

Fractal structure of training of a three-layer neural network

Bohdan Melnyk[†], Serhiy Sveleba[†], Ivan Katerynychuk^{*†}, Ivan Kuno[†] and Volodymyr Franiv[†]

Ivan Franko National University of Lviv, Universytetska St.1, 79000 Lviv, Ukraine

Abstract

The study of the fractal structure for a multilayer neural network is carried out in the work. The number of neurons in the input and hidden layers corresponded to the size of the input array. The software product was implemented in the Python programming environment for recognizing printed numbers. The sample of the input array for printed digits was 5 plus 4 variants of digit distortion with an error of $\approx 15\%$ for the 3x5 digit array and $\approx 10\%$ for the 4x7 digit array. The sample of the heterogeneous array was 8, and contained 3 options that did not correspond to any number. The study of the fractal structure was performed in three modes of multilayer neural network training, undertraining, satisfactory training, and retraining. It was established that the appearance of the fractal structure is caused by the retraining of neurons. Retraining neurons causes to local minima appears on the training error objective function. This leads to an increase in the error in the formation of the value of the correction function of the training weights. The process of transition of the neural network from the retraining mode to the chaotic mode is determined by the process of doubling the number of local minima on the objective function of the training error. Since the cause of the retraining and chaotic regimes are the same, the formation of the fractal structure in them is similar. Non-homogeneity of the input array negatively affects the formation of the fractal structure of the training process.

Keywords

the multilayer neural network, the fractal structure, Adam optimization method

1. Introduction

The process of formation of the objective function of the training error of a neuron is determined by the contribution of each neuron from the previous layer [1]. Such a process of formation of the neural network training error indicates that the objective function of

MoMLeT-2024: 6th International Workshop on Modern Machine Learning Technologies, May, 31 - June, 1, 2024, Lviv-Shatsk, Ukraine

* Corresponding author.

[†] These authors contributed equally.

✉ bohdan.melnyk@lnu.edu.ua (B. Melnyk); serhiy.sveleba@lnu.edu.ua (S. Sveleba);
ivan.katerynychuk@lnu.edu.ua (I. Katerynychuk); ivan.kuno@lnu.edu.ua (I. Kuno);
volodymyr.franiv@lnu.edu.ua (V. Franiv)

ORCID 0000-0001-6399-6317 (B. Melnyk); 0000-0002-0823-910X (S. Sveleba); 0000-0001-8877-8324 (I. Katerynychuk); 0000-0001-6092-7949 (I. Kuno); 0000-0001-9856-1962 (V. Franiv)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the training error should be considered as a set of periodic functions that determine the existence of neural network training modes. Namely, the mode of undertraining, satisfactory training, and retraining. It is known [2] that the process of retraining is associated with the appearance of local minima on the objective error function, and causes an increase in the value of the training error. The appearance of local minima, according to the work [2], is described in the first approximation by the logistical function of doubling their number from the training step. Doubling the number of existing local minima ultimately leads to the emergence of a chaotic mode of neural network training. In [2], the process of training a neural network with the dynamics of an incommensurate superstructure is compared. In particular, in this work it was noted that in the sinusoidal mode of an incommensurate superstructure is characterized by the absence of harmonics. An increase in the magnitude of the anisotropic interaction is accompanied by the appearance of harmonics of an incommensurate superstructure, which leads to the soliton regime of an incommensurate superstructure. A further increase in the anisotropic interaction causes the appearance of a block structure characterized by different periodicity, and the average value of the wave vector of an incommensurate superstructure for a given ensemble can take an incommensurate value. At the same time, the formation of a chaotic phase can be traced. A similar picture can be traced in the process of training a neural network. So, with an increase in the training step, in the mode of retraining individual neurons, the appearance of local minima can be traced. An increase in the number of local minima in the first approximation is described by the process of doubling their number with an increase in the training step. This process is described by a recurrence function on the form:

$$x_{n+1} = \alpha - x_n - x_n^2$$

This is confirmed by the results of the Fourier analysis of the objective function of training error and the appearance of branching diagrams [2].

The fractality of an incommensurate superstructure is determined both by the processes of harmonic occurrence and by the processes of nucleation and annihilation of solitons. A number of works are devoted to the analysis of fractality of an incommensurate superstructure. The fractality of the neural network training process may be determined by the appearance of local minima on the objective function of the training error. That is, the process of retraining a neural network may be fractal in nature. Thus, the study of the fractal structure of the training error function in different modes of neural network training will confirm or identify new mechanisms for the formation of training error.

The study of fractal structure in stochastic systems is important for understanding and predicting their dynamics. It can assist in explaining complex structures and interactions in chaotic systems, as well as in developing effective methods for modeling and controlling such systems. Therefore, the task of this work is to establish a picture of the neural network training process, and the features of the formation of training error in the retraining mode.

2. Methodology

Since the appearance of local minima is described by the process of doubling their number by the training error function, the reflection of fractality will be carried out with the help of a logistic function that describes the doubling process, that is, a complex function on the form:

$$z_{n+1} = -z_n - z_n^2$$

Taking into account the peculiarities of this mapping, namely, that the variable z_n is a complex quantity in which the real part is the training step (*alpha*), and the imaginary part is the value of the weight correction function (w). The image of the fractal structure was performed in the coordinates: the value of the correction of the training weights w , the training step *alpha*, is the speed of moving away from the solution of points that are not included in the solution of this system is represented by different colors on figures.

