computing@computingonline.net
www.computingonline.net

Print ISSN 1727-6209
On-line ISSN 2312-5381
International Journal of Computing

APPROACH TO THE SYNTHESIS OF NEURAL NETWORK STRUCTURE DURING CLASSIFICATION

Avazjon R. Marakhimov, Kabul K. Khudaybergenov

National University of Uzbekistan, Vuzgorodok 4, Tashkent, 100174, Uzbekistan
avaz.marakhimov@yandex.com, kabul85@mail.ru

Paper history:
Received 27 August 2019
Received in revised form 13 February 2020
Accepted 21 February 2020
Available online 31 March 2020

Keywords:
Neural network;
hidden layer;
MLP model;
sigmoidal function;
classification;
pattern recognition.

Abstract: Evaluating the number of hidden neurons necessary for solving of pattern recognition and classification tasks is one of the key problems in artificial neural networks. Multilayer perceptron is the most useful artificial neural network to estimate the functional structure in classification. In this paper, we show that artificial neural network with a two hidden layer feed forward neural network with d inputs, d neurons in the first hidden layer, 2d+2 neurons in the second hidden layer, k outputs and with a sigmoidal infinitely differentiable function can solve classification and pattern problems with arbitrary accuracy. This result can be applied to design pattern recognition and classification models with optimal structure in the number of hidden neurons and hidden layers. The experimental results over well-known benchmark datasets show that the convergence and the accuracy of the proposed model of artificial neural network are acceptable. Findings in this paper are experimentally analyzed on four different datasets from machine learning repository.

1. INTRODUCTION

The past two decades have seen an enormous change in the field of artificial neural networks and their applications. Especially, there has been a considerable growth of progress in a multi-layer feed-forward neural network [1-4]. The application areas of artificial neural network include wide fields in computer science, biology, decision science, medicine, finance, engineering, etc. While solving applied problems, one can use various kinds of architectures of neural network with some appropriately chosen activation functions. If a constructed artificial neural network model is not capable to solve a specific problem or to give acceptable precision, they increase hidden units (neurons) in the layers and/or increase hidden layers on the network model. However, in applications it is necessary to define or indicate how many neurons and how many layers one should take in the neural network architecture. Of course, the larger the number of neurons and hidden layers is, the larger the probability of the network to give more precise results is. Unfortunately, practicality decreases with the increase in the number of neurons and the number of hidden layers in the neural network model.

Researchers of artificial neural network have proposed and experimented various kinds of neural network models exploiting multi layered architectures, for instance, convolutional neural networks [5], cascade neural networks [6], deep neural networks [7,8], graph neural networks [9], recurrent neural networks [10], etc.

These above proposed models have found a broad application to challenging issues in various fields, for instance, in health services [11], natural language processing [12], neural signal processing [13], robotics and instrumentation [14, 15], and data mining [16, 17], classification [25].

However, from a theoretical view, problems of the lower and upper bounds of neurons in hidden layers and the bounds of hidden layers to classification and pattern recognition architectures have not been completely studied yet. For instance, one of the current existing findings is related to binary neural networks [18]. They defined a structure named a most isolated samples in the Boolean field and proved that at least 2^{n-1} hidden neurons are necessary in binary neural network. However, these kinds of findings are restricted to specific situations and do not provide a fundamental theory for studying the computational capabilities of neural network in general. Also, there can't be made any extension of those existing results to feed-forward neural network, such as layered feed-forward neural networks with some activation functions.

In addition, previous works on evaluating the number of hidden layer neurons related with the capabilities of the feed-forward networks [19-23]. In [24, 26], authors proposed to apply the singular value decomposition method in neural networks to estimate the number of hidden neurons. However, algorithms for computing the singular value decomposition is time consuming process. In addition, in [27] a new approach was proposed to fix the number of hidden neurons of multi-layer perceptron (MLP) architecture. They proposed an approach that consists of the post training employment of the singular value decomposition and principal component analysis (PCA) method to adjust parameters of network. The number of hidden neurons is then fixed to the number of singular values or eigenvalues, which is obtained from operations over matrices.

