DEEP LEARNING WITH KERAS WORKSHOP

MUHAMMAD RAJABINASAB @ TARBIAT MODARES UNIVERSITY

Chapter 9 :

Recurrent Neural Networks

WHAT ARE THE RECURRENT NEURAL NETWORKS

- In traditional neural networks, data passes sequentially through the network from the input layer, and through the hidden layers to the output layer
- Information passes through the network once and the outputs are considered independent of each other, and only dependent on the inputs to the model
- However, there are instances where a particular output is dependent on the previous output of the system

USE CASE EXAMPLES

- Consider the stock price of a company as an example, the output at the end of any given day is related to the output of the previous day
- Similarly, in Natural Language Processing (NLP), the final words in a sentence are highly dependent on the previous words in the sentence if the sentence is to make grammatical sense
- A special type of neural network, called a Recurrent Neural Network (RNN), is used to solve these types of problems where the network needs to remember previous outputs

SEQUENTIAL MEMORY AND SEQUENTIAL MODELING

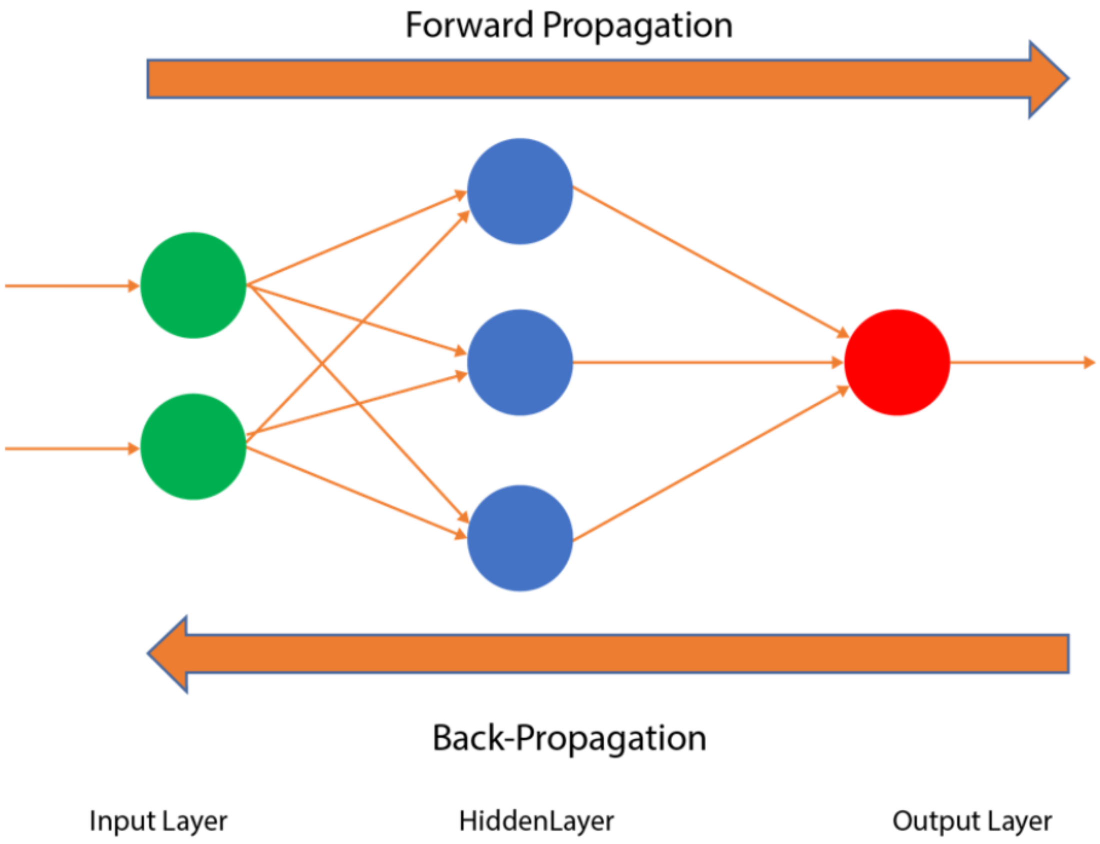
- If we analyze the stock price of Alphabet for the past 6 months, we can see that there is a trend
- To predict or forecast future stock prices, we need to gain an understanding of this trend and then do our mathematical computations while keeping this trend in mind

