Application of Artificial Neural Network in Mathematics Education

Abstract

Artificial neural networks, which have shown their high value in many applications today, are based on the biological model of the animal brain. These networks are in fact an information data processing system that has special performance characteristics such as animal neural networks that have emerged from the generalization of their mathematical models. Artificial neural networks are a branch of artificial intelligence that is a suitable method for detecting unknown patterns in data, one of the applications of which is prediction. For prediction using artificial neural network, there are two programs, one is the training stage and the other is the prediction is taught based on data. The post-diffusion algorithm is also used for training and the neural network is used to predict demand. In this research, we intend to have an overview of the use of artificial neural networks, their function and method of implementation in mathematics education.

Keywords: artificial neural network, mathematics education, post-diffusion algorithm, information data processing system

1. Introduction

Although not more than 60 years have passed since the beginning of computational methods based on artificial neural networks, these networks still have a significant place in complex issues such as pattern recognition, clustering, modeling due to features such as parallel processing, intelligence and flexibility. , Has opened estimates and forecasts for itself [1].

The idea for the neural network was introduced in 1943 by McClutch and Pizza. The two created a simple model of a neuron that could build AND and OR gates. This model was then developed by other scientists and evolved from a single neuron to a network of neurons. To use a neural network, you must first train the network. In recent years, various structures for neural networks have been proposed [2].

The human brain, according to many scientists, is a more complex system that has been observed and studied throughout the universe. But this most complex system has neither galactic dimensions nor more components than today's supercomputer processors. The mysterious complexity of this unique system is due to the many connections between its elements. What sets the human brain apart from all other systems is the 1,400-gram brain [3].

The conscious and unconscious processes that take place within the geographical boundaries of the human body are all controlled by the brain. Some of these processes are so complex that no computer or supercomputer in the world can process and perform them. However, research shows that the building blocks of the human brain are about a million times slower in terms of operating speed than the transistors used in CPU silicon chips [4].

The very high speed and processing power of the human brain is due to the very large connections that exist between the cells that make up the brain, and basically, without these links, the human brain is reduced to a normal system, which is definitely a possibility will not have the current [5].

After all, the excellent performance of the brain in solving all kinds of problems and its high efficiency, simulation of the brain and its capabilities, has become the most important goal of hardware and software architects. In fact, if the day comes (which, apparently, is not too far away) when we can build a computer the size of the human brain, there will certainly be a great revolution in science, industry, and of course human life [6].

In order to simulate the computational behavior of the human brain, since the last few decades, when computers made it possible to implement computational algorithms, research work has begun by computer scientists, engineers, and mathematicians, whose work, in A branch of artificial intelligence, and sub-branch of computational intelligence, is classified as "Artificial Neural Networks" (ANNs). In the field of artificial neural networks, several mathematical and software models inspired by the human brain have been proposed, which are used to solve a wide range of scientific, engineering and applied problems in various fields [7].

2. Research background:

Avellaneda & Lee (2010) presented eight regressions and classification methods using machine learning and deep learning to predict the price of the last transaction and to predict the movement of stock prices in the weekly forecast. They studied the 50 NIFTY index values of the Indian Stock Exchange for the 4-year period 2015-2018. They used a deep learning regression model using catalysis neural network to increase the predictive power of movement patterns of the index [8].

Baker et al. (2011) used a hybrid artificial neural network model to predict the Vietnam stock market index. In addition, they compared the performance of hybrid models, artificial neural network models, and autonomous moving average models to predict the Vietnam stock market, and the results indicate that the ARIMA model is suitable for the linear time series and the ANN model is suitable for the nonlinear time series. Hybrid models, on the other hand, cover both time series features [9].

In their study, Bali et al (2011) proposed a hybrid approach of combining the error propagation neural network and the autoregressive conditional heterogeneity variance model to predict the closed share price of Kuichu Mottai and the Shanghai stock market price index. In this research, a multilayer neural network model after error propagation is used to predict the closed price of stocks. Due to the significant difference between the actual values and the predicted values, in order to obtain more accurate prediction results, the error bias was optimized with the GARCH model is in the forecast [10].

Bogomolov (2013) in their article showed that stock price fluctuations can be predicted by daily data. The country surveyed is India and used the four main indices NIFTY, S & P500, FTSE and DAX for processing during the sample period from January 1996 to March 2015 [11].

Boyer et al (2010) predicted milestones of 30 shares using 15 technical indicators (including Williams's% R.RSIMACD) and ANN neural network. Dow Jones 15 In recent research, in-depth learning is used to predict stock prices and capital market indices, including the following research. Fisher et al. (2018) used an LSTM to predict the S&P 500 stock trend between 1992 and 2015 [12].