In [28], authors investigated to fix the bounds of the number of hidden neurons in a special type of network, which is called multi-valued multi-threshold neural networks. They studied neural network properties with q-valued function defined on some space. The obtained results can find implementations to build learning algorithms with training set of finite numbers.

In other words, neural networks are not always effective if the number of neurons in the hidden layer and the number of layers are prescribed without theoretical background. In the current paper, we show that there exists optimal architecture for neural networks with two hidden layers. We want to give that a two hidden layer neural network with infinitely differentiable sigmoidal activation function σ, d neurons in the first layer, $2d+2$ neurons in the second layer, k outputs and the ability to solve classification and pattern recognition problems within given arbitrary precision.

A single hidden layer neural network model with r neurons in its hidden layer and the input $x=\left(x_1,\dots,x_d\right)$ evaluates a function of the following form

$$\sum_{i=1}^{r} c_i\sigma\left(w^i\cdot x-\theta_i\right),$$

where the weights w^j are vectors in $\mathbb{R}^d$, θ_i – threshold values and c_i- the coefficients in real numbers and σ- is a univariate function, which is called activation function in neural network literature.

2. PROBLEM STATEMENT

A two hidden layer neural network architecture is defined by iteration of hidden layers. In general, artificial neural network with two hidden layers, input $x=\left(x_1,\dots,x_d\right)$, r neurons in the first hidden layer, s neurons in the second hidden layer and output $y=\left(y_1,\dots,y_k\right)$ is as follows:

$$\sum_{p=1}^{s} d_p^{(i)}\sigma\left(\sum_{q=1}^{r} c_{pq}\sigma\left(w_{pq}\cdot x-\theta_{pq}\right)-\gamma_p^{(i)}\right),$$

where $d_p=\left(d_p^{(s)}\right)$, $\lambda_p=\left(\gamma_p^{(s)}\right)$, c_{pq}, θ_p are real numbers, w_{pq} is a real matrix and σ is a fixed univariate function. In neural network literature, such functions are called sigmoidal activation functions [33].

In the current study, we are going to give proof that there exist two hidden layer neural network with infinitely differentiable sigmoidal activation function σ, d neurons in the first hidden layer, $2d+2$ neurons in the second hidden layer, k outputs and with the ability to solve classification and pattern recognition problems within an arbitrary accuracy.

Definition 1. A univariate function σ defined on $\mathbb{R}$ is called a sigmoidal activation function if the following conditions are met:

$$\lim_{x\to+\infty}\sigma\left(x\right)=1,\quad \lim_{x\to-\infty}\sigma\left(x\right)=0.$$

The sigmoidal activation function is a special type of function, which plays a significant role in the research of artificial neural networks, especially in classification and pattern recognition problems.

In this current paper, we use the sigmoidal activation function, which is defined as follows:

$$\sigma\left(x\right)=\frac{1}{1+e^{-\alpha x}},\ x\in\mathbb{R},$$

where $\alpha\in\mathbb{R}$ is a constant.

3. THE MAIN RESULT

Our result is based on the capabilities of the universal approximation property of the neural networks. Neural network can do approximation of any continuous function on any compact subset of $\mathbb{R}$ with a given arbitrary precision. Kurkova's results [29, 30] showed that an arbitrary continuous function can be approximated arbitrarily quite well off by two hidden layer neural network with a univariate sigmoidal activation function. However, according to these findings, the number of neurons (units) in the hidden layers needed to implement the approximation is exceedingly large. The problem of fixing the number of neurons in the hidden layers was first solved by Mairov and Pinkus [32]. They showed that there exists a two hidden layer network with a sigmoidal activation function for which it is sufficient to do approximation arbitrarily enough of any continuous multivariate function.

