International Journal of Multidisciplinary and Current Research

Research Article

ISSN: 2321-3124
Available at: http://ijmcr.com

Performance Analysis of Conjugate descent learning Rule of Feed Forward Neural Networks for Pattern Classification

Shivpuje Nitin Parappa[†] and Manu Pratap Singh[‡]

[†]Shri Venkateshwara University, Gajraula , Amroha (Uttar Pradesh), India.
‡Department of Computer Science, Institute of Engineering & Technology, Dr. B. R. Ambedkar University, Khandari, Agra (U. P.), India

Accepted 12 July 2015, Available online 17 July 2015, **Vol.3** (**July/Aug 2015 issue)**

Abstract

Conventional Backpropagation learning algorithm is frequently used for multilayer feed forward neural networks for pattern mapping. This learning rule uses first derivative of instantaneous local square error with respect to the current weight vector in the weight space. Therefore for every different presented pattern, the network exhibits the different local error and the weights modify in order to minimize the current local error. Therefore it frequently traps in the local minimum. The determination of the optimal weights is possible only when the global error is supposed to minimize. In this paper, we are providing the conjugate descent method to obtain the optimal weight vector. The conjugate descent method determines the second derivative of the error surface with respect to weight vector in the weight space. The proposed method indicates that the performance of feed forward neural network improves for pattern classification of handwritten English words.

Keywords: *Conjugate descent, Feed forward neural networks, back propagation learning, gradient descent method, Pattern Classification.*

1. Introduction

It has been found that the derivatives of an error function with respect to the weights can be obtained by the propagation of errors backwards through the network. This technique can also be applied to the calculation of other derivatives. The evaluation of the Jacobin matrix can be consider in this respect [1], whose elements are given by the derivatives of the network outputs with respect to the inputs. The Jacobin matrix provides a measure of the local sensitivity of the outputs to change in each of the input variables. Thus, the network mapping represented by a trained neural network will be non-linear, and so the elements of the Jacobin matrix will not be constant but depends on the particular input vector used.

The second derivatives of the error can also obtained by Hessian Matrix [2]. This derivative involves the important aspects in the neural computing [3]. Thus, due to the several applications of the Hessian matrix, there is various approximation schemes have been used to evaluate it. However, the Hessian can also be calculated exactly using an extension of the back propagation technique for evaluating the first derivatives of the error function. There are various approaches like the diagonal approximation [4] outer product approximation and inverse Hessian [5] have been used for its evaluation. The exact evaluation of the Hessian matrix has also proposed, which is valid for a network of arbitrary feed forward topology. This method is based on an extension of the technique of the back propagation used to evaluate the first derivatives, and shares the many desirable features of it [6, 7, 8].

In this paper, we are proposing the modified conjugate descent method [9] to provide an alternative approach for minimizing the instantaneous local square error with respect to the current weight vector in the weight space for improving the performance of feed forward neural network for generalized pattern classification. The handwritten English words are considered as the data samples to evaluate the performance feed forward neural network trained with our proposed learning technique i.e. modified conjugate descent learning. The simulation results are indicating the better performance and good correct classification rate for the proposed method with respect to other methods of second derivatives.

The next section presents the implementation of the neural network architecture with modified conjugate descent method. The experimental results and discussion are presented in section 3. Section 4 contents the conclusion of this paper.

2. Modified conjugate descent method

The conjugate descent of instantaneous local square error can obtain by considering the second derivative of effort surface with respect to each weight vector of output layer and hidden layers. It can realize that second derivative of instantaneous local square error finds the weight modification for obtaining the minima point of every local error. Hence to obtain the optimal weight vector for the feed forward neural network, the weight modification should perform for the global minimum point among the various local error minima points. Thus, the network would train for the every presented input-output pattern pair, and if the global error point has identified then the weights will further modify in order to minimize the global error point. Thus, here we have the modification in the weight vector in the iteration for minimizing the local and further to minimize the global error. The weight modification with proposed method can represent for output layer as:

$$\frac{\partial^2 E^l}{\partial W_{kj}^2} = \frac{\partial}{\partial W_{kj}}[\frac{\partial E^l}{\partial W_{kj}}] = \frac{\partial}{\partial W_{kj}}[\frac{\partial E^l}{\partial y_k}.\frac{\partial y_k}{\partial W_{kj}}] \qquad (1)$$