SEQUENTIAL MEMORY AND SEQUENTIAL MODELING

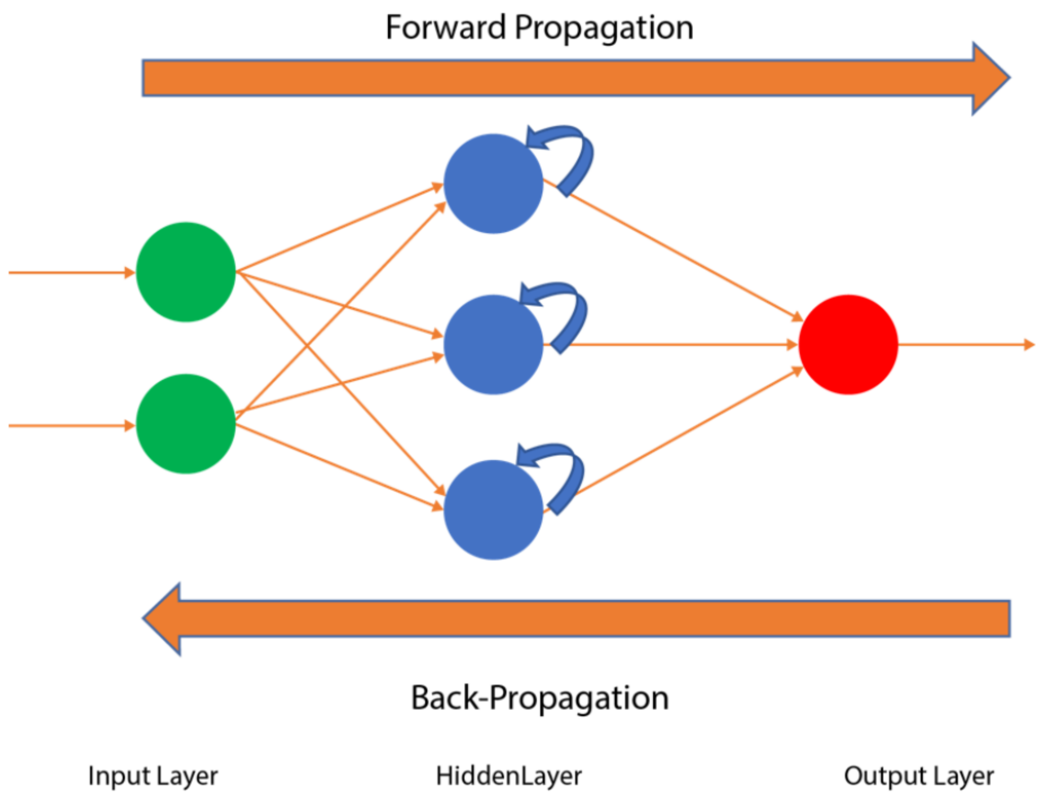
RECURRENT NEURAL NETWORKS

- RNNs are a class of neural networks that are built on the concept of sequential memory
- Unlike traditional neural networks, an RNN predicts the results in sequential data
- Currently, an RNN is the most robust technique that's available for processing sequential data

