Prerequisites for the analysis of the neural networks functioning in terms of projective geometry

Yelizaveta S Vitulyova¹, Akhat S Bakirov¹,²*, Dina B Shaltykova³ and Ibragim E Suleimenov²

¹ Almaty University of Power Engineering and Telecommunications, Almaty, Republic of Kazakhstan
² National Academy of Engineering of the Republic of Kazakhstan, Almaty, Republic of Kazakhstan
³ Institute of Information and Computational Technologies, Almaty, Republic of Kazakhstan

axatmr@mail.ru

Abstract. It is shown that during training an artificial neural network built on elements with a threshold activation function, the weighting coefficients of each neuron can be selected from a certain discrete set, the nature of which depends on the number of neurons inputs. An example of a neuron with three inputs is considered in detail. It is shown that there are only 8 non-trivial combinations of weight coefficients that specify the work of neurons of this variety: any other combination of weight coefficients that describe the work of a non-degenerate neuron with this number of inputs can be replaced by one of the above 8 combinations without changing the character of the neuron functioning in the network. A method has been developed that allows one to determine relevant combinations of weights for neurons with an arbitrary number of inputs. It is shown that the proposed approach creates the prerequisites for assessing the “information power” of neural networks based on data on the nature of the connections between the neurons that form it.

1. Introduction
There is an extensive class of tasks related to the classification of some objects according to one or another group of attributes, for the solution of which neural networks are used [1-5]. Such tasks de facto come down to finding an empirically determined rule, according to which a certain sequence of variables taking logical values assigned to an object from a set of images represented in digital form. We emphasize that this rule is found empirically in the process of training a neural network. Moreover, the existing training procedures for neural networks are guided by a continuous change in weighting coefficients. This, inter alia, is due to the fact that for the implementation of existing training procedures, as a rule, one has to use differentiable neuron activation functions.

At the same time, those neural networks that are used to solve the problems of image classification, in principle, could be solved using neural networks using threshold activation functions, which follows from the very nature of the problem (mapping an object described by a set of binary variables from one set, to a similar object from another set).

In this paper, it is shown that analysis of neural networks’ operation using the threshold activation function can be significantly simplified by using the ideas of projective geometry. It is shown that the set of weighting factors that specify the operation of a neuron with a threshold activation function can
actually be selected from some discrete set, the nature of which is also established in this paper. This creates the prerequisites for a significant simplification of training procedures for artificial neural networks (ANN) that solve classification problems, as well as other tasks whose nature allows the use of ANNs whose elements have a threshold activation function.

There is also a well-defined problem, usually interpreted as “the neural networks logical opacity” [6]. In particular, in the vast majority of cases, the weights of a neural network are formed empirically in the process of its training using specific images. Algorithms that make it possible to judge why the weight coefficients formed during the training take exactly such and not other values, as a rule, remain unknown.

This does not create significant problems from the point of view of the applied use of ANNs, however, their logical opacity does not allow solving many problems that arise when trying to use the neural networks theory in other areas of knowledge.

So, the neural networks theory, in principle, can be used to interpret such concepts as “mentality”, “Collective unconscious” [7] and other phenomena that describe society [8]. Indeed, let us consider two people entering into conversation with each other. It is customary to say that in this case two individuals communicate. However, this is nothing more than an approximation, and very rude. In fact, we are talking about the fact that the neurons that make up the brain of each interlocutor exchange signals with each other. From the point of view of the neural network functioning, it is absolutely not important what kind of nature those signals exchanged by neurons possess. Therefore, we can conclude that as a result of the dialogue, a certain common neural network is formed, consisting of two fragments, each of which is localized within the brain of two communicating individuals. Continuing this logic, it is easy to conclude that there is a global communication network that can be identified with the noosphere, understood in the spirit of V.I. Vernadsky’s point of view [9]. Moreover, there is every reason to believe that information that can only be stored indirectly in connection with the information recorded by the memory of individuals can be stored in an extensive neural network. It can be assumed that it is this information that determines the nature of what is called “collective unconscious”, “dictate of the environment” in humanitarian literature, etc.

Today, statements of this kind are nothing more than philosophical conclusions arising from the logic of dialectical positivism [10-12]. Their verification by means of the theory of neural networks, however, creates quite certain prerequisites for the transition to a quantitative description of several processes taking place in human society. However, in order to proceed to quantitative conclusions, first of all, it is necessary to solve the problem of neural networks logical opacity.

The results of this work, in which it is shown that the theoretical description of neural networks with threshold activation functions can be significantly simplified, allows us to make a certain step in this direction.