Later, the best result was obtained by Vugar [31] for two hidden layer feed-forward networks. Namely, he showed that a two hidden layer neural network with d inputs, d neurons in the first hidden layer, $2d+2$ neurons in the second hidden layer and with a specifically constructed sigmoidal and infinitely differentiable activation function can do approximation of any continuous multivariate function with a given (arbitrary) precision.

We extend the above mentioned approximation properties to multivariate output neural networks. Namely, we show that artificial neural network with a two hidden layer neural network with d inputs, d neurons in the first hidden layer, $2d+2$ neurons in the second hidden layer, k outputs and with a sigmoidal and infinitely differentiable activation function can solve classification and pattern recognition problems with arbitrary accuracy.

We continue this section with the definition of a λ-increasing (λ-decreasing) function which is given in [31]. Let $\lambda > 0$ be any real number on $\mathbb{R}$. A real function f defined on $(a;b)$ interval is called λ-increasing (λ-decreasing) if there exists a function $u:(a,b)\to\mathbb{R}$ which is also an increasing (decreasing) on $(a;b)$ interval such as $\left|f(x)-u(x)\right| \le \lambda$, for $\forall x \in (a,b)$. If a function u is strictly increasing (decreasing), then the above function f is also called a λ-strictly increasing ($\lambda -$ decreasing) function.

On the k-dimensional space $\mathbb{R}^k$ consider a Euclidean norm

$$\|x\| = \sqrt{\sum_{i=1}^{k} |x_i|^2},\ x = (x_1,\ldots,x_d).$$

Theorem 3.1. Let Q be a compact set in $\mathbb{R}^d$. For any positive real numbers α and λ there exists a $\mathbb{C}^\infty(R)$, sigmoidal activation function $\sigma:\mathbb{R}\to\mathbb{R}$ which is strictly increasing on $(-\infty,\alpha)$, λ-strictly increasing on $[\alpha,+\infty)$, and satisfies the following properties: For any $f \in \mathbb{C}(\mathbb{Q},\mathbb{R}^k)$ and $\varepsilon > 0$ there exists k-dimensional real vectors $d_p = \left(d_p^{(s)}\right)$, $\lambda_p = \left(\gamma_p^{(s)}\right)$, c_{pq}, θ_p and real matrix w_{pq} for which

$$\left\| f(x) - \sum_{i=1}^{k}\sum_{p=1}^{2d+2} d_p^{(i)} \sigma\left(\sum_{q=1}^{d} c_{pq}\sigma\left(w_{pq}\cdot x - \theta_{pq}\right) - \gamma_p^{(i)} \right) \right\| < \varepsilon$$

for all $x = (x_1,\ldots,x_d) \in Q$.

Proof. For each $i = 1\ldots k$ define projection map $\pi_i:\mathbb{R}^k\to\mathbb{R}$ as follows

$$\pi_i(x) = x_i,\ x = (x_1,\ldots,x_d).$$

Then for each $i = 1\ldots k$ a function $f\circ\pi_i:\mathbb{R}^d\to\mathbb{R}$ satisfies all conditions of Theorem 2.1 from [31]. Then by Theorem 2.1 for each $i = 1\ldots k$ there exist k-dimensional real vectors $d_p = \left(d_p^{(s)}\right)$, $\lambda_p = \left(\gamma_p^{(s)}\right)$, c_{pq}, θ_p and real matrix w_{pq}

$$\left\| \pi_i(f(x)) - \sum_{p=1}^{2d+2} d_p^{(i)} \sigma\left(\sum_{q=1}^{d} c_{pq}\sigma\left(w_{pq}\cdot x - \theta_{pq}\right) - \gamma_p^{(i)} \right) \right\| < \frac{\varepsilon}{k}. \quad (1)$$