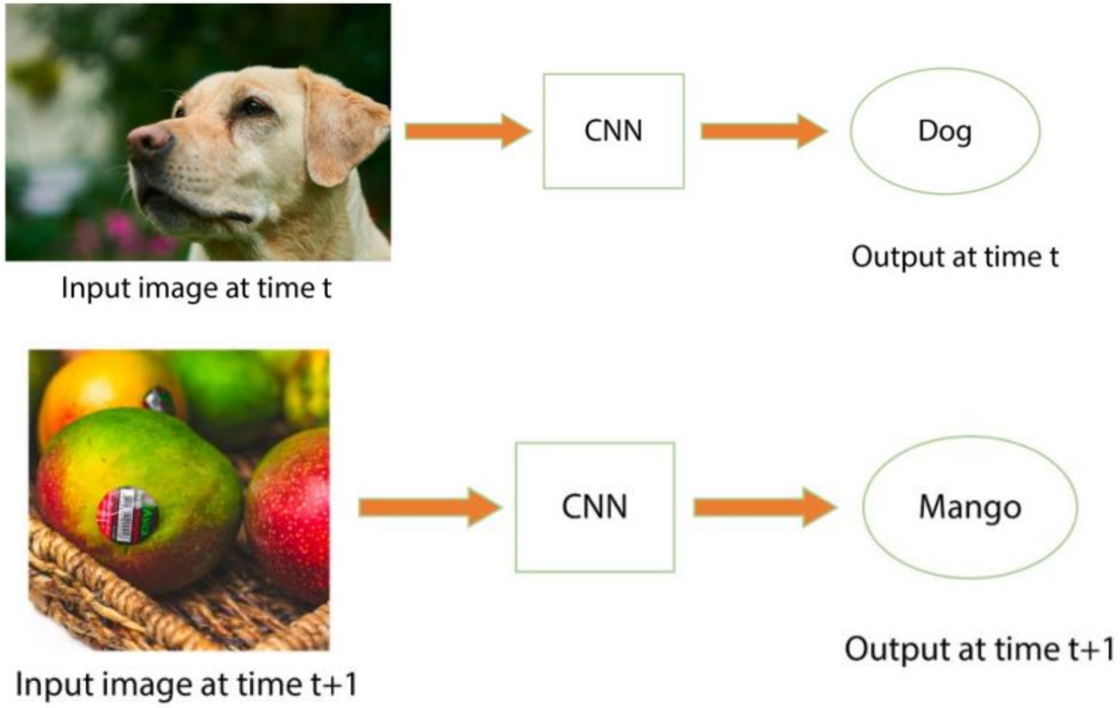
TRADITIONAL NEURAL NETWORKS



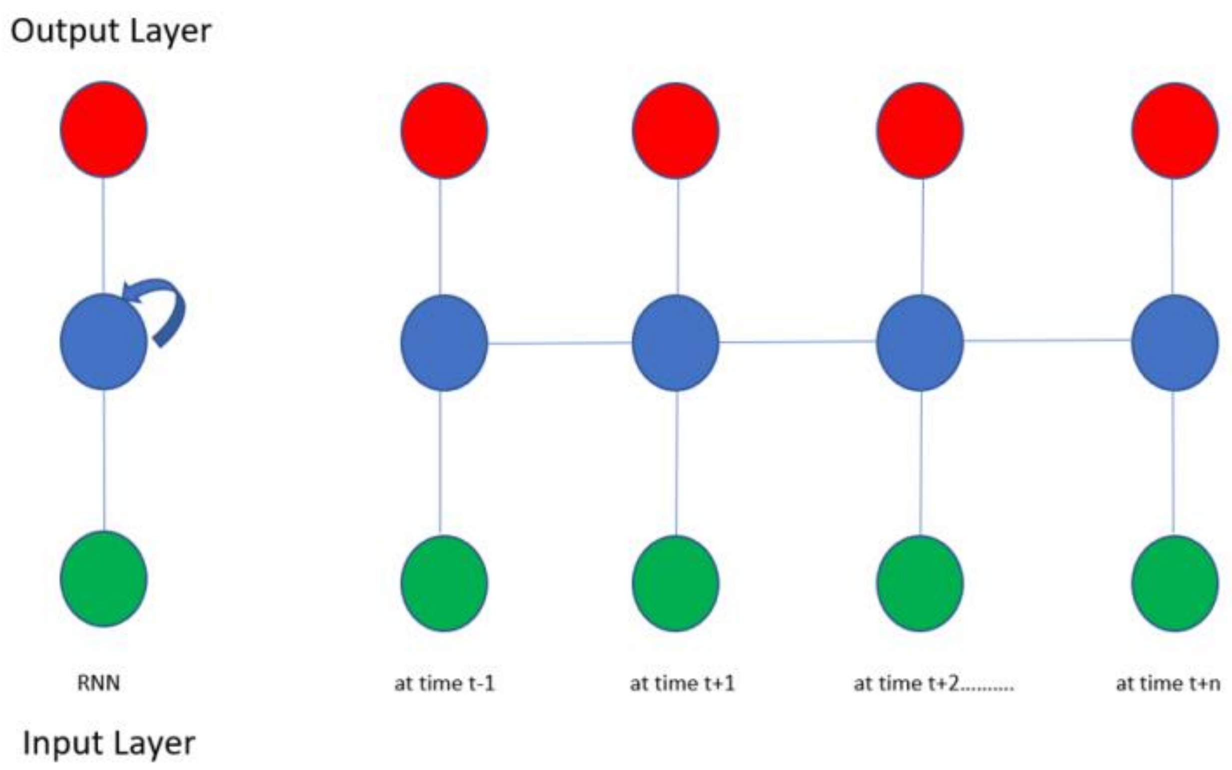
RECURRENT NEURAL NETWORKS



DEPENDING ON TIME



DEPENDING ON TIME



PROBLEMS WITH THE RECURRENT NEURAL NETWORKS

- There are some issues that you may face when training RNNs that are related to the unique architecture of RNNs
- They concern the value of the gradient because, as the depth of the RNN increases, the gradient can either vanish or explode