The dynamics of the fractal structure was investigated depending on the training parameters of the neural network, in particular on the number of iterations and the dimension of the input array and its homogeneity, the training step and the parameters for optimizing the training process. The study of the fractal structure was performed for a multilayer neural network. The number of neurons in the input and hidden layer corresponded to the size of the input array. The program was written in Python and performed the recognition of printed numbers. The training data for printed digits were 5 representations of the digit plus 4 variants with a distortion of the digit with an error of 15% for the 3x5 figure setting array and >10% for the 4x7 digit setting array. We used the optimization method of training Adam [3], which is characterized by a monotonous process of training a neural network. On the basis of this architecture, a study of the fractal structure was carried out when applying this method of training optimization to the objective function of the training error. When studying the fractal structure from the neural network training parameters, in particular from the number of iterations and the size of the input array and its homogeneity, and the training step, the optimization parameters for this optimization method were selected according to the works [4] and were, for $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

3. Homogeneous training data

According to the Fourier spectra of the training error function, the retraining process begins to be traced in the vicinity of the training step *alpha* = 0.45. Therefore, let's consider the formation of a fractal structure when changing the *alpha* training step in the range of 0.1 - 0.7. Fig.1 shows the view of the fractal structure from the number of iterations, provided that the optimization method of Adam training is applied with the optimization parameters $\beta_1=0.9$ and $\beta_2=0.999$. Under the condition of 10, 100, 500 iterations of the fractal structure image for the number "0" (Fig.1) in the retraining mode of the neural network, a complex boundary is demonstrated, which gradually reveals smaller and smaller recursive details when zoomed in. A set boundary is made up of

smaller versions of the basic form, so the fractal property of self-similarity refers to the entire set, not just a part of it. With an increase in the number of iterations, there is a change in the fractal picture. Namely, the appearance of smaller recursive details can be traced. It is known [5] that with an increase in the number of iterations, the process of retraining can be traced, which may prompt a change in the picture of fractality when the number of iterations changes. Similar fractal structures have been obtained for other printed figures. Thus, the presence of a fractal structure indicates that in the mode of retraining the neural network, and taking into account the Fourier spectra of the error function, there is an increase in the number of local minima, and in the first approximation this process is described by the function of doubling their number. The mechanism of transition to the chaotic mode of neural network training is described by the process of doubling the number of local minima.

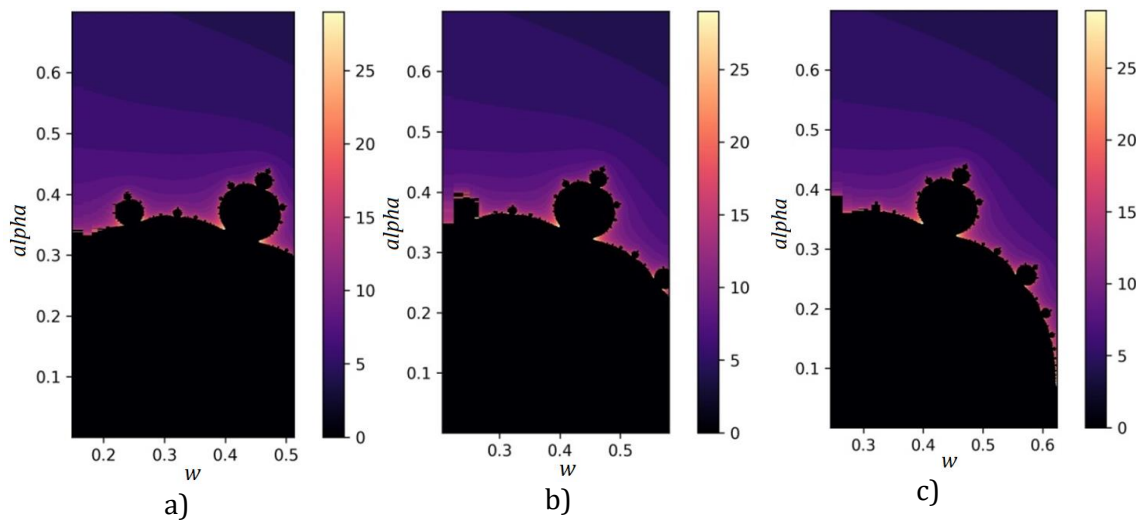


Figure 1: Fractal structure of training of a three-layer neural network when recognizing printed digit "0" given by a 3x5 array depending on the training step, with the number of iterations: a) 10, b) 100, c) 500.

It is known that for this multilayer neural network, according to Fourier spectra, the process of retraining begins to be traced in the vicinity of the value of the training step 0.45. With a further increase in the training step, the chaotic training mode of the neural network begins to manifest itself. In order to identify the manifestation of the neural network retraining process in the fractal structure, a study of the fractal structure from the size of the training step was carried out. Fig.2 shows the view of the fractal structure when changing the step α training. Starting from $\alpha = 0.3$, the recurrence system describes the magnitude of the training error from the training step and demonstrates the absence of a solution to the system. Under the condition $\alpha = 0.4$, this system has a single solution, and is characterized by almost no retraining process. That is, this mode of neural network training demonstrates a satisfactory training process. A further change in the training step leads to the appearance of two, four, and so on stable solutions, followed by a transition to the mode described by the doubling process and a transition to a chaotic

training mode. Under these conditions, this mode of training is described by a fractal structure ($\alpha=0.4\div 0.7$). According to the obtained fractal structure in Fig.2, the chaotic training mode of the neural network is described by the appearance of additional small details of different levels. Thus, the process of transition from the retraining mode to the chaotic training mode of the neural network is accompanied by an increase in the number of local minima, and therefore the appearance of additional small details of the fractal structure. Similar results were obtained for other figures.

