

Scan to know paper details and author's profile

Statistical Evaluation of the Performance of the Neural Network

Sargsyan Siranush,Hovakimyan Anna

ABSTRACT

In this paper the problem of evaluating the performance of the neural network, based on a study of the probabilistic behavior of the network is considered. Direct propagation network consisted of layer of input nodes, hidden layer and output layer is examined. To evaluate the network performance the mathematical expectation and dispersion of weight at the input of the output layer are considered. For such networks the estimates for some of the statistical characteristics of the neural network in the case of two recognized classes were obtained.

Keywords: neural network, the weight of the neuron, recognition of the stimulus, mathematical expectation, dispersion.

Classification: C.2.6 Language: English



LJP Copyright ID: 683211 ISBN 10: 1537631683 ISBN 13: 978-1537631684

London Journal of Research in Computer Science and Technology



Volume 17 | Issue 1 | Compilation 1.0

© 2017. Sargsyan Siranush,Hovakimyan Anna. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 Unported License http://creativecommons.org/licenses/by-nc/4.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.



Statistical Evaluation of the Performance of the Neural Network

Sargsyan Siranush^a & Hovakimyan Anna^o

I. ABSTRACT

In this paper the problem of evaluating the performance of the neural network, based on a study of the probabilistic behavior of the network is considered. Direct propagation network consisted of layer of input nodes, hidden layer and output layer is examined. To evaluate the network performance the mathematical expectation and dispersion of weight at the input of the output layer are considered. For such networks the estimates for some of the statistical characteristics of the neural network in the case of two recognized classes were obtained.

Keywords: neural network, the weight of the neuron, recognition of the stimulus, mathematical expectation, dispersion.

Author α σ : Department of Programming and Information Technologies Yerevan State University, Yerevan, Armenia.

II. INTRODUCTION

Artificial neural networks are often used for a variety of applications. For the successful application of artificial neural networks it is necessary to choose the right network architecture, to pick up its parameters, thresholds elements, activation function, and others [1, 2, 4]. Currently neural computation of different level 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 of 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 and to establish of associative connections in the input data. Typically, depending on the task, it is 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].

III. PROBLEM DESCRIPTION

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 neural network technologies an important role particularly play the selection of the network architecture and the values of its parameters that affect its efficiency.

A particular interest the study of the probabilistic behavior of the neural networks is presented [3, 4]. 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 with 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 are referred to as the input, hidden and output elements, respectively [1, 2, 4]. Neurons have unidirectional links and do not contain links between elements within a layer and feedbacks between the layers.

Signal (causative agent, incentive, stimulus) supplied to the input layer of neural network corresponds to external stimuli and is modeled by the vector whose coordinates take the values 0 and 1, depending on whether or not the corresponding neuron of the input layer is excited. The output signal from each neuron of the input layer is set to 1 if the neuron is excited, and 0 otherwise. The output signal generated by the input layer, is transferred to the hidden layer. Each neuron of the hidden layer implements a threshold function η as it inputs [4]:

$$\eta_{k}(i) = \begin{cases} 1, & \text{if } F_{k}(i) \ge \theta \\ 0, & \text{if } F_{k}(i) < \theta \end{cases}$$

 \div where σ_n and σ_m are numbers of excited and inhibited bonds from excitatory and m inhibitory connections respectively, θ is threshold of hidden element (integer number). Excitatory and inhibitory connections between the input and hidden layers are assumed to be randomly and uniformly distributed. The outputs of all elements of the hidden layer are passed as an input to the output layer that implements the function R:

$$R = \begin{cases} 1, & \text{if } \sum_{k=1}^{N_A} \eta_k V_k > \theta_R \\ 0, & \text{if } \sum_{k=1}^{N_A} \eta_k V_k < \theta_R \\ \text{not determined,} & \text{if } \sum_{k=1}^{N_A} \eta_k V_k = \theta_R \end{cases}$$

where η_k is the activity the k-th hidden element, v_k is the weight of the k-th hidden element, N_A is the

number of elements in the hidden layer, and θ_R is the threshold of the output element.

The structure of the neural network allows to correct errors during incorrect responses by means of gradual complication of decision rules. This is done by changing the weight vectors.

Let us determine the weight of the hidden element after training.

The activity $\eta_k(i)$ of k-th hidden element when the stimulus ξ_i is submitted to the network is determined by the formula

$$\eta_{k}(i) = \begin{cases} 1, & \text{if } F_{k}(i) \ge \theta \\ 0, & \text{if } F_{k}(i) < \theta \end{cases}$$

where $F_k(i)$ is the value of signal of the k-th hidden element, and θ is the threshold of hidden element.