2. The essence of the proposed approach
Let’s start from the expression connecting the variable characterizing the output state of neuron $Y$ with the variables characterizing the state of inputs $X_i$.

$$Y = \theta(w_1X_1 + w_2X_2 + \cdots + w_NX_N), \quad (1)$$

where $w_i$ - weighting factors,

$$\theta(x) = \begin{cases} -1, & x \leq 0 \\ +1, & x > 0 \end{cases} \quad (2)$$

Relation (1) can be used to analyse the operation of neural networks of the type under consideration without loss of generality, since for a similar formula with a constant term $a$

$$Y = \theta(w_1X_1 + w_2X_2 + \cdots + w_NX_N + a) \quad (3)$$

one can formally increase the number of considered inputs per unit and put $w_{N+1} = a, X_{N+1} = 1$.

Without loss of generality, we can also assume that all variables $X_i$ can only take discrete values of -1 and +1. (Often the case is considered when these variables acquire the values 0, + 1, but it reduces to the considered elementary change of variables.)
Formula (1) is easy to interpret from a geometric point of view. Indeed, it actually defines a section of a hypercube in N-dimensional Euclidean space by a hyperplane passing through the origin

\[ w_1 X_1 + w_2 X_2 + \cdots + w_N X_N = 0 \]  

Relation (1) indicates that the variable Y takes the value +1 if the product of the normal vector to a given hyperplane and the radius vector of one of the vertices of the hypercube is positive, and -1 in the opposite case. In other words, the hyperplane (4) divides the N-dimensional space into two regions, and the value of the variable Y depends on which one the corresponding vertex of the hypercube falls into.

Expression (4) can also be considered from the perspective of projective geometry [13], that uses the concept of dual space. Namely, record (4) is symmetric with respect to the sets of quantities \((w_1, w_2, \ldots, w_N)\) and \((X_1, X_2, \ldots, X_N)\), i.e. it can be considered both as an equation of a hyperplane in the original space (then the set of quantities \((w_1, w_2, \ldots, w_N)\) is considered as specifying the orientation of this hyperplane), and as an equation of a hyperplane in a dual space. The orientation of this hyperplane in dual space is set, respectively, by a set of quantities \((X_1, X_2, \ldots, X_N)\).

Accordingly, the set of quantities \((w_1, w_2, \ldots, w_N)\) has the meaning of the coordinates of a point in dual space. A point in the original space becomes a hyperplane in dual space, and a hyperplane becomes a point. For the case \(N = 3\), this is illustrated by Fig. 1, which shows a cube in three-dimensional space, i.e. a set that exhausts the possible states of the inputs of a neuron described by function (1), and the figure into which it passes in dual space (regular octahedron). Six vertices of an octahedron correspond to six faces of a cube, and eight vertices of a cube correspond to eight faces of an octahedron.

![Figure 1. In dual space, a cube transforms into an octahedron.](image)

Thus, a set of weighting factors that specify the operation of a neuron with a threshold activation function is represented by a point in a dual N-dimensional space. The set of vertices of an N-dimensional hypercube in a dual space becomes a set of hyperplanes, each of which also divides the dual space into two parts.

We say that this set of weights \(\{w_i\}\) belongs to the set \(\Gamma^+\left(\{X_i\}\right)\) if the condition is met

\[ \sum_i^N w_i X_i > 0 \]  

and, conversely, it belongs to the set \(\Gamma^-\left(\{X_i\}\right)\) if the condition is met

\[ \sum_i^N w_i X_i \leq 0 \]  

For a hypercube of dimension N there exist exactly \(2^{N-1}\) sets \(\Gamma^+\left(\{X_i\}\right)\) that do not coincide with any of the sets \(\Gamma^-\left(\{X_i\}\right)\), and exactly \(2^{N-1}\) of such sets \(\Gamma^-\left(\{X_i\}\right)\). This is due to the fact that for each vertex defined by the set of binary variables \(X_i \in \{-1, 1\}\), i.e. vector
\[ \vec{n} = (X_1, X_2, X_3, ..., X_N) \]  

(8)

It is possible to specify a vertex that is characterized by an oppositely directed vector

\[ \vec{n} = -(X_1, X_2, X_3, ..., X_N) \]  

(9)

Let’s form sets

\[ \Lambda(\vec{\beta}) = \bigcup_{\vec{\beta}} \mathcal{F}^N([X_i]^{\pm N}) \]  

(10)

where the multi-index \( \vec{\beta} \) denotes the sequence of the choice of signs of the sets \( \mathcal{F}^N([X_i]^{\pm N}) \) of length \( 2^{N-1} \).

