

The Performance of Neural Network - Statistical Evaluation

Hovhannisyan H.G.
Russian University
of Economics, Armenia
hovhannisyan@list.ru

Verlinski S.V.
National Polytechnic
University of Armenia
sergeyverl@mail.ru

Ciudin R.
"Lucian Blaga" University
of Sibiu, Romania
rodica.ciudin@ulbsibiu.ro

Cioca, M.
"Lucian Blaga" University
of Sibiu, Romania
marius.cioca@ulbsibiu.ro

Abstract—The problem of evaluating the performance of the neural network, based on a study of the probabilistic behavior of the network is considered in this paper. Direct propagation network consisted of layer of input nodes, hidden layer and output layer is examined. For such networks estimates for some of the statistical characteristics of the neural network in the case of two recognized classes were obtained. The problem of constructing and realization a self-organizing neural network-type map for solving the problem of data classification is also considered. The neural network receives on input the input vector, make preprocesses in separation of significant components in the input data, and then decides the classification problem.

Index Terms—neural network, the weight of the neuron, recognition of the stimulus, self-organizing map.

I. INTRODUCTION

Artificial neural networks are often used for a variety of applications, such as pattern recognition. For the successful application of artificial neural networks is necessary to choose the right network architecture, to pick up its parameters, thresholds elements, activation function, and others [1], [2], [6], [7], [9]. Currently, neural computation of different levels of implementation, from specialized hardware to the neural network software packages are becoming more widely used.

They are successfully used to solve a number of tasks such as forecasting economic and financial indicators, the prediction of complications in patients in the postoperative period, biometric identification based on various characteristics, image processing, and others.

All this is possible for the neural networks because of their ability to learn, the establishment of associative connections in the input data. Typically, depending on the task, required to resort to different methods to transform input data, allowing a correct judgment about the laws and special features in the data that reflect their quality characteristics [4].

There is a problem data filtering and display them in the space of smaller dimension.

II. PROBLEM DESCRIPTION

The neural network technologies particularly important role played by the selection of the network architecture and the values of its parameters that affect its efficiency.

Artificial neural network is a mathematical model of biological neuron. It consists of relating to each other neurons. A neural network is exposed to learning by providing it with

the input information in the form of numerical sequences. The training set up internal connections between neurons through which the network is endowed ability to recognize unfamiliar images.

There are different types of neural networks that differ in both the topology and the learning algorithms. In the problems of classification of some of the most effective are self-organizing maps, which unsupervised learning algorithm. They consists of two layers: the input and functional. The functional layer neurons are located on the grid composed of cells. Each neuron takes a single cell and is connected to all the neurons of the input layer (Fig.1).

Self-organizing map operates on the principle of "winner takes all" [3]. That is, when a sequence of input vectors presented to the network, all the neurons of the functional layer cause activation function, which reflects the relationship between the input data and weight characteristics of these neurons. The winner is the neuron whose activation function (for example, function of the Euclidean distance between the vector of the input data and the weight vector of the neuron) takes the optimum value, such as the maximum or minimum.

If $V_1, V_2, V_3, \dots, V_n$ - input vector, a functional layer is self-organizing maps with k neurons, the winner will be the one neuron for which where W_{ij} is weight coefficient i -th neuron functional layer associated with the j -th neuron of the input layer.

$$\sqrt{\sum_{j=1}^n (V_j - V_{ij})^2} = \min, i = 1, \dots, k \quad (1)$$

At the stage of network training presented training sequence data. The essence of the training is that at each step is determined by the neuron-winner, and neurons, located in a neighborhood of the neuron-winner more close to him. The result of training the neurons closer to the destination areas formed in a grid defined training sequence.

The radius of the neighborhood of the neuron-winner varies according to the formula:

$$\sigma(t) = \sigma_0 \exp\left(\frac{t}{\lambda}\right), t = 1, 2, 3, \dots \quad (2)$$