Britz (2015) in a study using artificial neural network tried to predict the closing price of the next day. In this study, three different portfolios are formed for comparison between them. In the first portfolio, judgmental prediction was used and in the second portfolio, neural network was used for prediction and in the third portfolio, CAPM model optimized by artificial neural network was used. The results of his research showed that the third portfolio, which was based on the CAPM model optimized by artificial neural network, has achieved higher returns than the other two portfolios [13].

Abarbanell (1997) conducted a study comparing neural network prediction methods with classical GARCH and ARMA prediction methods; the evaluation criteria are MSE and R2. The results show the superiority of neural networks over GARCH and ARMA samples [14].

Table 1: A summary of recent work related to this research

Author (year)	Data type (number of input features × latency)	Case study	Sampling period (frequency)	Method	Result
Enke and Mehdiyev (2013) [15]	US S&P Index (1 × 20)	Stock price	1980-2010 (Daily)	Feature selection + fuzzy clustering + fuzzy ANN	Predicting market prices through fuzzy neural network based on the findings of the previous step and optimizing it using the gradient method. This method performed better than the leading neural network.
.Boll erslev, T. M. (2014.). [16]	KOSPI index 200 (1 × 20)	Market direction (up or down)	1994-2008 (daily)	Select feature + ANN	Better performance of neural network model than logit model. The ANN model can dramatically increase revenue compared to a buy and hold strategy.
Adebiyi, A. A [17]. (2014)	Indian Indicator	Stock price	2003-2012 (Daily)	SVR+ ANN	According to the results, the random forest method performed better than the ANN prediction model.
.Aggarwal, C. C. (2018) [18]	TAIEX Taiwan (20 × 27)	Market trends	1989-2004	Dimension reduction + pattern matching	Comparison of genetic algorithm, artificial intelligence and neural network in predicting the index. The neural network had the best performance.
Chong, E. C. (2017) [19]	Global Stock Indices	Trading signal (stock price)	2008-2010 (Daily)	Optimization +ANN	Comparison of deep learning methods and ARIMA for index prediction. Deep learning performance has been better.

Qiu, M. S. (2016) [20]	General ASE Index of Greece	Portfolio composition	1996-2012	Fuzzy	The results show that regardless of the transaction fee, the efficiency of the fuzzy model is much higher than the buying and maintaining strategy. This system can be considered conservative because compared to the maintenance buy strategy, it makes less profit in bear markets (boom) and less loss in cattle markets (recession).

3- Neural networks

Artificial neural networks are mathematical models inspired by the human nervous system and the brain. The Artificial Neural Network is called ANN for short. Inspired by the neural network of the human brain, this network seeks to develop information processing. In fact, the neural network helps us train our computer (programming) to respond appropriately to events, instead of dictating what needs to be done. Each neuron in this network is a processing element and solves different problems along with other processing elements.

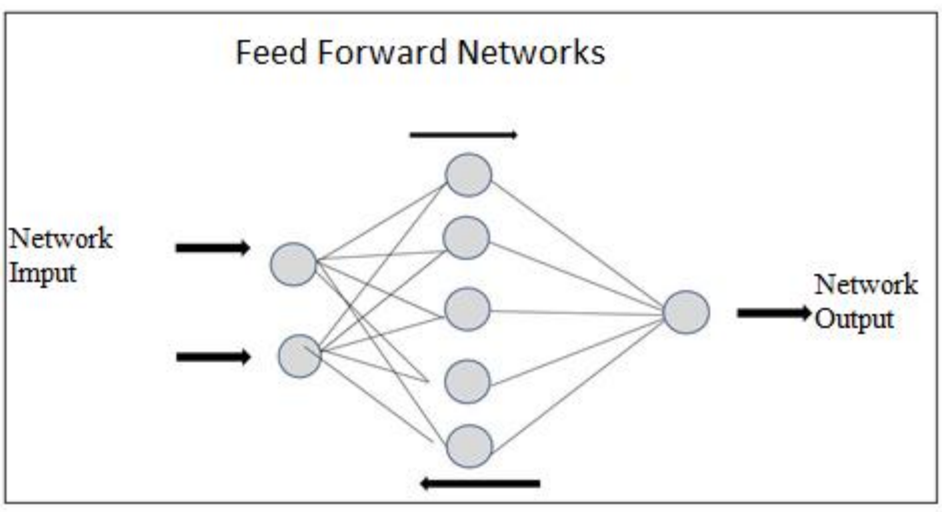


Figure 1: Artificial neural network

In recent decades, due to the emergence of new issues, the advancement of science, the commonality and breadth of various issues, interdisciplinary and interdisciplinary fields have flourished. With the entry of researchers in one field into other fields, it is possible to examine new dimensions of the issues ahead. For example, scientists in physics and mathematics have tried to apply their knowledge in the field of finance with the cooperation of economics, finance and programming experts, and by designing prediction models and artificial intelligence programs, to achieve relationships between variables and be able to have a more accurate forecast in the shortest time [21].

