COMPUTATIONAL PROBLEMS OF ELECTRICAL ENGINEERING

Vol. 11, No. 2, 2021

MULTILAYER NEURAL NETWORKS – AS DETERMINED SYSTEMS

Sergii Sveleba[1], Volodymyr Brygilevych[2], Ivan Katerynchuk[1], Ivan Kuno[1], Ivan Karpa[1], Ostap Semotiuk[3], Yaroslav Shmyhelskyy[1], Nazar Sveleba[1]

[1] Ivan Franko National University of Lviv, Lviv, Ukraine

[2] State Higher School of Technology and Economics in Jarosław, Poland

[3] Ukrainian Academy of Printing, Lviv, Ukraine

sergii.sveleba@lnu.edu.ua, vbrygilevych@pwste.edu.pl, ivan.katerynchuk@lnu.edu.ua, ivan.kuno@lnu.edu.ua, incomlviv@gmail.com, yaroslav.shmyhelskyy@lnu.edu.ua

https://doi.org/10.23939/jcpee2021.02.026

Abstract: The study of the influence of learning speed (η) on the learning process of a multilayer neural network is carried out. The program for a multilayer neural network was written in Python. The learning speed is considered as a constant value and its optimal value at which the best learning is achieved is determined. To analyze the impact of learning speed, a logistic function, which describes the learning process, is used. It is shown that the learning error function is characterized by bifurcation processes that lead to a chaotic state at $\eta > 0.8$. The optimal value of the learning speed is determined. The value determines the appearance of the process of doubling the number of local minima, and is $\eta = 0.62$ for a three-layer neural network with 4 neurons in each layer. Increasing the number of hidden layers (3 ÷ 30) and the number of neurons in each layer (4 ÷ 150) does not lead to a radical change in the diagram of the logistic function (x_n, η), and hence, in the optimal value of the learning speed.

Key words: multilayer neural network, learning speed, logistic function, chaotic state.

1. Introduction (Style Header1)

Most tasks that use neural networks are decision tasks. The increase in the volume of the information being processed leads to the use of multilayer neural networks (NN) and convolutional NN. NN are also used for cognitive-type tasks, the most effective NN are those in which the method of teaching with a teacher with the inverse propagation of error is used. This method involves the selection of weights, which are initially set randomly. In this paper, we will consider a multilayer NN, and we will consider a wrapped NN in the next work. In a multilayer NN weight correction is performed in each hidden layer. In the learning process, the correction of weights in them is carried out from one epoch to another until the specified accuracy of training is achieved. The learning process is influenced by a number of parameters: the speed of learning, the number of epochs, the number of input and output values, etc. In particular, increasing the number of input and output values increases the time for learning and decision-making. Significant training time, and hence decision-making is one of the negative characteristics of NN. Decreasing the learning time can be achieved by reducing the number of epochs required for learning. One of the effective ways to reduce the number of epochs (N) required to achieve a given accuracy is to increase the speed of learning. There are [1–3] several effective algorithms for selecting the speed of learning to optimize the learning process. Since the speed of learning (η) depends on the number of epochs N ($\eta = A / N$) [1] and η is a decreasing function of time, the optimization can be done in the same way as it is done in stochastic methods of optimization and adaptation [2]. This procedure can be performed until the values of the controlled variables are not stabilized, or until the error is reduced to an acceptable level. It should be noted that this procedure is characterized by a slow rate of convergence and the possibility of getting into the local minima of the target functional [2]. Also, this is a decrease and subsequent increase in the learning speed [3]. And also, this is the multiplication of the learning parameter by some coefficient depending on the number of input parameters. These are methods of conjugate gradients that are based on determining the direction in which the objective function decreases most rapidly.

In a multilayer NN in each layer there shall be its own algorithm for selecting the learning parameter. When approaching the global minimum, there is an increase in the number of local minima, which also affect the algorithm for learning NN. An increase in the number of local minima when approaching the global minimum can lead to the appearance of indeterminate states (chaotic states) of NN. The values of the local and global minima of the error function become identical, and the number of such local minima increases. That is, in the first approximation of NN in the vicinity of the global minimum, it can be considered as a system characterized by uncertainty. The degree of this uncertainty goes to infinity in the vicinity of the global minimum. This mode of NN can be considered stochastic with the appearance of stationary and chaotic solutions. This mode is typical for both a single-layer and a multilayer NN. It is known [4] that an effective way to overcome local minima is to select such a learning parameter for the systems so that the number of the epochs required for learning can be reduced.

