Chapter 2

Literature Review

A literature review is a type of review article that contains the sufficient description about the thesis. A literature review is a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic like definition of the various terms, Mathematical background, technology that have applied for implementing work, Algorithm etc.

Stock Market:

The stock market refers to the collection of markets and exchanges where the issuing and trading of equities or stocks of publicly held companies, bonds, and other classes of securities take place. This trade is either through formal exchanges or over-the-counter (OTC) marketplaces which also known as the equity market. It provides companies with access to capital in exchange for giving investors their be worth ownership.

*Companies Sell Stocks:*

Stocks are how companies manage capital to grow larger. Usually, when one wants to start a business, they pay for it with loans or even their own credit. Once they grow the company enough, they can get loans, or also float their bonds to individual or group of investors.

Eventually, they'll need a lot of capital to take the business to the next phase. At that time, they will sell the first stocks, called an initial public offering after that no single person owns the company because they have sold it to the stockholders.

*Invest in the Stock Market:*

Stock market investing is one of the best way to achieve returns that beat inflation over time. There are several benefits like stock ownership takes advantage of a growing economy. Best of all, one can make money in several ways like some investors prefer to let their stock appreciate in value over time or others prefer stocks that pay dividends to provide a steady income stream. Instead of buying individual stocks, one could buy them as part of an index fund which is called mutual fund.

*Relation to modern financial system:*

The financial system in most modern countries has gone a remarkable evaluation. A portion of the capital involved in saving and financing, flows directly to the financial markets instead of being routed via the traditional bank lending and deposit operations and thus the general public interest in investing in the stock market, either directly or indirectly has been an important component of this process. Statistics show that in recent years, shares have made up an increasingly large proportion of households' financial assets in many areas. A second evaluation is the move to electronic trading to replace human trading of listed companies.

Artificial Neural Networks:

Artificial neural networks are one of the main tools used in artificial intelligence. As the “neural” part of their name suggests, they are animal brain-inspired systems which are intended to replicate the way that animal learn. Neural networks consist of input and output layers, as well as (in most cases) a hidden layer consisting of units that transform the input into something desired that the output layer can use. They are excellent techniques for finding patterns which are far too complicated or numerous for a human programmer to extract and teach the machine to recognize.

*Biological and Artificial Neurons:*

An ANN is based on a collection of connected units or nodes called artificial neurons which is analogous to biological neurons in an animal brain. Each connection between artificial neurons can pass a signal from one to another. The artificial neuron that receives the signal can process it and then transmit the signal to the artificial neurons connected to it.

A biological neuron is an electrically excitable cell that processes and transmits information through electrical and chemical signals where a chemical signal occurs via a synapse, a specialized connection with other cells. Neurons connect to each other to form neural networks and obviously neurons are the core components of the nervous system, which includes the brain, spinal cords, and peripheral ganglia.

An artificial neuron is a mathematical function conceived as an abstraction of animal neurons. The artificial neuron receives one or more inputs (representing the dendrites) and sums them to produce an output (representing the axon) where the sums of each node are weighted, and the sum is passed through a non-linear function known as an activation function.