3-1 Artificial neural network

There are two general types of artificial neural networks that are determined by the type of signal sent by neurons:

Feed Forward Networks Neural networks are made up of neural layers. In this type of network, the signal is only allowed to move in one direction and cannot return to the neurons of the previous layer. The response path of neurons in the feeder network is like a one-way street and is always forward.

The feed neural network, also known as the MLP, is an example of a simple neural network. Each input neuron is connected to the next hidden layer neurons via a w_{ki} weight matrix. The network has three sections: input layer, hidden layer and output layer. Artificial neurons are those located in the hidden and output layers. Each of these neurons receives input from the previous layer. Network neurons are not connected to similar layer neurons, but to next layer neurons. The equation for the activation function of a secret neuron is as follows:

Equation 1:

$$h_i = f(u_i) \; = f(\sum_{k=0}^{k} w_{ki} x_k)$$

h_i: i The secret of hidden neurons, $f(u_i)$: The nonlinear link function that shows the relationship between the input layer and the hidden, w_{ki}: The input weight in the weight matrix (K X N), x_k: k is the input value:

Equation 2:

$$y_i = f(u_j) = f(\sum_{\iota=1}^{\acute{}k} w_{\iota j}^{\acute{}} h_\iota)$$

y_i: j is the output value

Feed Networks: (Feed Back Networks) as you can see in the English name of this type of network, there is feedback in it. Feedback means that the output of a neuron depends on its previous output in addition to the current input. There is at least one return signal in feedback networks.

3-2 Machine learning

Scientific machine learning is that computers function without being explicitly programmed. Over the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a much better understanding of the human genome. Many researchers also think that this is the best way to advance artificial intelligence at the human level. Machine learning focuses on the development of computer programs that can access and use data. The learning process begins with observations or data, to look for patterns in the data based on the patterns we present, and to make better decisions in the future. The main goal is to allow computers to automatically learn and adjust actions based on human intervention.

3-3 Deep learning

Deep learning is an artificial intelligence (AI) function that mimics the function of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in artificial intelligence that has networks that are capable of learning without monitoring unstructured or unlabeled data. Also known as deep neural learning or deep neural network.

3-4 Multilayer perceptron models

One of the simplest and at the same time most efficient architectures proposed for use in real nerve modeling is the multilayer perceptron (MLP) model, which consists of an input layer, one or more hidden layers and an output layer. In this structure, all single-layer neurons are connected to all next-layer neurons.

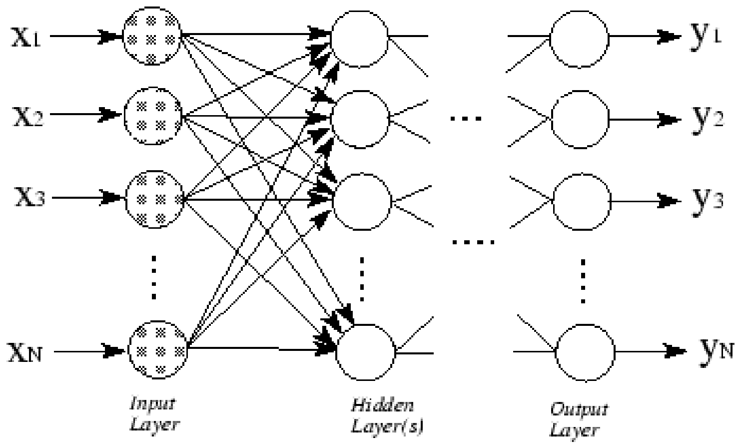


Figure 2: Multilayer Neural Network (MLP)

This arrangement forms a so-called network with complete connections [21]. And shows a multilayer perceptron network. It can be easily inferred that the number of neurons in each layer is independent of the number of neurons in the other layers. It is important to note that in any accumulated circle, there is a summation and thresholding operation (passing through the nonlinear sigmoid function). In fact, each solid circle in this figure is a model of the summation and threshold block in Figure 2, which is shown in this form for ease of display. According to the figure, the output of the i-th nerve (in the last layer) can be shown as follows [22]:

Equation 3:

$h_1 = \delta_1(W_1 u + b_1)$

$h_2 = \delta_2(W_2 u + b_2)$

$\vdots\vdots\vdots$

$h_L = \delta_L(W_L u + b_L)$

L: Number of layers

δ: Transfer function (activation)

Weights = $(W_1, W_2, \ldots, W_L)$

Bias = $(b_1, b_2, \ldots, b_L)$

The method to achieve the desired relation $\{h_n\}_{n=1}^{L}$ is as follows: First, the inputs and goals of the network are defined in the form $\{U^n, \tau^n\}_{n=1}^{N}$Each time the network is trained, an output value of the error function $\varepsilon(y^n, \tau^n)$ where the difference between the predicted output $\{h_n\}_{n=1}^{L}$ and the network target $\{\tau^n\}_{n=1}^{L}$ obtains t at any time.