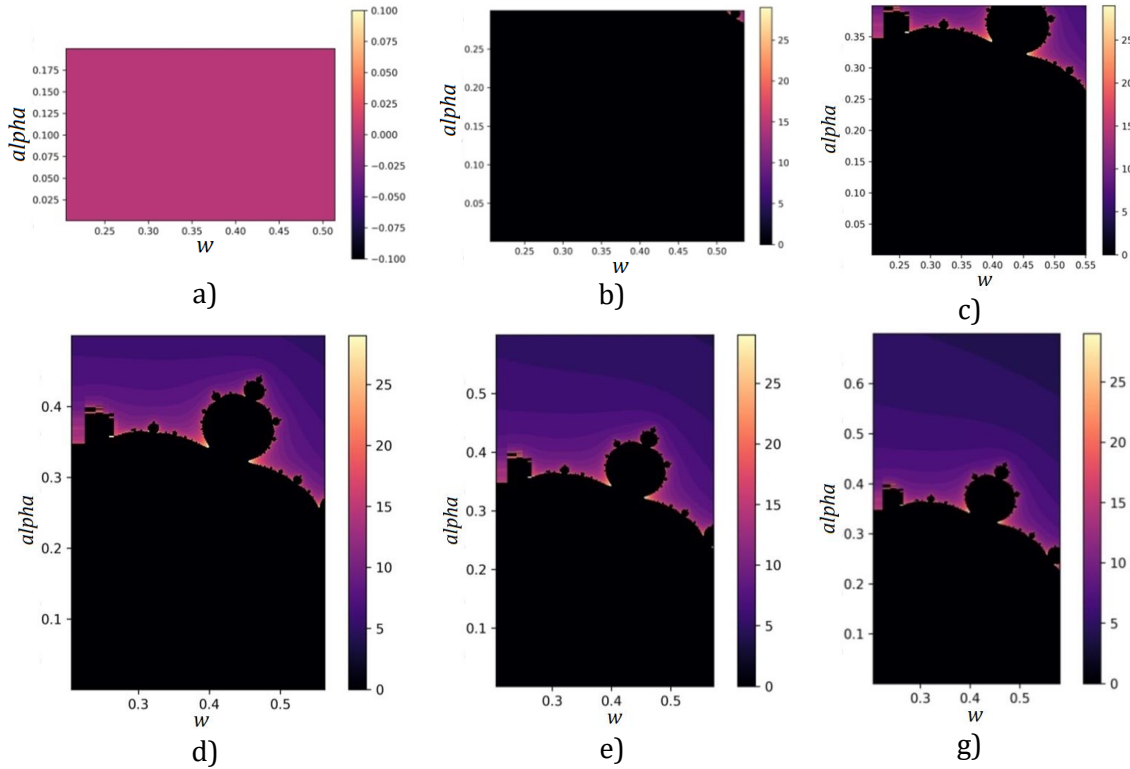


Figure 2: Fractal structure of training of a three-layer neural network when recognizing printed digit "0" given by a 3x5 array depending on the training step: a) 0.2, b) 0.3, c) 0.4, d) 0.5, e) 0.6, g) 0.7, 100 iterations.

The key quantity that describes a fractal quantitatively is the "fractal dimension". However, in different sources, this term is understood as different quantities: the Minkowski dimension, the Hausdorff-Bezikovich dimension, the self-similarity dimension. The Hausdorff-Bezikovich dimension DH is a measure of the division of an object into parts of size r , followed by counting the number of $N(r)$ parts covering the object under study [6]. Fig.3 shows the results of the fractal dimension calculated by the Hausdorff-Bezikovich method when changing the training step. The dependence of the fractal dimension on the training step also indicates the emergence of a satisfactory neural network training mode in the vicinity of $\alpha = 0.3$, and the appearance of the neural network retraining process at $\alpha > 0.3$. A further increase in the training step leads to

an increase in the fractal dimension, which indicates the transition to a chaotic mode [7] of neural network training. Similar dependencies were obtained for other figures.

It is known that an increase in the number of iterations can also lead to overtraining of the neural network. To confirm this statement, consider the effect of the number of iterations on the shape of the fractal structure. For this purpose, at the $\alpha=0.3$ training step, a study of the impact of the number of iterations on the neural network training process was conducted. The training step was chosen close to the value at which the process of retraining the neurons of the neural network begins to be traced. That is, the fractal structure was investigated at the training step $\alpha = 0.3$, provided that the number of iterations increased. At a given value of the training step of a multilayer neural network, a certain number of iterations are required to achieve the retraining mode. Fig.4 shows the fractal structure for the digit "0" given by a 3x5 array, provided $\alpha = 0.3$, when changing the number of iterations and applying the Adam optimization method to the training error function.

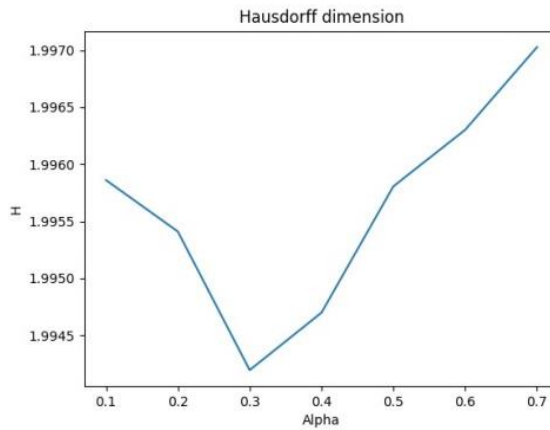


Figure 3: Dependence of the fractal dimension on the training step α under the condition of 100 iterations, for the digit "0" given by a 3x5 array.

According to Fig.4, under the condition of 10 iterations, the mode of undertraining can be traced. At 100 iterations, a satisfactory training mode can be traced, with the emergence of a retraining mode. At 500 iterations and above, the mode of retraining the neural network with the formation of a fractal structure is clearly manifested. Thus, the emergence of a fractal structure is due to the process of retraining the neural network. Similar patterns of fractal structure from the number of iterations were obtained for other figures. That is, with an increase in the number of iterations, the formation of a fractal structure can be traced, which indicates that starting with a certain number of iterations, the process of retraining can be traced, and this process is associated with the process of the emergence of local minima and doubling of their number.

Fig.5 shows the dependence of the fractal dimension calculated by the Hausdorff-Bezikovich method on the number of iterations, the training step $\alpha = 0.3$, for the figure "0" given by a 3x5 array. According to Fig.5, the dependence of the fractal dimension on

the number of iterations is characterized by a minimum of 100 iterations, and then increases with an increase in the number of iterations. Taking into account the results given in Fig.4, in the vicinity of 100 iterations, this system is characterized by a satisfactory training mode with no retraining mode of the neural network. With an increase in the number of iterations, there is a retraining of neurons, with the formation of a fractal structure (Fig. 4), and an increase in the fractal dimension (Fig. 5).