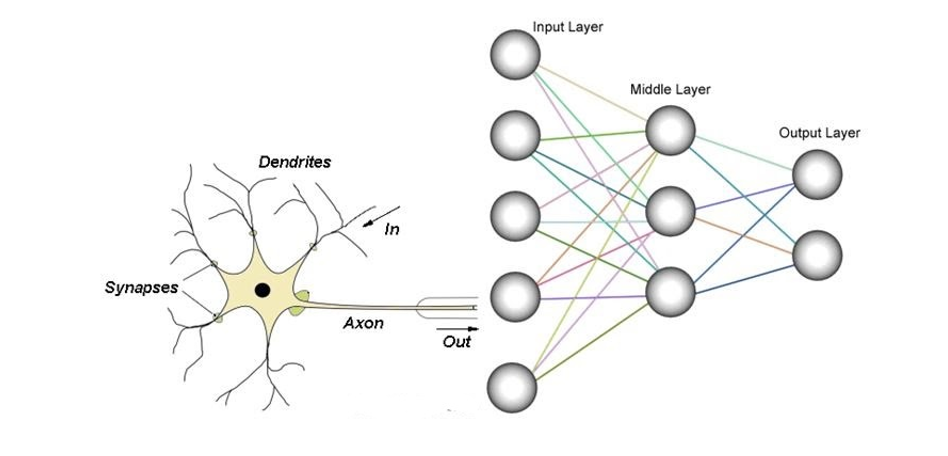


Fig: biological and artificial neurons

The original goal of the ANN approach was to solve problems in the same way that the biological neurons would do where attention focused on matching specific mental abilities, leading to deviations from biology. ANNs have been used on a variety of tasks including computer vision, speech recognition, time series forecasting, financial analysis, machine translation, social network filtering, playing board and video games and medical diagnosis.

*Layers at Artificial Neural Networks:*

There are three different layers in artificial neural networks. They are input layers, hidden layers, output layers.