Set of network parameters:

$\theta = \{W_1, \cdots, W_L, b_1, \cdots, b_L\}$

These parameters are obtained by minimizing the sum of errors. [23]

Equation 4:

$$min_{\theta}\left[j = \sum_{n=1}^{N} \varepsilon(y^{n}, \tau^{n})\right]$$

The approach to solving this problem is to minimize the gradient descent [24]

3-6 Torsional or convergent neural networks

This network, abbreviated CNN or ConvNet, is a class of deep networks that are commonly used for visual or speech analysis in machine learning.

Torsional neural networks use a variety of multilayer perceptron networks to minimize preprocessing. These networks are sometimes referred to as immutable neural networks with transmission or invariant with space. This naming is based on the structure of this network, which we will refer to later.

The structure of torsional networks is inspired by the biological processes of the cat's visual cortex. This structure is such that the neurons respond to stimulation only in a limited region, which is called the acceptance region. The receptor regions of the different neurons overlap slightly; in such a way that they cover the whole field of view.

Torsional neural networks use less preprocessing than other image classification approaches. This means that the network adopts criteria that were learned manually in previous approaches. This independence from prior knowledge and human manipulation of torsional neural networks is a fundamental advantage.

So far, various applications have been proposed for neural networks, including in computer vision, suggestion systems, and natural language processing. A torsional neural network consists of an input layer, an output layer, and a number of hidden layers. Hidden layers are usually torsional or cumulative or complete.

• Torsional layers

Twisting layers apply a twisting action to the input, and then pass the result to the next layer. This twist actually simulates a techno neuron’s response to a visual stimulus.

Each torsion neuron processes data only for its own reception area. Mesh torsion networks allow them to correct transmission or rotation.

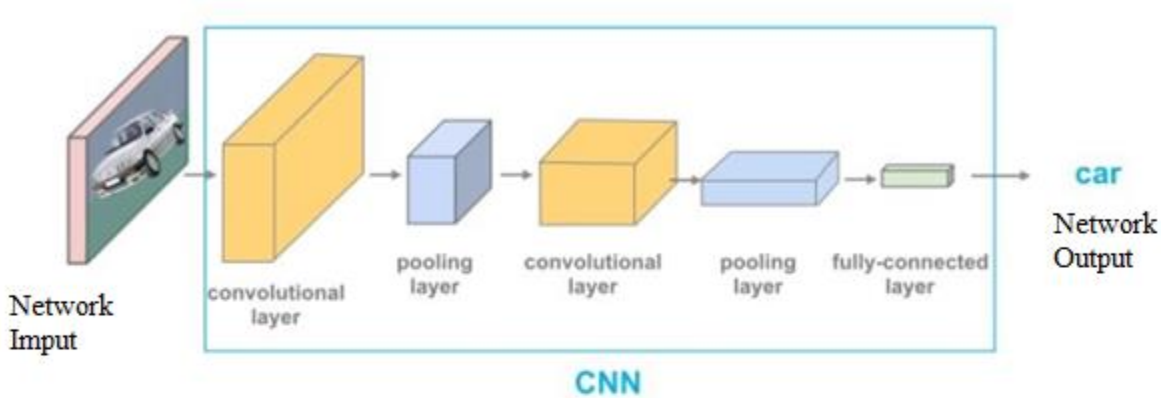
Although fully connected feed neural networks can be used to learn properties and classify data, this architecture does not apply to images. In this case, even a shallow network requires a large number of neurons. Twisting is a solution to this situation that reduces the number of free parameters to deepen the network.

• Integration layers

Torsional neural networks may consist of local or global fusion layers that combine the outputs of neuronal clusters in one layer into a single neuron in the next layer. The maximal aggregation method, for example, uses the maximum value between neuronal clusters in the anterior layer. Another example is aggregation averaging, which uses the average value of neural clusters in the anterior layer.

Torsional neural networks share weights in torsional layers, resulting in minimal memory and maximum efficiency.

Figure 3: View of the CNN neural network



3-7 Recursive neural network

Recursive neural network (RNN), also known as repetitive neural network, is a type of artificial neural network used in speech recognition, natural language processing (NLP), and sequential data processing. Many deep networks, such as CNN, are feed networks, meaning that the signal in these networks only travels in one direction from the input layer to the hidden layer and then to the output layer, and previous data is not stored in memory. But recursive neural networks (RNNs) have a feedback layer in which the network output, along with the next input, is returned to the network. RNN can remember its previous input due to its internal memory and use this memory to process a sequence of inputs. Simply put, recursive neural networks consist of a recursive loop that ensures that information we have gained from previous moments is not lost and remains in the network.

