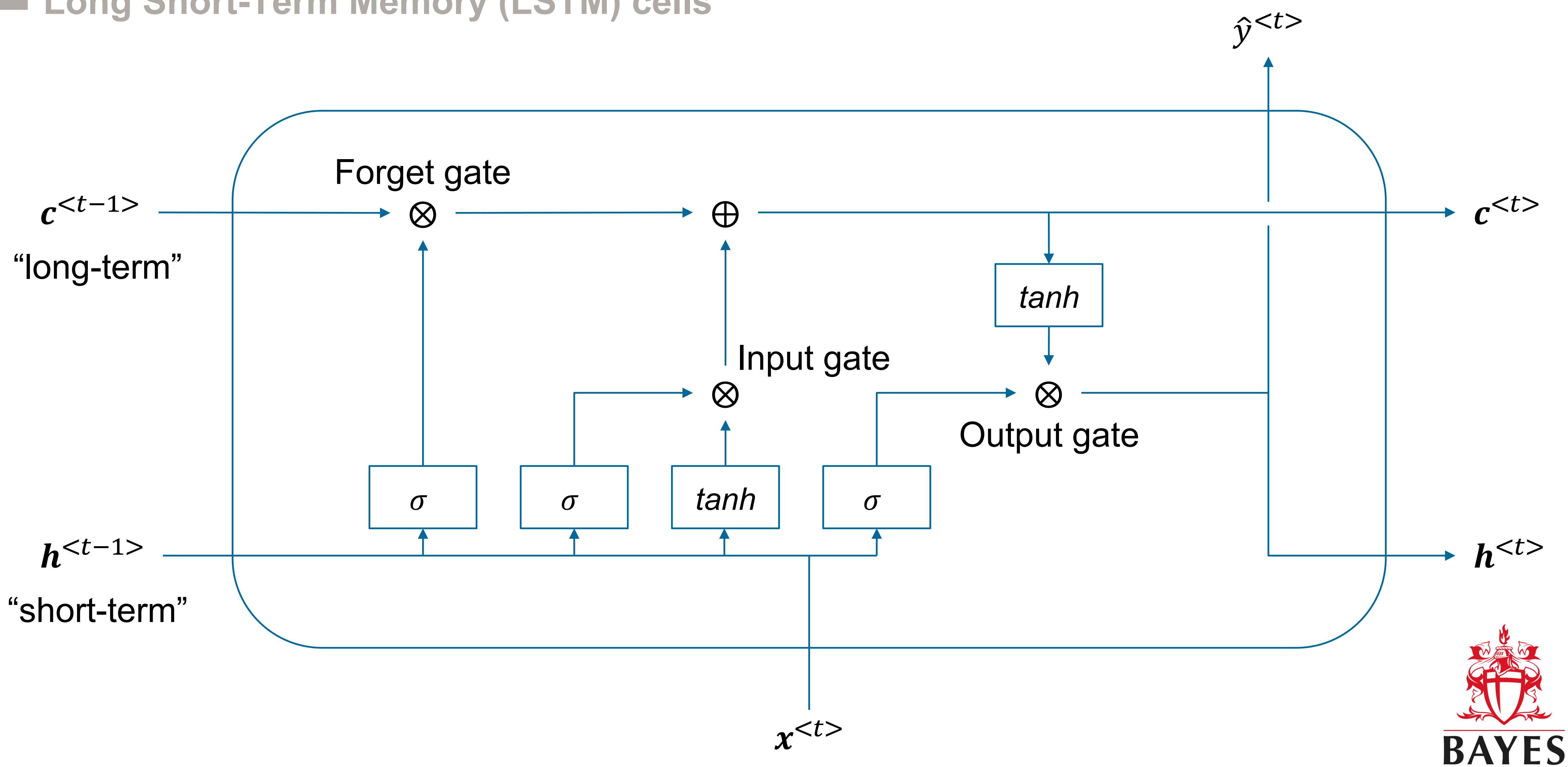
General idea

- Knowledge acquired so far as a state that is managed

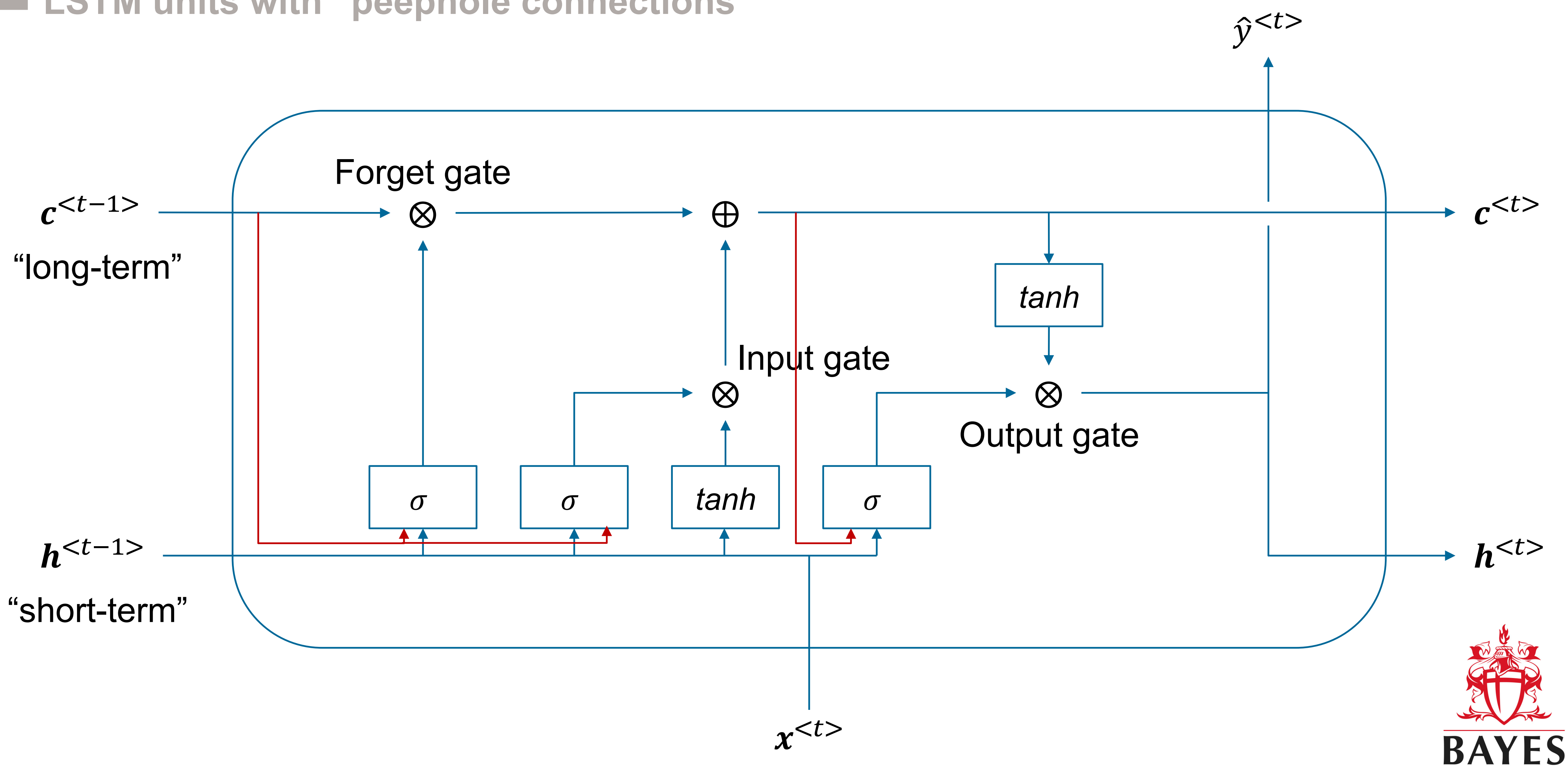
- Use “gates” to add or remove information in each recurrent unit
  - Remove information that is no longer relevant
  - Selectively add information from current input that will be relevant later down the line