2. Main

Therefore, the task of this work is to establish an algorithm for automatic selection of learning speed for NN. To achieve this goal, in the Python environment we have developed a program for a multilayer neural network, which provides settings of the number of hidden layers and the number of neurons in them, and changing the learning parameter in the range of 0.001 ÷ 10. This interval of changes η was chosen taking into account the data of [5], where the influence of learning speed on the learning process in a multilayer NN is studied. We will consider each layer of the given NN as a separate deterministic system for which we will investigate the branching diagram with the use of the image function:

$$x_{n+1} = \eta - x_n - x_n^{\;2},$$

where n is the step, η is the parameter that determines the learning speed.

Its fixed points:

$$x_{1,2} = -1 \pm (\eta + 1)^{1/2},$$

eigenvalues, which can be calculated as follows:

$$\rho_1 = 1 - 2\,(\eta + 1)^{1/2}.$$

The choice of this logistic mapping is due to the fact that it describes the process of doubling the oscillation frequency when approaching chaos [6]. In our case, this process will be determined by the process of local minima when approaching the global minimum.

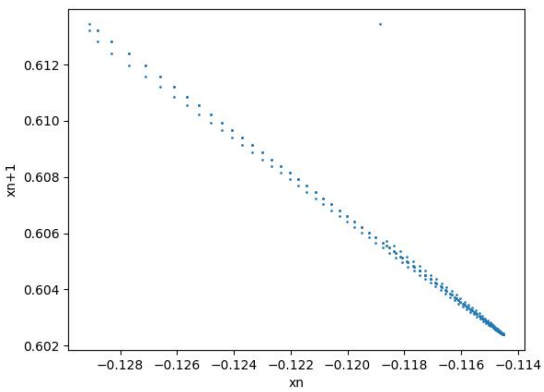
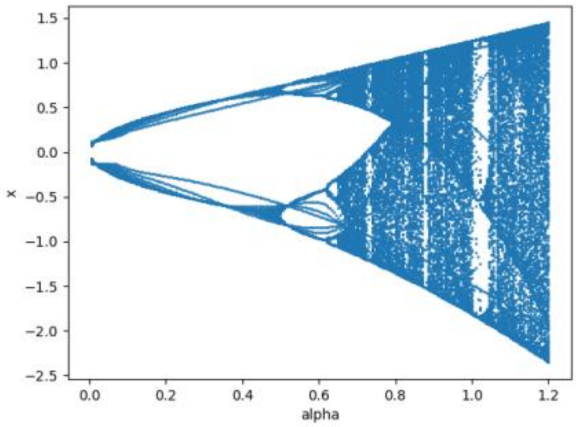
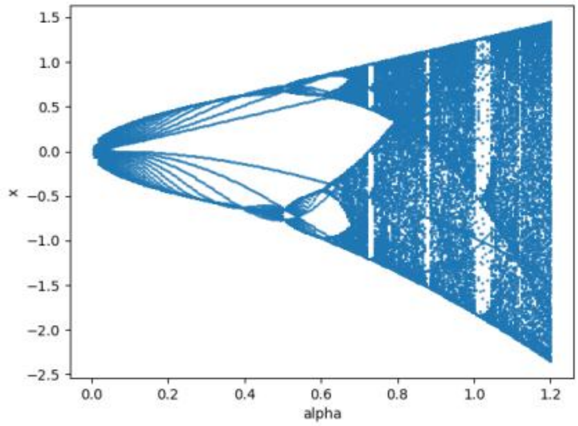


Fig. 1. Diagram of the logistic function after the first layer of a three-layer NN under the condition: η = 0.5, n = 10 neurons in the layer, N = 100 epochs.