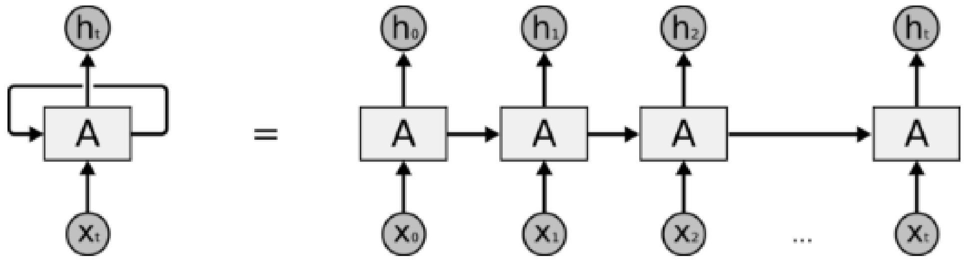


Figure 3: Recursive neural network

As you can see in the figure, it takes $X_0$from the input sequence, output $h_0$with X_1 is the input of the next step. In the next step, output $h_1$and $X_2$are the input of the next step. This way, the network will be able to recall previous inputs during training.

3-8 Long term short term memory

LSTM is a special type of RNN. These networks specialize in learning about long-term dependencies. LSTM was introduced by Hochreiter and Schmidhuber in 1997. These networks are clearly designed to avoid long-term dependency errors, but remembering information for long periods of time is their normal behavior. The figure presents an image of the LSTM cell. LSTM has a different structure than other neural networks. Conventional RNN has a very simple neural network with a feedback loop, but LTSM consists of a memory block or cells instead of a single artificial network layer. Each cell or block has three gates and a cell state that wants to modulate the flow of data information through the cells. In figure c_ (t-1) is the previous state of the cell, c_{t-1} is the state of the cell, h_{t-1} is the output of the previous cell, h_t is the output of the current cell, $i_t$is the input gate layer, f_t is the forget gate gate, o_t is the output sigmoid gate layer

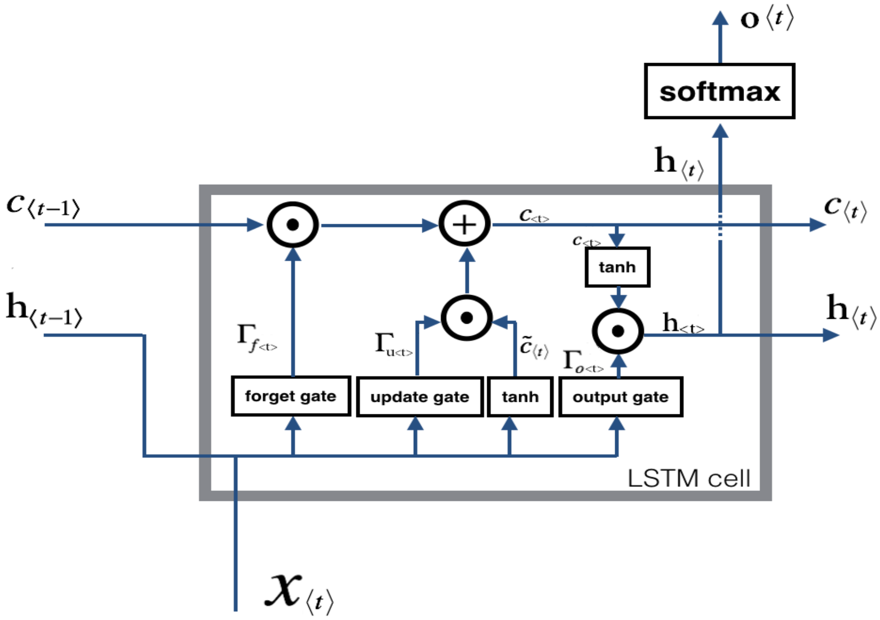


Figure 4: LSTM neural network

In the figure, the horizontal line passing through the top of the diagram is known as the cell position (c_t) and acts as a conveyor drawn over the entire grid. This tape carries information from the previous cell to the current cell as well as to the next. The decision to store information in the cell's state is made by the forget-me-not (f_t) layer, also known as the sigmoid layer. The output of the forget-me-not gate is added to the cell state using the reverse multiplication operation. Then there is the entrance gate, which consists of a sigmoid layer (i_t) and a tanh layer. The input layer combines the two as a cell state. Here are the new C values created by the tanh layer. The output (h_t) is formed by the inverse multiplication of the sigmoid gates o_t and tanh.