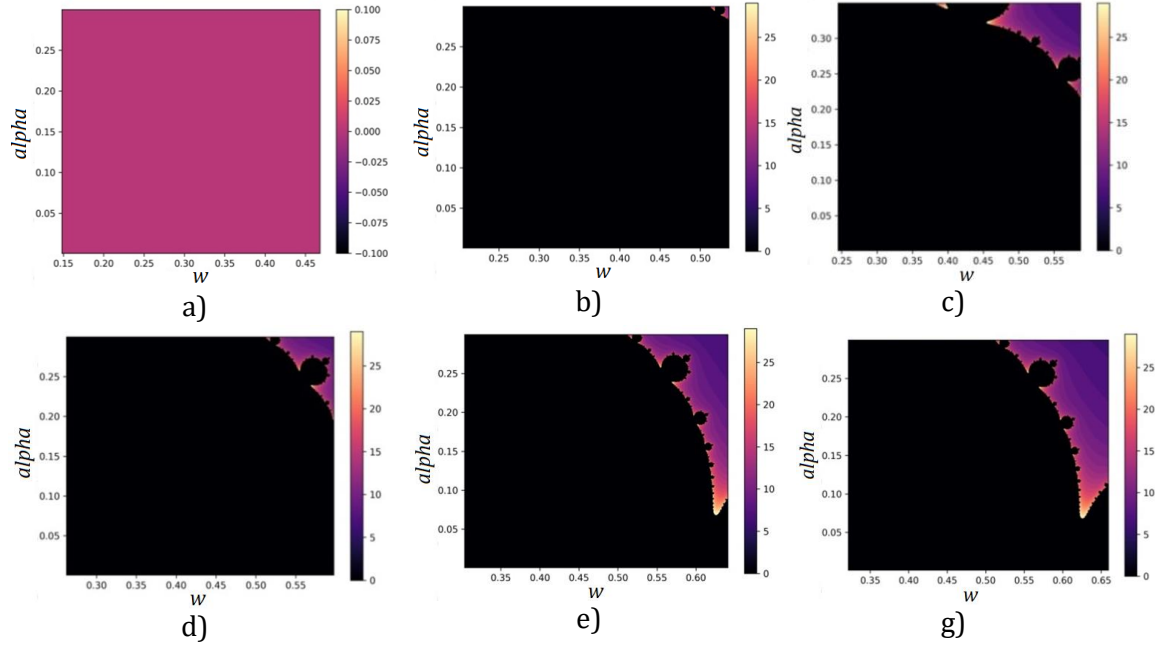


Figure 4: Fractal structure of training of a three-layer neural network when recognizing printed digit "0" given by a 3x5 array depending on the number of iterations: a) 10, b) 100, c) 500, d) 1000, e) 5000, g) 10000, the maximum training step $\alpha = 0.3$.

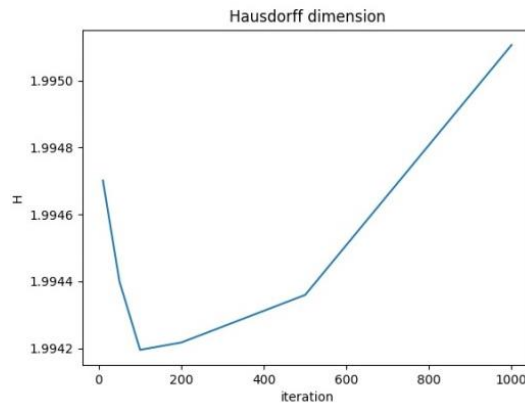


Figure 5: Dependence of fractal dimension on the number of iterations, $\alpha=0.3$, for the figure "0" given by a 3x5 array.

4. Heterogeneous training data

Let consider the figure of the influence of the parameter β_2 on the formation of the fractal structure. Figure 6 shows pictures of the fractal structure depends the parameter β_2 , which characterizes the degree of attenuation of the previous values of the square of the gradient of the objective function of the training error.

According to [10], this parameter is decisive in the process of training a neural network, and its optimal value is equal to 0.999. Fig.6 shows pictures of the fractal structure when changing the optimization parameter β_2 in the range of $0.1 \div 0.9999$. The obtained fractal structures in Fig.6 do not undergo a significant change when the β_2 parameter changes. Although it should be noted that with an increase in the value of β_2 , the picture of the fractal structure becomes richer in the presence of smaller fragments. The values of the fractal dimension given in Table 1 also demonstrate the above pattern. That is, there are no qualitative changes in the vicinity of the value of $\beta_2 = 0.999$. It is possible that the tendencies of changes in the training error from the value β_2 , which were declared in [10], can be manifested with larger sizes of the input array, or for more heterogeneous arrays.

Table 1.

The dependence of the fractal dimension on the value of the optimization parameter β_2 of the Adam optimization method with the optimization parameter $\beta_1=0.9$, provided that the maximum training step is $\alpha=0.5$, for the digit "0" given by a 3×5 array.

β_2	Hausdorff dimension
0.1	1.99584
0.9	1.99548
0.99	1.99628
0.999	1.99584
0.9999	1.99592

The dynamics of the fractal structure for this input array from the training step (Fig.7) demonstrates similar dynamics of changes as for the array with inhomogeneity $\approx 15\%$ (Fig.2). In the vicinity of $\alpha=0.3$, a homogeneous training process can be traced with almost no retraining mode of the neural network. A further increase in the training step causes the emergence of a retraining mode, which subsequently makes the transition to a chaotic training mode. According to Fourier studies of the spectra of the training error function [2], already at $\alpha > 0.5$, the emergence of a chaotic training mode of the neural network can be traced. According to Fig.8, the fractal structure does not show changes when changing the training mode from retraining to chaotic.

