Kernel Support Vector Regression with imprecise output *

Emilio Carrizosa, José Gordillo†
Universidad de Sevilla (Spain)
{ecarrizosa,jgordillo}@us.es

Frank Plastria
Vrije Universiteit Brussel (Belgium)
Frank.Plastria@vub.ac.be

27th January 2008

Abstract

We consider a regression problem where uncertainty affects to the dependent variable of the elements of the database. A model based on the standard ϵ-Support Vector Regression approach is given, where two hyperplanes need to be constructed to predict the interval-valued dependent variable. By using the Hausdorff distance to measure the error between predicted and real intervals, a convex quadratic optimization problem is obtained. Non-linear regressors are introduced via the use of kernels and several numerical experiments are performed to test our methodology.

Keywords: Support Vector Regression, Kernels, Interval Data, Quadratic Programming

*This work has been partially supported by the grants MTM2005-09362-103-01 of MEC, Spain, and FQM-329 of Junta de Andalucía, Spain. Their financial support is gratefully acknowledged.

†Corresponding author.

1 Introduction

In certain situations, the most suitable way to express some variables in a database is not via single feature vectors but via intervals. Intervals appear for expressing ranges, such as the cost of certain items in the set of shops of a town (since there will be some variations in the price between the different shops), or for summarizing the performances of a certain measurement to the same individual, for instance, the weight of a newborn during his/her first week. Other examples of interval data appear whenever imprecision or vagueness affect the database of our problem.

In this work, we study the regression problem in which the predictive variables are single-valued, but the dependent variable is affected by uncertainty. In the bibliography, a first problem related to ours involves the concept of interval regression analysis, which is the simplest version of possibilistic regression analysis, introduced by Tanaka et al. (see [19, 22, 23]). Given a database with crisp input and output, the aim of interval regression analysis is to predict the dependent variable via an interval by using the predictor variables. For this, the coefficients of the model used for the regression are also intervals. Each coefficient is expressed via its center and its radius.
In the original model, a linear programming formulation is given to solve the problem, where the objective is minimizing the sum of radii of the predicted outputs, with the constraint that the real value of the dependent variable must be included in the predicted output (see [22]). Later, in [23], a quadratic formulation is given to include in the objective function a term to minimize the sum of the squared distances from the center of the predicted output to the real value of the dependent variable.
Other improvements have been performed to study the role of outliers in the regression process. In [19], two regression models are built for each database by using quantile techniques, and two interval outputs are given for each observation, with the smaller one included in the larger one. The first model is built with a given proportion of the data (this way, we can study the general behaviour of the data, without containing outliers), whereas the second model is built with all the observations. Then, given a database, two intervals will be assigned as a prediction, and this can be seen as a trapezoidal fuzzy output. Support Vector Machines have been applied to that problem to build the two models (see [15]) and to the general interval regression analysis case. In [17], an ϵ-SVR is solved (with $\epsilon = 0$) to obtain an initial crisp value of the output, which will be the center of an interval output with radius equal to a value ϵ, computed by using the obtained regression errors. This interval output will be given as initial seed to two Radial Basis Function networks which identify the upper and lower sides of the output. In [16], the quadratic formulation of [23] is integrated with the standard ϵ-SVR approach.

SVMs have also been applied to fuzzy multiple linear regression models (see [13, 14]). Two different models have been studied in these works: when the predictor and the dependent variables are symmetric triangular fuzzy numbers (fuzzy input - fuzzy output) and when the predictor variables are crisp and the dependent variable is a triangular fuzzy number (crisp input - fuzzy output). The latter model is clearly connected to our work. The standard ϵ-SVR methodology is applied in both cases by imposing that the mode and the extremes of the intervals must satisfy the usual constraints. In the crisp input - fuzzy output case, nonlinear regressors are introduced via kernel methods.

In our paper, we study the regression problem with single-valued input and interval-valued output. We adapt the standard ϵ-Support Vector Regression methodology to our problem, building two hyperplanes to give an interval output for each element of the database. The dual of the problem is also studied to allow the introduction of non-linear regressors via kernels.

The structure of the paper is the following. In Section 2, after an introduction on the standard ϵ-SVR approach and ϵ-SVR for interval data, we formulate the optimization problem to solve in the case of interval output. The dual formulation is obtained in Section 3, and kernels are introduced in Section 4 to be able to model non-linear relations in the dataset. Some computational experiments are performed with the primal and dual formulations in Section 5. Section 6 finishes the paper with some discussion.