- Output based on the state and the input

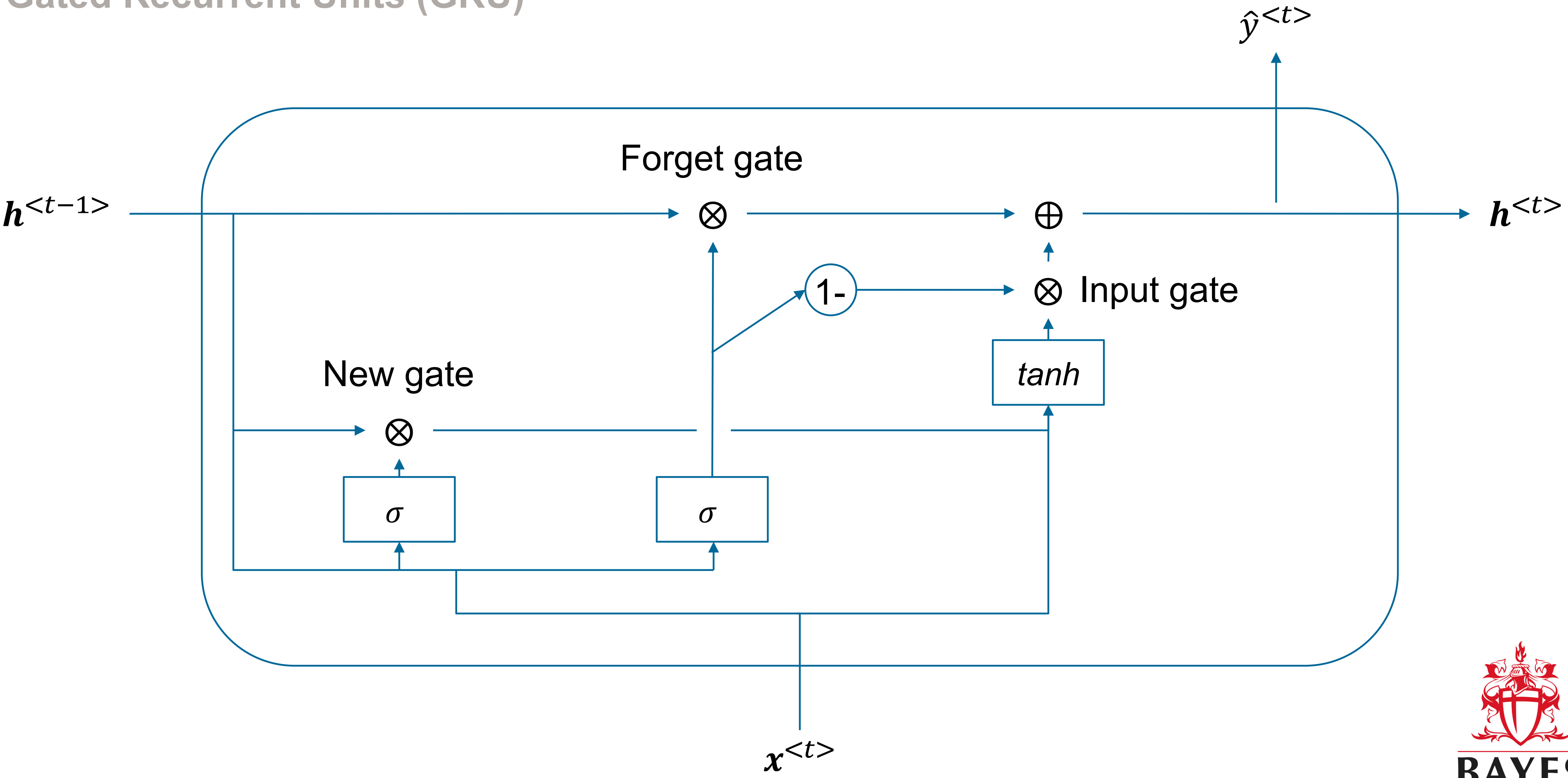
Long Short-Term Memory (LSTM) cells



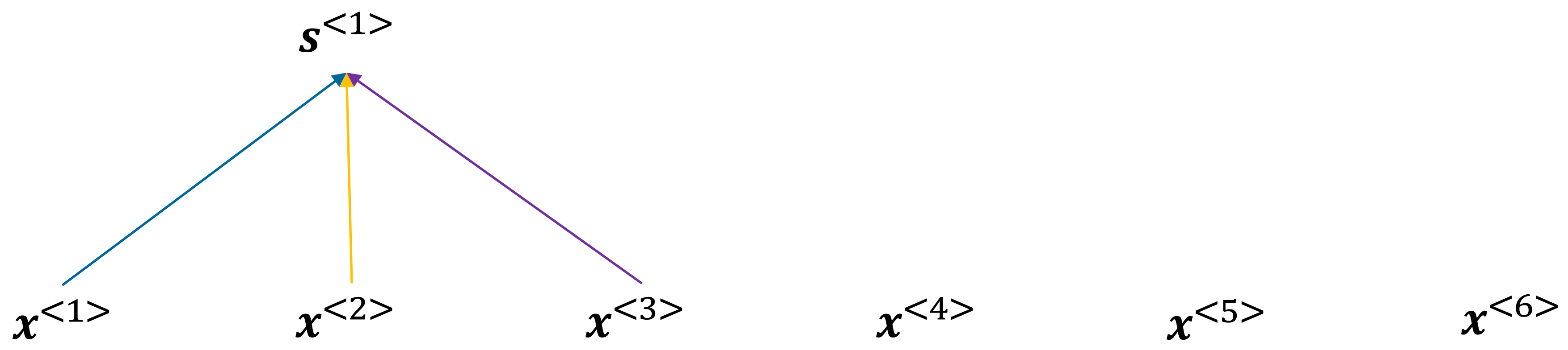
LSTM units with “peephole connections”