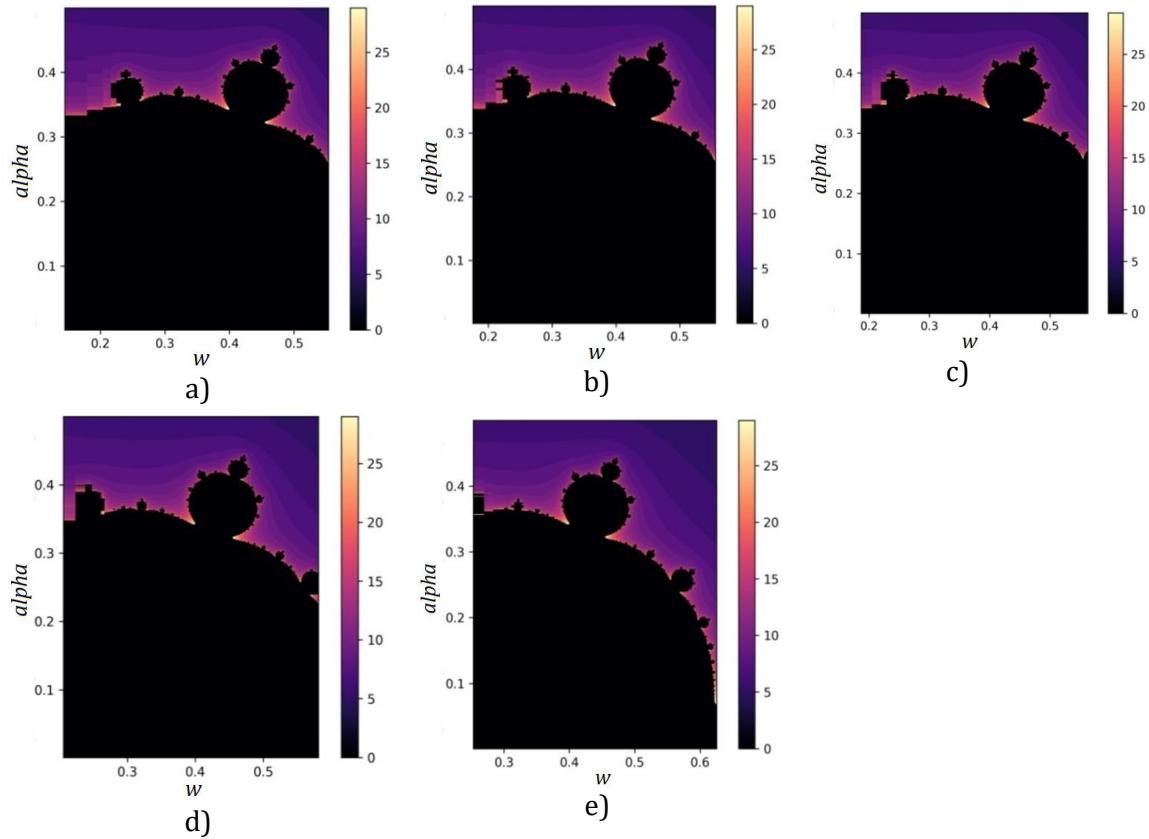


Figure 6: Fractal structure of training of a three-layer neural network in recognition of printed numbers depending on the value of β_2 : a) 0.1, b) 0.9, c) 0.99, d) 0.999, e) 0.9999, provided that the maximum training step is $alpha = 0.5$, 100 iterations, for the digit "0" given by the array and the heterogeneity of the input array $> 15\%$.

Although the magnitude of the fractal dimension from the training step, in this range of changes $alpha$ is characterized by an increase. This indicates an increase in the randomness of the training mode at $alpha > 0.4$. It is possible that the transition to a chaotic mode of training is associated with a change in the smaller fractal structure. To confirm or refute this assumption, let us consider the influence on the fractal structure of such training parameters as the number of iterations and the optimization parameter β_2 . An increase in the number of iterations at the beginning leads to a transition to the retraining mode of neural networks (100 iterations), and subsequently to a chaotic training mode. The transition from the retraining mode to the chaotic mode is accompanied by an increase in the fine structure of the lower (first and second) level.