Let the functioning of a neuron with N inputs be described by the activation function (2). Then, the result of applying to the inputs of any sequences N binary variables for two different activation functions described by row vectors \( \vec{w}^{(1)} = (w_1^1, w_2^1, w_3^1, ..., w_N^1) \) and \( \vec{w}^{(2)} = (w_1^1, w_2^1, w_3^1, ..., w_N^1) \) will be the same if \( \vec{w}^{(1)} \in \Lambda(\vec{\beta}), \vec{w}^{(2)} \in \Lambda(\vec{\beta}) \), i.e. these vectors belong to the same set \( \Lambda(\vec{\beta}) \). In other words, neurons connected to the network with weights \( \vec{w}^{(1)} \) and \( \vec{w}^{(2)} \) work indistinguishably.

This follows from the fact that for two vectors from this set their scalar products by vectors (10) corresponding to any possible combination of binary variables supplied to the inputs of a neural network will have the same sign.

This proves the following theorem.

**Theorem 1**

*In the formula describing the activation of a neuron (2) when applying to its input a set of binary signals capable of taking values -1 and +1, the vector of weighting coefficients \( (w_1, w_2, w_3, ..., w_N) \) belonging to the set \( \Lambda(\vec{\beta}) \) can be replaced by any other vector from this set.*

Formally, such sets can be formed \( 2^{2^{N-1}} \), i.e. there are no more than \( 2^{2^{N-1}} \) different vectors describing possible variants of sets of weight coefficients that are de facto distinguishable. In reality, the very nature of function (2) leads to the fact that there are much fewer possible options.

We show this by the example of a neuron with three inputs.

For the case of a neuron with three inputs, the activation function has the following form.

\[ Y = \theta(w_1X_1 + w_2X_2 + w_3X_3) \]  

(11)

Based on this function, it can be argued that all possible areas in the projective space, which will differ in the nature of the reaction of the neuron to the state of the input variables, are determined by the position of the following four planes.

\[ w_1 + w_2 + w_3 = 0 \]  

(12)

\[ w_1 + w_2 - w_3 = 0 \]  

(13)

\[ w_1 - w_2 + w_3 = 0 \]  

(14)

\[ w_1 - w_2 - w_3 = 0 \]  

(15)

In other words, these four planes, formulas (12)-(15) divide the entire dual space into well-defined subdomains. If two sets of weights belong to the same area, then, as shown above, the corresponding neurons will behave identically.

In other words, as follows from what was said above, there are as many possible variants of the implementation of neurons as there are subdomains into which the dual space is divided by planes defined by equations (12)-(15).

For the three-dimensional case under consideration formally, there should be 16 such regions. In reality, there are exactly fourteen such subdomains; this is due to the fact that some intersections of the planes defined by equations (12) - (15) are empty sets.
It is possible to choose the following values of vectors characterizing the subdomains into which the dual space splits.

\[
\begin{align*}
\vec{w}_1 &= \pm(1,1,1), \quad \vec{w}_2 = \pm(1,1,-1), \quad \vec{w}_3 = \pm(1,-1,1), \quad \vec{w}_4 = \pm(1,-1,-1) \\
\vec{w}_5 &= \pm(1,0,0), \quad \vec{w}_6 = \pm(0,1,0), \quad \vec{w}_7 = \pm(0,0,1), \quad \vec{w}_8 = \pm(0,0,0)
\end{align*}
\] (16) (17)

In the list of such vectors in formulas (16) - (17), a vector \(\vec{w}_8\) is also formally added, all of whose components are zero. This subdomain is a degenerate subdomain that is formed when the intersection of some sets \(\Gamma = \left\{ X_i \right\}_{i=1}^{\Xi^N} \) defined above is empty. In other words, there are no sixteen transfigurations, as follows from the formal calculation given above, but fourteen. Two regions are degenerate and correspond to the fact that the neuron is actually absent, i.e. all components of the vector are equal to zero.

The nature of the vectors set defined by formulas (16)-(17), which can be called basic, is illustrated by Figure 2.

![Figure 2: Correspondence of vectors from the set (16), (17) to the cuboctahedron faces](image)

Further, as follows directly from formulas (17), eight out of 16 variants are actually degenerate. Indeed, if only one weighting factor in the considered formula is nonzero, then the neuron de facto does not have three inputs, but only one. The cases corresponding to formulas (17) also correspond to situations when two of the three inputs are not connected to the circuit.