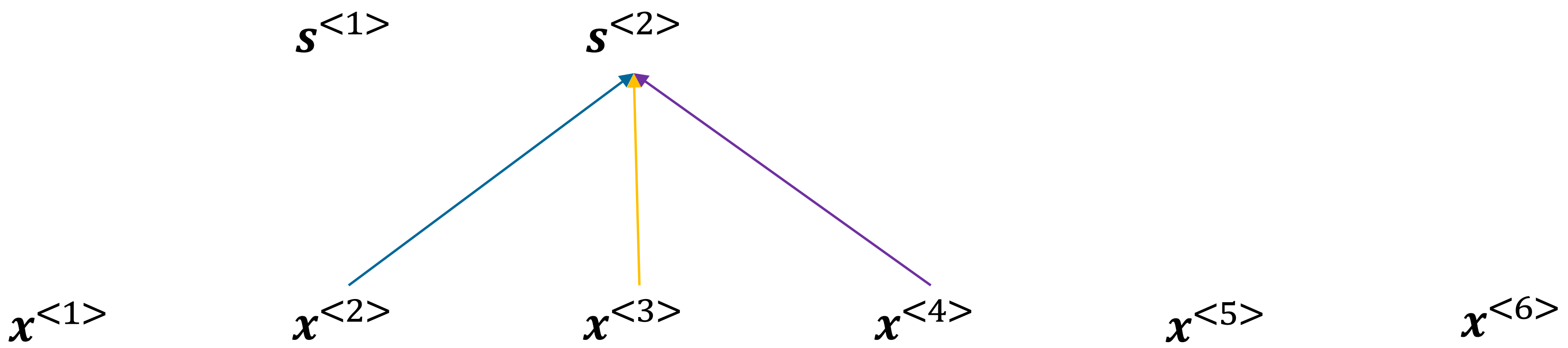
Gated Recurrent Units (GRU)