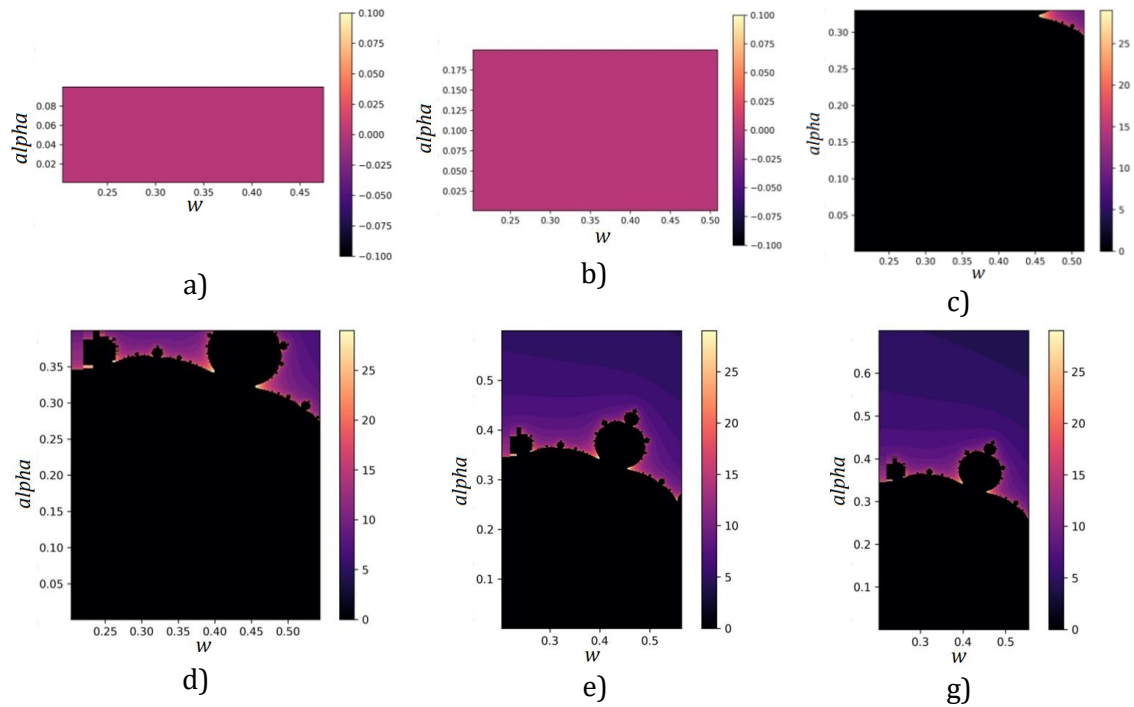


Figure 7: Fractal structure of three-layer neural network training when recognizing printed numbers given by a 3x5 array depending on the training step: a) 0.1, b) 0.2, c) 0.33, d) 0.5, e) 0.6, g) 0.7, 100 iterations, for the digit "0" and the heterogeneity of the input array $\approx 40\%$.

Comparing the obtained fractal structure under the condition of non-homogeneity of the input array $\approx 40\%$ with the obtained fractal structure with the structure under the condition of non-homogeneity of the input array $\approx 15\%$ (Fig. 4), it can be argued that an increase in the heterogeneity of the input array leads to a decrease in the number of small details on the fractal structure.

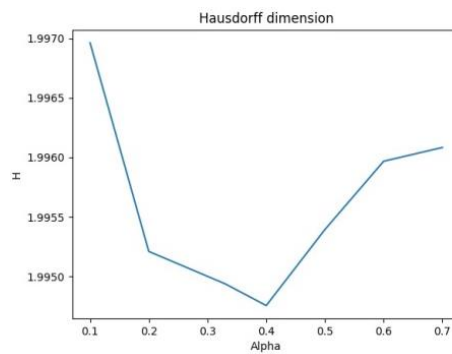


Figure 8: Dependence of the fractal dimension on the training step alpha under the condition of 100 iterations, for the digit "0" given by a 3x5 array, with non-homogeneity of the input array $\approx 40\%$.

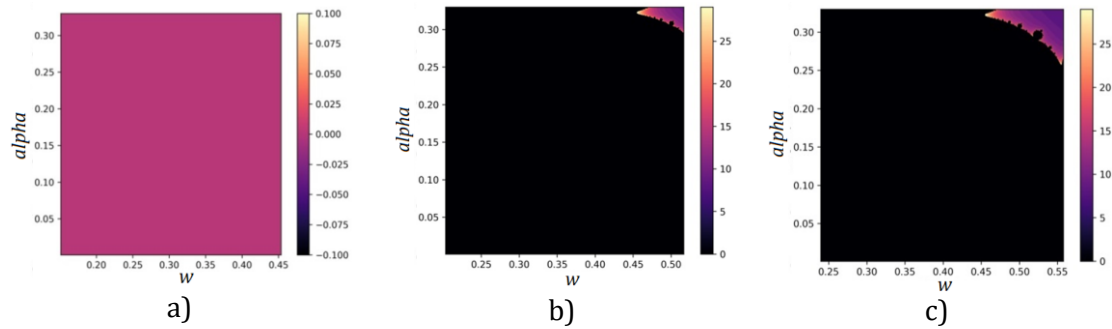


Figure 9: Fractal structure of training of a three-layer neural network when recognizing printed numbers given by an array of 3x5, depending on the number of iterations: a)10, b)100, c)500, the maximum training step $\alpha = 0.3$, for the digit "0" and the heterogeneity of the input array $\approx 40\%$.

It is known [8] that the parameter β_2 affects the degree of attenuation of the previous values of the square of the gradient of the objective function of the training error, and is characterized by the minimum value of the training error at $\beta_2 = 0.999$. Fig.10 shows fractal structures at different values of the optimization parameter β_2 , provided that the maximum training step is $\alpha = 0.5$, 100 iterations, for the number "0" and the input array is not homogeneous $\approx 40\%$.

When the parameter β_2 is changed in the range of $0.1 \div 0.99$, no changes in the fractal structure can be observed. At $\beta_2 = 0.999$, a change in the small details of the fractal structure begins to be traced. Namely, the spatial areas of their existence are beginning to increase. A further change in the β_2 parameter leads to a more pronounced picture of such changes. The calculated fractal dimension according to the Hausdorff-Bezikovitch method (Table 2) shows a similar dynamics from the optimization parameter β_2 . Namely, in the interval of change of the optimization parameter $\beta_2 = 0.1 \div 0.999$, there is a decrease in the value of the fractal dimension, reaching the smallest value at $\beta_2 = 0.999$. A further change in β_2 leads to a sharp increase in the fractal dimension.

Table 2.

The dependence of the fractal dimension on the value of the optimization parameter β_2 of the Adam optimization method with the optimization parameter $\beta_1=0.9$, provided that the maximum training step is $\alpha=0.5$, for the digit "0" given by a 3x5 array with the inhomogeneity of the input array $\approx 40\%$.