2 Modeling the problem

2.1 ϵ-Support Vector Regression for points and intervals

In the standard ϵ-Support Vector Regression, ϵ-SVR for short (see e.g. [6, 8, 10, 21, 24, 25]), a database $\Omega \subseteq \mathbb{R}^d \times \mathbb{R}$ is given, with elements $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$, where x_i is the vector of predictor variables and y_i is the dependent variable, whose value is to be predicted from the value of x_i.
The aim of ϵ-SVR is to find a linear function $f(x) = \omega^\top x + \beta$, with $\omega \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$, to approximate the values of the dependent variable y, such that the deviation between the real value y_i and the predicted value $f(x_i)$, for each instance $i \in \Omega$, is at most ϵ. Thus, the constraints of this problem are

$$|\omega^\top x_i + \beta - y_i| \leq \epsilon, \ \forall i \in \Omega. \tag{1}$$

As done in Support Vector Machines (see [2, 5, 6]), one seeks $f(x)$ such that the norm of ω is as small as possible. Then, the optimization problem to solve, as stated e.g. in [21], is the following

$$\begin{array}{ll} \min\limits_{\omega,\beta} & \dfrac{1}{2}\sum\limits_{j=1}^{d} \omega_j^2 \\ \text{s.t.} & y_i - \omega^\top x_i - \beta \leq \epsilon, \ \forall i \in \Omega \\ & \omega^\top x_i + \beta - y_i \leq \epsilon, \ \forall i \in \Omega. \end{array} \tag{2}$$

A Soft-Margin version, with slack variables ξ_i, ξ_i^* in the constraints and the corresponding penalty term in the objective function, has the following form

$$\begin{array}{ll} \min\limits_{\omega,\beta,\xi,\xi^*} & \dfrac{1}{2}\sum\limits_{j=1}^{d} \omega_j^2 + C\sum\limits_{i\in\Omega}(\xi_i + \xi_i^*) \\ \text{s.t.} & y_i - \omega^\top x_i - \beta \leq \epsilon + \xi_i, \ \forall i \in \Omega \\ & \omega^\top x_i + \beta - y_i \leq \epsilon + \xi_i^*, \ \forall i \in \Omega \\ & \xi_i, \xi_i^* \geq 0, \ \forall i \in \Omega, \end{array} \tag{3}$$

with C and ϵ constants of the model.
In this paper, we consider an interval-valued database $\Omega \subset \mathbb{R}^d \times \mathbb{R}$ with elements $i = (x_i, Y_i) \in \Omega$, where $x_i \in \mathbb{R}^d$ is the vector of single-valued predictive variables and Y_i is an interval in $\mathbb{R}$, $Y_i = [\tilde{l}_i, \tilde{u}_i]$, with $\tilde{l}_i \leq \tilde{u}_i$ (that is, the dependent variable is affected by uncertainty).
This problem can be approached via the model proposed in [3] by the authors. In that paper, the more general case where both the dependent and the predictive variables are intervals is studied.
As a particular case, we also studied the problem where only the dependent variable is affected by vagueness. In that case, Problem (3) was adapted by using a distance in the constraints of the problem to measure the deviation between the real interval and the predicted one.
Two different distances are studied in [3]: the maximum distance d_{max} and the Hausdorff distance d_H, defined on intervals $[\underline{a}, \overline{a}]$, $[\underline{b}, \overline{b}]$ as

$$\begin{aligned} d_{max}([\underline{a}, \overline{a}], [\underline{b}, \overline{b}]) &= \max\{|a - b| : \ a \in [\underline{a}, \overline{a}], \ b \in [\underline{b}, \overline{b}]\} \qquad (4) \\ &= \max\{|\underline{a} - \overline{b}|, |\overline{a} - \underline{b}|\} \\ d_H([\underline{a}, \overline{a}], [\underline{b}, \overline{b}]) &= \max\{\max_{a \in [\underline{a}, \overline{a}]} \min_{b \in [\underline{b}, \overline{b}]} |a - b|, \min_{a \in [\underline{a}, \overline{a}]} \max_{b \in [\underline{b}, \overline{b}]} |a - b|\} \qquad (5) \\ &= \max\{|\underline{a} - \underline{b}|, |\overline{a} - \overline{b}|\}. \end{aligned}$$

