

# PARAMETER SELECTION IN LEAST SQUARES-SUPPORT VECTOR MACHINES REGRESSION ORIENTED, USING GENERALIZED CROSS-VALIDATION

## SELECCIÓN DE PARÁMETROS EN MÍNIMOS CUADRADOS-MÁQUINAS DE VECTORES DE SOPORTE ORIENTADAS A REGRESIÓN, UTILIZANDO VALIDACIÓN CRUZADA GENERALIZADA

ANDRÉS M. ÁLVAREZ MEZA

*Eng., Signal Processing and Recognition Group, Universidad Nacional de Colombia, Sede Manizales, amalvarezme@unal.edu.co*

GENARO DAZA SANTACOLOMA

*PhD., Universidad Antonio Nariño, Sede Bogotá, ensamblegl@gmail.com*

CARLOS D. ACOSTA MEDINA

*PhD., Scientific Computing and Mathematical Modeling Group, Universidad Nacional de Colombia, Sede Manizales, cdacostam@unal.edu.co*

GERMÁN CASTELLANOS DOMÍNGUEZ

*PhD., Signal Processing and Recognition Group, Universidad Nacional de Colombia, Sede Manizales, cgcastellanosd@unal.edu.co*

Received for review November 5<sup>th</sup>, 2010, accepted October 31<sup>th</sup>, 2011, final version November, 21<sup>th</sup>, 2011

**ABSTRACT:** In this work, a new methodology for automatic selection of the free parameters in the least squares–support vector machines (LS-SVM) regression oriented algorithm is proposed. We employ a multidimensional generalized cross-validation analysis in the linear equation system of LS-SVM. Our approach does not require prior knowledge about the influence of the LS-SVM free parameters in the results. The methodology is tested on two artificial and two real-world data sets. According to the results, our methodology computes suitable regressions with competitive relative errors.

**KEYWORDS:** parameter selection, least squares–support vector machines, multidimensional generalized cross validation, regression

**RESUMEN:** En este trabajo, se propone una metodología para la selección automática de los parámetros libres de la técnica de regresión basada en mínimos cuadrados máquinas de vectores de soporte (LS-SVM), a partir de un análisis de validación cruzada generalizada multidimensional sobre el conjunto de ecuaciones lineales de LS-SVM. La técnica desarrollada no requiere de un conocimiento a priori por parte del usuario acerca de la influencia de los parámetros libres en los resultados. Se realizan experimentos sobre dos bases de datos artificiales y dos bases de datos reales. De acuerdo a los resultados obtenidos, se concluye que el algoritmo desarrollado calcula regresiones apropiadas con errores relativos competentes.

**PALABRAS CLAVE:** selección de parámetros, mínimos cuadrados–máquinas de vectores de soporte, validación cruzada generalizada multidimensional, regresión

### 1. INTRODUCTION

Solving machine learning problems requires for one to suitably fix the needed free parameters of the system in order to obtain reliable results according to the given application such as: data preprocessing, feature extraction, classification, and regression [6,17,18]. Particularly, in order to solve a regression problem, it is necessary to generate a methodology that analyzes, interprets, and discerns patterns, finding the relationships between the outputs and inputs of the system. In this sense, some

algorithms have been developed based on statistical models and artificial neural networks (ANNs) [1,2]. Nonetheless, in most cases these techniques overfit the regression system due to the large number of parameters to fix, and the little prior user knowledge about the relevance of the inputs in the analyzed problem [3].

This is why support vector machines (SVMs) have been developed as an alternative that avoids such limitations. Their practical successes can be attributed to solid theoretical foundations based on VC-theory [4]. The SVM computes

globally optimal solutions, unlike those obtained with ANNs, which tend to fall into local minima. However, many SVM application studies are performed by expert users having a good understanding of the SMV methodology [5]. Therefore, the quality of SMV models depends on a proper setting of a considerable number of parameters. Moreover, the SVM algorithm demands a high-computational load due to the form of its optimization problem.

In this sense, the least squares–support vector machines (LS-SVM) method is proposed in [6], which is a reformulation of the traditional SVM algorithm. The LS-SVM uses a regularized least squares function with equality constraints, leading to a linear system which meets the Karush-Kuhn-Tucker (KKT) conditions for obtaining an optimal solution. Consequently, the regression problem can be solved by a linear equation system rather than quadratic programming, as in SVM.

Although LS-SVM simplifies the SVM procedure, the regularization parameter and the kernel parameters play an important role in the regression system. Therefore, it is necessary to establish a methodology for properly selecting the LS-SVM free parameters, in such a way that the regression obtained by LS-SVM must be robust against noisy conditions, and it does not need priori user knowledge about the influence of the free parameters values in the problem studied.