Or, $\Delta W_{kj} = -\eta \sum_k \dot{s}_k^l [\dot{s}_k^l - 2\delta_l^k (1-2s_k^l)].(s_j^l)^2$

(2)

Similarly the weight modification can represent for single hidden layer as:

$$\Delta W_{ji} = -\eta[\Delta^l (a_i^l)^2 - \eta \sum_k \dot{s}_k^l [\dot{s}_l^k - \delta_l^k (1-2s_k^l)].W_{kj}.\dot{s}_j^l.(a_i^l)^2 \qquad (3)$$

Thus, to present an input–output pattern pair for the training the weight vector will modify itself with the learning equation i.e. $W(t+1) = W(t) + \Delta W$. This process will continue for the each presented input- output pattern pair either as new pattern or as the repeating pattern.

3. Simulation Design and Result

In this simulation design we are analyzing the performance of multilayer feed forward neural network architecture trained with proposed conjugate descent learning method for correct pattern classification of the handwritten English words of three letters. Thus, the training set consists with English alphabets in binary form as input pattern with corresponding binary output pattern information. The 100 test sample words are presented to the vertical segmentation program which is designed in MATLAB and based on portion of average height of the words. These segmented characters are clubbed together after binarization to form training patterns for neural network. Since out of 100 samples words 80 words are used for training purpose and remaining 20 words are used for the testing purpose. The target pattern set corresponding to each English alphabet is coded with 5 bits. Thus the target pattern set consists with 5 X 80. The network is designed to learn its behavior by presenting each one of the samples 100 times. Thus simulation considers the 800 trails. To accomplish the simulation work we consider the feed forward neural network system which consists of 150x35x5 neurons in input, hidden and output layers respectively. The segmented characters are resized onto 15x10 binary matrixes and then reshaped in to the 150 X 1 matrix to represent a single pattern vector for training set consist with 150 X 80. Thus the training set cons and are exposed to 150 input neurons. The 5 output neurons are used to code the 26 letters of English alphabet. The feature extraction method of vertical and horizontal segmentation for the sample images of three letters words scanned the images and convert them into the gray level image then in the binary images after fitting the images into a fixed size rectangle box as shown in figure 1 to calculate the width and height of the image [10].

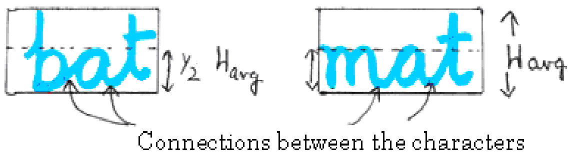

Figure 1: Scanned images are in rectangle box

Here, we are considering the $\frac{1}{2} * H_{avg}$ (Average of height) for deciding segmentation points. Each word is traced vertically after converting the gray scale image into binary matrix. This binarization is done using logical operation on gray intensity level as: I = (I >= Level) Here 0<= Level <= 1 is the threshold parameter. This Level is based on the gray-scale intensity of the text in document. More intensity leads to the more threshold value as shown in figure 2.

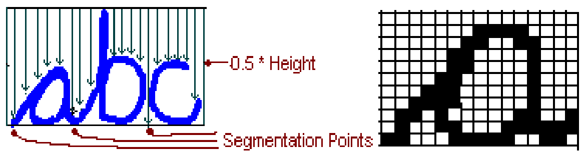

Figure 2: Binary image of a character after segmentation

The neural network has been trained with conventional descent gradient method and from the proposed modified conjugate descent learning method for obtaining the second derivative of instantaneous local square error method. The performance of the network has been analysis. The values of conventional second derivative methods proposed modified conjugate descent learning method are computed for each trail of learning and the mean value of all the trails of their performance for the classification error has been used as the final result for the representation as shown in figure 3.

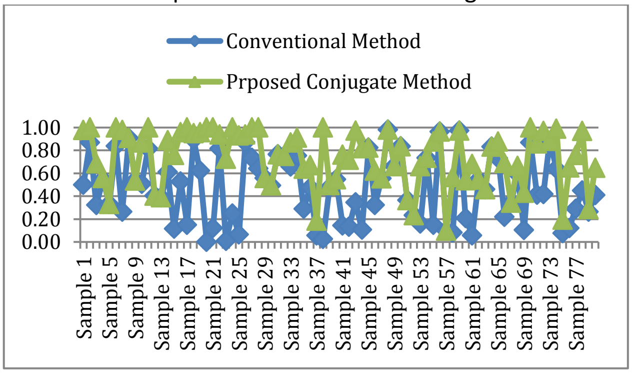

Figure 3: Comparison between the performances for classification

The results in the figure 3 are representing the accuracy in classification after the training in the network for conventional method of second derivatives and proposed modified conjugate descent method. The results shown here are the mean of all the trials. The proposed modified conjugate descent method for the handwritten words recognition is showing the remarkable enhancement in the performance.

The results are also indicating that for some of the samples the accuracy is near to 1 i.e. 100%. Thus on average the proposed method reflects the accuracy near to 90%. This shows the remarkable improvement in the performance of feed forward neural network for pattern classification.

Conclusion

The paper is highlighting the performance of proposed modified conjugate descent method for classification of English alphabets from the handwritten three letters words of English language. The proposed method is based on the approach of second derivative of instantaneous local square error with respect to each weight vector of the network in the weight space correspondingly to the presented input-output pattern pairs from the given training set. The weights for the feed forward network modified in order to minimize the each local error. The weight modification had consider for the hidden & output layer and inputs & hidden layer, beside this the second derivative of the instantaneous local square error has obtained with respect to both the weights i.e. hidden and output layer in combination. The simulated results are indicating that the proposed method of conjugate descent exhibits more accuracy for the pattern classification with respect to the conventional approach of second derivatives of local square error.

References

[1]. Jacobs, R., Jordan, M., Nowlan, S., Hinton, G. , Adaptive mixtures of local experts, Neural Computation, 3, 79-87, 1991.
[2]. Bishop, C. M., Exact calculation of the Hessian matrix for the multilayer perceptron, Neural Computation 4(4), 494–501, 1992.
[3]. Bishop, C. M., A fast procedure for retraining the multilayer perceptron, International Journal of Neural Systems 2(3), 229–236, 1991 b.
[4]. S. Backer and Y. Le Cun, Improving the convergence of back-propagation learning with second order methods, Proceedings of the 1988 Connectionist Models Summer School, pp 2937, San Mateo, Morgan Kaufmann, 1989.
[5]. Hassibi, B and stork, D.G. second order derivatives for network pruning: optimal brain surgeon, in Advances in Neural Information processing systems, 5, 164-171, 1993, San Mateo, CA: Morgan Kaufman.
[6]. Buntine, W.L. and weigend A.S., Computing second derivatives in feed-forward networks: a review. IEEE Transactions on Neural Networks, 5(3), 480-488, 1993.
[7]. Mandic, D. P., Hanna, A. I. and Razaz, M. A normalized gradient descent algorithm, for nonlinear adaptive filters using a gradient adaptive step size. IEEE Signal Processing Letters, 8 (11), 1-3, 2001.
[8]. Haykin S., Neural Networks: A Comprehensive Foundation, New York: Macmillan College Publishing Company Inc., 1994.
[9]. Singh M. P., Kumar S., Sharma N. K., Mathematical formulation for the second derivative of Backpropagation error with Non-liner output function in Feed forward Networks, International Journal of Information and Decision Sciences, Journal of INDERSCIENCE, 2(4), 352-373 (2010).
[10]. Dhaka V. S., Singh M. P., Handwritten Character Recognition using Modified Gradient Descent Technique of Neural Networks and Representation of Conjugate Descent for training Patterns, International Journal of Engineering, IJE Transaction A: Basics 22 (2) 145-158 (2009).