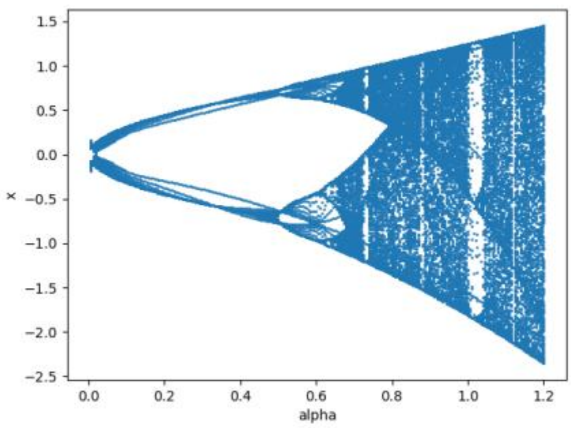
In the beginning consider a three-layer neural network and investigate it as a deterministic system. Figure 1 shows the diagram of a logistic function at a constant value of the parameter η. The given diagram of the logistic function derives its adequate dependence under the condition that the logistic mapping is given only by the iterative formula for the bifurcation of the doubling.



layer 1



layer 2

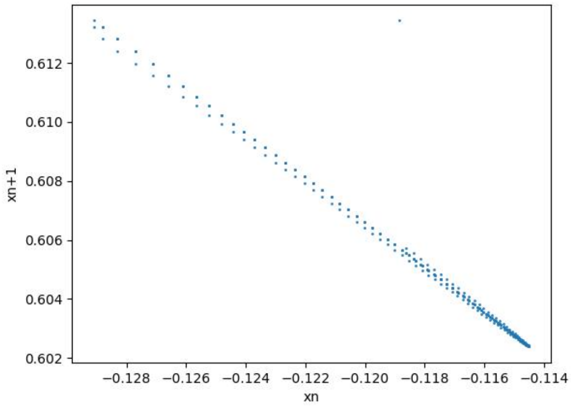


layer 3

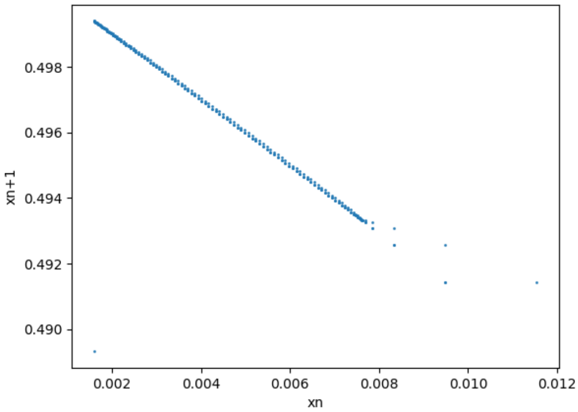
Fig. 2a. Branching diagram for each layer when changing the parameter η under the condition: n = 10 neurons in the layer, N = 100 epochs.

First of all, we note that this logistic function most adequately describes this system, and suggests that the process of formation of local minima can be described in the first approximation as a process of doubling their number.

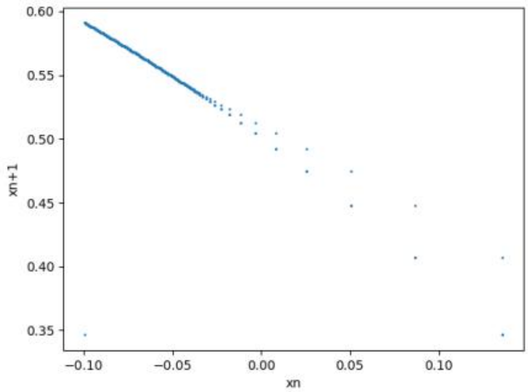
The branching diagrams shown in Fig. 2, for each layer of a three-layer NN, indicate that each layer can be considered as a separate system characterized by the existence of chaotic and stationary solutions.