β_2	Hausdorff dimension
0.1	1.99689
0.9	1.99670
0.99	1.99608
0.999	1.99539
0.9999	1.99608

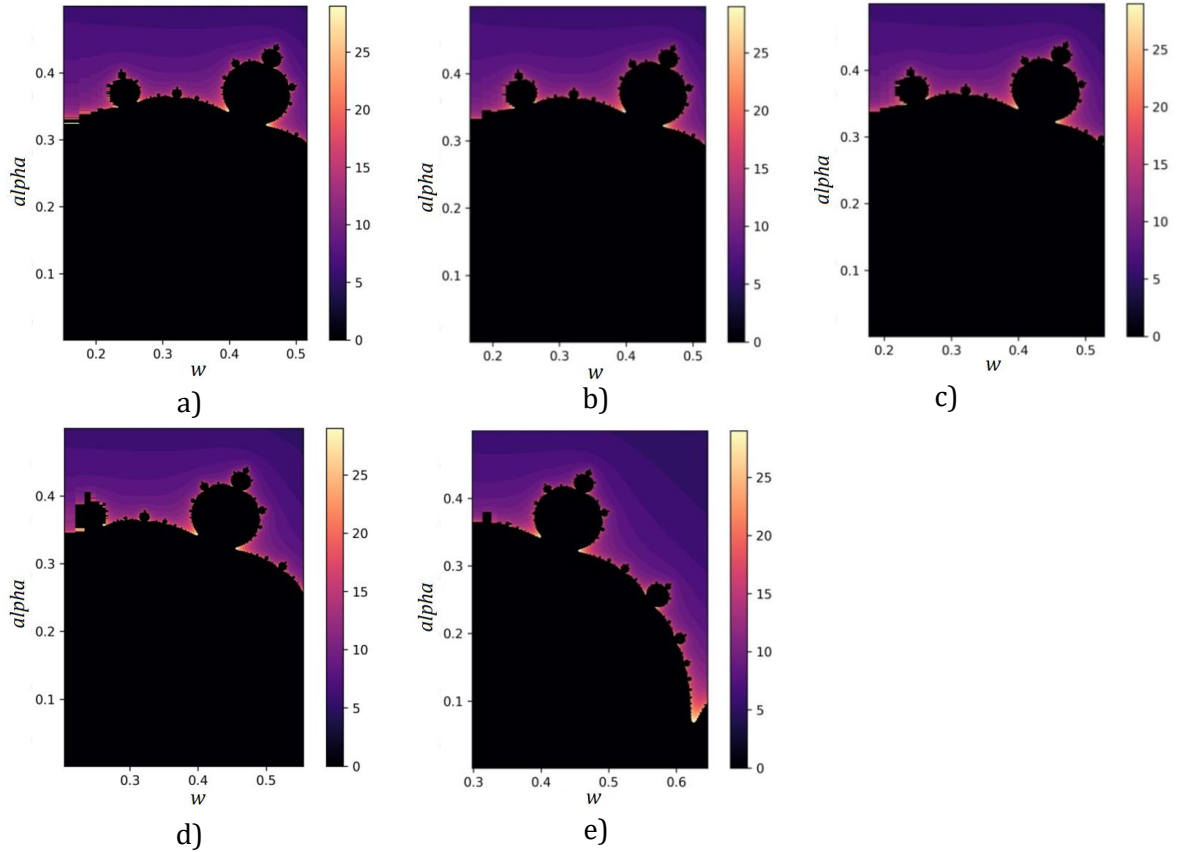


Figure 10: Fractal structure of training of a three-layer neural network when recognizing printed digits given by a 3x5 array, depending on the value of β_2 : a) 0.1, b) 0.9, c) 0.99, d) 0.999, e) 0.9999, provided that the maximum training step $\alpha = 0.5$, 100 iterations, for the number "0" and the heterogeneity of the input array $\approx 40\%$

Let's consider the effect of the size of the input array on the fractal structure of the neural network. Fig.11 shows the fractal structure with a different number of iterations for the array of setting the figure 4x7, and the heterogeneity of the sample $\approx 10\%$ (Fig.11,a) and $\approx 40\%$ (Fig.11,b), from the change in the training step $\alpha = 0.01 \div 0.7$, and with the optimization parameter $\beta_2=0.999$. An increase in the number of iterations is accompanied by a slight change in the fractal structure due to its shift along the real axis. The real axis in our case corresponds to the change in the training step. Therefore, a shift to the region of higher values of the real part may indicate a redistribution of contributions of a particular mode of neural network training. As you know, with an increase in the number of iterations, there is an increase in the role of the retraining mode, and in the future, the role of the chaotic training mode. An increase in sample heterogeneity ($\approx 40\%$) does not lead to a change in the overall picture of the fractal structure (Fig. 11, b). As noted above, an increase in the heterogeneity of the figure sample leads to a decrease in changes in the fractal pattern of the structure. Comparing samples with heterogeneity of $\approx 10\%$ and $\approx 40\%$, a similar pattern can be noted.