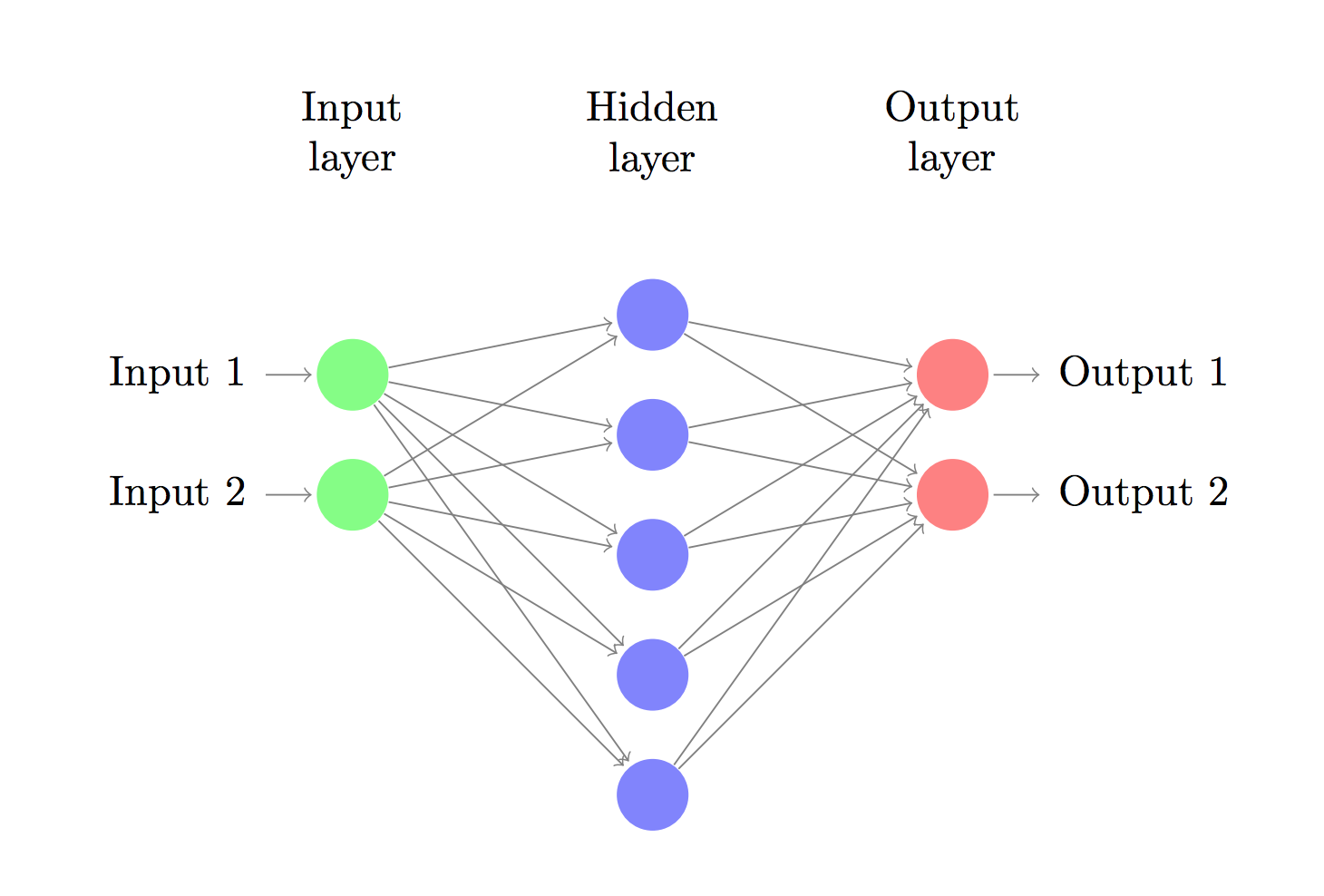
**

Fig: artificial neural networks

Input layers:

The Input layer communicates with the external environment to deal with all the inputs only. This input gets transferred to the hidden layers or output layers. The input layer should represent the features for which we are training the neural network.

Hidden layers:

The hidden layer is the collection of neurons which has activation function applied on it and it is an intermediate layer found between the input layer and the output layer and has no connection with outer state. Its job is to process the inputs obtained by its previous layer and so it is the layer which is responsible extracting the required features from the input data. There can be no, single or multiple hidden layers in a Neural Network.

Output layers;

The output layer of the neural network collects and transmits the information accordingly in way it has been designed to give the results. The number of neurons in output layer should be directly related to the type of work that the neural network was performing give us our desired result.

Deep learning:

Deep learning is a subset of machine learning where machine learning is a subset of artificial intelligence that uses deep layered artificial neural networks to deliver accuracy in tasks such as object detection, speech recognition, language translation, time series analysis, economical analysis and others.