3-9 Methods of minimizing the error cost function

There are two approaches to solving this minimization problem, one is the classical method and the other is the use of intelligent optimization methods [25]

3-10 Reduction gradient method

First the gradient of the cost function (error) is calculated, then it moves in the opposite direction and this can be continued until the cost function becomes very small. The backslash method is actually a way to find the gradient of the Q (W) function. Now suppose we want to get the gradient of the function Q (W) relative to the weight w_{pc} t to do this, we need a chain rule in derivation. The chain rule works like this: If we have a function called f that depends on three inputs u, v, w, and each of these three inputs in turn depends on t, the derivative of f to t to this figure is calculated:

Equation 5

$$\frac{\partial f(u(t),v(t),w(t))}{\partial t}=\frac{\partial f}{\partial u}*\frac{\partial u}{\partial t}+\frac{\partial f}{\partial v}*\frac{\partial v}{\partial t}+\frac{\partial f}{\partial w}*\frac{\partial w}{\partial t}$$

Using this chain rule, we follow the return method as follows:

Equation 6

$$\delta_c=\frac{\partial Q}{\partial a_c}\ a_c=\sum_p w_{pc}*b_{pc}\ \ b_c=\theta_c(a_c)$$

Equation 7

$$\partial_c=\frac{\partial Q}{\partial a_c}=\frac{\partial Q}{\partial b_c}\frac{\partial b_c}{\partial a_c}=\frac{\partial Q}{\partial b_c}*\theta_c'(a_c)=(\sum_n\frac{\partial Q}{\partial a_n}\frac{\partial a_n}{\partial b_n})*\theta_c'(a_c)=\sum_n w_{cn}\delta_n)*\theta_c'(a_c)$$

Equation 8

$$\frac{\partial Q}{\partial w_{pc}}=\frac{\partial Q}{\partial a_c}\frac{\partial a_c}{\partial w_{pc}}=\delta_c b_p$$

As we saw in the previous line, we need two values to get the gradient relative to w_{pc}, the input to the neuron c is easily obtained from the neuron p, which is b_p, and δ_c, which is obtained by the recursive method, depending on the δ_i of the next layer. Cell c is attached to them, more precisely

Equation 9

$$\delta_c=(\sum_n w_{cn}\,\delta_n)*\theta_c'(a_c)$$

The recursive method for obtaining δ works by first calculating δ for the output layer cells, then lowering the layers in turn, and calculating for each cell δ by combining the δ layers of its top layers according to the formula. Calculating δ for the output layer is easy and is obtained directly by deriving from Q.

Supervised learning: A type of machine learning in which input and output are specified and there is a so-called observer who provides information to the learner, and thus the system tries to assign a function from input to Inclusive output.

Unsupervised learning: Unsupervised learning, unlike supervised learning, does not already have specific data and is not intended to communicate input and output, but only to categorize them, and it is the learner who must look for a specific structure in the data.

Semi-supervised learning: Semi-supervised learning is a type of learning that uses both classified (labeled) and non-classified (unlabeled) data simultaneously to improve learning accuracy.

Feature extraction: A process in which data is mapped in a high-dimension space to a lower-dimension space. This mapping can be linear (like the principal component analysis method) or nonlinear. The number of variables that must be measured for each observation is called the data dimension. The term "variable" is more commonly used in statistics, while in computer science and machine learning the terms "attribute" or "adjective" are more commonly used. Multidimensional data platforms pose many computational challenges despite the opportunities they create. One of the problems with large data sets is that most of the time all the properties of the data are not critical to finding the knowledge that lies in the data. For this reason, in many areas, reducing the size of the data remains a significant issue. [22].

4. Mathematical model of a nerve cell

When modeling nerves, their complexity is ignored and only the basic concepts are valued, because otherwise the modeling approach will be very difficult. At a simple glance, the model of a nerve should include inputs that act as synapses. These inputs are multiplied by weights to determine the signal strength. Finally, a mathematical operator decides whether the neuron is activated or not, and if the answer is yes, determines the output. Therefore, the artificial neural network processes information using a simplified model of the real nerve. Given these explanations, a simple model can be proposed to describe a neuron (a node in an artificial neural network).

Among other things, the weights of a neural network, which transmits the amount of output, can be positive or negative. On the other hand, the functions used for thresholding can be very diverse. Among the most famous of these functions are functions such as Arxin, Arctan and Sigmoid. These functions must be continuous and smooth and derivable. The number of input nodes can also vary. However, as the number of these nodes increases, it becomes more difficult to clearly determine the weights. So we have to look for ways to solve this issue. The process of determining the optimal weights and adjusting their values is mainly done in reverse. For this purpose, the network is trained using rules and data and using the network learning capability, various algorithms are proposed, all of which try to bring the output produced by the network closer to the ideal and expected output [25].