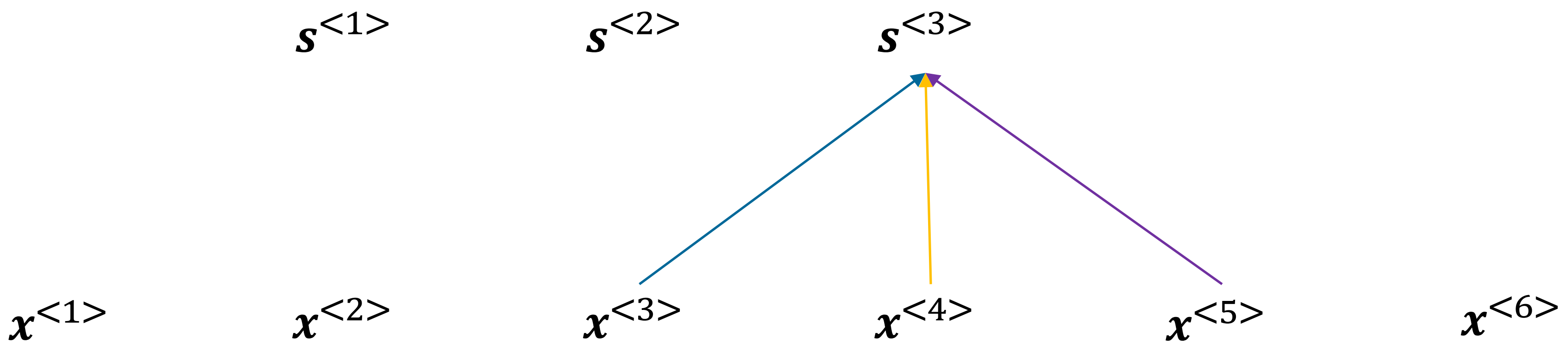
1D convolutional layers



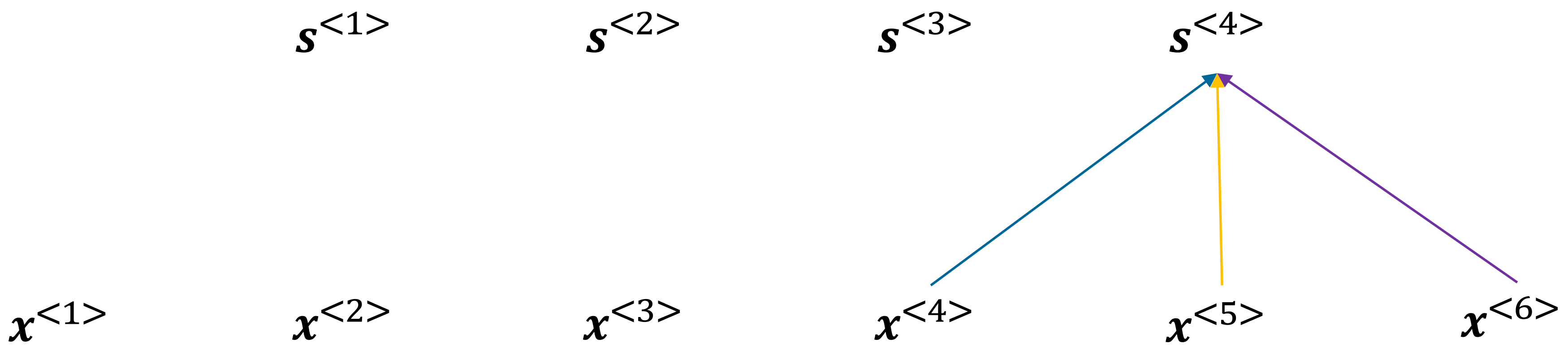
1D convolutional layers



1D convolutional layers



1D convolutional layers

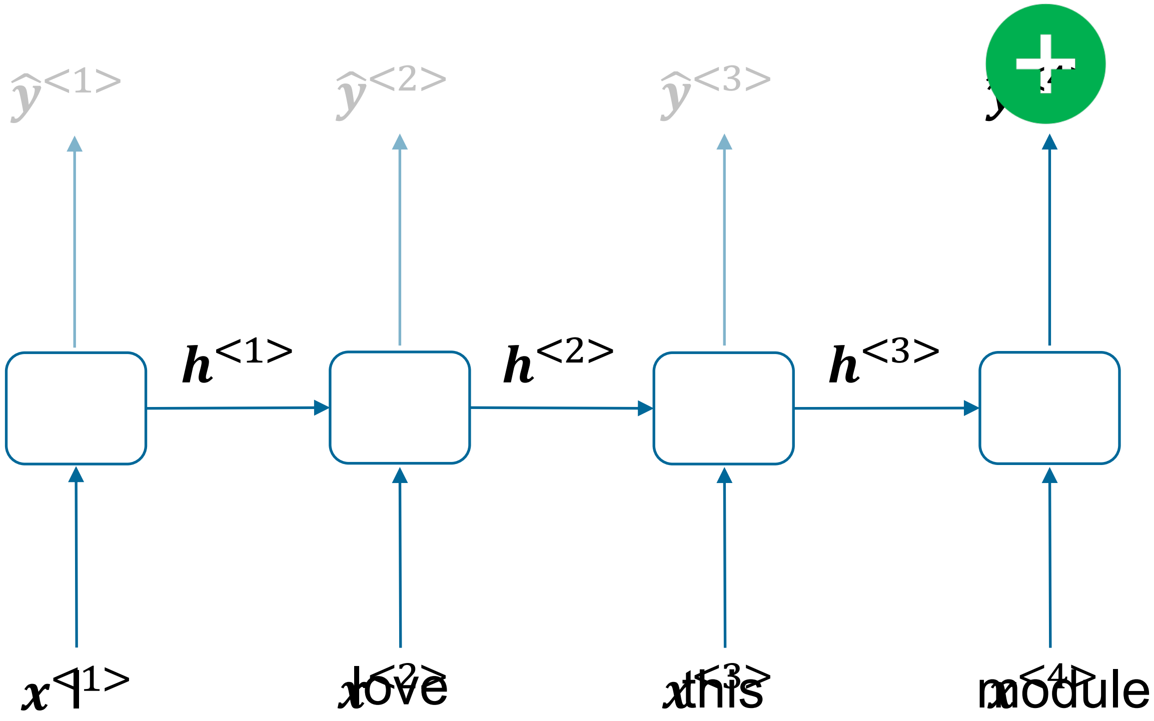


BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Long-term memory in practice