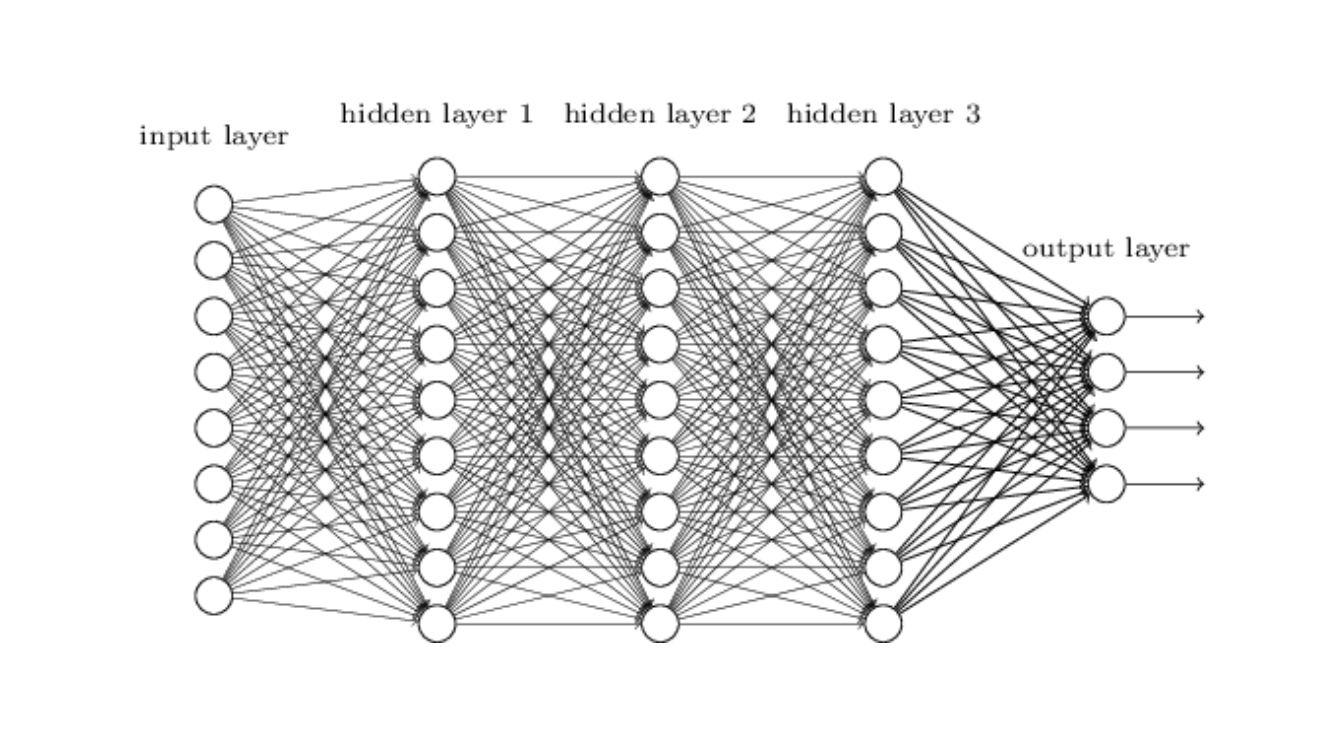


Fig: deep neural networks

*Why Deep Neural Networks:*

Deep networks have achieved accuracies that are far beyond that of classical ML methods in many domains including speech, natural language, computer vision, and playing games, time series forecasting, economical analysis [deep learning]. As better computation power and with lots of data available, deep learning is outclassed and outperformed machine learning in every possible way. In many tasks, classical ML can’t even compete and deep neural networks blows classical ML out of the water.

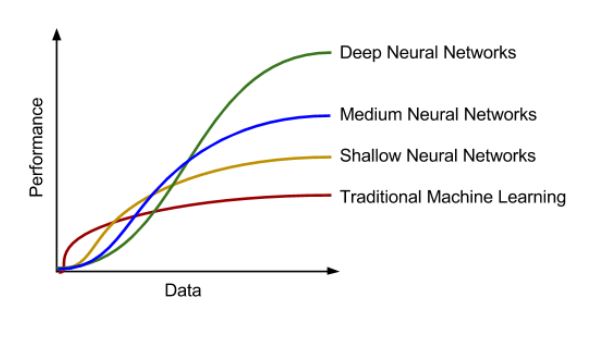


Fig: deep neural networks and others learning

Deep networks scale much better with more data than classical ML algorithms. The figure above is a simple yet effective illustration on how performance increases as the data volume is increased.

Classical ML algorithms often require complex coding for feature extraction as a deep dive exploratory data analysis is first performed on the dataset. After that dimensionality reduction might be done for easier and faster processing and then the best features must be carefully selected to pass over to the machine learning algorithm. As a deep neural networks learn its feature all by itself there’s no need for this when using a deep network as one can just pass the data directly to the input layer and usually achieve better performance. This totally eliminates the big and challenging feature extraction stage of the whole process and reduce lots of complex programming part.

*Supervised and unsupervised deep learning:*

Supervised deep learning is a learning in which we teach or train the networks using data which is well labeled that means some data is already tagged with correct answer. After that, networks is provided with new set of testing data so that supervised deep learning algorithm analyses the training data and produces an correct outcome from labeled data.

Unsupervised deep learning is the training of networks using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of networks is to group unsorted information according to patterns, similarities and differences without any prior training of data.

Table: supervised and unsupervised deep learning

|  |  |  |
| --- | --- | --- |
| Supervised  Deep Learning | Multilayer Perceptron | Used for Regression & Classification |
| Convolutional Neural Networks | Used for Computer Vision |
| Recurrent Neural Networks | Used for Time Series Analysis |
| Unsupervised  Deep Learning | Self-Organizing Maps | Used for Feature Detection |
| Deep Boltzmann Machines | Used for Recommendation Systems |
| Auto Encoders | Used for Recommendation Systems |

Recurrent Neural Network:

The idea behind recurrent neural network is to make use of sequential information. In a traditional artificial neural network all inputs and outputs are independent of each other. But for many tasks that’s a very bad idea like when we try to predict stock market because prices at current time is dependent on previous prices history. RNNs are called *recurrent* because they perform the same task for every element of a sequence, with the output being depended on the previous computations. Another way to think about RNNs is that they have a “memory” which captures information about what has been calculated so far. In theory RNNs can make use of data in arbitrarily long sequences but practically they are limited to looking back only a few steps because of vanishing gradient problem. The idea behind the recurrent neural network are explained with the help of colah’s blog [colah’s blog].