When using the maximum distance, d_{max}, it is shown in [3] (Remark 3.2) that the Soft-Margin formulation to use is the following convex quadratic program with linear constraints,

$$\begin{array}{ll} \min\limits_{\omega, \beta, \xi, \xi^*} & \frac{1}{2} \sum\limits_{j=1}^{d} \omega_j^2 + C \sum\limits_{i \in \Omega} (\xi_i + \xi_i^*) \\ \text{s.t.} & \tilde{u}_i - \sum\limits_{j=1}^{d} \omega_j x_{ij} - \beta \leq \epsilon + \xi_i, \ \forall i \in \Omega \\ & \sum\limits_{j=1}^{d} \omega_j x_{ij} + \beta - \tilde{l}_i \leq \epsilon + \xi_i^*, \ \forall i \in \Omega \\ & \xi_i, \xi_i^* \geq 0, \ \forall i \in \Omega, \ j = 1, \ldots, d. \end{array} \qquad (6)$$

Here $\sigma_j = \max\{0, \omega_j\}$ and $\tau_j = \max\{0, -\omega_j\}$, for $j = 1, \ldots, d$.
Moreover, when using the Hausdorff distance, d_H, the following convex quadratic problem

with linear constraints must be solved (see [3], Remark 4.3),

$$
\begin{array}{lll}
\min\limits_{\omega,\beta,\xi,\xi^*,\eta,\eta^*} & \frac{1}{2}\sum\limits_{j=1}^{d}\omega_j^2 + C\sum\limits_{i\in\Omega}(\xi_i+\xi_i^*+\eta_i+\eta_i^*) & \\
\text{s.t.} & \tilde{u}_i - \sum\limits_{j=1}^{d}\omega_j x_{ij} - \beta \leq \epsilon + \xi_i,\ \forall i \in \Omega & \\
& \sum\limits_{j=1}^{d}\omega_j x_{ij} + \beta - \tilde{u}_i \leq \epsilon + \xi_i^*,\ \forall i \in \Omega & (7)\\
& \tilde{l}_i - \sum\limits_{j=1}^{d}\omega_j x_{ij} - \beta \leq \epsilon + \eta_i,\ \forall i \in \Omega & \\
& \sum\limits_{j=1}^{d}\omega_j x_{ij} + \beta - \tilde{l}_i \leq \epsilon + \eta_i^*,\ \forall i \in \Omega & \\
& \xi_i, \xi_i^*, \eta_i, \eta_i^* \geq 0,\ \forall i \in \Omega,\ j = 1, \ldots, d. &
\end{array}
$$

The model proposed in this work can be considered as further research from our model in [3]. In this paper, the single-valued input and interval-valued output case is studied in more detail and we give a formulation allowing the use of kernel structures, which could not be introduced directly in the interval-valued input and output problem.

2.2 SVR for single-valued input and interval-valued output

In [3] (Remarks 3.2 and 4.3), one can find formulations (6)-(7) for the case in which the uncertainty only affects to the dependent variable Y_i, and the predictive variables are single-valued.
Given a new element, the predicted output is a singleton instead of an interval,

$$x \hookrightarrow f(x) := \omega^\top x + \beta. \tag{8}$$

However, if interval values are allowed as output, a better performance (in the sense of predicted intervals closer to the ones in the dataset) may be expected.
Hence, another model is proposed here, where two hyperplanes for approximating the lower and upper bounds of the dependent variable are used. Our aim will be to seek the pairs of coefficients (ω_L, β_L) and (ω_U, β_U) such that

$$d([\omega_L^\top x_i + \beta_L, \omega_U^\top x_i + \beta_U], [\tilde{l}_i, \tilde{u}_i]) \leq \epsilon, \forall i \in \Omega, \tag{9}$$

for a given distance measure defined on the set of intervals.
Given a new element x, the predicted interval output will be built as follows,

$$x \hookrightarrow f(x) := [\omega_L^\top x + \beta_L, \omega_U^\top x + \beta_U]. \tag{10}$$