RNN variants and their applications

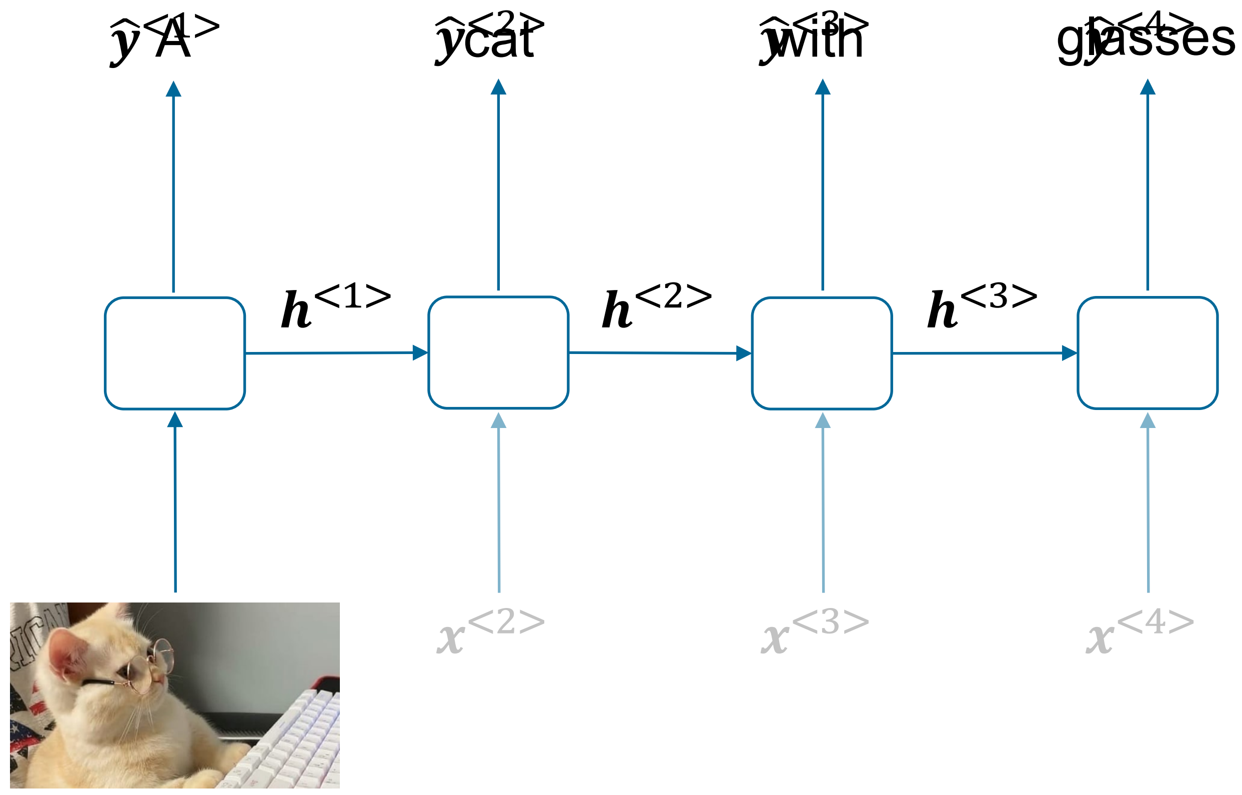
Sequence-to-vector networks



For example:

- Video activity recognition
- DNA sequence probing
- Sentiment classification

Vector-to-sequence networks

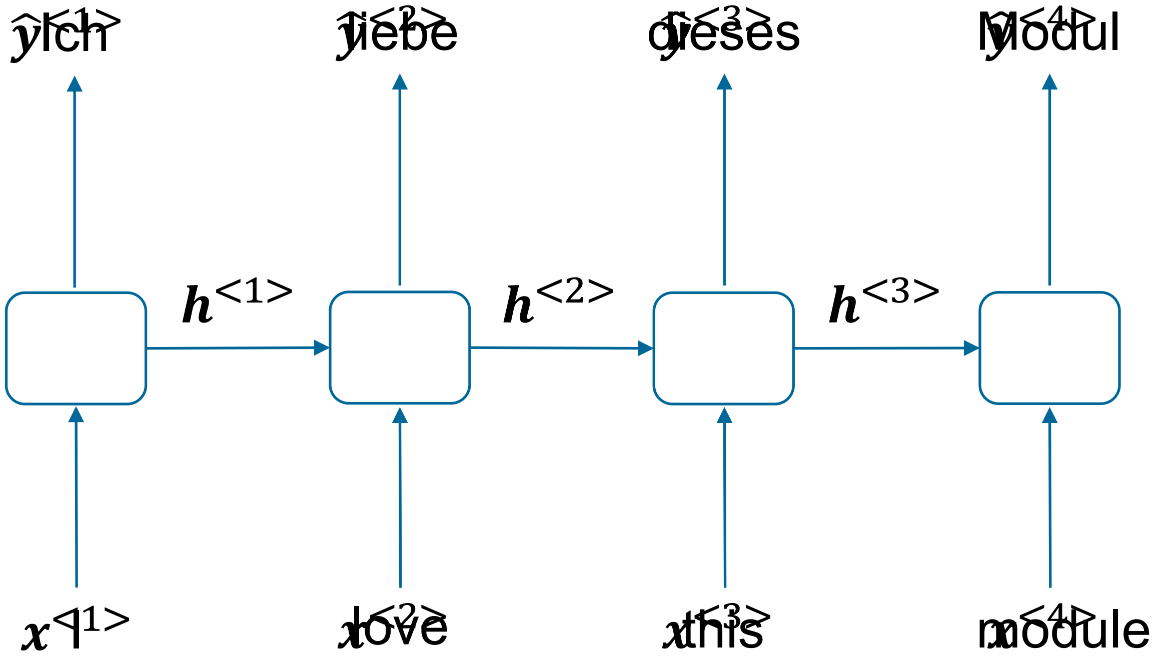


For example:

- Text generation
- Music generation
- Image captions

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Sequence-to-sequence networks