The initial value of the radius of the neighborhood is taken

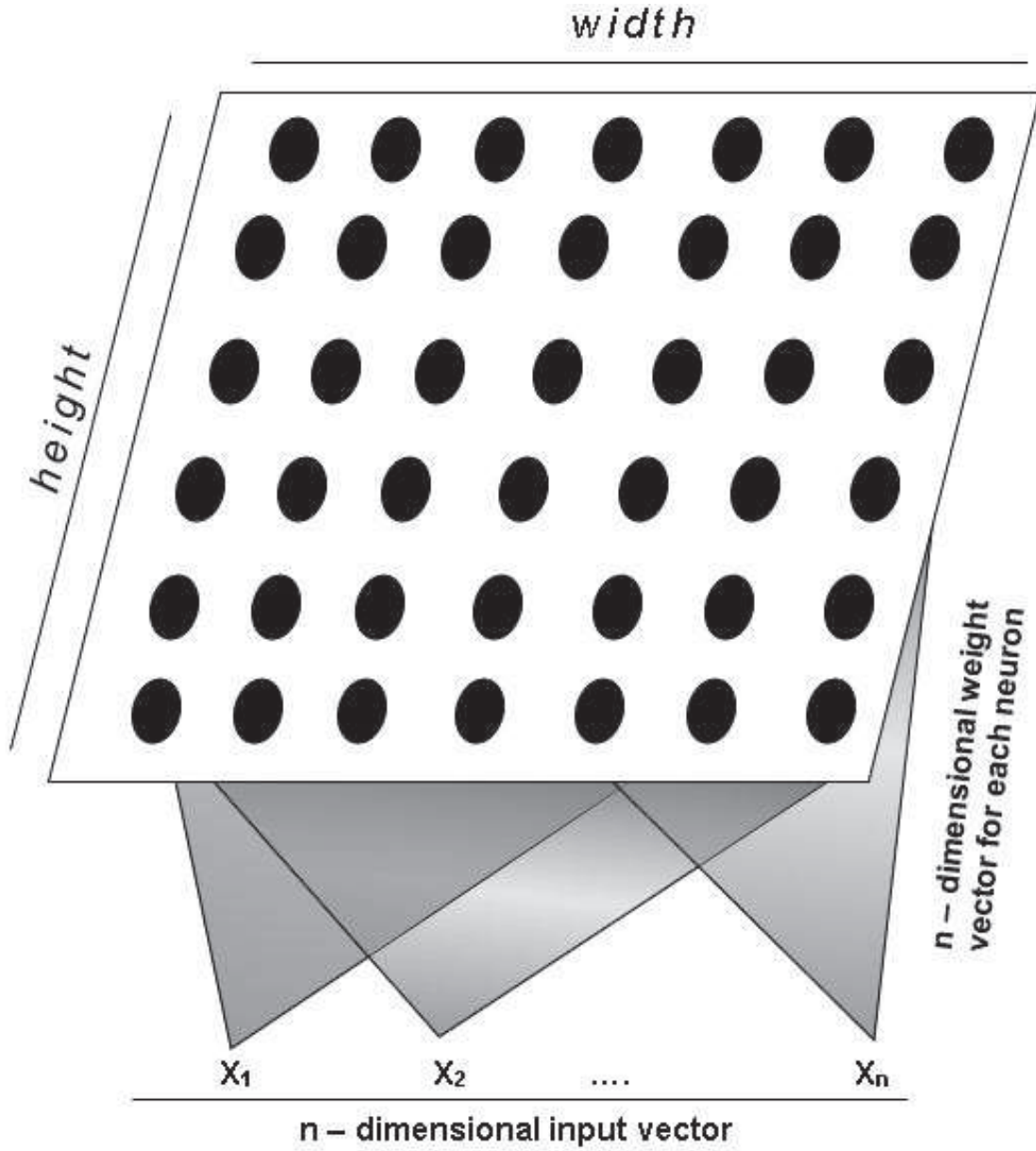


Fig. 1. The structure of self-organizing map.

$$\sigma_0 = \frac{\max(\text{width}, \text{height})}{2} \quad (3)$$

where width is width of the grid of the functional layer, height is height, λ is time constant is calculated by the formula:

$$\lambda = \frac{n}{\log \sigma_0} \quad (4)$$

where n is number of training iterations.

During the training, a network of neurons, neurons of the winner and his σ -neighborhood change their weights as follows:

$$W(t+1) = W(t) + \theta(t)L(t)(V(t) - W(t)) \quad (5)$$

where t is a iteration number,

$$L(t) = L_0 \exp\left(-\frac{t}{\lambda}\right), t = 1, 2, 3, \dots \quad (6)$$

$$\theta(t) = \exp\left(-\frac{\text{dist}^2}{2\sigma^2(t)}\right), t = 1, 2, 3, \dots \quad (7)$$

dist - the distance between the neuron and the neuron-winner on the grid of the functional layer. $L(t)$ function is called learning coefficient, $\theta(t)$ reflects the impact of the location of the neuron on the grid. Obviously, the most changed the weight of the winning neuron (Fig. 2).