THE VANISHING GRADIENT PROBLEM

- The gradient is usually a small number between 0 and 1
- The product of two numbers between 0 and 1 gives you an even smaller number
- The deeper your network is, the smaller the gradient is in the initial layers of the network
- In some cases, it reaches a point that is so small that no training happens in that network; this is the vanishing gradient problem

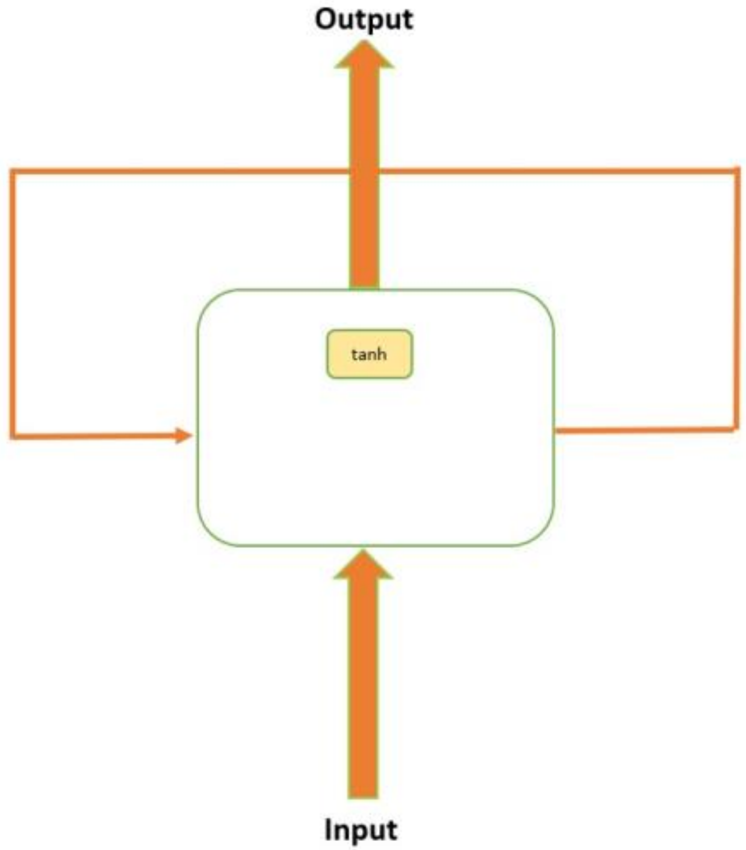
THE EXPLODING GRADIENT PROBLEM

- If instead of the weights being small, the weights are greater than 1, then the subsequent multiplication will increase the gradient exponentially; this is known as the exploding gradient problem
- The exploding gradient is simply the opposite of the vanishing gradient as in the case of the vanishing gradient, the values become too small, while in the case of the exploding gradient, the values become very large
- As a result, the network suffers heavily and is unable to predict anything
- But how to fix these issues?