4-1- General structure of artificial neural network

A neural network consists of the components of layers and weights. Network behavior also depends on communication between members. In general, there are three types of neural layers in neural networks:

4-1-1 Data entry layer:

In this layer, the basic information, which is the same values related to the independent and dependent variables that enter the network. The number of input layer cells is the number of inputs. In practice, it is tried to consider all the parameters that affect the response, but it should be borne in mind that unused information input makes the network more difficult, because although the neural network to noise (error data) It is resistant, but in any case, if the amount of noise is too high, the network may not be able to converge.

4.2 Artificial neural networks

Artificial neural networks, although not comparable to the natural nervous system, have features that make them useful in some applications such as pattern separation, robotics, control, and generally wherever a linear or non-linear mapping is required. Be linear, excel. These features are as follows:

4-2-1 Learning Ability

Extracting analytical results from a nonlinear mapping marked with a few examples is not an easy task. The neuron is a nonlinear system, and as a result a neural network formed from the community of these neurons will also be a completely complex and nonlinear system. In addition, the nonlinear properties of the processing elements are distributed throughout the network. Implementing these results with a conventional, non-learning algorithm requires great care and attention. In such a case, a system that can extract this relationship itself seems very useful. In particular, adding possible examples in the future to a system with learning capability is much easier than doing it in a system without such capability, because in the latter system, adding a new example replaces all previous work [26].

Learning capability means the ability to adjust network parameters (synaptic weights) over time as the network environment changes and the network experiences new conditions, with the aim that if the network is trained for a particular situation and a small change in its environmental conditions (special state) It turned out that the network could be effective with brief training for new conditions, and that information was stored in neural networks at synapses, and that each neuron in the network was potentially affected by the total activity of the other neurons. It is not separate but is affected by the whole network [27].

4-2-2 Dispersion of information "Information processing as text"

What the network learns (information or knowledge) is hidden in the synaptic weights. There is no one-to-one relationship between the inputs and the synaptic weights. In other words, each neuron in the network is affected by the total activity of the other neurons; as a result, the information is processed by the neural networks as text 5. Accordingly, if part of the network cells are deleted or malfunction there is also the possibility of reaching the correct answer, although this possibility has been reduced for all inputs but has not been eliminated for any of them [28].

4-2-3 Generalizability

After the initial examples are taught to the network, the network can be confronted with an untrained input and provide a suitable output. This output is based on the generalization mechanism, which is nothing more than an interpolation process. In other words, the network learns the function, learns the algorithm, or obtains a suitable analytical relation for a number of points in space [29].

4-2-4 Parallel processing

When the neural network is implemented in the form of hardware, cells that are in a level can respond to the inputs of that level at the same time. This feature increases the processing speed. In fact, in such a system, the overall processing task is distributed among smaller, independent processors [30].

4-2-5 being resilient

In a neural network, each cell operates independently, and the overall behavior of the network is the result of the local behaviors of multiple cells. This feature keeps local errors out of sight of the final output. In other words, cells in a process of cooperation correct each other's local errors. This feature increases the resilience (fault tolerance) in the system [30].

5. Discussion and Conclusion

Today, the use of intelligent systems, especially artificial neural networks, has become so widespread that these tools can be classified as basic and common tools in the line of basic mathematical operations. Because there are few academic disciplines that need to analyze, decide, estimate, predict, design and build, and it does not use the subject of neural networks. The following is an incomplete list, but it illustrates to a large extent the applications of artificial neural networks.

Table 2: Applications of artificial neural networks

General context	Application
Computer science	• Classify documents and information in computer networks and the Internet • Development of monitoring software and antivirus
Computer science	• Reverse engineering and systems modeling • Predicting the consumption of electric charge • Troubleshooting industrial and technical systems • Design a variety of control systems • Design and optimization of technical and engineering systems • Optimal decision making in engineering projects
Technical and engineering sciences	• Predicting test results • Evaluate and estimate the correctness of hypotheses and theories • Modeling of complex physical phenomena
Medical Sciences	• Modeling of biomedical processes • Diagnosis of diseases according to the results of medical tests and imaging • Predict treatment and surgery results • Implementation of patient-specific treatment devices and models
Experimental and biological sciences	• Modeling and forecasting of biological and environmental phenomena

	• Predicting time series with application in environmental sciences • Classification of findings from experimental observations • Identify hidden and repetitive patterns in nature
Economics and Finance	• Stock price forecast and stock index • Classification of stock signs and symbols • Risk analysis and assessment • Allocation of capital and credit
Social Sciences and Psychology	• Classification and clustering of individuals and groups • Modeling and predicting individual and social behaviors
Art and literature	• Predicting the success and public acceptance of works of art • Extracting basic components from literary texts and works of art • Classify and explore literary texts
Military Sciences	• Targeting and pursuing missile weapons • Implementation of smart defense systems • Predict the behavior of the attacking force and the enemy • Implementation of attacks and defense systems in electronic warfare (forest)