Cherkassky et al. [5] present a methodology to choose the regularization value in SVM from an analytic analysis over the regression function, which is similar to the LS-SVM one. Moreover, they employ a Gaussian kernel to train the system. However, this approach does not consider the direct possible influence of the band-width kernel parameter, which is manually fixed. In this sense, the user must infer the kernel parameter value according to his/her prior knowledge about the problem, over-fitting the regression system. Again, in Zhou et al. [7] a multi-parameter selection in LS-SVM is proposed. Even though this technique computes a competitive regression, it requires the assumption of some parameter values for the quantum-behaved particle swarm optimization (QPSO) algorithm, which can be unsuitable. Besides, they just test the proposed multi-parameter selection technique using a single database, which is perturbed with Gaussian noise. As a result, it is not possible to ensure reliable performance over different data sets.

In this paper, a new methodology for choosing the

regularization and Gaussian kernel band-width parameters in LS-SVM is proposed. We analyze the LS-SVM linear system using the generalized cross-validation (GCV) technique [8,9] in order to simultaneously infer the free parameters. Our approach does not require a prior knowledge about the influence of the LS-SVM parameters in the regression results.

The proposed algorithm is experimentally verified on two artificial and two real-world data sets. The regression quality is measured using the relative error between the target and the predicted sample.

This paper is organized as follows: Section 2 gives a brief introduction to the LS-SVM algorithm and the GCV methodology. Section 3 describes the algorithm proposed to simultaneously select the LS-SVM free parameters (regularization parameter and Gaussian kernel band-width). Section 4 presents the experimental conditions and shows the regression results obtained. Finally, the discussion and conclusions are given in Sections 5 and 6.

## 2. BACKGROUND

### 2.1 Least squares-support vector machines LS-SVM

Let  $\mathbf{X}$  be the  $n \times p$  input data matrix, and  $\mathbf{y}$  the  $n \times 1$  output vector. Given the  $\{\mathbf{x}_i, y_i\}_{i=1}^n$  training data set, with  $\mathbf{x}_i \in \mathfrak{R}^p$ , and  $y_i \in \mathfrak{R}$ , the LS-SVM goal is to construct the function  $f(\mathbf{x}) = y$ , which represents the dependence of the output  $y_i$  on the input  $\mathbf{x}_i$ . This function is formulated as

$$f(\mathbf{x}) = \mathbf{w}^\top \varphi(\mathbf{x}) + b, \quad (1)$$

where  $\mathbf{w}$  and  $\varphi(\mathbf{x}): \mathfrak{R}^p \rightarrow \mathfrak{R}^n$  are  $n \times 1$  column vectors, and  $b \in \mathfrak{R}$ . The LS-SVM algorithm [6] computes the function (1) from a similar minimization problem found in the SVM method [4]. However, the main difference is that LS-SVM involves equality constraints instead of inequalities, and it is based on a least square cost function. Furthermore, the LS-SVM method solves a linear problem while conventional SVM solves a quadratic one. More precisely, the optimization problem and the equality constraints of LS-SVM are defined as follows:

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{e}, b} J(\mathbf{w}, \mathbf{e}, b) &= \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \frac{1}{2} \mathbf{e}^\top \mathbf{1} \\ \text{s.t. } y_i &= \mathbf{w}^\top \varphi(\mathbf{x}_i) + b + e_i \end{aligned} \quad (2)$$

where  $\mathbf{e}$  is the  $n \times 1$  error vector,  $\mathbf{1}$  is an  $n \times 1$  vector with all entries 1, and  $C \in \mathfrak{R}^+$  is the tradeoff parameter between the solution size and training errors. From (2) a Lagrangian is formed, and differentiating with respect to  $\mathbf{w}, b, \mathbf{e}, \mathbf{a}$  ( $\mathbf{a}$  : Lagrangian multipliers), we obtain

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & -\mathbf{Z}^\top \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{1}^\top \\ \mathbf{0} & \mathbf{0} & C\mathbf{I} & -\mathbf{I} \\ \mathbf{Z} & \mathbf{1} & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ b \\ \mathbf{e} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{y} \end{bmatrix}, \quad (3)$$

where  $\mathbf{I}$  represents the identity matrix and  $\mathbf{Z} = [\varphi(\mathbf{x}_1), \varphi(\mathbf{x}_2), \dots, \varphi(\mathbf{x}_n)]^\top$ .