Before learning network converts the input vectors on the allocation of the main components, whereby the dimension

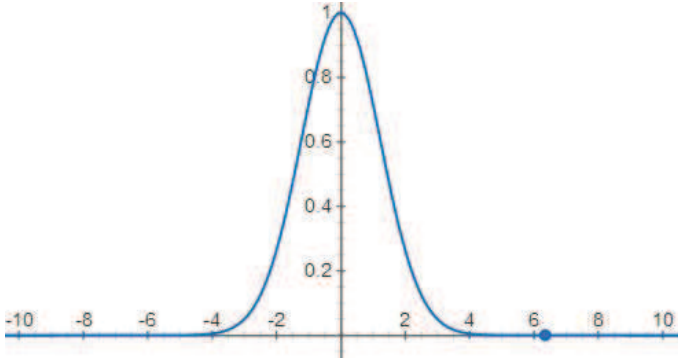


Fig. 2. Effect of the distance from the neuron-winner at changing the balance

of the input space decreases and, consequently, the number of neurons in the input layer of the network. The trained network is tested for control of the input data and evaluate the effectiveness of its work. Training and tested neural network capable of recognizing data, similar to those on which it is trained.

Of particular interest is the study of the probabilistic behavior of the neural networks [8]. The neural network of direct distribution, which consists of a layer of input nodes, hidden layer and output layer is investigated. Neurons have a one-way communication, do not contain links between elements within the layer and backward linkages between the layers. The neurons of the input layer are connected to the hidden layer neurons excitatory and inhibitory connections randomly.

The outputs of all neurons in the hidden layer neurons are connected to the output layer. Neurons in each layer is referred to as the input, hidden and output elements, respectively [2], [7]. In addition, we note that the software engineers can develop software and implement the type of self-organizing neural network card for data classification solutions, and develop a library of special Java-classes, providing parallel processing.

III. PROBLEM

If each class has, respectively, l_1 and l_2 representatives, so after training sequence $L = l_1 + l_2$ two classes activators, so in k class hidden items accumulated weight:

$$V_k = \sum \delta_i \sum \eta_k(j) + V_0, (i = 1, 2), (j = 1, l_i) \quad (8)$$

where V_0 - initial weight of k - th hidden element. When at the entrance to a network of the pathogen ξ_i the input of the output layer is fed weight ($k = 1, N_A$)

$$U_t = \sum V_k \eta_k(t) = \sum \left(\sum \delta_j \sum \eta_k(j) + V_0 \right) \eta_k(t) \quad (9)$$

where

$$U_t = \sum \sum \delta_i \sum \eta_k(j, t) \quad (10)$$

$$\eta_k(j, t) = \begin{cases} 1, & \text{if } k\text{-th hidden element is active} \\ & \text{for } \xi_j \text{ and } \xi_t \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

The pathogen belonging of pathogen ξ_t to one of two classes is determined by comparing the weight U_t with a threshold output element R . When $U_t > \theta_R$ so we have first class. When $U_t < \theta_R$ - second class. When $U_t = \theta_R$ - refusal of recognition.

For the given network, a predetermined training sequence and a predetermined reference stimulus ξ_t , the weight U_t has a certain value. However, for the class of neural networks the U_t is a random variable. To determine the probability, that the network selected from a class, correctly classifies stimulus ξ_x , is required to calculate the probability characteristics of the random variable of weight U_x , supplied to the input of the output layer.

According to the modified Chebyshev inequality [5], for any random variable z with mathematical expectation $M_z = \mu$ with any dispersion $D_z = \sigma^2$ we have following relations:

$$P(z > 0) \geq 1 - \frac{1}{\left(\frac{\mu^2}{\sigma^2}\right)} = 1 - \frac{\sigma^2}{\mu^2}, \quad \text{when } \mu > 0 \quad (12)$$

$$P(z < 0) \geq 1 - \frac{1}{\left(\frac{\mu^2}{\sigma^2}\right)} = 1 - \frac{\sigma^2}{\mu^2}, \quad \text{when } \mu < 0 \quad (13)$$

where P - the probability of the corresponding event, σ - average quadratic deviation. Equations (12,13) can be used to estimate the probability of a correct response network with $\theta_R = 0$ on ξ_x in the case of two recognized classes.

From the (12,13) follows that the probability of correct recognition is increases, if σ^2/μ^2 attitude is tends to zero.