Given that neural networks have two basic features of learning or mapping based on the presentation of empirical data (power and ability to be generalizable) and parallel structurability, these networks are for complex systems that model these systems or are possible. Artificial neural networks are more accurate and valid in mathematics education due to the use of proven formulas with minimal errors. Much has been done to address the shortcomings of these networks. One of the reasons that limit the use of neural networks is the training phase. The training of neural networks always requires a large group of input information. Network training parameters are very difficult and require experience, as well as problems such as local minimums and convergence to an inappropriate response, memorizing input information instead of learning it, and most importantly the need for a lot of time to learn about network problems is nervous.

References

[1]. Baek, Y., & Kim, H. Y. (2018). ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Systems with Applications*, *113*, 457-480.

[2]. Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to financial forecasting. *Applied intelligence*, *6*(3), 205-213.

[3]. Adhikari, R. (2015). A mutual association based nonlinear ensemble mechanism for time series forecasting. *Applied Intelligence*, *43*(2), 233-250.

[4]. Akita, R., Yoshihara, A., Matsubara, T., & Uehara, K. (2016, June). Deep learning for stock prediction using numerical and textual information. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)* (pp. 1-6). IEEE.

[5]. Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014, March). Stock price prediction using the ARIMA model. In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation* (pp. 106-112). IEEE.

[6]. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654-669.

[7]. Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques–Part II: Soft computing methods. *Expert systems with applications*, *36*(3), 5932-5941.

[8].Avellaneda, M., & Lee, J. H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, *10*(7), 761-782.

[9]. Baker, M., Bradley, B., & Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal*, *67*(1), 40-54.

[10]. Bali, T. G., Cakici, N., & Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, *99*(2), 427-446.

[11]. Bogomolov, T. (2013). Pairs trading based on statistical variability of the spread process. *Quantitative Finance*, *13*(9), 1411-1430.

[12]. Boyer, B., Mitton, T., & Vorkink, K. (2010). Expected idiosyncratic skewness. *The Review of Financial Studies*, *23*(1), 169-202.

[13].Britz, D. (2015). Recurrent neural networks tutorial. *URL: http://www. wildml. com/2015/09/recurrentneural-networks-tutorialpart-1-introduction-to-rnns/[accessed June 28, 2019]*.

[14].Abarbanell, J. S. (1997). Fundamental analysis, future earnings, and stock prices. . *Journal of accounting research,35(1)1-2*.

[15].Enke, D. &. (2013). Stock market prediction using a combination of stepwise regression analysis, differential evolution-based fuzzy clustering, and a fuzzy inference neural network. *Intelligent Automation & Soft Computing,* (4) *19,* 636-648.

[16].Boll erslev, T. M. (2014.). Stock return predictability and variance risk premia: statistical inference and international evidence. Journal of Financial and Quantitative Analysis. 49(3),633-661.

[17].Adebiyi, A. A. (2014). Comparison of ARIMA and artificial neural networks models for stock price prediction. . *Journal of Applied Mathematics*.

[18].Aggarwal, C. C. (2018). *Neural networks and deep learning.* . Cham: Springer International Publishing..*Artificial neural network*. wikipedia.org: https://en.wikipedia.org

[19].Chong, E. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications. 205, 183-187.*

[20].Qiu, M. S. (2016). Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. *Chaos, Solitons & Fractals.85, 1-7.*

[21].CireşAn, D. M. (2012). Multi-column deep neural network for traffic sign classification. *Neural networks.32, 333-338.*

[22].Malkiel, B. G. (1970). The journal of Finance. *Efficient capital markets: A review of theory and empirical work. 25(2), 383-417.*

[23].Moons, B. D. (2019). *Digital and Mixed-Signal Always-On Binary Neural Network Processing." Embedded Deep Learning.* . Cham: BINAREYE: Springer.

[24].Patel, J. S. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications. 42(4), 2162-2172.*

[25].*Poisson distribution*. wikipedia: https://en.wikipedia.org

[26].Rather, A. M. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications. 42(6), 3234-3341*.

[27].Skansi, S. (2018). *Introduction to Deep Learning: from logical calculus to artificial intelligence.* Springer.

[28].*t-Student distribution*. wikipedia: https://en.wikipedia.org

[29].Zhong, X. a. (2017). Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications. 67* 126-139.

[30].Clegg, M., & Krauss, C. (2018). Pairs trading with partial cointegration. *Quantitative Finance*, *18*(1), 121-138.