From rows one and three in (3)  $\mathbf{w} = \mathbf{Z}^\top \mathbf{a}$  and  $C\mathbf{e} = \mathbf{a}$ . Then, by defining the kernel matrix  $\mathbf{K} = \mathbf{Z}\mathbf{Z}^\top$ , and the parameter  $\lambda = C^{-1}$ , the conditions for optimality lead to the following overall solution:

$$\begin{bmatrix} \mathbf{0} & \mathbf{1}^\top \\ \mathbf{1} & \mathbf{K} + \lambda\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix}. \quad (4)$$

In this work, we consider the Gaussian Kernel, which is defined as

$$\mathbf{K}_\sigma(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (5)$$

We can obtain the solution of the linear equation system presented in (4) as

$$\begin{aligned} \mathbf{a}_{(\sigma, \lambda)} &= \tilde{\mathbf{A}}_{(\sigma, \lambda)}^{-1} (\mathbf{y} - \mathbf{1}b_{(\sigma, \lambda)}) \\ b_{(\sigma, \lambda)} &= \frac{\mathbf{1}^\top \tilde{\mathbf{A}}_{(\sigma, \lambda)}^{-1} \mathbf{y}}{\mathbf{1}^\top \tilde{\mathbf{A}}_{(\sigma, \lambda)}^{-1} \mathbf{1}} \end{aligned}, \quad (6)$$

with  $\tilde{\mathbf{A}}_{(\sigma, \lambda)} = \mathbf{K}_\sigma + \lambda\mathbf{I}$ . Hence, Eq. (1) can be rewritten as a function of the Lagrangian multipliers

$$f(\mathbf{x}) = \sum_{i=1}^n \mathbf{a}_{(\sigma, \lambda)_i} \mathbf{K}_\sigma(\mathbf{x}_i, \mathbf{x}) + b_{(\sigma, \lambda)}. \quad (7)$$

Taking into account Eq. (7), the LS-SVM performance depends of two free parameters:  $\lambda = C^{-1}$  and  $\sigma$ . In this sense, it is necessary to develop a methodology for finding suitable values of the LS-SVM free parameters.

We use the generalized cross-validation (GCV) method for analyzing the linear equation system (4) to fix the free parameters of LS-SVM. Next, a brief description of GCV is presented.

### 2.2 Generalized Cross-Validation (GCV)

For dealing with ill-conditioned matrices  $\mathbf{A}$ , the regularization techniques are based on approximations of the form  $\mathbf{u}_\gamma^\epsilon = \mathbf{A}_\gamma^\dagger \mathbf{r}^\epsilon$ , where  $\gamma > 0$  is the regularization parameter,  $\mathbf{r}^\epsilon$  is a column vector with the estimated measures,  $\mathbf{u}_\gamma^\epsilon$  is a column vector containing the calculated solutions, and  $\mathbf{A}_\gamma^\dagger$  is a stable, easy to compute approximation of the generalized inverse of  $\mathbf{A}$ .

The GCV algorithm [8,9] looks for a  $\gamma$  value that allows for one to obtain a suitable balance between the regularization error and the perturbation in the solution. In this sense, the GCV method calculates the  $\gamma$  value that minimizes

$$\min_\gamma \{G(\gamma)\} = \frac{\|\mathbf{A}\mathbf{u}_\gamma^\epsilon - \mathbf{r}^\epsilon\|^2}{\text{tr}(\mathbf{I} - \mathbf{A}\mathbf{A}_\gamma^\dagger)^2} \quad (8)$$

### 3. LS-SVM FREE PARAMETER SELECTION

In this work, we relate the linear equation system (4) with a problem of the form  $\mathbf{A}\mathbf{u} = \mathbf{r}^\epsilon$ , in order to fix the free parameters of LS-SVM ( $\lambda = C^{-1}, \sigma$ ) using the GCV method. Nevertheless, it should be noted that the original GCV algorithm method (8) was designed for the selection of a single parameter. For this reason, it is necessary to formulate the inverse problem of LS-SVM, in order to select its two free parameters simultaneously. Other similar approaches of the GCV method for choosing multiple parameters can be found in [10,11]. Based on the linear system presented in (4), we propose to set the relationships between LS-SVM and GCV as

$$\begin{aligned} \mathbf{A} &= \mathbf{K}_\sigma, \quad \mathbf{A}_\gamma^\dagger = \tilde{\mathbf{A}}_{(\sigma, \lambda)}^{-1} = (\mathbf{K}_\sigma + \lambda\mathbf{I})^{-1}, \\ \mathbf{u}_\gamma^\epsilon &= \mathbf{a}_{(\sigma, \lambda)}, \quad \mathbf{r}^\epsilon = \mathbf{y} - \mathbf{1}b_{(\sigma, \lambda)}. \end{aligned} \quad (9)$$

Note that  $\mathbf{K}_\sigma + \lambda\mathbf{I}$  is positive definite. Hence, the GCV function to be considered is

$$\min_{\sigma, \lambda} \{G(\sigma, \lambda)\} = \frac{\|\mathbf{K}_\sigma \mathbf{a}_{(\sigma, \lambda)} - (\mathbf{y} - \mathbf{1}b_{(\sigma, \lambda)})\|^2}{\text{tr}(\mathbf{I} - \mathbf{K}_\sigma (\mathbf{K}_\sigma + \lambda\mathbf{I})^{-1})^2} \quad (10)$$

s.t.  $\sigma, \lambda > 0$ .