layer 1

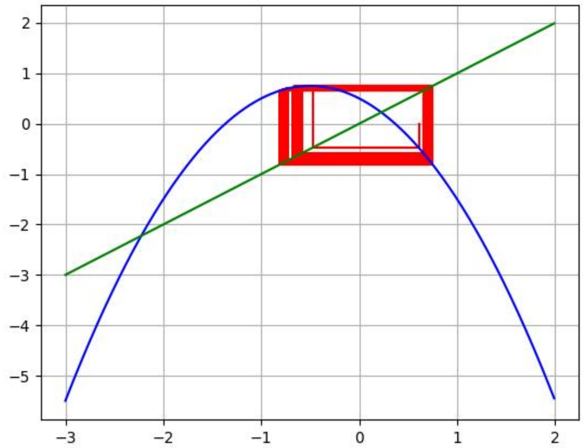


layer 2

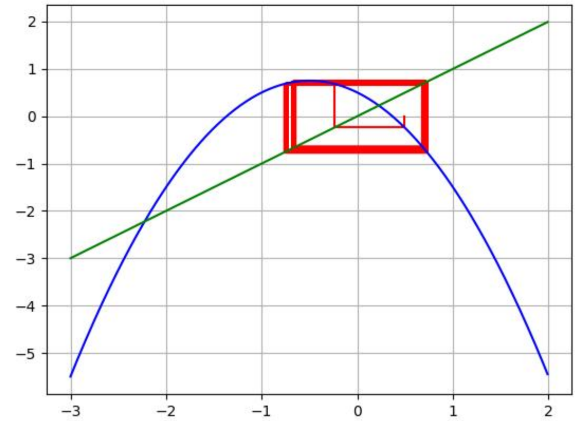


layer 3

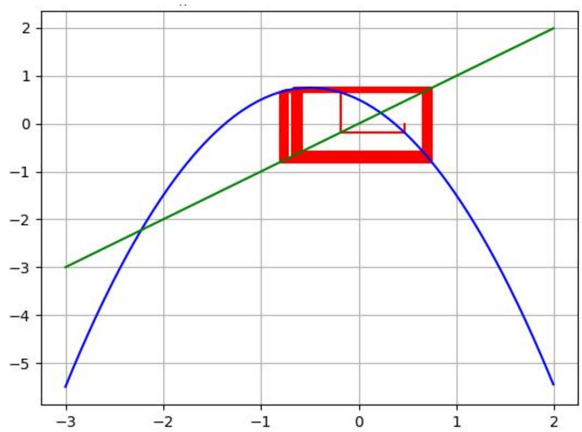
Fig. 2b. Logistic function for each layer when changing the parameter η under the condition: n = 10 neurons in the layer, N = 100 epochs.



layer 1 x_0=0.608



layer 2 x_0=0.494



layer 3 x_0=0.469

Fig. 2c. Diagram of logistic function $x_{n+1}=\eta - x_n - x_n^2$ *for each layer when changing the parameter η under the condition: n = 10 neurons in the layer, N = 100 epochs, η=0.501.*

Depending on the value of the parameter η, the mapping has a different number of fixed points. At $0 < \eta < 0.4$ the number of fixed points does not change, but both fixed points are unstable. Since the mapping is limited, the absence of a point attractor means the formation of a more complex attractor of the limit cycle type. Although the map itself does not have stable fixed points, its square may have such stable fixed points. Therefore, the bifunction diagram in this area shows the branching of the lines. When the values of the parameter η > 0.4 (Fig. 2, *a*), the limit cycle loses stability. With such parameter values, stable fixed points must be sought in higher-order mappings.

The situation with the period of the limit cycle of mappings of higher orders exists in a certain area of the parameter η, and then changes – there is an increase in local minima, etc.

When the values of the parameter K > 0.78 (Fig. 2, *a*), none of the critical cycles has stability. From this value the chaotic behavior of the system begins. The value of x_n does not change periodically. Such chaos is called deterministic because there is a clear, strictly defined law by which the value of a variable can be determined at any iteration, starting from the chosen initial value, but there is no periodicity in its behavior. If in the region of stability of boundary cycles the behavior of reflections was weakly dependent on the initial point, then in the region of chaos a small change in the initial value leads to a significant change in the value of the n-th iteration [7].

In the region of chaos, there are "windows of transparency", when at certain values of the parameter η the reflection behavior becomes regular (Fig. 2, *a*). The fractality of this map is due to the similarity of the process of doubling the boundary cycles. Therefore, Fig. 3, *b*–*d*, on an enlarged scale shows the doubling of the limit cycle. According to these dependencies, this process takes place in a single scenario.