Further

$$\left\| f(x) - \sum_{i=1}^{k}\sum_{p=1}^{2d+2} d_p^{(i)} \sigma\left(\sum_{q=1}^{d} c_{pq}\sigma\left(w_{pq}\cdot x - \theta_{pq}\right) - \gamma_p^{(i)} \right) \right\| =$$

$$\sqrt{\sum_{i=1}^{k} \left| \pi_i(f(x)) - \sum_{p=1}^{2d+2} d_p^{(i)} \sigma\left(\sum_{q=1}^{d} c_{pq}\sigma\left(w_{pq}\cdot x - \theta_{pq}\right) - \gamma_p^{(i)} \right) \right|} \overset{(1)}{<}$$

$$\sqrt{\sum_{i=1}^{k} \frac{\varepsilon^2}{k^2}} = \sqrt{k\frac{\varepsilon^2}{k^2}} < \varepsilon.$$

The proof is completed.

Remark 1. In some papers, a single layer network is defined as function

$$\sum_{i=1}^{r} c_i \sigma\left(w^i \cdot x - \theta_i\right) - c_0.$$

A two hidden layer network then takes the form as follows

$$\sum_{p=1}^{s} d_p^{(i)} \sigma\left(\sum_{q=1}^{r} c_{pq} \sigma\left(w_{pq} \cdot x - \theta_{pq}\right) - \gamma_p^{(i)}\right) - d_0. \quad (2)$$

From the proof of Theorem 3.1, for networks of type (2) the theorem is valid if we take $2d+1$ neurons in the second layer, instead of $2d+2$.

4. SIMULATION RESULTS AND DISCUSSION

We have investigated the performance results of the training/testing process using the obtained criteria from Theorem 3.1, on four types of classification datasets: (a) Iris datasets, (b) Australian, (c) Wine and (d) Spambase, which are well-known benchmark datasets from the UCI Machine Learning database (https://archive.ics.uci.edu/ml/index.php). We constructed the neural network architecture as stated in the problem statement section: a two hidden layer feed-forward neural network with infinitely differentiable sigmoidal activation function σ, which has d neurons in the first hidden layer, $2d+2$ neurons in the second layer and k outputs. The neural network model for each dataset was trained and tested to guarantee that it is sufficiently trained. The standard back-propagation method was performed using a batch learning.

Figure 1 shows the training process for the Australian dataset, while Table 1 shows the training and testing results, and the proposed number of the hidden neurons and the number of hidden layers using criteria which are given in Section 3.

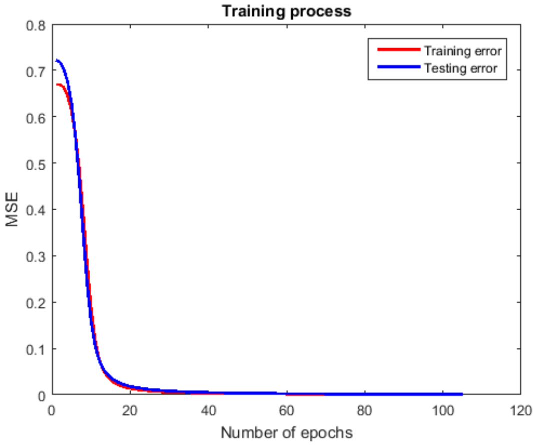

Figure 1 – Training process (Australian data set)

Table 1. Accuracy results for pattern recognition

Data set	Input size	Output size	Number of neurons in the hidden layers	Number of epochs	Classification accuracy on training set (%)	Classification accuracy on testing set (%)
Iris	4	3	4x10	72	100.0	100.0
Australian	14	2	14x30	254	100.0	100.0
Wine	13	3	13x28	104	100.0	98.75
Spambase	57	2	57x116	1287	98.25	97.50

Figures 2-5 show training and testing accuracies for selected benchmark datasets, and the proposed number of hidden neurons and hidden layers using criteria which are given in Section 3. The “*Number of neurons*” axis in Figures 2-5 depicts the number of the first hidden layer (d-neurons) and similarly, the number of the second hidden layer neurons is calculated according to the first layer neurons amount. From the simulation results, increasing the number of hidden neurons in the first hidden layer and in the second hidden layer more than d and $2d+2$, respectively, we can see that there is no need for that.