The above optimization problem is generally referred to as a constrained nonlinear optimization. It can be solved using the active-set optimization algorithm, which uses a sequential quadratic programming (SQP) method [12,13]. We will use its implementation in the

*fmincon* Matlab routine.

To avoid over-fitting in the initialization of the unknown variables in (10), we propose the following procedure: First, we choose the initial value of  $\sigma$  according to Sylverman's rule [14]:

$$\sigma_s = 0.9 \min(\overline{\text{std}(\mathbf{X})}, \frac{1}{1.34} \overline{\text{iqr}(\mathbf{X})}) n^{-1/5} \quad (11)$$

where  $\overline{\text{iqr}(\cdot)}$  computes the average interquartil range and  $\overline{\text{std}(\cdot)}$  calculates the average standard deviation. Then, we select the initial value of  $\lambda$  ( $\lambda_1$ ), minimizing Eq. (10) with  $\sigma = \sigma_s$ . We fix the bounds of  $\lambda = C^{-1}$  according to [8]:

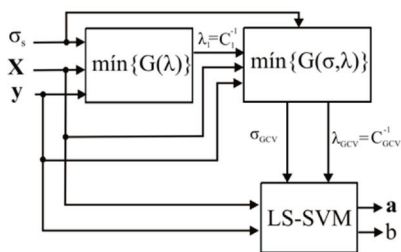
$$\begin{aligned} \lambda_{\min} &= \min(\nu_{\mathbf{K}_\sigma}) \\ \lambda_{\max} &= \max(\nu_{\mathbf{K}_\sigma}), \end{aligned} \quad (12)$$

where  $\nu_{\mathbf{K}_\sigma}$  contains the eigenvalues of  $\mathbf{K}_{\sigma_s}$  greater than zero.

Later, we use  $\lambda_1$  and  $\sigma_s$  as initial values to minimize (10), setting the bounds of  $\lambda$  as in (12) and the bounds of  $\sigma$  as

$$\begin{aligned} \sigma_{\min} &= 0.1\sigma_s \\ \sigma_{\max} &= \max\{\overline{\text{std}(\mathbf{X})}, \overline{\text{iqr}(\mathbf{X})}\}. \end{aligned} \quad (13)$$

Finally, the optimal values  $\lambda_{\text{gcvt}} = C_{\text{gcvt}}^{-1}$  and  $\sigma_{\text{gcvt}}$  that minimize (10) are used for training the regression system base on the LS-SVM algorithm. The proposed methodology can be summarized as presented in Fig. 1.



**Figure 1.** Proposed scheme for the LS-SVM parameter selection

## 4. EXPERIMENTS

Two artificial and two real-world data sets are tested. We employ a 10-fold cross validation analysis to determinate the experiment's generalization and robustness. For each fold we randomly select a training set, which is used to calculate the LS-SVM parameters according to the proposed approach, and it is also employed to train the LS-SVM regression algorithm.

The remaining data is used as test set. We compute the relative error (RE) over the test set according to Eq. (14), and the performance of the system is calculated as the mean relative error (MRE) for the 10-folds.

$$100 \frac{\|\mathbf{y}_{\text{test}} - f(\mathbf{X}_{\text{test}})\|^2}{\|\mathbf{y}_{\text{test}}\|^2} [\%] \quad (14)$$

### 4.1 Artificial data sets

The first artificial data set is the univariate *Sinc* function, which has been studied in [6,7]. This function is defined as

$$y_i = \begin{cases} \frac{\sin(x_i)}{x_i} & \forall x_i \neq 0 \\ 1 & x_i = 0 \end{cases}, \quad (15)$$

where  $\mathbf{x} \in \mathfrak{R}$ . We generate 300 observations. The vector  $\mathbf{x}$  is taken from a uniform grid in the  $[-10, 10]$  interval. We randomly select 150 samples as a training set and the remaining 150 conforms the test set. Moreover, the training output samples are perturbed with Gaussian noise ( $\delta \sim (0, \sigma_\delta)$ ). In Table 1 the MRE for the *Sinc* data set is shown for different noise conditions. Besides, the  $\sigma_s$ ,  $\sigma_{\text{gcvt}}$ ,  $\lambda_1 = C_1^{-1}$  and  $\lambda_{\text{gcvt}} = C_{\text{gcvt}}^{-1}$  parameter values are presented for the lowest relative error. Additionally, in Fig. 2 some regression results for the *Sinc* data set are presented.