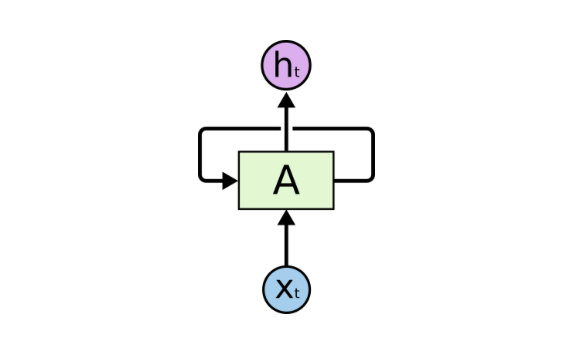


Fig: recurrent neural network

In the above figure, a neural network, ‘***A***’ looks at some input ***xt*** and outputs a value ***ht*** which is also the input of next step and thus a loop allows data to be passed from one step of the network to the next.

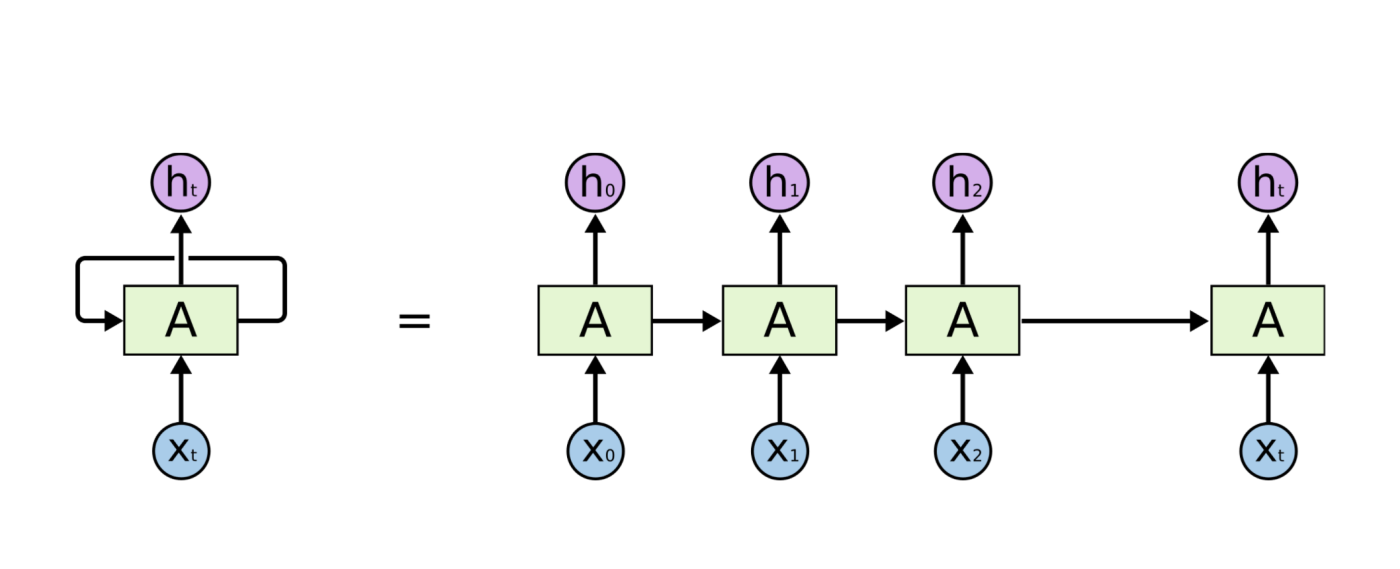


Fig: unrolling the rolled recurrent neural network

A recurrent neural network is called recurrent because of multiple copies of the same network, each passing a message to next network and that chain-like nature is the reason that recurrent neural networks are basically related to sequences and lists like stock market data. So recurrent neural networks have the natural architecture of neural network to use for extracting the hidden pattern of stock market.