From the experimental results on well-known benchmark datasets, the convergence and the accuracy of the proposed model of neural network is acceptable and even better than we have expected.

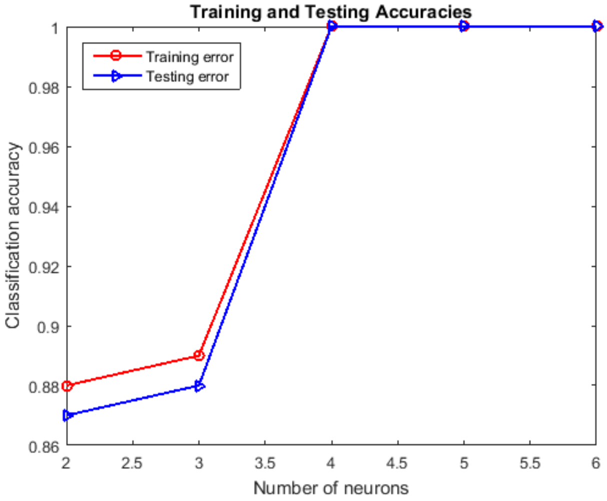

Figure 2 – Iris dataset

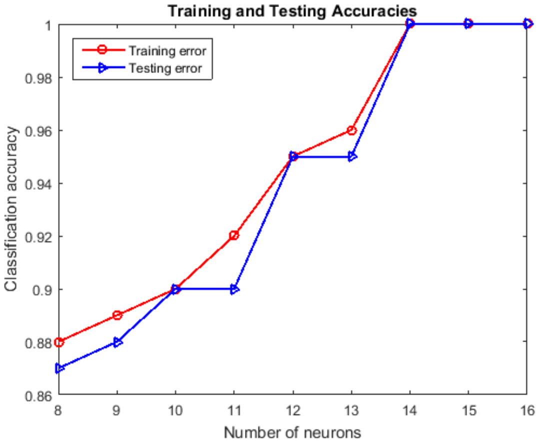

Figure 3 – Australian dataset

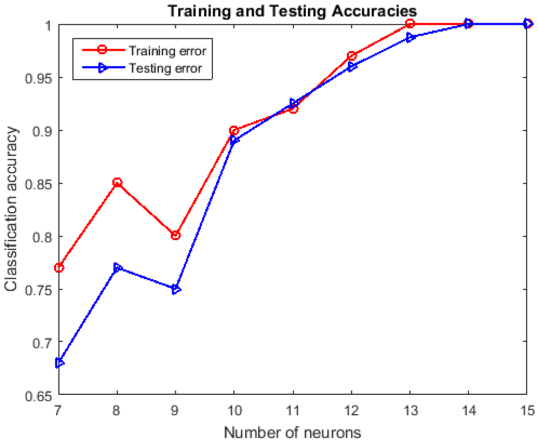

Figure 4 – Wine dataset

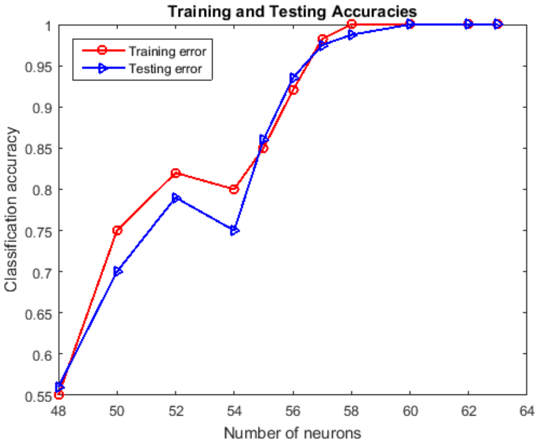

Figure 5 – Spambase dataset

5. CONCLUSION

In this paper, a theoretical result has been obtained on the constructing feed-forward neural networks for setting the number of hidden neurons and the number of hidden layers. The proposed result can be helpful to design a proper architecture of feed-forward neural networks (two hidden layers) with multivariate outputs. The theorem formulated in this paper provides a compact neural network architecture on the size of the feature space of input data to have a minimum degree of training complexity with sufficient number of parameters. In a nutshell, this can produce a compact classification/recognition model of high generalization with a relatively fast training time.

6. REFERENCES

[1] P. Baldi, R. Vershynin, "The capacity of feedforward neural networks," *Neural Networks*, vol. 116, pp. 288-311, 2019.

[2] V. E. Ismailov, "Approximation by neural networks with weights varying on a finite set of directions," *Journal of Mathematical Analysis and Applications*, vol. 389, issue 1, pp. 72-83, 2012.

[3] Z. Chen, F. Cao, "The approximation operators with sigmoidal functions," *Computers & Mathematics with Applications*, vol. 58, issue 4, pp. 758-765, 2009.

[4] A. Pinkus, "Approximation theory of the MLP model in neural networks," *Acta Numerica*, vol. 8, pp. 143-195, 1999.

[5] R. Yamashita, M. Nishio, R.K. Do, K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, issue 4, pp. 611-629, 2018.

[6] M.J. Diamantopoulou, V.Z. Antonopoulos, D.M. Papamichail, "Cascade correlation artificial neural networks for estimating missing monthly values of water quality parameters in rivers," *Water Resources Management*, vol. 21, issue 3, pp. 649-662, 2007.