The second artificial data set corresponds to the function *Sinc3D*, which is also analyzed in [5] and can be calculated as

$$y_i = \begin{cases} \frac{\sin(\|\mathbf{x}_i\|)}{\|\mathbf{x}_i\|} & \forall \|\mathbf{x}_i\| \neq 0 \\ 1 & \|\mathbf{x}_i\| = 0 \end{cases}, \quad (16)$$

where  $\mathbf{x}_i \in \mathfrak{R}^p$ . We compute 845 observations. The  $\mathbf{x}$  values are sampled on a uniform square lattice  $[-5, 5]^p$  ( $p = 2$ ). We randomly select 169 samples as training set and the remaining 676 as a test set. Besides, the training set is perturbed with Gaussian noise ( $\delta \sim (0, \sigma_\delta)$ ). In Table 2 and Fig. 3, the *Sinc3D* results are shown.

### 4.1 Real-world data sets

The first real-world data set corresponds to the concrete compressive strength (CCS) [15], which is a highly nonlinear function of the time and ingredients. Eight variables are measured: including cement, blast furnace slag, fly ash, water, super plasticizer, coarse aggregate,

fine aggregate, and time. There are 1030 observations ( $\mathbf{X} \in \mathfrak{R}^8$ ), and the goal is to predict the CCS ( $\mathbf{y} \in \mathfrak{R}$ ) for different input conditions. In this case, we train the regression system with 309 random samples, and the MRE is calculated on the remaining 721 data. In Table 3 and Fig. 4, the CCS data set results are presented.

Finally, the European Climate Assessment (ECA) real-world data set [16] is tested. This database is a daily weather summary of Berlin, Germany from between 2001 to 2004. Nine variables are measured: cloud cover, mean relative humidity, mean barometric pressure, snow depth, precipitation amount, sunshine, amount of rain, minimum air temperature, maximum air temperature, and mean air temperature. In our experiments, we analyze the relationships between the mean air daily temperature and the remaining meteorological features. Therefore, we have 1465 observations, where  $\mathbf{X} \in \mathfrak{R}^8$  and  $\mathbf{y} \in \mathfrak{R}$ . We randomly choose 439 samples for the training set, and 1026 for the test set. In Table 5 the ECA data set results are presented. For illustration, see Fig. 5.

## 5. DISCUSSION

According to the obtained results shown in Table 1 and Fig. 2, it is possible to notice that the proposed methodology for choosing the values of  $\lambda = C^{-1}$  and  $\sigma$  in the LS-SVM algorithm, allows for one to find suitable regression results for the *Sinc* database.

Our methodology improves the results presented in Zhou et al. [7], where a RE of 2.3141[%] is reported for a similar experiment ( $\sigma_\delta = 0.1$ ). Even when this technique computes competitive regression, it requires the assumption of some free parameter values for the QPSO algorithm, which can be undesirable when the user does not have prior knowledge about the phenomenon. Besides, no more experiments with different noise conditions are presented; so only limited conclusions can be reached. On the other hand, our methodology shows a suitable performance, even for different noise conditions.

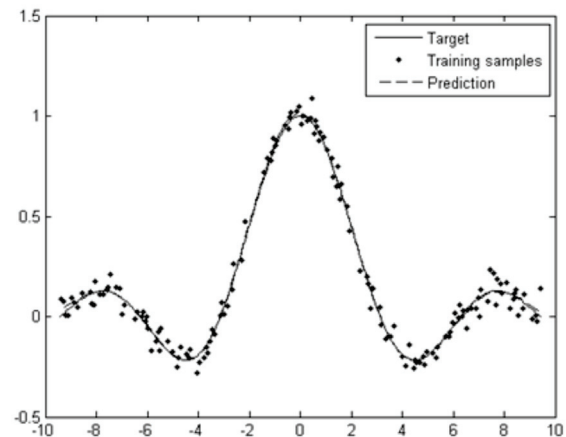
Furthermore, it can be seen in Table 1 how our algorithm controls the LS-SVM free parameters in the *Sinc* dataset. If  $\sigma_\delta$  increases, the  $C_{gcv}$  value is low, which prevents an over-fitting in the LS-SVM training.

Otherwise, if  $\sigma_\delta$  decreases, the  $C_{gcv}$  value is high, giving more weight to the training error of the LS-

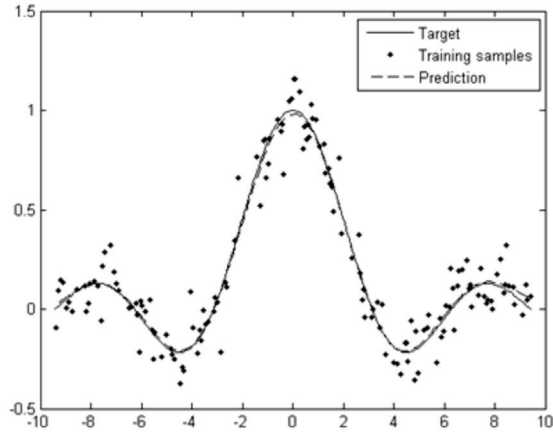
SVM optimization problem. Now, the lowest  $\sigma_{gcv}$  is calculated for the highest  $\sigma_\delta$ , which reveals that the proposed methodology analyzes the system with a low band-width when the output signal is highly perturbed.