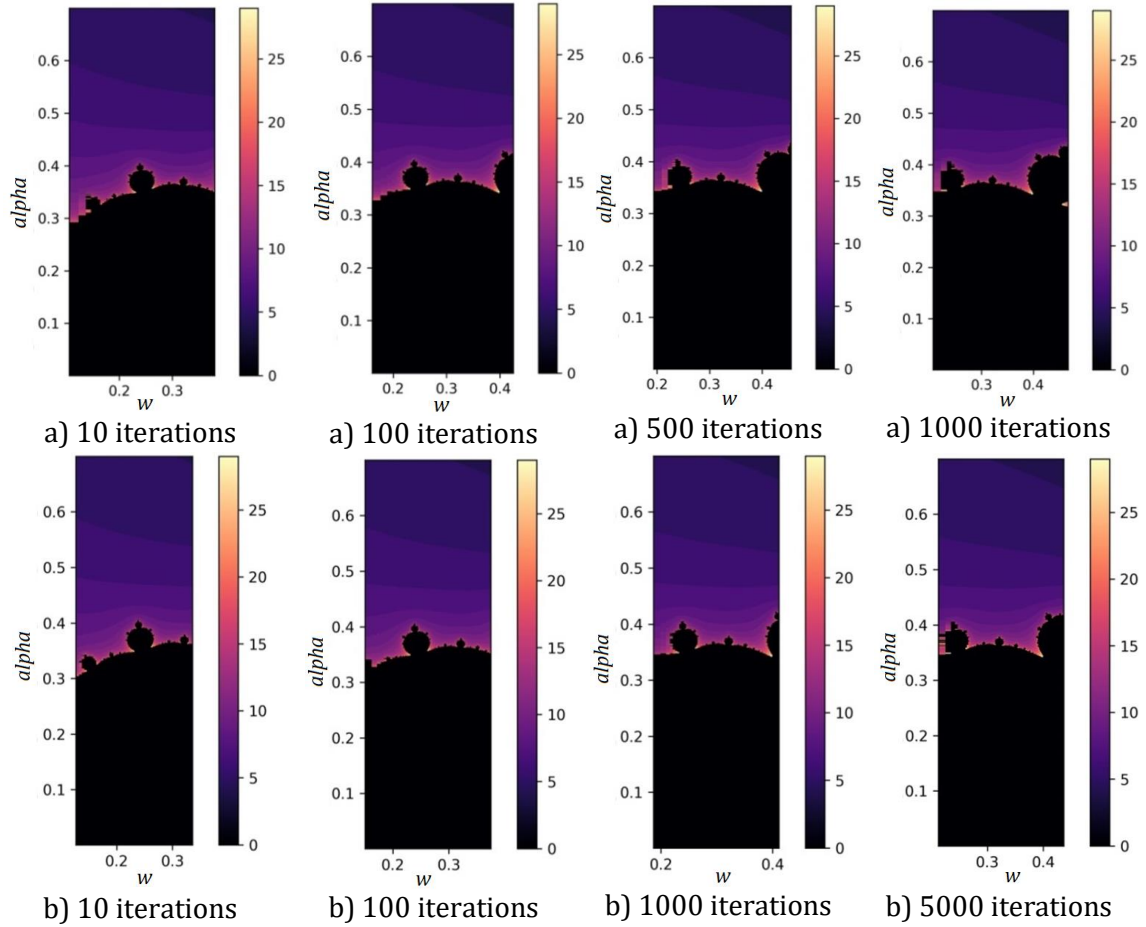


Figure 11: Fractal structure of training of a three-layer neural network when recognizing printed numbers depending on the training step, when representing a digit in a 4×7 array, with heterogeneity of the input array $\approx 10\%$ a) and $\approx 40\%$ b), for the digit "0".

Taking into account the above-mentioned dependencies of changes in the fractal structure on the training parameters (number of iterations, training step, optimization parameter β_2), a common pattern can be noted. An increase in the value of the training parameters causes a shift in the picture of the fractal structure in the interval of higher values of the actual part of its representation. In our opinion, this indicates that the retraining mode of the neural network and the chaotic mode are equally involved in the formation of the fractal structure. This is not surprising, since the reasons for the occurrence of retraining mode and chaotic mode are the same. If there are differences in fractal structures that describe the retraining mode and the chaotic mode, they are related to fine details.

5. Conclusions

So, summarizing the above, it can be noted that the fractal structure of the neural network training process is due to the retraining of neurons. Neuronal overtraining causes the

appearance of local minima on the target function of the training error. This leads to an increase in the error in the formation of the value of the correction function of the training weights. The process of transition of the neural network from the retraining mode to the chaotic mode is due to the process of doubling the number of local minima on the target function of the training error. Since the reason for the occurrence of the retraining mode and the chaotic mode are the same, the retraining mode of the neural network and the chaotic mode are equally involved in the formation of the fractal structure in the process of training the neuron. Heterogeneity of the input array has a negative impact on the formation of the fractal structure of the training process.

References

- [1] S. O. Subbotin, Neural networks: theory and practice: teaching. (Ed. O. O. Evenok), Zhytomyr, 2020, 184p.
- [2] B. Melnyk, S. Sveleba, I. Katerynychuk, I. Kuno and V. Franiv, Multilayer Neural Network Training Error when AMSGrad, Adam, AdamMax Methods Used. COLINS-2024: 8th International Conference on Computational Linguistics and Intelligent Systems, April 12–13 Lviv, Ukraine, 2024, pp. 232-254 URL: <https://ceur-ws.org/Vol-3664/paper17.pdf>
- [3] D. P. Kingma, J. B. Adam, A Method for Stochastic Optimization 3rd International Conference for Learning Representations, San Diego, 2015, URL: <https://doi.org/10.48550/arXiv.1412.6980>
- [4] J. Ma, D. Yarats, Momentum and Adam for deep learning, 7th ICLR: New Orleans, LA, USA, 2019, pp.19-21 URL: <https://arxiv.org/abs/1810.06801>
- [5] K. Kawaguchi, Effect of Depth and Width on Local Minima in Deep Learning Neural Computation. MIT Press Volume 31 Issue 7, 2019, pp. 1462-1498. URL: https://doi.org/10.1162/neco_a_01195
- [6] O. V. Kapustyan, V. V. Pichkur, V. V. Sobchuk, Theory of Dynamic Systems, Vezha-Druk, Lutsk, Ukraine, 2020, 348 p.
- [7] I. Y. Adashevska, O. O. Kraievska, Self-similarity as a characteristic property of fractal. Fractal (fractional) dimension of Hausdorff, Scientific achievements of modern society: abstr. of 4th Intern. Sci. and Practical Conf., Liverpool, United Kingdom, 4-6 December 2019, pp. 603-612. URL: http://sci-conf.com.ua/wp-content/uploads/2019/12/scientific-achievements-of-modern-society_4-6.12.19.pdf.
- [8] Y. Dokkyun , A. Jaehyun and J. Sangmin, An Effective Optimization Method for Machine Learning Based on ADAM, Appl. Sci. Volume 10, 2020, pp. 1073-1093, URL: <https://www.mdpi.com/2076-3417/10/3/1073>
- [9] X. Zeng, Z. Zhang and D. Wang, AdaMax Online Training for Speech Recognition, CSLT TECHNICAL REPORT-20150032, 2016, URL: http://www.cslt.org/mediawiki/images/d/df/Adamax_Online_Training_for_Speech_Recognition.pdf
- [10] S. J. Reddi, K. Satyen, & S. Kumar, On the Convergence of Adam and Beyond. 6th ICLR, Vancouver, BC, Canada, 2018, p.23-35 URL: <https://arxiv.org/abs/1904.09237>