When d is the Hausdorff distance (5), we obtain that constraint (9) is transformed into

$$\max\left\{|\tilde{u}_i - (\omega_U^\top x + \beta_U)|, |\tilde{l}_i - (\omega_L^\top x + \beta_L)|\right\} \leq \epsilon,\ \forall i \in \Omega, \tag{11}$$

and this is equivalent to say that

$$|\tilde{u}_i - (\omega_U^\top x + \beta_U)| \leq \epsilon, \quad \forall i \in \Omega \tag{12}$$

$$|\tilde{l}_i - (\omega_L^\top x + \beta_L)| \leq \epsilon, \quad \forall i \in \Omega. \tag{13}$$

A similar analysis can be done when d is the maximum distance (4), although we have chosen to study the model with the Hausdorff distance because the best results for the interval-valued input and output problem in [3] were obtained with this distance.
According to rule (10), we must also add that

$$\omega_L^\top x_i + \beta_L \leq \omega_U^\top x_i + \beta_U, \ \forall i \in \Omega. \tag{14}$$

With constraints (12)-(14), we can write the optimization problem where the sum of the squared Euclidean norms of ω_L and ω_U must be minimized (following the standard ϵ-SVR approach, [21]) as follows,

$$\begin{aligned}
\min_{\omega_L,\omega_U,\beta_L,\beta_U} \quad & \frac{1}{2}\sum_{j=1}^{d}(\omega_{Lj}^2 + \omega_{Uj}^2) \\
\text{s.t.} \quad & \tilde{u}_i - \sum_{j=1}^{d}\omega_{Uj}x_{ij} - \beta_U \leq \epsilon, \ \forall i \in \Omega \\
& \sum_{j=1}^{d}\omega_{Uj}x_{ij} + \beta_U - \tilde{u}_i \leq \epsilon, \ \forall i \in \Omega \\
& \tilde{l}_i - \sum_{j=1}^{d}\omega_{Lj}x_{ij} - \beta_L \leq \epsilon, \ \forall i \in \Omega \\
& \sum_{j=1}^{d}\omega_{Lj}x_{ij} + \beta_L - \tilde{l}_i \leq \epsilon, \ \forall i \in \Omega \\
& \sum_{j=1}^{d}\omega_{Lj}x_{ij} + \beta_L \leq \sum_{j=1}^{d}\omega_{Uj}x_{ij} + \beta_U, \ \forall i \in \Omega.
\end{aligned} \tag{15}$$

By adding slack variables, a Soft-Margin version, as the following convex quadratic formulation, is obtained,

$$
\begin{array}{rl}
\min\limits_{\omega_L,\omega_U,\beta_L,\beta_U,\xi,\xi^*,\eta,\eta^*} & \frac{1}{2}\sum_{j=1}^{d}(\omega_{Lj}^2+\omega_{Uj}^2)+C\sum_{i\in\Omega}(\xi_i+\xi_i^*+\eta_i+\eta_i^*) \\
\text{s.t.} & \tilde{u}_i-\sum_{j=1}^{d}\omega_{Uj}x_{ij}-\beta_U\le\epsilon+\xi_i,\ \forall i\in\Omega \\
& \sum_{j=1}^{d}\omega_{Uj}x_{ij}+\beta_U-\tilde{u}_i\le\epsilon+\xi_i^*,\ \forall i\in\Omega \\
& \tilde{l}_i-\sum_{j=1}^{d}\omega_{Lj}x_{ij}-\beta_L\le\epsilon+\eta_i,\ \forall i\in\Omega \\
& \sum_{j=1}^{d}\omega_{Lj}x_{ij}+\beta_L-\tilde{l}_i\le\epsilon+\eta_i^*,\ \forall i\in\Omega \\
& \sum_{j=1}^{d}\omega_{Lj}x_{ij}+\beta_L\le\sum_{j=1}^{d}\omega_{Uj}x_{ij}+\beta_U,\ \forall i\in\Omega \\
& \xi_i,\xi_i^*,\eta_i,\eta_i^*\ge 0,\ \forall i\in\Omega.
\end{array}
\tag{16}
$$

Before building the dual problem in the following section, which will help to derive an optimal solution of Problem (16), we consider the trivial case in which the objective function is equal to zero.