Again, in agreement with the results shown in Table 2 and Fig. 3, our approach computes suitable regression for the *Sinc3D* dataset. Moreover, the  $C_{gcv}$  and  $\sigma_{gcv}$  values decrease for high noise conditions, which allow for one to find a regression function that can deal with perturbed samples. The last statement can be especially corroborated by the regression results attained for  $\sigma_\delta = 0.4$  (Fig. 3 (b)) Note that our approach improves the *Sinc3D* results presented in [5], where an RE of 23.9393[%] and 2.0908[%] are reported for  $\sigma_\delta = 0.4$  and  $\sigma_\delta = 0.1$ , respectively. It is important to note that the methodology presented in [5] analytically chooses the  $C$  value, but it does not directly consider the influence of  $\sigma$ , which is manually fixed. This is why it is not possible to ensure suitable performance in several cases.

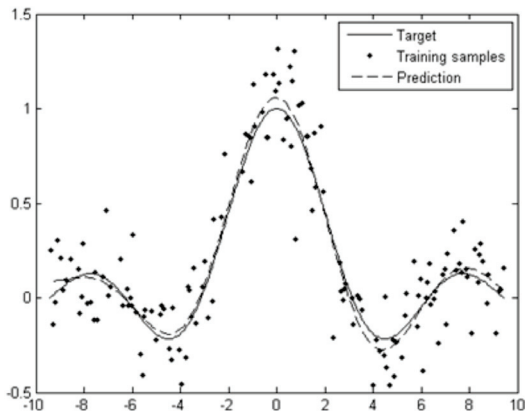
Regarding to the real-world experiments, our methodology calculates appropriate regressions with low ARE results (Tables 3 and 4, Fig. 4 and Fig. 5), which confirm its applicability in complex problems. According to the fixed  $C_{gcv}$  and  $\sigma_{gcv}$  values, it can be seen how our method aims to analyze the data with a low band-width Gaussian kernel, while a considerably high value for the tradeoff parameter is selected. Indeed, our approach fixed the highest  $C_{gcv}$  value for the ECA dataset (Table 4), which can be explained by the fact that ECA has a well defined dynamic that is suitably modeled by LS-SVM. On the other hand, CCS data contains more complex nonlinearities properties, which LS-SVM aim to compensate with low  $\sigma_{gcv}$  and  $C_{gcv}$  values.



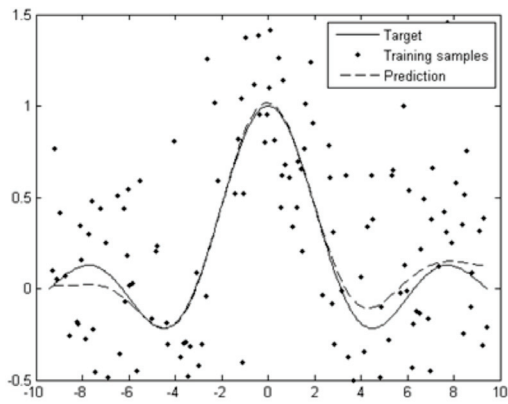
2(a)  $\sigma_\delta = 0.05$



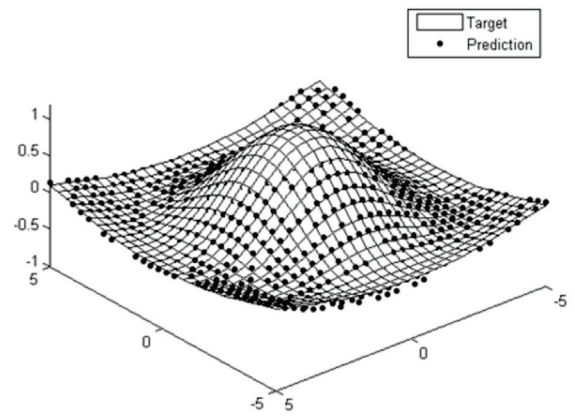
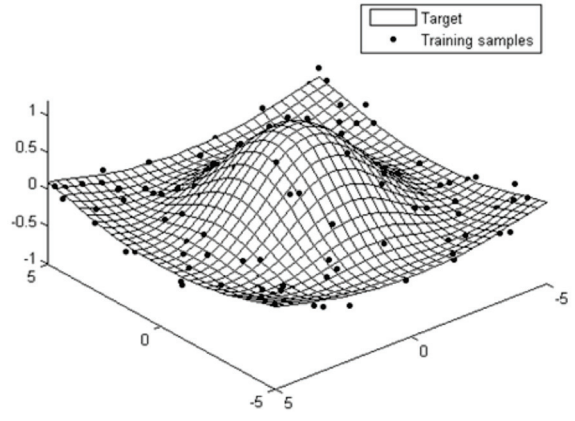
**2(b)**  $\sigma_\delta = 0.1$