Thus, we have proved that any non-degenerate neuron that has three inputs and a stepwise activation function can in fact be described through weighting factors, which also take only -1 and +1 values. In other words, this result admits a well-defined generalization, more precisely, it holds

**Theorem 2**

For any non-degenerate neuron with an odd number of inputs, the set of weighting coefficients can be replaced by one of the sequences of the form \(q_1, q_2, ..., q_N\), the elements of which can only take values -1 or +1.

By “non-degenerate” in this formulation we mean a neuron, none of which weights are equal to zero. The proof of the theorem is given in the same way as was used for the case \(N = 3\).

This theorem suggests that the weights of non-degenerate neurons with an odd number of inputs and a threshold activation function are de facto described by vectors from the same set as the sets of input variables. This set can also be associated with the vertices of a hypercube with edge 2 in an \(N\)-dimensional space.

The case of even \(N\) is somewhat more complicated for analysis, but even in this case it is possible to show that the real number of possible sequences of weighting coefficients is relatively small.
3. Conclusion

Thus, the results of this work show that instead of arbitrary combinations of weighting coefficients describing the operation of neural networks with threshold activation function, we can take combinations of weighting coefficients whose elements acquire discrete values taken from a well-defined discrete set.

Outside the scope of the work, there remains the question of training algorithms for neural networks with threshold activation functions and sets of weighting coefficients selected from certain fixed sets. It is solved using the method of "logical differentiation", the consideration of which is beyond the scope of this work.

The most significant conclusion of this work, however, is that the obtained results allow us to estimate the amount of information that is stored in the weight coefficients of the neural network. The possibility of going over to discrete sets clearly indicates that Shannon's formulas are applicable to its calculation. Thus, prerequisites are created in order to try, for example, compare the amount of information stored in a collective neural network (noosphere) with the amount of information directly related to the memory of individuals. Such a comparison, obviously, represents the first step towards building quantitative models that describe human society from the standpoint of the neural networks' theory. The relevance of creating such models is not in doubt due to the increasing role of telecommunications in society and the actual formation of complex human-machine systems.

4. References

[1] Basheer I A, Hajmeer M 2000 Artificial neural networks: fundamentals, computing, design, and application Journal of microbiological methods 43 (1) pp 3-31 DOI: https://doi.org/10.1016/S0167-7012(00)00201-3

[2] Ibragim S. et al. 2020 Example of the use of artificial neural network in the educational process Future of Information and Communication Conference, Springer, Cham pp 420-430 DOI: https://doi.org/10.1007/978-3-030-39445-5_31

[3] Romanovsky A V and Shokin Ya V 2010 Neuroeconomics and its integration into economic science. Economics 70 (9) pp 42-44

[4] Ciregan D, Meier U, Schmidhuber J 2012 Multi-column deep neural networks for image classification 2012 IEEE conference on computer vision and pattern recognition pp 3642-49 DOI: 10.1109/CVPR.2012.6248110

[5] Liang M, Hu X 2015 Recurrent convolutional neural network for object recognition Proceedings of the IEEE conference on computer vision and pattern recognition pp 3367-75

[6] Burrell J 2016 How the machine ‘thinks’: Understanding opacity in machine learning algorithms Big Data & Society 3(1) DOI: https://doi.org/10.1177/2053951715622512

[7] Gunderson K 1985 Mentality and machines University of Minnesota Press

[8] Suleimenov I and Bakirov A 2018 Semi-empirical Theory of Distributions of Users of Social Online Networks by the Number of Contacts MATEC Web Conf. 214 DOI: 10.1051/matecconf/201821402001

[9] Vernadsky V I 2013 A few words about the noosphere Noosphere studies 1 pp 6-17

[10] Suleimenov I, Massalimova A, Bakirov A and Gabrielyan O 2018 Neural Networks and the Philosophy of Dialectical Positivism MATEC Web Conf. 214 DOI: 10.1051/matecconf/201821402002

[11] Suleimenov I E, Gabrielyan O A, Bakirov A S and Vitulyova Ye S 2019 Dialectical Understanding of Information in the Context of the Artificial Intelligence Problems IOP Conference Series: Materials Science and Engineering 630 DOI: 10.1088/1757-899X/630/1/012007

[12] Gabrielyan O, Suleimenov I, Vituleva E 2019 Artificial intelligence in the context of noosphere studies SCTCMG 2019 International Scientific Conference «Social and Cultural Transformations in the Context of Modern Globalism» https://doi.org/10.15405/epsbs.2019.12.04.130

[13] Coxeter H S M 2003 Projective geometry Springer Science & Business Media