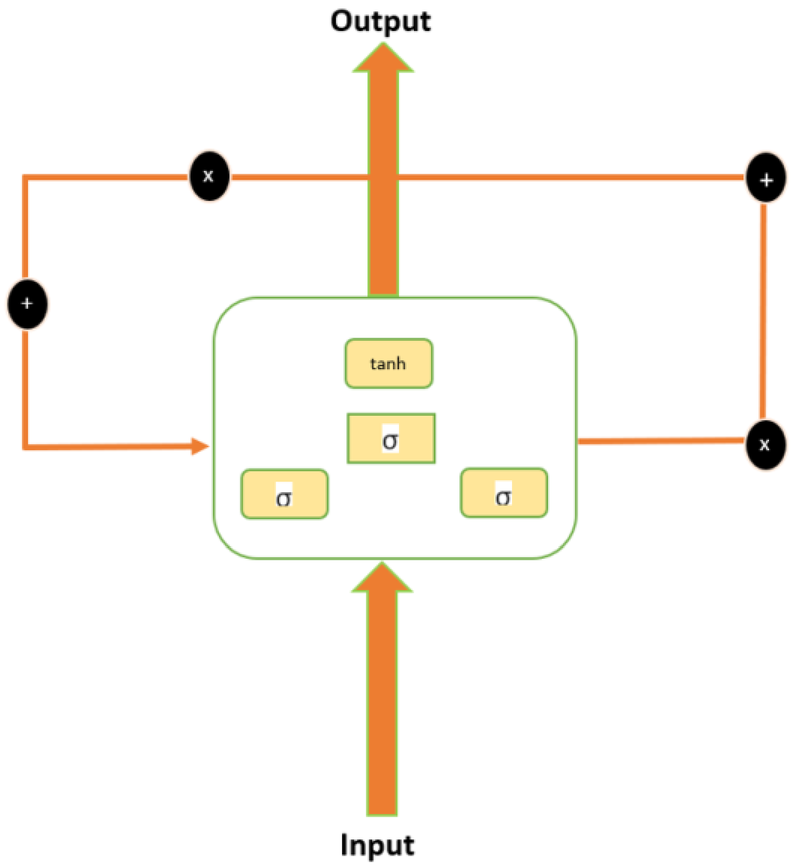
LONG SHORT-TERM MEMORY (LSTM)

- LSTMs are RNNs whose main objective is to overcome the shortcomings of the vanishing gradient and exploding gradient problems
- LSTM networks are a special kind of RNN that are capable of learning long-term dependencies
- They are designed to avoid the long-term dependency problem; being able to remember information for long intervals of time is how they are wired