*Vanishing gradient problem:*

The vanishing gradient problem is a difficulty found in training recurrent neural network with gradient based learning methods and Backpropagation through time (BPTT). In such methods, each of the neural network's weights receives an update proportional to the partial derivative of the error function with respect to the current weight in each iteration of training. The problem is that as the time of Backpropagation through time (BPTT) will increase, the gradient will be vanishingly small, effectively preventing the weight from changing its value. This may completely stop the weight updating of initial neural network. The problem was explored in depth by Hochreiter (1991) [german paper] and Bengio (1994) [Vanishing gradient problem] who found some main fundamental reasons about this problems.

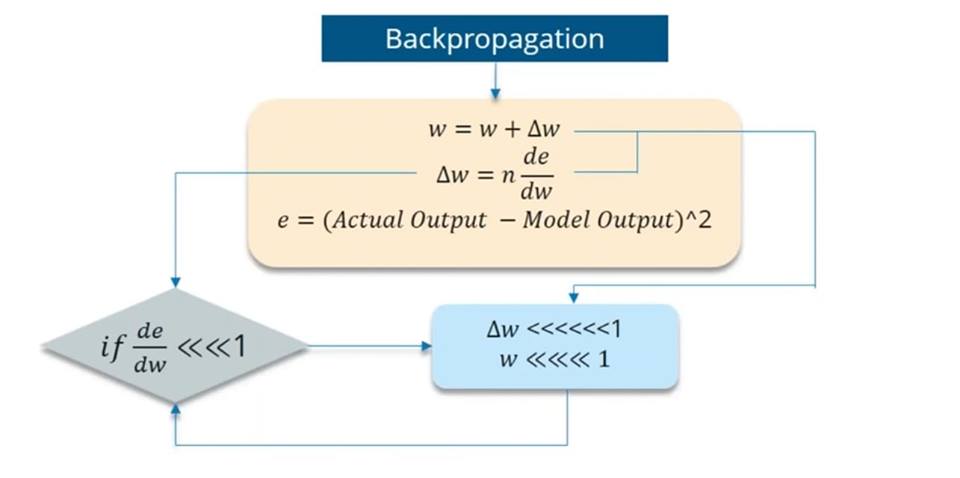


Fig: vanishing gradient problem

Gated Recurrent Unit (GRU) Neural Networks:

Gated Recurrent Unit networks usually just called “GRUs” – are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Cho, et al. (2014) [gru] .They work tremendously well on a large variety of problems and they are one of the most modern, powerful and effective neural networks. GRUs are explicitly designed to avoid the long-term dependency problem GRU has fewer parameter than LSTM and thus may train a bit faster or need less iterations to generalize. Writers of the paper ‘An Empirical Exploration of Recurrent Network Architectures’ showed that the GRU outperformed the LSTM on most tasks with the exception of language modeling [gru better].