With appropriate probabilistic characteristics of the network, we can estimate the probability of a correct response network to ξ_x . If the relation $\sigma^2(U_x)/\mu^2(U_x)$ can be made arbitrarily small, then for a selected network with $\theta_R = 0$ the probability that a stimulus ξ_x is classified correctly approaches unit.

Let N_A is number of hidden elements of the network, $L = l_i + l_j$, $l_i(l_j)$ is the length of the training sequence of i - th (j - th) class, $\delta_i(\delta_j)$ is increase of hidden element of weight when the stimulus from the i - th (j - th) grade pathogens is displaying. P_i is probability of excitation of hidden element when displaying a stimulus from the i - th class, P_{ij} - probability of excitation of hidden element when displaying incentives from the i - th and j - th grade. To find the weight of the expectation at the input of the output layer proved the following theorem [8].

Theorem. Let the a set pathogens $\Omega = \{\xi\}$, $\Omega = \Omega_1 \cup \Omega_2$, $\Omega_1 \cap \Omega_2 = \emptyset$, and training sequence $\xi_1, \xi_2, \dots, \xi_L$ are given. Then the expectation of the weight input in the output layer when the stimulation of the i - th class is equal to

$$\mu_i = N_A(\delta_i P_i l_i + \delta_j P_{ij} l_j), \quad i, j = 1, 2 \quad (14)$$

Having μ_i , let estimate dispersion of random variable U_x with appearance of the network at the entrance of the pathogen ξ_x .

For $j = 1, \dots, L$ and $r = 1, \dots, L$ will received:

$$\sigma^2(U_x) = N_A L^2 \sum \sum v_j v_r \delta_j \delta_r (P_{jrx} - P_{jk} * P_{rx}) \quad (15)$$

where P_{jx} - excitation probability of hidden element when displaying incentives ξ_j , ξ_x , P_{rx} - excitation probability of hidden element when displaying incentives ξ_r , ξ_x , P_{jrx} - excitation probability of hidden element when displaying incentives ξ_j , ξ_r , ξ_x .

Since the ratio $\sigma^2(U_x)/\mu^2(U_x)$ is not depend of the length of the training sequence, so any number of repetitions of the same training sequence does not change the characteristics of the system

Let consider the relation $\sigma^2(U_x)/\mu^2(U_x)$ for the control the stimulus ξ_x in cases, if $\xi_x \in \Omega_1$ and $\xi_x \in \Omega_2$.

IV. RESULTS

Assessing the probability P_{jrx} , P_{jx} , P_{rx} the expressions for the dispersion will be recieved.

Selected the following cases:

First case: assume that the control stimulus $\xi_x \in \Omega_1$. For the stimulus ξ_j and ξ_r in this case we obtain the following probabilities:

- a). if $\xi_j \in \Omega_1$ and $\xi_r \in \Omega_1$, so $P_{jrx} = P_{jx} = P_{rx} = P_1$;
- b). if $\xi_j \in \Omega_1$ and $\xi_r \in \Omega_2$, so $P_{jrx} = P_{12}$, $P_{jx} = P_1$, $P_{rx} = P_{12}$;
- c). if $\xi_j \in \Omega_2$ and $\xi_r \in \Omega_1$, so $P_{jrx} = P_{12}$, $P_{jx} = P_{12}$, $P_{rx} = P_1$;
- d). if $\xi_j \in \Omega_2$ and $\xi_r \in \Omega_2$, so $P_{jrx} = P_{12}$, $P_{jx} = P_{12}$, $P_{rx} = P_{12}$.

To calculating follow relationship $\frac{\sigma_1^2(U_x)}{\mu_1^2(U_x)}$ for these case (a,b,c,d):

$$\frac{\sigma_1^2(U_x)}{\mu_1^2(U_x)} = \frac{\sum_{j=1}^L \sum_{r=1}^L v_j v_r \delta_j \delta_r (P_1 - P_1^2)}{N_A \sum_{j=1}^L \sum_{r=1}^L v_j v_r \delta_j \delta_r P_1^2} = \frac{1 - P_1}{N_A P_1} \quad (16)$$

$$\frac{\sigma_1^2(U_x)}{\mu_1^2(U_x)} = \frac{\sum_{j=1}^L \sum_{r=1}^L v_j v_r \delta_j \delta_r (P_{12} - P_1 P_{12})}{N_A \sum_{j=1}^L \sum_{r=1}^L v_j v_r \delta_j \delta_r P_1 P_{12}} = \frac{1 - P_1}{N_A P_1} \quad (17)$$

$$\frac{\sigma_1^2(U_x)}{\mu_1^2(U_x)} = \frac{(P_{12} - P_1 P_{12})}{N_A P_1 P_{12}} = \frac{1 - P_1}{N_A P_1} \quad (18)$$

$$\frac{\sigma_1^2(U_x)}{\mu_1^2(U_x)} = \frac{(P_{12} - P_{12}^2)}{N_A P_{12}^2} = \frac{1 - P_{12}}{N_A P_{12}} \quad (19)$$

The probability of detection for the first class is increased by the maximum value of $P_1 - P_{12}$, called the characteristic function of the perceptron (CFP) [5]. Obviously, the function CFP seeks to a maximum value at $P_1 \rightarrow 1$.

