Neurocomputing 74 (2011) 2511–2519

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Application of error minimized extreme learning machine for simultaneous learning of a function and its derivatives

S. Balasundaram*, Kapil

School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India

ARTICLE INFO

Available online 8 May 2011

Keywords:
Derivatives approximation
Extreme learning machine
Feedforward neural networks
Function approximation
Incremental learning

ABSTRACT

In this paper a new learning algorithm is proposed for the problem of simultaneous learning of a function and its derivatives as an extension of the study of error minimized extreme learning machine for single hidden layer feedforward neural networks. Our formulation leads to solving a system of linear equations and its solution is obtained by Moore–Penrose generalized pseudo-inverse. In this approach the number of hidden nodes is automatically determined by repeatedly adding new hidden nodes to the network either one by one or group by group and updating the output weights incrementally in an efficient manner until the network output error is less than the given expected learning accuracy. For the verification of the efficiency of the proposed method a number of interesting examples are considered and the results obtained with the proposed method are compared with that of other two popular methods. It is observed that the proposed method is fast and produces similar or better generalization performance on the test data.

1. Introduction

Prediction by regression is an important method of solution for forecasting. In regression by analyzing the given input and their corresponding output values an approximation function that describes the underlying relationship between them is determined. For any unseen input example its output is predicted using this relationship.

The problem of regression estimation for a set of input examples given arises in a number of applications of practical importance like drug discovery [3], time series prediction [14,20], blind identification [18]. However in certain applications like linear circuits [17], Kalman filtering [2] it is necessary to estimate both the function and its derivatives. Recently Lazaro et al. [12] proposed a support vector regression (SVR) based approach for the simultaneous learning of a function and its derivatives that leads to solving a quadratic minimization problem. In this approach it is assumed that at each input example both the function and its derivative values are given together and to obtain the resulting optimal solution an iterative reweighted least squares (IRWLS) procedure is applied. In [10], the problem of simultaneous learning of a function and its derivatives is formulated using a regularized least squares SVR. This formulation allows the set of input examples at which the function values and the set where the derivatives to be given independent of each other. The main advantage of this formulation is that the solution is obtained by inverting a single matrix of order equals to the sum of the number of input examples at which the function and the derivative values are given. Finally for the study of simultaneous learning of a function and its derivatives using neural networks and/or its applications see [1,11,13,15].

In [6], Huang et al. proposed a new non-iterative learning algorithm for single hidden layer feedforward networks (SLFNs) architecture called extreme learning machine (ELM) which overcomes the problems caused by gradient decent based algorithms such as back propagation. In this algorithm the input weights and bias are randomly chosen. The ELM formulation leads to solving a system of linear equations in terms of the unknown weights connecting the hidden layer to the output layer and its solution is obtained using Moore–Penrose generalized pseudo-inverse [16]. Although ELM is a simple and an efficient learning algorithm, the number of hidden nodes of the SLFN is a parameter and its value is to be given at the beginning of the algorithm. Recently an error minimization based approach has been proposed in [4] to automatically determine the number of hidden nodes of the SLFN. In this approach the hidden nodes are allowed to grow one by one or group by group and once new hidden nodes are added the output weights are incrementally updated in an efficient manner. This process of adding more number of hidden nodes and determining its output weights is continued until the desired learning accuracy is achieved. Finally for the interesting work of an extension

* Corresponding author. Tel.: +91 11 26704724; fax: +91 11 26741586.
E-mail addresses: bala_jnu@hotmail.com, balajnu@gmail.com (S. Balasundaram), navkapil@gmail.com (Kapil).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.12.033

of ELM to support vector networks and the application of ELM for time series the interested reader is referred to [9,19].

In this paper, we extend the study of error minimized ELM (EM-ELM) [4] proposed for estimation of a function to the problem of simultaneous learning of a function and its derivatives. Though our method of solution is applicable to the problem of simultaneous learning of a function and its derivatives of higher order, for reason of simplicity, we consider only the problem of simultaneous learning of a function and its first-order derivatives. The main advantage of our approach is that our formulation will lead to solving a rectangular system of linear equations and its solution is obtained by Moore–Penrose generalized pseudo-inverse [16]. Finally to verify the effectiveness of our method a number of examples are considered. It is observed that the proposed method is fast and produces similar or better generalization performance clearly demonstrates its practical use.

The rest of the paper is organized as follows: in Section 2 we briefly discuss the ELM initially proposed by Huang et al. [6] for SLFNs. For the problem of simultaneous learning of a function and its derivatives, the approach proposed by Lazaro et al. [12] in formulating the problem as an extended ε-insensitive SVR problem and the regularized least squares approach of Jayadeva et al. [10] are briefly discussed in Section 3. Motivated by their work [10,12] we propose in Section 4 the extension of the study of EM-ELM for the problem of simultaneous learning of a function and its derivatives. Experimental results are reported in Section 5 and finally we conclude our paper in Section 6.

Throughout in this work all vectors are assumed as column vectors. For any two vectors x, y in R^n the inner product of the vectors is denoted by $x^t y$ where x^t is the transpose of the vector x. The 2-norm of a vector x is denoted by $\|x\|$.

2. Extreme learning machine method

Let $\{(x_i,y_i)\}_{i=1,2,\ldots,p}$ be a set of training examples given where for the input example $x_i=(x_{i1},\ldots,x_{in})^t \in R^n$ its corresponding observed value of the function being $y_i \in R$. Then for the randomly assigned values for the weight vector $a_s=(a_{s1},\ldots,a_{sn})^t \in R^n$ and the bias $b_s \in R$ connecting the input layer to the sth hidden node, the standard SLFNs with ℓ number of hidden nodes approximate the input examples with zero error if and only if there exists an output vector $w=(w_1,\ldots,w_\ell)^t \in R^\ell$ in which w_s is the weight connecting the sth hidden node to the output node such that the following condition:

$$y_i = \sum_{s=1}^{\ell} w_s G(a_s,b_s,x_i) \quad \text{for } i=1,\ldots,p$$

holds, where $G(a_s, b_s, x_i)$ is the output of the sth hidden node for the input x_i. This set of equations can be written in matrix form as

$$Hw = y, \tag{1}$$

where

$$H = \begin{bmatrix} G(a_1,b_1,x_1) & \ldots & G(a_\ell,b_\ell,x_1) \\ \cdot & \cdots & \cdot \\ G(a_1,b_1,x_p) & \ldots & G(a_\ell,b_\ell,x_p) \end{bmatrix}_{p\times\ell} \tag{2}$$

and

$$y=(y_1,\ldots,y_p)^t \in R^p. \tag{3}$$

For additive hidden node with activation function g: $R \to R$, $G(a_s, b_s, x)$ is given by

$$G(a_s,b_s,x)=g(a_s^t x+b_s),$$

where a_s and b_s are the weight vector and bias connecting the input layer to the sth hidden node. Similarly for radial basis function (RBF) hidden node with activation function g: $R \to R$, $G(a_s,b_s,x)$ is given by

$$G(a_s,b_s,x)=g(b_s\|x-a_s\|),$$

where a_s and $b_s>0$ are the center and impact factor of the sth RBF node.

For a given SLFN for which the activation function $g(\cdot)$ in any interval is infinitely differentiable and the p training examples are distinct with the number of hidden nodes ℓ equals to p, for any randomly chosen $a_s \in R^n$ and $b_s \in R$ from any intervals of R^n and R, respectively, according to any continuous probability distribution, it has been shown in [6] that with probability one the hidden layer output matrix H of the SLFN defined by (2) is invertible and $\|Hw-y\|=0$. However, in real applications $\ell<p$ is true and in this case for the randomly assigned values of the parameters $a_s \in R^n$ and $b_s \in R$, training the SLFN is equivalent to obtaining a least squares solution w of the linear system (1). In fact, w is determined to be the minimum norm least squares solution of (1) which can be explicitly obtained by [6]

$$w=H^\dagger y,$$

where $H^\dagger$ is the Moore–Penrose generalized inverse [16] of the matrix H. Also when rank$(H)=\ell$, we can write

$$H^\dagger=(H^t H)^{-1} H^t, \tag{4}$$

where H^t is the transpose of H. Finally by obtaining the solution $w \in R^\ell$, the regression estimation function $f(\cdot)$ for any input example $x \in R^n$ is determined to be

$$f(x)=\sum_{s=1}^{\ell} w_s G(a_s,b_s,x). \tag{5}$$

Remark 1. Once the values of the weight vector $a_s \in R^n$ and the bias $b_s \in R$ are randomly assigned at the beginning of the learning algorithm they remain fixed and therefore the matrix H is unchanged.

Remark 2. Since the sigmoidal, radial basis, sine, cosine and exponential functions are infinitely differentiable in any interval of definition they can be chosen as activation functions.

Finally, it is important to note that, according to ELM theory, ELM works for generalized feedforward networks which may not be neuron alike [5,7,8].

3. The SVR approach of Lazaro et al. [12] and Jayadeva et al. [10] with derivatives

3.1. Support vector regression with derivatives (SVRD)

In [12], Lazaro et al. proposed the regression approximation problem for the simultaneous learning of a function and its derivatives formulated as an extended ε-insensitive loss function based SVR problem. However in order to optimize the regression estimation problem they employed an IRWLS procedure. They studied initially the method for solving one-dimensional input space problems and later extended their work to the general case.

In this subsection we briefly discuss the model of Lazaro et al. [12] for one-dimensional input space problems and for the general case of multidimensional input spaces the interested reader is referred to [12].

Assume that a set of input examples $\{(x_i,y_i,y_i')\}_{i=1,2,\ldots,p}$ is given where x_i denotes the ith input example for which the observed values of the function and its derivative being $y_i \in R$ and $y_i' \in R$, respectively. Consider the problem of finding the regression