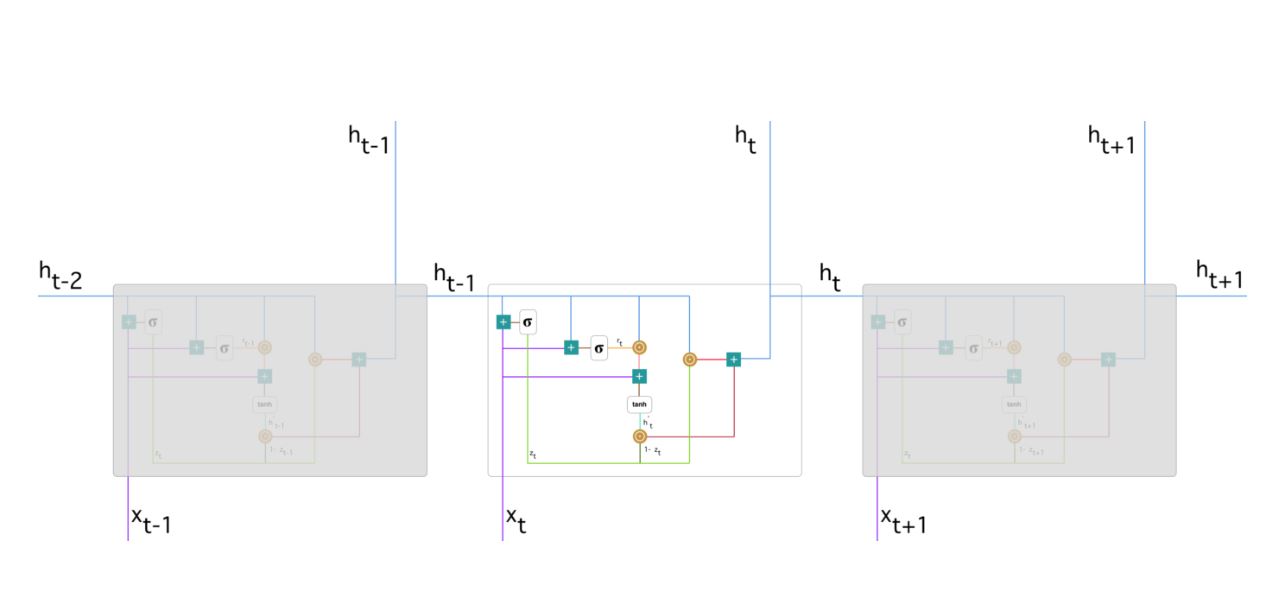


Fig: Gated Recurrent Unit (GRU) Neural Networks

*Internal Architecture of* *Gated Recurrent Unit (GRU) Neural Networks:*

GRU uses update gate and reset gate to solve the vanishing gradient problem of a standard RNN. Actually update gate and reset gate are two vectors which decide what information should be passed and what information should not be passed to the output. The best thing about them is that they can be trained to keep information from long ago, without washing it through time or remove information which is irrelevant to the prediction. With the help of [gru blog] we explain the mathematics behind a single cell of this networks.

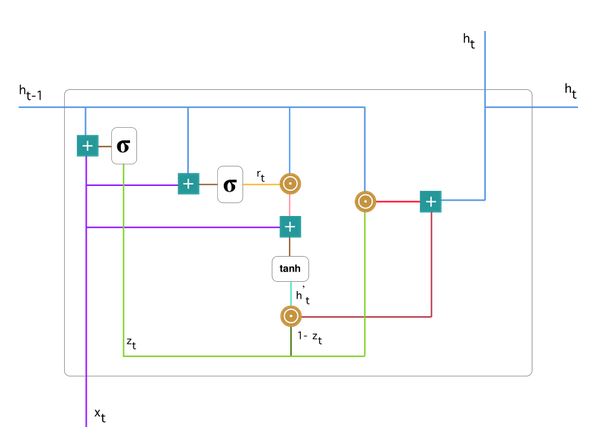




Fig: A single GRU cell

Update gate:

The update gate helps the network to control how much of the past information (from previous time steps) needs to be passed along to the future. That is really effective because the network can decide to copy all the information from the past and eliminate the risk of vanishing gradient problem.

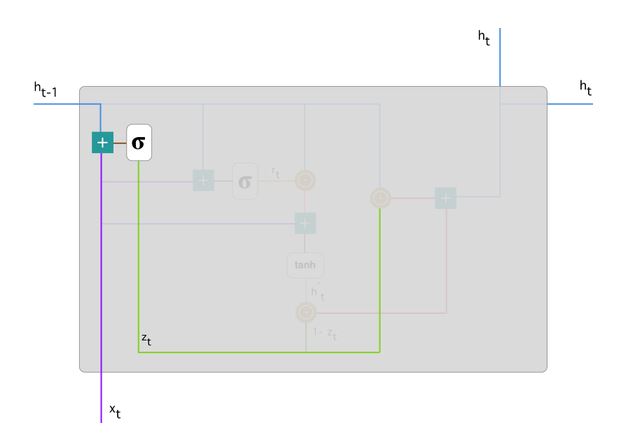


Fig: update gate

The formula for update gate is:

Here is the into the network unit, it is multiplied by its own weight. The same goes for which holds the information for the previous t-1 units and is multiplied by its own weight . Both results are added together and a sigmoid activation function is applied to squash the result between 0 and 1.

Reset gate:

The main purpose of reset gate in the network is basically to decide how much of the past information to forget.