At the stage of neural network training, a training sequence of stimuli is presented to network. For each stimulus a hidden element is tested for activity, and all the active elements of the hidden layer are encouraged with values δ_i (i = 1, 2) for the i-th class of stimuli.

$$\delta_i = \left\{ \begin{array}{ll} 1, & \mbox{ for the first class of stimuli} \\ -1\,, & \mbox{ for the second class of stimuli} \end{array} \right.$$

If each class has, respectively, l_1 and l_2 representatives, then after training via the sequence of two classes activators of the length L = $l_1 + l_2$, in the k-th hidden item the weight V_k is accumulated:

$$V_{k} = \sum_{i=1}^{2} \delta_{i} \sum_{j=1}^{l_{i}} \eta_{k}(j) + V_{0}$$
(1)

where $V_{\rm o}\text{-}$ is a initial weight of k –th hidden element.

When at the input to a network is the stimulus ξ_{t_1} , so the input of the output layer is fed weight (k=1, ...N_A)

$$\begin{split} U_{t} &= \sum_{k=1}^{N_{A}} V_{k} \eta_{k}(t) = \sum_{k=1}^{N_{A}} (\sum_{i=1}^{2} \delta_{i} \sum_{j=1}^{l_{i}} \eta_{k}(j) + V_{0}) \eta_{k}(t) \end{split} \tag{2} \\ \text{Let } V_{o} &= 0, l_{1} = l_{2} = l. \text{ Then} \end{split}$$

$$J_{t} = \sum_{k=1}^{N_{A}} \sum_{i=1}^{2} \delta_{i} \sum_{j=1}^{l} \eta_{k}(j, t)$$
(3)

 $\eta_k(j,t) = \begin{cases} 1, & \text{if the } k^{th} \text{ hidden element is active for } \xi_j \text{ and } \xi_{t'} \\ 0, & \text{otherwise} \end{cases}$ (4)

The belonging of pathogen ξ_t to one of two classes is determined by comparing of the weight U_t with a threshold output element R. When $U_t > \theta_R$ so we have first class. When $U < \theta_R$ - second class. When $Ut = \theta_R$ - 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 [6], for any random variable z with mathematical expectation $Mz=\mu$ with any dispersion $Dz=\sigma^2$ we have following relations:

 $P(z>0)>= 1-1/(\mu^2/\sigma^2))=1-\sigma^2/\mu^2$, when $\mu>0$ (5)

$$P(z<0) >= 1- 1/(\mu^2/\sigma^2)) = 1- \sigma^2/\mu^2, \text{ when } \mu < 0$$
 (6)

where P – is the probability of the corresponding event, σ - average quadratic deviation. Equations (5,6) can be used to estimate the probability of a correct response of network on ξ_x with θ_R =0 in the case of two recognized classes.

From the (5, 6) it follows that the probability of correct recognition increases, if attitude σ^2/μ^2 tends to zero.

With appropriate probabilistic characteristics of the network, we can estimate the probability of a correct response of network on 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 seeks to 1. 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 the increment of weight of hidden element when the stimulus from the i-th(j -th) class is presented. P_i is probability of excitation of hidden element when a stimulus from the i –th class is presented, P_{ij} - is probability of excitation of hidden element when incentives from the both i-th and j-th classes are presented. To find the expectation of the weight at the input of the output layer the following theorem is proved [3].`

Theorem. Let the a set of pathogens $\Omega = \{ \xi \}$, $\Omega = \Omega_1$ U Ω_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 stimulus is from the i –th class is equal to

$$\mu_{i} = N_{A} \left(\delta_{i} P_{i} l_{i} + \delta_{j} P_{ij} l_{j} \right); \quad i, j = 1, 2 i \neq j$$
(7)

Having μ_{i} , let's estimate dispersion of random variable U_x with appearance of the pathogen ξ_x at the entrance of the network.

For j = 1,..L, and r = 1,..L we receive:

$$\sigma^{2}(\mathbf{U}_{x}) = \mathbf{N}_{A} \mathbf{L}^{2} \sum \nu_{j} \nu_{r} \delta_{j} \delta_{r} (\mathbf{P}_{jrx} - \mathbf{P}_{jx} * \mathbf{P}_{rx}) \qquad (8)$$

where P_{jx} - is probability of excitation of hidden element when displaying incentives $\xi_{ij}\xi_{x_{ij}}$

 P_{rx} - is excitation probability of hidden element when displaying incentives $\xi_{r_{x}} \xi_{x}$,

 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 on the length of the training sequence, so any number of repetitions of the same training sequence does not change the characteristics of the system.

