ALGORITHM OF THE MULTILAYER ARTIFICIAL FEED-FORWARD NEURAL NETWORK LEARNING

Mukha V. S.
Department of Information Technologies of Automated Systems,
Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus
E-mail: mukha@bsuir.by

The algorithm of the multilayer feed-forward neural network learning is developed. The algorithm is realized the error back-propagation idea. The particularity of the algorithm is its multidimensional-matrix form, which provides its theoretical and algorithmic generality. The program realization of the algorithm is performed as the function of the Matlab programming language. In spite of the multidimensional-matrix form of the algorithm, this function is defined fully by the usual matrices. The validity of the algorithm is confirmed by the computer simulation on the instances of the different approximation problems including the problem of the pattern recognition.

Introduction

Nowadays, neural networks increasingly being used to solve various problems instead of traditional mathematical methods. The certain euphoria is observed in the popular literature especially in the student environment about the advantages of the artificial neural networks compared with the traditional methods (see, for example, [1]). It is explained by the apparent simplicity of use of the artificial neural network: let us choose the features, collect the learning sample and the artificial neural network will achieve everything else. On the other hand, there is the works in which the real comparative analysis of the classical methods and the artificial neural networks is performed for solving of the specific problems [2–5]. The main difficulty on this path is the developing of the learning algorithm of the artificial neural network. Therefore, ready-made software products which contain the learning algorithms of the artificial neural networks are usually used. However, it is advisable to have the clear and easy for the programming algorithms of the artificial neural networks learning for the expanding of the arsenal of the typical solutions in the field of the artificial neural networks.

The most popular method for the artificial neural networks learning is the so called error back-propagation method [6, 7]. There are numerous other descriptions of this method, but all of them are not brought to strict algorithm suitable for the error-free programming. In this paper, the solution of this problem is given. The developing algorithm bases on the idea of the error back-propagation method but this terminology is not used since the algorithm does not contain any error propagation in the literal sense. In the author understanding, it is the gradient method for finding the minimum of the loss function with specific but the natural way to calculate the gradient for this task.

I. Mathematical description of the multilayer artificial feed-forward neural network

The separate layer of the artificial neural network has the form represented in figure 1.

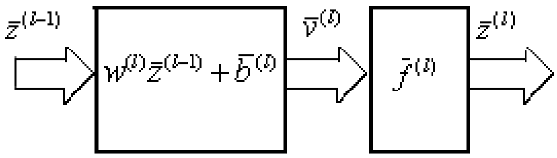

Рис. 1 – The separate layer of the artificial neural network

In figure 1, $\overline{z}^{(l-1)} = (z_1^{(l-1)}, z_2^{(l-1)}, \ldots, z_{s_{l-1}}^{(l-1)})$ is the input vector of the l-th layer, $\overline{v}^{(l)} = (v_1^{(l)}, v_2^{(l)}, \ldots, v_{s_l}^{(l)})$ is the output vector of the affine transformation of the l-th layer, $\overline{z}^{(l)} = (z_1^{(l)}, z_2^{(l)}, \ldots, z_{s_l}^{(l)})$ is the output vector of the activation function $\overline{f}^{(l)}$ of the l-th layer, $w^{(l)} = (w_{i,j}^{(l)})$, $i = 1,2,\ldots,s_l$, $j = 1,2,\ldots,s_{l-1}$, is the $(s_l \times s_{l-1})$-matrix of the weigth coefficients of the linear transformatin of the l-th layer, $\overline{b}^{(l)} = (b_1^{(l)}, b_2^{(l)}, \ldots, b_{s_l}^{(l)})$ is the bias vector of the l-th layer.

Such a layer is described mathematically by the following expression:

$$\overline{z}^{(l)} = \overline{f}^{(l)}(w^{(l)} z^{(l-1)} + \overline{b}^{(l)}),$$

where $\overline{f}^{(l)} = (f_1^{(l)}, f_2^{(l)}, \ldots, f_{s_l}^{(l)})$ is the vector function maps the vector $\overline{v}^{(l)} = w^{(l)} z^{(l-1)} + \overline{b}^{(l)}$ to the vector $\overline{z}^{(l)}$. This function represents the set of the independent scalar functions of the scalar variables, i.e. $z_i^{(l)} = f_i^{(l)}(v_i^{(l)})$, $i = 1,2,\ldots,s_l$.

The multilayer artifical feed-forward neural network is the consecutive connection of the separate layers. The number of the layers of the multilayer artifical neural network we will denote L, and the variable l in the previous expressions take the values $l = 1,2,\ldots,L$. The general recurrent expression describing the artifical neural network with the arbitrary number of the layers L has the following form:

$$\overline{z}^{(l)} = \overline{f}^{(l)}(\overline{v}^{(l)}) = \overline{f}^{(l)}(w^{(l)}\overline{z}^{(l-1)} + b^{(l)}), l = 1,2,\ldots,L,$$

where $\overline{z}^{(0)} = \overline{x}$.