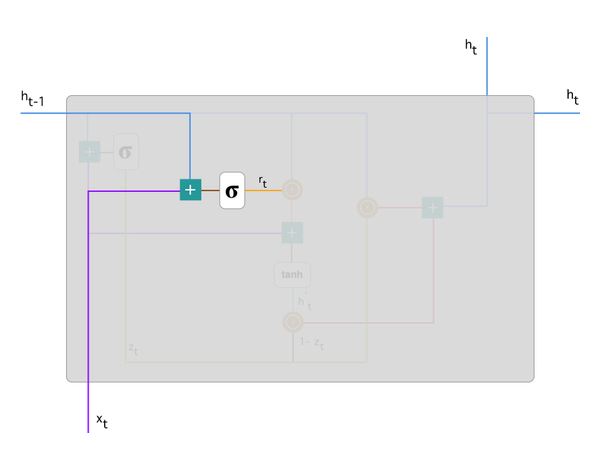


Fig: reset gate

The formula for reset gate is:

We plug in previous output as input and , multiply them with their corresponding weights, sum the results and apply the sigmoid function.

Current memory content:

This memory content will use the reset gate to store the relevant information from the past.

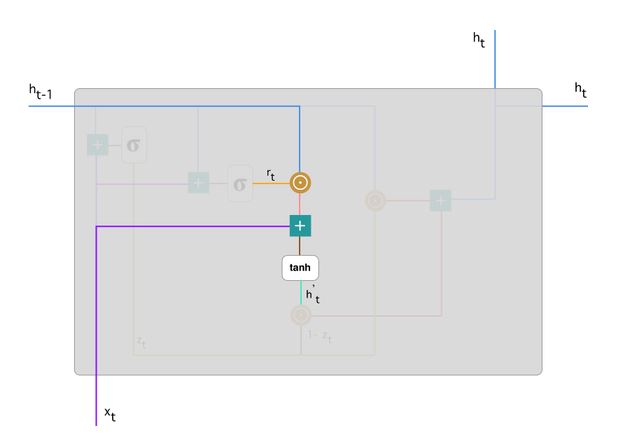


Fig: current memory content

The formula for current memory content is:

Here an element-wise multiplication happened between and line and then sum the result with the input finally, tanh is used to produce

Final memory at current time step:

Finally the network needs to determine which is the output of current unit and passes it down for next unit. In order to do that the update gate is needed which control what to collect from the current memory content  and what from the previous steps    .

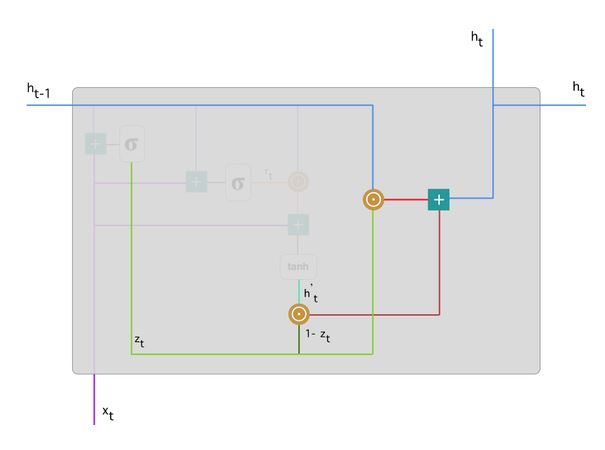


Fig: Final step

The formula this content is:

Software:

The programming language we used here is Python 3.7 [python]. Python is a powerful high-level, object-oriented programming language created by Guido van Rossum. It has simple easy-to-use syntax, making it the perfect language for someone trying to develop something creative as it comes with lots of modern packages.

For data analysis and data structures purpose we use Pandas which is an open source BSD-licensed library. We use NumpPy library for adding support for large, multi-dimensional arrays, matrices and for scientific calculation along with a large collection of high level mathematical functions to operate on these arrays [NumPy]. For collecting historical data at real time we use Pandas-datareader. For feature scaling we use Scikit-learn which is a machine learning library [Scikit-learn]. For developing our deep GRU neural networks we use Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow [Keras]. To visualize the result we use Matplotlib which is a Python 2D plotting library which produces quality figures in a variety of interactive environments across platforms [Matplotlib].