IV. RESULTS

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

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 $\xi_v \in$ Ω_1 . For the stimuli ξ_i and ξ_r in this case we obtain the following probabilities:

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

Let's calculate follow relationship $\overline{\mu_1^2(Ux)}$ for these case (a,b,c,d):

$$\frac{\sigma_1^2(\mathrm{Ux})}{\mu_1^2(\mathrm{Ux})} = \frac{\sum_{j=1}^L \sum_{r=1}^L v_j v_r \delta_j \delta_r(\mathrm{P}_1 - \mathrm{P}_1^2)}{N_A \sum_{j=1}^L \sum_{r=1}^L v_j v_r \delta_j \delta_r \mathrm{P}_1^2} = \frac{1 - \mathrm{P}_1}{N_A \mathrm{P}_1} \tag{9}$$

$$\frac{\sigma_1^2(\text{Ux})}{\mu_1^2(\text{Ux})} = \frac{\sum_{j=1}^L \sum_{r=1}^L \nu_j \nu_r \delta_j \delta_r (P_{12} - P_1 * P_{12})}{N_A \sum_{j=1}^L \sum_{r=1}^L \nu_j \nu_r \delta_j \delta_r P_1 * P_{12}} = \frac{1 - P_1}{N_A P_1}$$
(10)

$$\frac{\sigma_1^2(Ux)}{\mu_1^2(Ux)} = \frac{(P_{12} - P_{12} * P_1)}{N_A P_1 P_{12}} = \frac{1 - P_1}{N_A P_1}$$
(11)

$$\frac{\sigma_1^2(\text{Ux})}{\mu_1^2(\text{Ux})} = \frac{(\text{P}_{12} - \text{P}_{12}^2)}{\text{N}_A \text{P}_{12}^2} = \frac{1 - \text{P}_{12}}{\text{N}_A \text{P}_{12}}$$
(12)

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) [4]. Obviously, the function CFP seeks to a maximum value at $P_1 \rightarrow 1$.

In assessing the value of $\frac{\sigma_1^2(Ux)}{\mu_1^2(Ux)}$ again see that $\sigma_1^2(Ux)$

 $P_1 \rightarrow 1$ when $\overline{\mu_1^2(Ux)} \rightarrow 0$ (in most of the discussed cases). For the first class of stimuli the condition $\sigma_1^2(Ux)$

 $\overline{\mu_1^2(Ux)} \rightarrow 0$ is received at which execution CFP \rightarrow max, i.e. the probability of correct recognition for the first class stimulus increases.

Second case: assume that the control stimulus $\xi_x \in$ Ω_{2} .

For the stimulus ξ_i and ξ_r in discussed case the following probability will be obtained:

a). if $\xi_i \in \Omega_2$ and $\xi_r \in \Omega_2$, so $P_{irx} = P_{ix} = P_{rx} = P_2$; b). if $\xi_j \in \Omega_1$ and $\xi_r \in \Omega_2$, so $P_{jrx} = P_{12}$, $P_{jx} = P_{12}$, $P_{rx} =$ P₂; c). if $\xi_j \in \Omega_2$ and $\xi_r \in \Omega_1$, so $P_{jrx} = P_{12}$, $P_{jx} = P_2$, $P_{rx} =$ P₁₂; d). if $\xi_j \in \Omega_1$ and $\xi_r \in \Omega_1$, so $P_{jrx} = P_{12}$, $P_{jx} = P_{12}$, $P_{rx} = P_{12}$ P₁₂.

$$\frac{\sigma_2^2(Ux)}{\mu_2^2(Ux)},$$

Similarly, calculating , for the second class will be received the condition for correct recognition:

$$\frac{\sigma_2^2(Ux)}{\mu_2^2(Ux)} \rightarrow 0$$

Therefore we received new conditions to increase of the right recognition for the stimulus of the i-th)) class:

$$\frac{\sigma_i^2(\mathrm{Ux})}{\mu_i^2(\mathrm{Ux})} \rightarrow 0 \qquad (i = 1; 2) \tag{13}$$

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 neural networks. Moreover, a new condition for improving accuracy of recognition for the stimulus i-th class is received.

REFERENCES

- 1. Panteleev S.V. Development, research the use of neural network algorithms, M: 2001, 496 p. (in Russian).
- Barsky A.B. Neural networks: 2. the recognition, management, decision-making. M .: Finance and Statistics, 2004, 176 p. (in Russian).
- Sargsyan Determination 3. S.G. of the probability characteristics of adaptive recognition system, Trans. of Intern. Conf. software, Kishinev, Adaptable 1990.pp. 46-51. (in Russian).

- 4. Ivakhnenko A.G. Perceptron pattern recognition system, Naukova Dumka, Kiev, 1975.p. 426. (in Russian).
- 5. Hovakimyan A., Sargsyan S., Nazaryan A. Self-Organizing Map Application for Iris Recognition . Journal of Commun & Comput. Eng.ISSN 2090-623, www.m-sciences.com, Volume 3, Issue 2. 2013. PP.10-13.
- 6. Feller V., Introduction to probability theory and its applications, M., Mir, 1984, (in Russian).

Statistical Evaluation of the Performance of the Neural Network

This page is intentionally left blank