3. Conclusions

Thus, in a multilayer neural network, a chaotic state can occur due to an increase in the speed of learning. The resulting chaotic state is sensitive to changes in the parameters of NN. A slight change in the parameters of NN causes significant changes in the chaotic state. The appearance of local solutions (local minima) due to the increase in the speed of learning leads to the appearance of bifurcations on the dependence of the learning error on the number of epochs. Such dynamics of the magnitude of the error from the number of epochs at a constant rate of learning is the cause of a chaotic state, and shows that it is impossible to obtain the magnitude of the zero error. That is, there is a limit value of the error that can be achieved. Based on the studies of the branching diagram, this error value is almost independent of the configuration of the multilayer NN. One way to solve this problem (to avoid chaotic NN states) is to automatically determine the minimum number of solutions in the diagram of the logistic function, and hence the value of the speed at which the number of local minima is doubled.

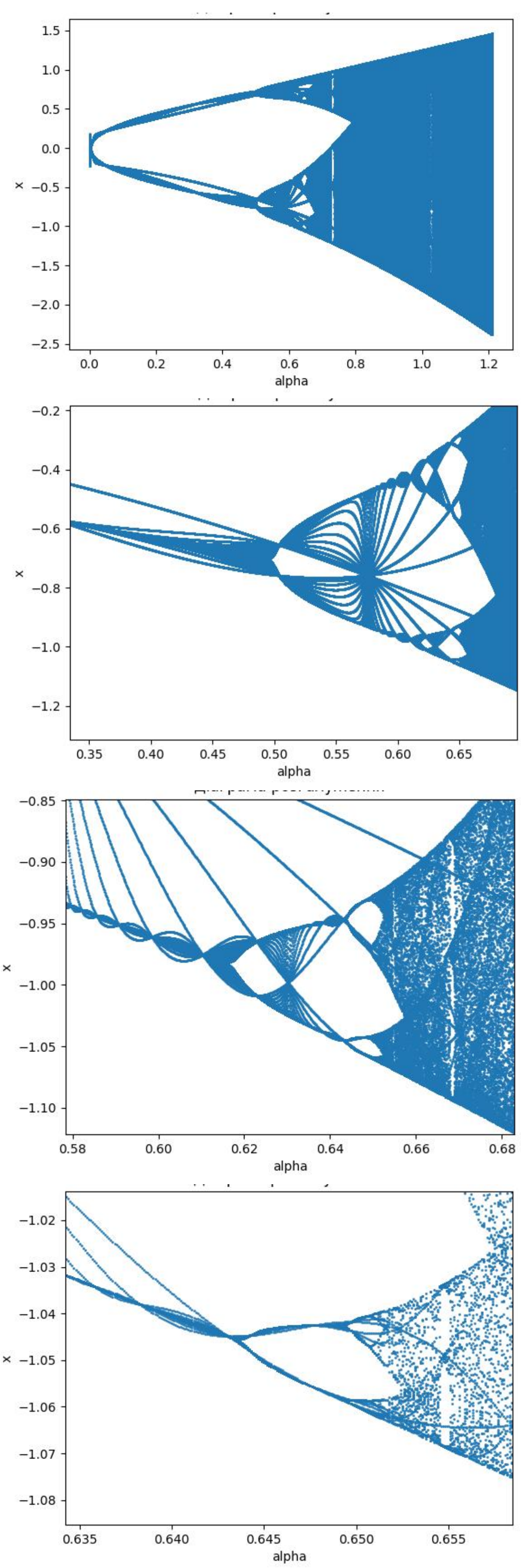


Fig. 3. The thin structure of the branching diagram when changing the parameter η in the third layer provided: n = 10 neurons in the layer, N = 200 epochs.

The algorithm for solving this problem is to determine the number of solutions on the diagram of the logistic function at a given value of the learning speed, which is the process of increasing the number of bifurcations by two or four times, and determines the optimal value of learning speed and optimal error.