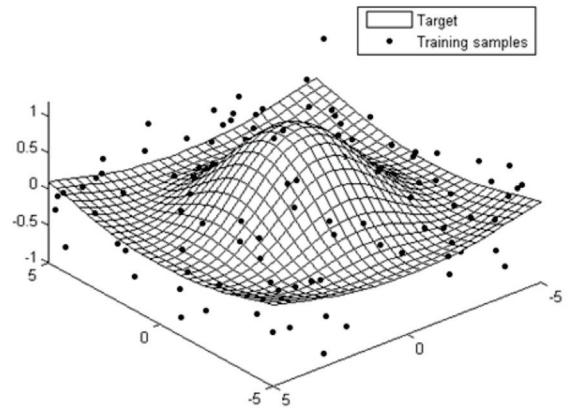
**2(c)**  $\sigma_\delta = 0.2$



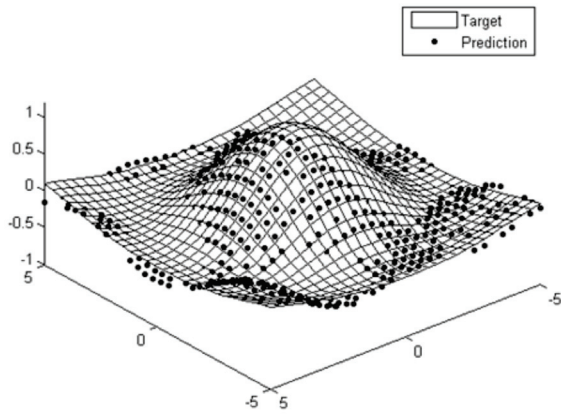
**2(d)**  $\sigma_\delta = 0.5$



**3(a)**  
 $\sigma_\delta = 0.1$



**Figure 2.** *Sinc* results



3(b)  $\sigma_\delta = 0.4$

Figure 3. *Sinc3D* results

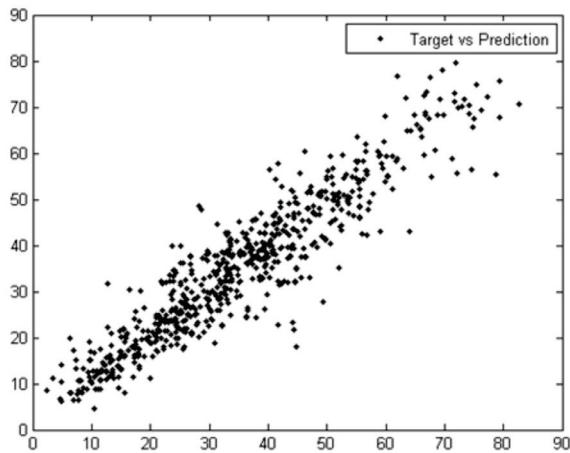


Figure 4. Target vs. prediction (CCS)

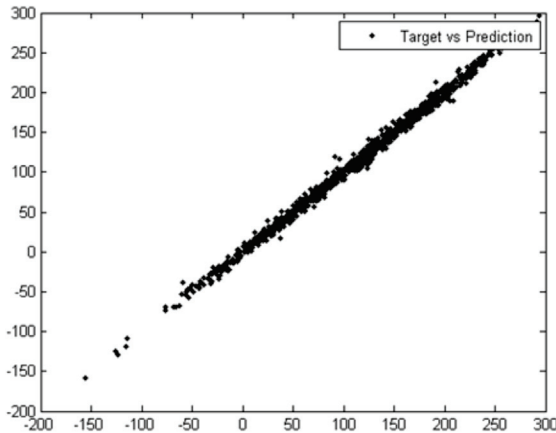


Figure 5. Target vs. prediction (ECA)

Table 1. *Sinc* results

$\sigma_\delta$	MRE [%]	$\sigma_s$	$C_1$	$\sigma_{gcv}$	$C_{gcv}$
0.5	7.8386±4.2507	1.8337	0.4152	2.0501	0.4156
0.2	1.5261±0.7471	1.8337	1.4220	2.9484	10.5242
0.1	0.4865±0.1721	1.8337	2.5433	2.6602	9.5901
0.05	0.0966±0.0488	1.8337	7.6568	2.8957	43.1230

Table 2. *Sinc3D* results

$\sigma_\delta$	MRE [%]	$\sigma_s$	$C_1$	$\sigma_{gcv}$	$C_{gcv}$
0.4	18.3059±5.2096	0.8169	0.6034	1.1157	0.5277
0.1	1.6655±0.3226	0.8169	2.5947	2.8516	20.7017

Table 3. CCS results MRE [%]

$\sigma_s$	$C_1$	$\sigma_{gcv}$	$C_{gcv}$
3.7873±1.1222	1.4027	21.4226	3.4043
			214.2257

Table 4. ECA results

MRE [%]	$\sigma_s$	$C_1$	$\sigma_{gcv}$	$C_{gcv}$
0.1611±0.0143	0.6117	188.6043	10.224	1886.0436

## 6. CONCLUSIONS

In this paper, a methodology for automatic parameters choice in the LS-SVM algorithm is proposed. It selects simultaneously suitable values for the parameters  $C$  and  $\sigma$  using the GCV method, formulating a scheme that relates the LS-SVM optimization to an inverse problem. According to the experiments, our technique computes suitable regression results even in several noise conditions.