In assessing the value of $\frac{\sigma_1^2(U_x)}{\mu_1^2(U_x)}$ again see that $P_1 \rightarrow 1$ when $\frac{\sigma_1^2(U_x)}{\mu_1^2(U_x)} \rightarrow 0$ (in most of the discussed cases). For the first class of stimuli received condition $\frac{\sigma_1^2(U_x)}{\mu_1^2(U_x)} \rightarrow 0$, at which execution $CFP \rightarrow \max$, i.e. the probability of correct recognition for the first class stimulus is increases.

Second case: assume that the control stimulus $\xi_x \in \Omega_2$. For the stimulus ξ_j and ξ_r in discussed case the following probability will obtained:

- a). if $\xi_j \in \Omega_2$ and $\xi_r \in \Omega_2$, so $P_{jrx} = P_{jx} = P_{rx} = P_2$;
- b). if $\xi_j \in \Omega_1$ and $\xi_r \in \Omega_2$, so $P_{jrx} = P_{12}$, $P_{jx} = P_{12}$, $P_{rx} = P_2$;
- c). if $\xi_j \in \Omega_2$ and $\xi_r \in \Omega_1$, so $P_{jrx} = P_{12}$, $P_{jx} = P_2$, $P_{rx} = P_{12}$;
- d). if $\xi_j \in \Omega_1$ and $\xi_r \in \Omega_1$, so $P_{jrx} = P_{12}$, $P_{jx} = P_{12}$, $P_{rx} = P_{12}$;

Similarly, calculating $\frac{\sigma_2^2(U_x)}{\mu_1^2(U_x)}$, for second class will be received the condition correct recognition:

$$\frac{\sigma_1^2(U_x)}{\mu_1^2(U_x)} \rightarrow 0. \quad (20)$$

Hence

$$\frac{\sigma_1^2(U_x)}{\mu_1^2(U_x)} \rightarrow 0, (i = 1, 2). \quad (21)$$

V. CONCLUSIONS

The resulting estimates for certain statistical characteristics of a neural network in the case of two recognized classes have shown efficiency in the training of neuron networks. Moreover, a new condition for improving accuracy of recognition for the stimulus i-th class is received.

REFERENCES

- [1] Anil K. Jain, Jianchang Mao, K. M. Mohiuddin. Artificial Neural Networks: A Tutorial. Computer - Special issue: neural computing: companion issue to Spring 1996 IEEE Computational Science & Engineering, Volume 29, Issue 3, March 1996.
- [2] Barsky A.B. Neural networks: the recognition, management, decision-making. M.: Finance and Statistics, 2004, 176 p. (in Russian)
- [3] Buraga, S.C., Cioca, M., Cioca, A. Grid-based decision support system used in disaster management, Studies in Informatics and Control, 16(3), 2007
- [4] Cioca, M. Cioca, L.I., Buraga, S.C. Using semantic web technologies to improve the design process in the context of virtual production systems, WSEAS Transactions on Computers, 4(12), pp. 1788-1793, 2005
- [5] Ivakhnenko A.G. Perceptron pattern recognition system, Naukova Dumka, Kiev, 1975. p. 426. (in Russian)
- [6] Neural networks. Matlab 6, Dialog-MIFI. 2002, 496 p. (in Russian)
- [7] Pantelev S.V. Development, research the use of neural network algorithms, .: 2001, 496 p. (in Russian)
- [8] Sargsyan S.G. Determination of the probability characteristics of adaptive recognition system, Trans. of Conf. Adaptable software, Kishinev, 1990. pp. 4651. (in Russian)
- [9] Vesna Rankovic, Jasna Radulovic, Nenad Grujovic, Dejan Divac, Neural Network Model Predictive Control of Nonlinear Systems Using Genetic Algorithms, INT J COMPUT COMMUN, 7(3), ISSN 1841-9836, pp. 540-549, 2012