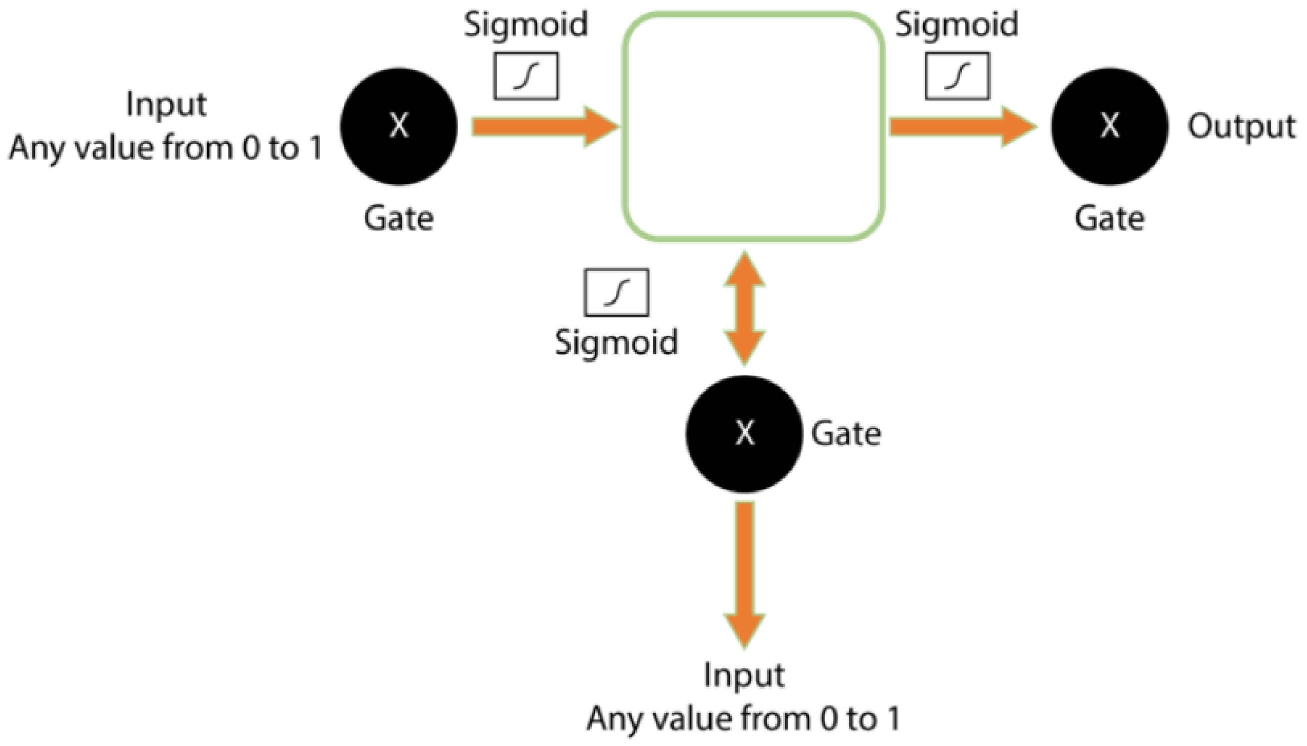
SIMPLE RNN ARCHITECTURE



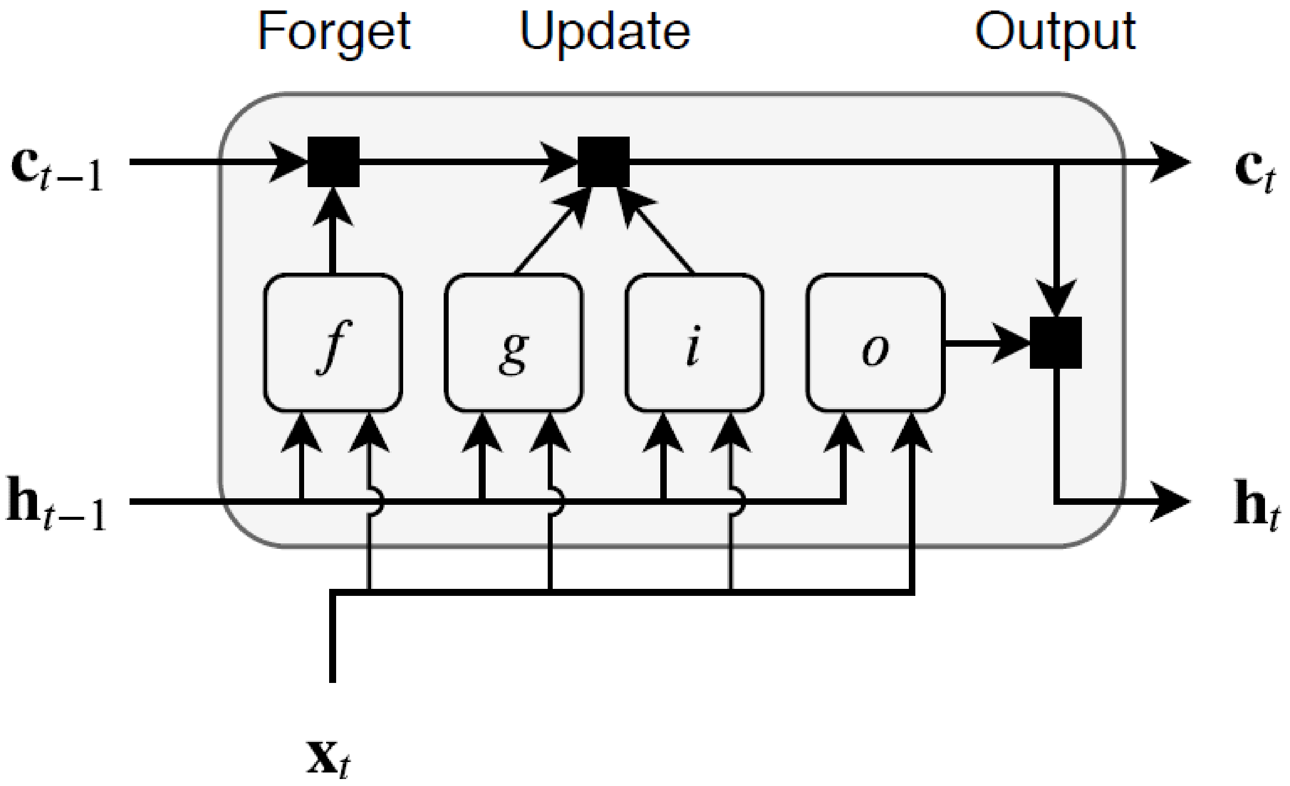
LSTM ARCHITECTURE



AN LSTM IN DETAIL



AN LSTM IN EVEN MORE DETAIL



GATED RECURRENT UNITS

- Gated recurrent units (GRUs) are a gating mechanism in recurrent neural networks
- The GRU is like a long short-term memory (LSTM) with a forget gate but has fewer parameters than LSTM, as it lacks an output gate
- GRU's performance on certain tasks of polyphonic music modeling, speech signal modeling and natural language processing was found to be similar to that of LSTM
- GRUs have been shown to exhibit better performance on certain smaller and less frequent datasets