[7] Q. Zhang, L.T. Yang, Z. Chen, P. Li, "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146-157, 2018.

[8] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11-26, 2017.

[9] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P.S. Yu, "A comprehensive survey on graph neural networks," 2019. ArXiv, abs/1901.00596.

[10] B. Chandra, R.K. Sharma, "On improving recurrent neural network for image classification," *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, 2017, pp.

1904-1907.

[11] P. Doupe, J.H. Faghmous, S. Basu, "Machine learning for health services researchers," *Value in Health*, vol. 22, issue 7, pp. 808-815, 2019.

[12] M.D. Becker, S. Kasper, B. Böckmann, K. Jöckel, I. Virchow, "Natural language processing of German clinical colorectal cancer notes for guideline-based treatment evaluation," *International Journal of Medical Informatics*, vol. 127, pp. 141-146, 2019.

[13] W. Tam, Z. Yang, "Neural parallel engine: A toolbox for massively parallel neural signal processing," *Journal of Neuroscience Methods*, vol. 301, pp. 18-33, 2018.

[14] P.R. Kurka, A.A. Salazar, "Applications of image processing in robotics and instrumentation," *Mechanical Systems and Signal Processing*, vol. 124, pp. 142-169, 2019.

[15] A.K. Maier, C. Syben, T. Lasser, C. Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift für Medizinische Physik*, vol. 29, issue 2, pp. 86-101, 2019.

[16] J. Zhang, S.O. Williams, H. Wang, "Intelligent computing system based on pattern recognition and data mining algorithms," *Sustainable Computing: Informatics and Systems*, vol. 20, pp. 192-202, 2018.

[17] J. Gola, D. Britz, T.M. Staudt, M. Winter, A.S. Schneider, M. Ludovici, F. Mücklich, "Advanced microstructure classification by data mining methods," *Computational Materials Science*, vol. 148, pp. 324-335, 2018.

[18] Y. Lu, J. Yang, Q. Wang, Z. Huang. "The upper bound of the minimal number of hidden neurons for the parity problem in binary neural networks," *Science China Information Sciences*, vol. 55, issue 7, pp. 1579-1587, 2012.