II. Learning of the artificial neural network

The learning of the artificial neural network consists of the selection of the weight matrices $w^{(1)},w^{(2)},\ldots,w^{(L)}$ and the bias vectors $\overline{b}^{(1)},\overline{b}^{(2)},\ldots,\overline{b}^{(L)}$ providing the best performance by the neural network of its functions. The learning set (the learning sample) contains the learning pairs $(\overline{x}_k,\overline{y}_k)$, $k = 1,2,\ldots,n$, where $\overline{x}_k = (x_{k,1},x_{k,2},\ldots,x_{k,s_0},)$ is the input vector of the artificial neural network and $\overline{y}_k = (y_{k,1},y_{k,2},\ldots,y_{k,s_L},)$ is the required output vector (the target vector) of the artificial neural network corresponding to the input vector $\overline{x}_k$, n is the size of the learning sample. For instance, The vector $\overline{x}_k = (x_{k,1},x_{k,2},\ldots,x_{k,s_0},)$ is the vector of the features of the pattern in the patterns recognition problem. The loss function C is introduced for the neural network learning which depends in general case from the set $\overline{\overline{y}} = (\overline{y}_1,\overline{y}_2,\ldots,\overline{y}_n)$ of the target vectors $\overline{y}_k$ and the set $\overline{\overline{z}}^{(L)} = (\overline{z}_1^{(L)},\overline{z}_2^{(L)},\ldots,\overline{z}_n^{(L)})$ of the output vectors $\overline{z}_k^{(L)}$ of the neural network: $C = C(\overline{\overline{y}},\overline{\overline{z}}^{(L)})$. For instance, it is possible to use the quadratic loss function:

$$C = \sum_{k=1}^{n} {}^{0,1}(\overline{z}_k^{(L)} - \overline{y}_k)^2, \qquad (1)$$

where ${}^{0,1}(\overline{z}_k^{(L)} - \overline{y}_k)^2$ means the (0,1)-rolled square of the vector $\overline{z}_k^{(L)} - \overline{y}_k$.

The proposing neural network learning algorithm supposes the calculation of the derivatives of the loss function on the vector and matrix arguments. Since such a differentiation and the differentiation of the superposition are not defined in classical vector-matrix mathematical approach we use the multidimensional-matrix differentiation theory [8].

III. Computer modelling

The learning algorithm of the multilayer artificial feed-forward neural network was developed on the base of the results above and do not present here. The developed algorithm is realised programmatically in the programm system Matlab as the single m-file-function. The quadratic loss function, linear activation function for the last layer and logistic activation function for all other layers apart the first were choose. The algorithm for three-layers neural network was checked in the tasks of the approximation of the three degree scalar polynomial of two variables and the discriminant function for the gaussian patterns recognition with two features. The following parameters of the neural network are used: $s_0 = s_1 =_2 s = 2$, $s_3 = 1$. The algorithm has shown its efficienty on these tasks. However, it was not possible to obtain the satisfactory approxomation of the polynomial by the neural network. At the same time, the problems typical for any search system have been identified. They are the difficulty of shoosing of the search step size, of the initial values of the parameters, low convergence speed of the learning algororithm and the specific for the artifical nueral networks problems with shoosing the number of the layers, sizes of layers, activation functions [9].

IV. Conclusion

The main result of this work is the algorithm of the multilayer feed-forward neural network learning. The algorithm was realized programmatically in the Matlab language as the single function. This program software confirmed its efficiency on the task of the approximation of the polynomial of many variables and discriminant functions in patterns recognition. It is suitable for the deep learning of neural networks with an arbitrary number of layers of arbitrary size [9, 10]. The disadvantages were exposed such as in traditional search algorithms and specific advantages of the neural networks.

V. References

1. Golovinov A.O. Advantages of neural networks over traditional algorithms / A.O. Golovinov, E.N. Klimova // Experimental and theoretical reserch in modern science / Coll. of art. based on the materials of the 5th intern. sc.-pract. conf. No 5 (5). – Novosibirsk: «СибАК», 2017. – Pp. 11–15. In russian.
2. C. A. Mitrea, C. K. M. Lee and Z. Wu. A Comparison between Neural Networks and Traditional Forecasting Methods. A Case Study // International Journal of Engineering Business Management. 2009, Vol. 1, No. 2, pp. 19–24.
3. Jock A. Blackard 1, Denis J. Dean. Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables // Computers and Electronics in Agriculture, 24 (1999), pp. 131–151.
4. Eze, Chinonso Michael, Ugwuowo, Ifeanyi Fidelis, Asogwa, Oluchukwu. A comparative analysis of vector autoregressive model and neural networks // EPH – International Journal of Mathematics and Statistics. 2018, Vol. 4, Issue 8, pp. 1–13.
5. Fahima Charef, Fethi Ayachi. A Comparison between Neural Networks and GARCH Models in Exchange Rate Forecasting // International Journal of Academic Research in Accounting, Finance and Management Sciences. Vol. 6, No.1, January 2016, pp. 94–99.
6. Rumelhart D., Hinton G., Williams R. Learning internal representations by errors propagation // Parallel distributed processing. – Cambridge, 1986, pp. 318–362.
7. Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer Science+Business Media, LLC, 2006. 738 p.
8. Mukha V.S. Analysis of multidimensional data. – Minsk.: Technprint, 2004. – 368 p. In russian
9. Golovko V.A. Neural network: learning, organization and application: tutorial. – M.: IPRJP, 2001. – 256 p. In russian.
10. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. Neural Networks. (2015). 61: 85–117.