References

[1] O. Rudenko and E. Bodianskyy, *Artificial neural networks.* Kharkiv, Ukraine: SMIT Company, 2006. (Ukrainian)

[2] S. Subotin, *Neural networks: theory and practice* Zhytomyr, Ukraine: Publisher O. O. Evenok, 2020. (Ukrainian).

[3] V. Kruglov and V. Borisov *Artificial neural networks. Theory and practice.* Moscow, Russia: Hotline – Telecom, 2001. (Russian)

[4] S. Subotin and A. Olijnyk *Neural networks: tutorial* Zaporizhzhia, Ukraine: ZNTU, 2014. (Ukrainian)

[5] Yurij Olenych, Sergiy Sveleba, Ivan Katerynchuk, Ivan Kunio, and Ivan Karpa, "Features of deep study neural network" *https://openreviewhub.org/lea/paper-2019/features-deep-study-neural-network#*

[6] Yu Taranenko, "Information entropy of chaos", *https:// habr.com/ru/post/447874/*

[7] A. Kuznetsov, *Dynamical systems and bifurcations. Saratov*, Russia: LLC Publishing Center "Science", 2015. (Russian)

БАГАТОШАРОВІ НЕЙРОННІ МЕРЕЖІ – ЯК ДЕТЕРМІНОВАНІ СИСТЕМИ

Сергій Свелеба, Володимир Бригілевич, Іван Катеринчук, Іван Куньо, Іван Карпа, Остап Семотюк, Ярослав Шмигельський, Назар Свелеба

В роботі досліджено вплив швидкості навчання (η) на процес навчання багатошарової нейронної мережі. Програма для багатошарової нейронної мережі була написана мовою Python. Швидкість навчання розглядалась як постійна величина і визначалась її оптимальна величина, за якої досягалось найкраще навчання. Для аналізу впливу швидкості навчання використовувалась логістична функція, яка описує процес навчання. Показано, що функція похибки навчання характеризується біфуркаційними процесами, які призводять до хаотичного стану, якщо η>0,8. Визначено оптимальне значення швидкості навчання, яке визначає появу процесу подвоєння кількості локальних мінімумів, і становить для тришарової нейронної мережі з 4 нейронами в кожному шарі η=0,62. Збільшення кількості прихованих шарів (3÷30), та кількості нейронів у кожному шарі (4÷150) не приводить до кардинальної зміни діаграми логістичної функції (x_n,η), а отже, і оптимальної величини швидкості навчання.

Sergii Sveleba – Doctor, Professor of Optoelectronics and Information Technologies Department of Ivan Franko National University of Lviv.

Research interests: complex systems, mathematical modeling, neuron networks.

Volodymyr Brygilevych – PhD, Associate Professor of the Institute of Technical Engineering at the State Higher School of Technology and Economics in Jarosław, Poland.

Research interests: mathematical modeling of processes in mechatronic systems, diagnosis of electronic systems.

Ivan Katerynchuk – PhD, Associate Professor of Optoelectronics and Information Technologies Department of Ivan Franko National University of Lviv.

Research interests: complex systems, mathematical modeling, FPGA.

Ivan Kuno – Ph.D, Deputy Head of Research and Development Department at Ivan Franko National University of Lviv, Ukraine.

Research interests: micro and nano-sized dielectric crystals, dynamics of incommensurate structures, dimensional effects, nano- and microelectronics, mathematical modeling.

Ivan Karpa – PhD, Associate professor of Optoelectronics and Information Technologies Department of Ivan Franko National University of Lviv.

Research interests: complex systems, mathematical modeling.

Yaroslav Shmyhelskyy – Associate professor of Radiophysics and Computer Technologies Department of Ivan Franko National University of Lviv.

Research interests: theory of nonlinear electronic circuits, complex systems, mathematical modeling.

Ostap Semotiuk - PhD, Associate professor of Ukrainian Academy of Printing, Lviv, Ukraine.

Research interests: phase transitions, optical properties of dielectric crystals with an incommensurate phase.

Nazar Sveleba – student of Faculty of Applied Mathematics and Informatics of Ivan Franko National University of Lviv.

Research interests: complex systems, mathematical modeling, neuron networks.

Received: 12.07.2021. Accepted: 25.10.2021