Besides, our algorithm does not need prior knowledge about the influence of the LS-SVM parameters in the phenomenon studied. The proposed method seems to be appropriated for real-world regression tasks. It is important to note that due to the nonconvex characteristic of the proposed optimization problem for the LS-SVM free parameter selection, our approach can not ensure the computation of the optimal values for  $C$  and  $\sigma$ . However, our initialization procedure allows for one to work in a suitable domain for minimizing the proposed objective function, which can be confirmed by the results attained.

As future work, we are interested in testing more complex regression problems and forecasting procedures.

## ACKNOWLEDGMENTS

This research was carried out under grants provided by the project 1115-470-22055 funded by the Research Center for Excellence ARTICA, Medellín, Colombia, and by projects 20201006599, 20201006570, and 20201006594 funded by the *Universidad Nacional de Colombia Sede Manizales*. Moreover, GDS was supported by project #20110108-PI/UAN-2011-510gb UAN.

## REFERENCES

- [1] Methaprayoon, K., Lee, W. J., Rasmiddatta, S., Liao, J. and Ross, R., Multi-stage artificial neural network short-term load forecasting engine with front-end weather forecast, *IEEE Trans. Ind. Appl.*, pp. 1410–1416, 2007.
- [2] Maier, H. and Dandy, G., Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications, *Environmental Modeling and Software*, vol. 15, pp. 101–124, 2000.
- [3] Leng, X. and Miller, H.-G., Input dimension reduction for load forecasting based on support vector machines, *IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies (DRPT2004)*, 2004.
- [4] Vapnik, V., *The nature of statistical learning*, second edition, Springer, 1999.
- [5] Cherkassky, V. and Ma, Y., Practical Selection of SVM Parameters and Noise Estimation for SVM regression. *Neural Networks*, vol., 17, pp. 113-126, 2004.
- [6] Suykens, J. A. K., Gestel, V. T., Brabanter, J. D., Moor, B. D. and Vandewalle, J. *Least squares support vector machines*, World Scientific, 2002.
- [7] Zhou, L., Yang, H. and Liu, C., QPSO-based hyper parameters selection for LS-SVM regression. *Fourth International Conference on Natural Computation*. Jinan, China, 2008.
- [8] Hansen, C., Nagy, J. and O'leary, D., *Deblurring Images: Matrices, Spectra, and Filtering*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2006.
- [9] Golub, M. and Wahba, G., Generalized Cross Validation as a method for choosing a good ridge parameter, *Technometrics*, vol. 21, pp. 215–223, 1979.
- [10] Nguyen, N., Milanfar, P., Member, S. and Golub, G., Efficient generalized cross validation with applications to parametric image restoration and resolution enhancement, *IEEE Transactions on Image Processing*, vol. 10, 2001.
- [11] Peiliang, X., Iterative generalized cross-validation for fusing heteroscedastic data of inverse ill-posed problems, *Geophys. J. Int.*, vol. 179, pp. 182-200, 2009.
- [12] Boggs, P.T. and Tolle, J.W., Sequential quadratic programming for large-scale nonlinear optimization, *Journal of Computational Application Mathematics*, vol. 124, pp. 123-137, 2000.
- [13] Powell, M.J.D., *A Fast Algorithm for Nonlinearly Constrained Optimization Calculations*, Numerical Analysis, Lecture Notes in Mathematics, Springer Verlag, Vol. 630, 1978.
- [14] Sheather, S.J., Density Estimation, *Statistical Sci.* 19 (2004) 588–597.
- [15] Cheng, Y., Modeling of strength of high performance concrete using artificial neural networks. *Cement and Concrete Research*, vol. 28, pp. 1797-1808, 1998.
- [16] Tank, A. M., coauthors, Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, *Journal of Climatology* 22, pp. 1441–1453, 2002.
- [17] Soto, C. and Jiménez, C., Supervised learning for fuzzy discrimination and classification, *Revista DYNA*, vol. 78, pp. 26-36, 2011.
- [18] Pulgarín, J., Acosta, C. and Castellanos, G., Multiscale analysis by means of discrete mollification for ECG noise reduction, *Revista DYNA*, vol. 76, pp. 185-191, 2009.