GATED RECURRENT UNITS

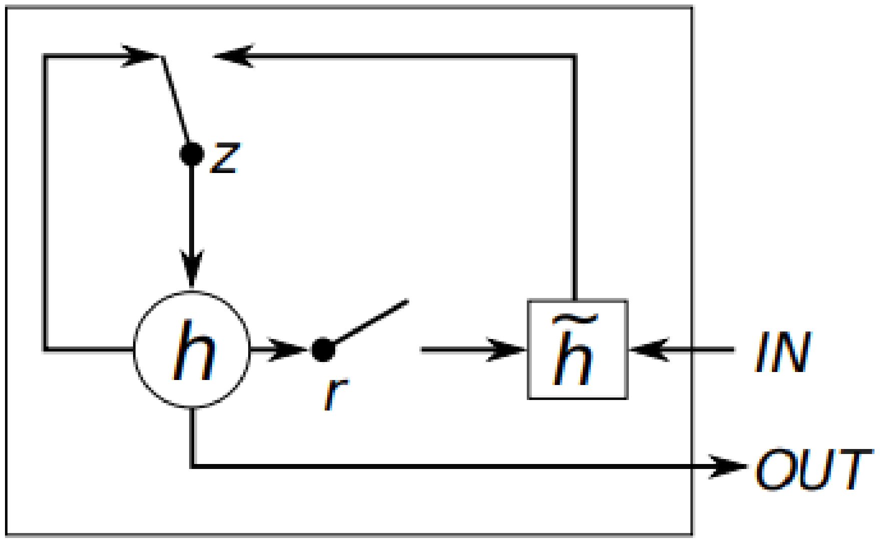


Fig: Gated Recurrent Unit

EXERCISE 9.01: PREDICTING THE TREND OF ALPHABET'S STOCK PRICE USING AN LSTM WITH 50 UNITS

- In this exercise, we will examine the stock price of Alphabet over a period of 5 years— that is, from January 1, 2014, to December 31, 2018
- In doing so, we will try to predict and forecast the company's future trend for January 2019 using RNNs
- We have the actual values for January 2019, so we will be able to compare our predictions with the actual values later

ACTIVITY 9.01: PREDICTING THE TREND OF AMAZON'S STOCK PRICE USING AN LSTM WITH 50 UNITS

- In this activity, we will examine the stock price of Amazon for the last 5 years—that is, from January 1, 2014, to December 31, 2018
- In doing so, we will try to predict and forecast the company's future trend for January 2019 using an RNN and LSTM
- We have the actual values for January 2019, so we can compare our predictions to the actual values later

EXERCISE 9.02: PREDICTING THE TREND OF ALPHABET'S STOCK PRICE USING AN LSTM WITH 100 UNITS

- In this exercise, we will examine the stock price of Alphabet over a period of 5 years— that is, from January 1, 2014, to December 31, 2018
- In doing so, we will try to predict and forecast the company's future trend for January 2019 using RNNs
- We have the actual values for January 2019, so we will be able to compare our predictions with the actual values later

ACTIVITY 9.02: PREDICTING AMAZON'S STOCK PRICE WITH ADDED REGULARIZATION

- In this exercise, we will examine the stock price of Alphabet over a period of 5 years— that is, from January 1, 2014, to December 31, 2018
- In doing so, we will try to predict and forecast the company's future trend for January 2019 using RNN and LSTM
- We have the actual values for January 2019, so we will be able to compare our predictions with the actual values later
- Here, we will also add dropout regularization and compare the results with Activity 9.01

ACTIVITY 9.03: PREDICTING THE TREND OF AMAZON'S STOCK PRICE USING AN LSTM WITH AN INCREASING NUMBER OF LSTM NEURONS

- In this exercise, we will examine the stock price of Alphabet over a period of 5 years— that is, from January 1, 2014, to December 31, 2018
- In doing so, we will try to predict and forecast the company's future trend for January 2019 using RNNs
- We have the actual values for January 2019, so we will be able to compare our predictions with the actual values later

SUMMARY

- In this chapter, we learned about sequential modeling and sequential memory
- We learned how sequential modeling is related to RNNs, as well as how RNNs are different from traditional feedforward networks
- We learned about the vanishing gradient problem
- In detail and how using an LSTM is better than a simple RNN to overcome the vanishing gradient problem
- We applied what we learned to time series problems by predicting stock trends
- We also briefly learned about GRUs

THANK YOU FOR YOUR ATTENTION!