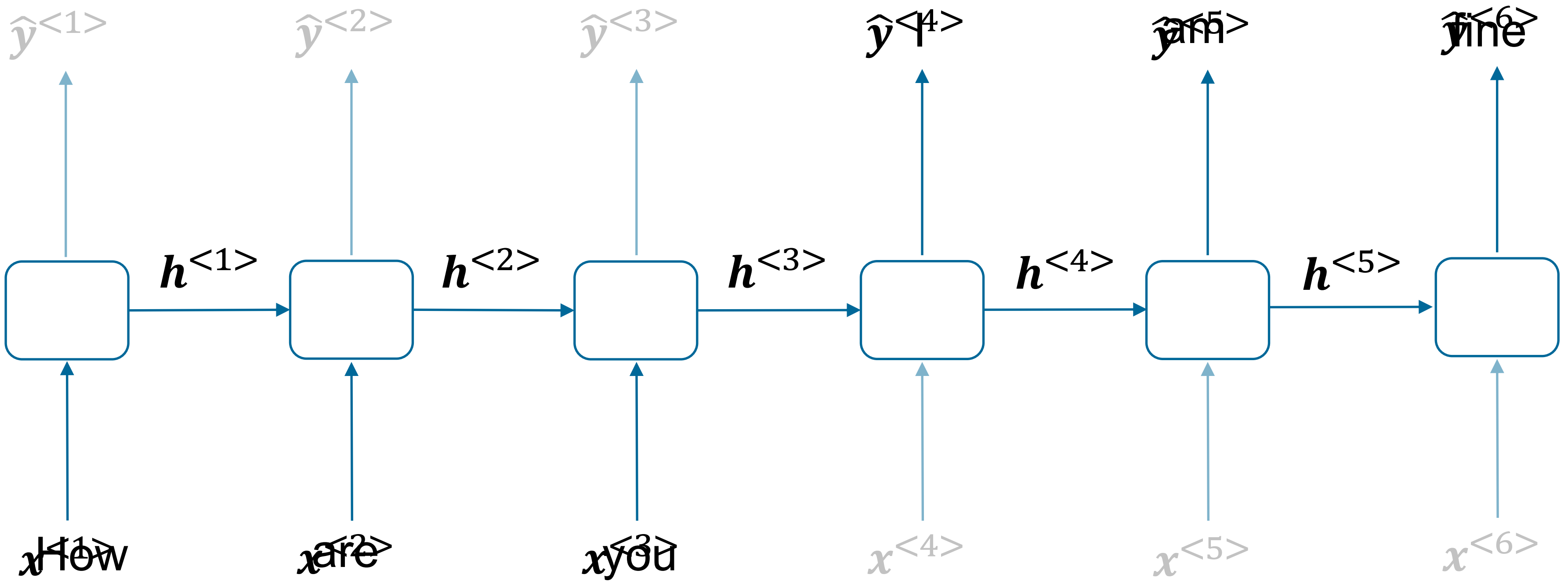


For example:

- Speech recognition
- Price predictions
- Translations

Let's use a sequence-to-sequence network to forecast multiple time steps ahead

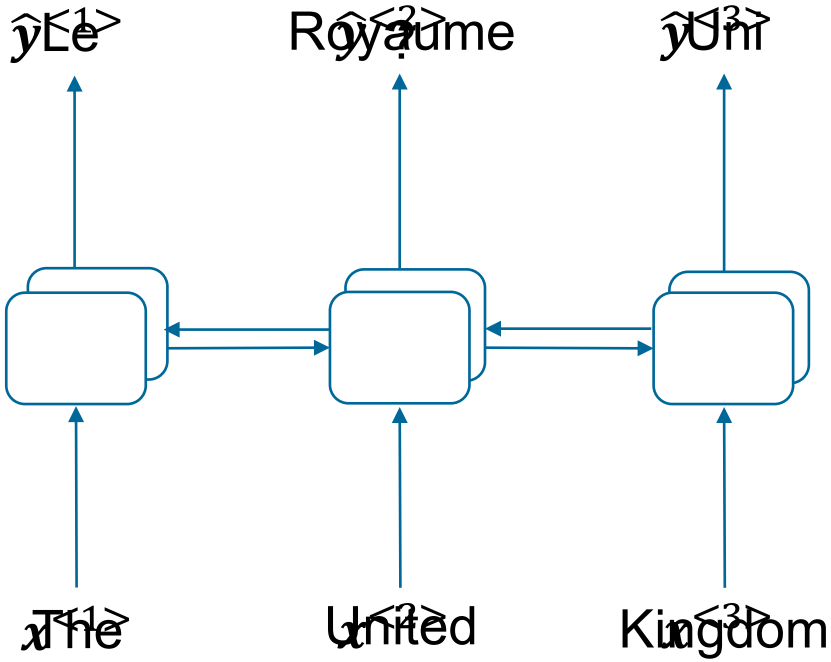
Encoder-decoder networks



For example:

- Translations
- Dialogue

Bidirectional RNNs – looking into the future



For example:

- All sorts of NLP
- Also, in combination with the previous

Data handling with TensorFlow

The data problem with deep learning

Complex data preprocessing

Instead of a “normal” dataset, work with TensorFlow’s Data API

The TensorFlow Data API

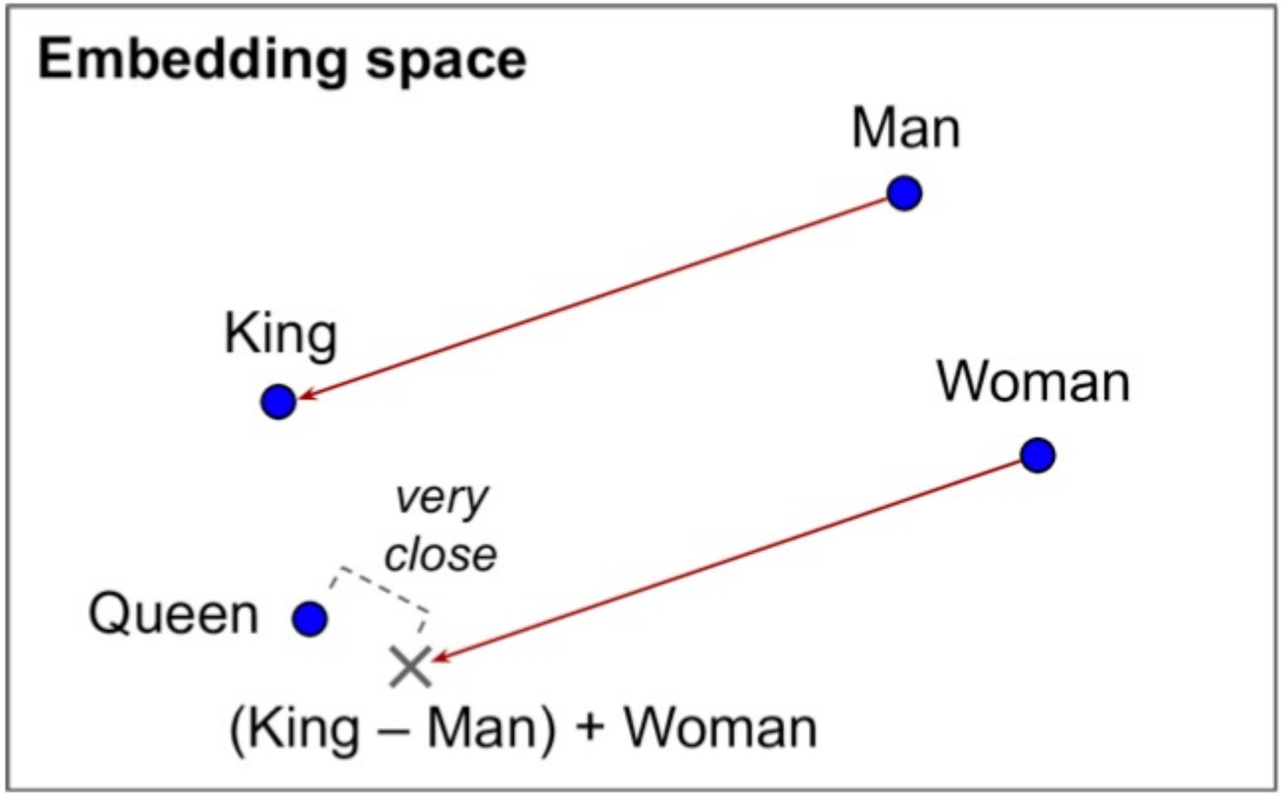
- Go through part 2 of the notebook "*ADL_Week 9_Recturrent Neural Networks.ipynb*"

- We introduce some of the key functionalities of the TensorFlow Data API

- This is useful for models in general, but it is particularly important for RNNs, since we need to do a lot of data manipulation

Creating music with RNNs – time permitting

A note on embeddings



Let's take a look: https://projector.tensorflow.org

Source: Géron

Let’s be creative

- Go through part 3 of the notebook “*ADL_Week 9_Recturrent Neural Networks.ipynb*”

- We develop a model that learns to predict chorales written by Joghann Sebastian Bach… and can create new ones in the same style!

- In principle, this is an extension of what we have learned so far about RNNs. However, there are a few details on data wrangling that make it an important exercise

See you next week!

Sources

- DeepLearning.AI, n.d.: deeplearning.ai
- Garnelo, 2020, Lecture 6: Sequences and Recurrent Networks: https://storage.googleapis.com/deepmind-media/UCLxDeepMind_2020/L6%20-%20UCLxDeepMind%20DL2020.pdf
- Géron, 2019, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow
- Goodfellow, Bengio, Courville, 2016, The Deep Learning Book: http://www.deeplearningbook.org
- Liang, 2016, Introduction to Deep Learning: https://www.cs.princeton.edu/courses/archive/spring16/cos495/
- Soleimany, 2022, Deep Sequence Modeling: http://introtodeeplearning.com/slides/6S191_MIT_DeepLearning_L2.pdf

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON