Available online at www.sciencedirect.com

SciVerse ScienceDirect

Physics Procedia 24 (2012) 2227 – 2230

Physics Procedia

2012 International Conference on Applied Physics and Industrial Engineering

Least Squares Support Vector Machine Regression with Equality Constraints

Kun Liu[a],Bing-Yu Sun[b]

[a]Information Management Center,Anhui Normal University,Wuhu, Anhui, 241000, China
[b]Institute of Intelligent Machines,Chinese Academy of Sciences,Hefei, Anhui, China

Abstract

In regression problems sometimes we may know a prior that some data is noiseless or the approximated function must passes through several points. In order to resolve these problems, the LS-SVM regression with equality constraints is proposed. The experimental results proved that our method is effective.

keywords : Support Vector Machine ,Equality Constraints ,Function Regression.

1.Introduction

The Support Vector Machine (SVM) has been introduced by Vapnik [1] as a method for classification and for function approximation. In this paper, we will be concerned with function approximation only. The SVM is typically based on an ε -insensitive cost function, meaning that approximation errors smaller than ε will not increase the cost function value. This results in a quadratic convex optimization problem. Instead of using an ε -insensitive cost function, a quadratic cost function can be used. This approach results in so-called least squares support vector machines (LS-SVM), which was introduced by Suykens [2] and are closely related to regularization networks [3]. With the quadratic cost function, the optimization problem reduces to finding the solution of a set of linear equations, which is computationally attractive.

To a regression problem we can not know the very expression of the function that we will approximate, but sometimes we can known a prior that some data can be regarded as noiseless or this function must pass through certain points. However, in traditional LS-SVM regression this prior knowledge can not be utilized, so in this paper we will incorporate it into regression problems and the equality constraints is added to the traditional LS-SVM regression.

 doi:10.1016/j.phpro.2012.02.327

2.LS-SVM for Function Approximation

Given a training data set of N points $\{\mathbf{x}_k, y_k\}_{k=1}^N$ with input data $\mathbf{x}_k \in R^n$ and output data $y_k \in R$, one considers the following optimization problem in primal weight space:

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \frac{1}{2}\gamma \sum_{k=1}^{N} e_k^2 \qquad (1)$$

$$\text{s. t: } y_k = w^T \varphi(\mathbf{x}_k) + b + e_k, \quad k = 1,\ldots,N$$

with $\varphi(.): R^n \to R^{n_h}$ a function which maps the input space into a so-called higher dimensional(possibly infinite dimensional) feature space, weight vector $w \in R^{n_h}$ in primal weight space, error variables $e_k \in R$ and bias term b. Note that the cost function J consists of a sum squared error (SSE) and a regularization term, which is also a standard procedure for the training of MLP's[4]. The relative importance of these terms is determined by the positive real constant γ.

In primal weight space one has the model

$$y(\mathbf{x}) = w^T \varphi(\mathbf{x}) + b \qquad (2)$$

The weight vector w can be infinite dimensional, which makes a calculation of w form (1) impossible in general. Therefore, one computes the model in the dual space instead of the primal space. One defines the Lagrangian

$$L(w,b,e,\alpha) = J(w,e) - \sum_{k=1}^{N} \alpha_k \{w^T \varphi(\mathbf{x}_k) + b + e_k - y_k\} \qquad (3)$$

with Lagrange multipliers $\alpha_k \in R$.

According to Mercer's theory, one can choose a kernel function K(.,.) such that

$$K(\mathbf{x}_k, \mathbf{x}_j) = \varphi(\mathbf{x}_k) \cdot \varphi(\mathbf{x}_j) \qquad (4)$$

then from (3) we can get the resulting LS-SVM model for function estimation:

$$y(\mathbf{x}) = \sum_{k=1}^{N} \alpha_k K(\mathbf{x}, \mathbf{x}_k) + b \qquad (5)$$

where α, b are the solution to (3).

3.LS-SVM Regression with Equality Constraints

In traditional LS-SVM all the data in the training set have the same contribution to the training of LS-SVM. If it is known a prior that some training data are free of noise or the function must pass through certain points according to our need, we can add equality constraints of the form $w^T \varphi(\mathbf{x}_k) + b = y_k$ to the conventional LS-SVM. Supposing there are m points that the function must pass through or free of noise, it leads to the optimization problem:

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \frac{1}{2}\gamma \sum_{k=1}^{N-m} e_k^2 \qquad (6)$$

$$\text{s.t: } y_k = w^T \varphi(\mathbf{x}_k) + b + e_k, \quad k = 1,2,\ldots,N-m \qquad y_k = w^T \varphi(\mathbf{x}_k) + b, \ k = N-m+1, N-m+2,\ldots,N$$

Then the Lagrangian becomes:

$$L(w,b,e,\alpha) = J(w,e) - \sum_{k=1}^{N-m} \alpha_k \{w^T \varphi(\mathbf{x}_k) + b + e_k - y_k\} - \sum_{k=N-m+1}^{N} \alpha_k \{w^T \varphi(\mathbf{x}_k) + b - y_k\} \quad (7)$$

with Lagrange multipliers $\alpha_k \in R$. The condition for optimality is given by (8).

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \to w = \sum_{k=1}^{N} \alpha_k \varphi(\mathbf{x}_k) \\ \frac{\partial L}{\partial b} = 0 \to \sum_{k=1}^{N} \alpha_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 \to \alpha_k = \gamma e_k \quad k = 1,2,\ldots,N-m \\ \frac{\partial L}{\partial \alpha_k} = 0 \to \begin{array}{l} w^T \varphi(\mathbf{x}_k) + b + e_k - y_k = 0, k = 1,2,\ldots,N-m \\ w^T \varphi(\mathbf{x}_k) + b - y_k = 0, k = N-m+1, N-m+2,\ldots,N \end{array} \end{cases}$$

After elimination of w,e one obtains the KKT system:

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (9)$$

where $y = [y_1, y_2, \ldots, y_N]$, $1_v = [1,1,\ldots,1]$, $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]$, $\Omega_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$,for $i,j = 1,2,\ldots,N$,and $V = diag\{c_1, c_2, \ldots, c_N\}$,

for $c_1, c_2, \ldots, c_{N-m} = \frac{1}{\gamma}$, $c_{N-m+1}, c_{N-m+2}, \ldots, c_N = 0$.

Then corresponding LS-SVM model for regression can be derived through (9).

Sometimes the training data are free of noise or the signal-to-noise ratio is so low that the noise can be neglected; our aim is to approximate the underlying function. Then (6) is transferred to:

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w \quad (10)$$

$$\text{s. t:} \quad y_k = w^T \varphi(\mathbf{x}_k) + b, \; k = 1,2,\ldots,N.$$

4.Experimental Results

In order to verify the efficient of our method, we will estimate a sinc function from noisy data. The sinc function had been used to test the generalization of SVM regression in [1] [2] and it has been proved that SVM has good generalization performance. This paper will show that its performance can be further improved by adding equality constraints to the traditional LS-SVM. In our experiment the training data are contaminated by strong outliers and zero mean Gaussian noise except several data which is regarded as noiseless. Given is a training data set of N = 300 data points as shown in figure 1.

The simulation results of traditional LS-SVM and LS-SVM with equality constraints is shown in Figure.2. During the training the RBF kernel function

$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 / \sigma^2)$$

is used and corresponding parameters are determined by means of cross-validation on the training data.

Figure.2 is the regression results with three equality constraint. The points that compel the approximated function to pass through are denoted by the circle. From this figures it can be found that the performance of LS-SVM regression can be further improved if equality constraints is used during the training of LS-SVM in the area near the points denoted by circles. Both the traditional LS-SVM and our method use the same training data set, which proved that our method is efficient.

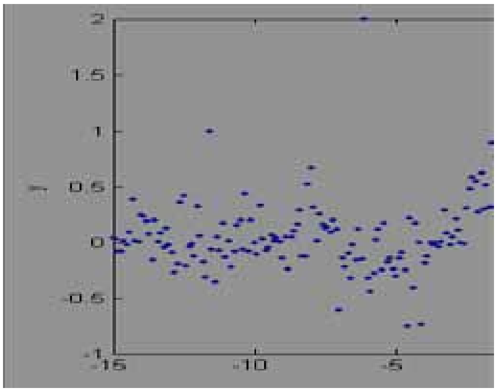

Fig .1. Training data $f(x) = \sin c(x)$

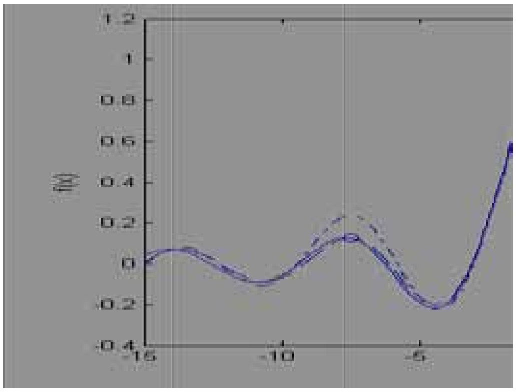

Fig.2. Comparison between traditional LS-SVM and LS-SVM with 3 equality constraints (EC)

5.Conclusions

A method that adds equality constraints to traditional LS-SVM is proposed. This method is very useful when we known a prior that the approximated function pass through some points or some data in training set is noiseless. The experiments on simulated problems showed that our method is effective.

References

[1] Vapnik V, The nature of Statistical Learning Theory, New York: Wiley, 1998.

[2] Suykens J A K, :Weighted least squares support vector machines: Robustness and sparse approximation, Neurocomput. 48(2002), pp85-105

[3] Smola. A J : Learning with Kernels, Ph.D. dissertation, GMD, Birlinghoven,Germany, 1998.

[4] Golub G H, Matrix Computations, Johns Hopkins University Press, Baltimore,MD,1989.