[19] S. Huang, Y. Huang, "Bounds on number of hidden neurons of multilayer perceptrons in classification and recognition," *Proceedings of the IEEE International Symposium on Circuits and Systems*, New Orleans, LA, USA, May 1-3 1990, pp. 2500-2503

[20] S. Tamura, "Capabilities of a three-layer feedforward neural network," *Proceedings of the IEEE International Joint Conference on Neural Networks*, Singapore, Singapore, November 18-21, 1991, pp. 2757-2762.

[21] S. Tamura, M. Tateishi, "Capabilities of a four-layered feedforward neural network: Four layers versus three," *IEEE Transactions on Neural Networks*, vol. 8, issue 2, pp. 251-255, 1997.

[22] G. Huang, H.A. Babri, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions," *IEEE Transactions on Neural Networks*, vol. 9, issue 1, pp. 224-229, 1998.

[23] G. Huang, "Learning capability and storage capacity of two-hidden-layer feedforward networks," *IEEE Transactions on Neural Networks*, vol. 14, issue 2, pp. 274–281, 2003.

[24] E.J. Teoh, C. Xiang, K.C. Tan, "Estimating the number of hidden neurons in a feedforward network using the singular value decomposition," *IEEE Transactions on Neural Networks*, vol. 17, issue 6, pp. 1623-1629, 2006.

[25] S. Omatu, "Classification of mixed odors using a layered neural network," *International Journal of Computing*, vol. 16, issue 1, pp. 41-48, 2017.

[26] D.C. Psichogios, L.H. Ungar, "SVD-NET: An algorithm that automatically selects network structure," *IEEE Transactions on Neural Networks*, vol. 5, issue 3, pp. 513-515, 1994.

[27] J.D. Santos, G.D. Barreto, C.M. Medeiros, "Estimating the number of hidden neurons of the MLP using singular value decomposition and principal components analysis: a novel approach," *Proceedings of the 2010 Eleventh Brazilian Symposium on Neural Networks*, Sao Paulo, Brazil, October 23-28 2010, pp. 19-24.

[28] N. Jiang, Z. Zhang, X. Ma, J. Wang, "The lower bound on the number of hidden neurons in multi-valued multi-threshold neural networks," *Proceedings of the 2008 Second International Symposium on Intelligent Information Technology Application*, Shanghai, China, January 6'2008, pp. 103-107.

[29] V. Kurkova, "Kolmogorov's theorem is relevant," *Neural Computation*, vol. 3, pp. 617-622, 1991.

[30] V. Kurkova, "Kolmogorov's theorem and multilayer neural networks," *Neural Networks*, vol. 5, issue 3, pp. 501-506, 1992.

[31] V.E. Ismailov, "On the approximation by neural networks with bounded number of neurons in hidden layers," *Journal of Mathematical Analysis and Applications*, vol. 417, issue 2, pp. 963-969, 2014.

[32] V. Maiorov, A. Pinkus, "Lower bounds for approximation by MLP neural networks,"

Neurocomputing, vol. 25, pp. 81-91, 1999.

[33] Ch. Zhixiang, C. Feilong, "The approximation operators with sigmoidal functions," *Computers & Mathematics with Applications*, vol. 58, issue 4, pp. 758-765, 2009.

Avazjon R. Marakhimov, *Doctor of Sciences, a Professor, Department of Mathematics, National University of Uzbekistan. Rector of the National University of Uzbekistan. His research areas are: intelligent control systems and information technology.*

His last researches are devoted to the problems of Neural networks, Data mining, fuzzy modeling, synthesis fuzzy control system and Intelligent Transportation Systems. He has more than 70 published research papers and 3 monographs.

Kabul K. Khudaybergenov *is pursuing a PhD degree at the National University of Uzbekistan. His research interests include issues related to theoretical foundations of Neural networks, Data mining, Machine learning.*