Proposition 2.1. *Problem (16) has optimal solution equal to zero iff the parameter $\epsilon > 0$ satisfies*

$$2\epsilon \ \ge\ \max_{i\in\Omega}\{\tilde{u}_i\}-\min_{i\in\Omega}\{\tilde{u}_i\} \tag{17}$$

$$2\epsilon \ \ge\ \max_{i\in\Omega}\{\tilde{l}_i\}-\min_{i\in\Omega}\{\tilde{l}_i\}. \tag{18}$$

Proof.
The objective function of Problem (16) is the sum of non-negative elements, thus, the value of the objective function must be greater than or equal to zero. Since our aim is to minimize that objective function, the minimum value which could be reached is zero. Then, we must study under which assumptions this situation is possible.
The objective function is equal to zero iff $\omega_{Uj} = \omega_{Lj} = 0$, $j = 1, \ldots, d$, and all the slack variables ξ_i, ξ_i^*, η_i, η_i^* are also zero. Then, the set of constraints of Problem (16) remains as follows,

$$
\begin{array}{l}
\left.\begin{array}{r}\tilde{u}_i-\beta_U\le\epsilon,\quad\forall i\in\Omega\\ \beta_U-\tilde{u}_i\le\epsilon,\quad\forall i\in\Omega\end{array}\right\}\ \leftrightarrow\ \tilde{u}_i-\epsilon\le\beta_U\le\tilde{u}_i+\epsilon,\ \forall i\in\Omega\\
\left.\begin{array}{r}\tilde{l}_i-\beta_L\le\epsilon,\quad\forall i\in\Omega\\ \beta_L-\tilde{l}_i\le\epsilon,\quad\forall i\in\Omega\end{array}\right\}\ \leftrightarrow\ \tilde{l}_i-\epsilon\le\beta_L\le\tilde{l}_i+\epsilon,\ \forall i\in\Omega\\
\qquad\beta_L\le\beta_U
\end{array}
$$

or equivalently,

$$\max_{i\in\Omega}\{\tilde{u}_i\} - \epsilon \quad \leq \beta_U \leq \quad \min_{i\in\Omega}\{\tilde{u}_i\} + \epsilon \tag{19}$$

$$\max_{i\in\Omega}\{\tilde{l}_i\} - \epsilon \quad \leq \beta_L \leq \quad \min_{i\in\Omega}\{\tilde{l}_i\} + \epsilon \tag{20}$$

$$\beta_L \leq \beta_U. \tag{21}$$

Observe that we can always choose $\beta_L \leq \beta_U$, since $\tilde{l}_i \leq \tilde{u}_i$, $\forall i \in \Omega$. For example, one can select $\beta_L = \max_{i\in\Omega}\{\tilde{l}_i\} - \epsilon \leq \max_{i\in\Omega}\{\tilde{u}_i\} - \epsilon = \beta_U$.

Then, the solution is feasible iff the intervals where β_U and β_L must be contained, according to expressions (19)-(20), are non-empty, that is,

$$\begin{aligned}
\max_{i\in\Omega}\{\tilde{u}_i\} - \epsilon \quad &\leq \quad \min_{i\in\Omega}\{\tilde{u}_i\} + \epsilon \\
\max_{i\in\Omega}\{\tilde{l}_i\} - \epsilon \quad &\leq \quad \min_{i\in\Omega}\{\tilde{l}_i\} + \epsilon,
\end{aligned}$$

and these constraints are equivalent to expressions (17)-(18). □

Remark 2.1. *Observe that if the optimal solution of Problem (16) is equal to zero, all the elements of the database will have the same predicted interval output,*

$$x \hookrightarrow f(x) := [\beta_L, \beta_U].$$

3 Building the dual problem

In this section, the dual formulation of Problem (16) is obtained. We first construct the dual program, obtaining an optimal solution of (16) in the standard input space. Later, we extend these results to the case where the data are mapped to a higher dimensional feature space.

3.1 Dual formulation

Below, we build the dual formulation of Problem (16). The dual program can also be used to obtain an optimal solution and allows us to introduce a kernel structure in the objective function.

Theorem 3.1. *Problem (16) has a finite optimal solution iff the following concave quadratic maximization problem has a finite optimal solution,*

$$\begin{array}{ll}
\max\limits_{\lambda,\lambda^*,\mu,\mu^*,\nu} & -\dfrac{1}{2}\sum\limits_{i,l\in\Omega}\left[(\lambda_i - \lambda_i^* + \nu_i)(\lambda_l - \lambda_l^* + \nu_l) + (\mu_i - \mu_i^* - \nu_i)(\mu_l - \mu_l^* - \nu_l)\right] x_i^\top x_l \\
& -\epsilon\sum\limits_{i\in\Omega}(\lambda_i + \lambda_i^* + \mu_i + \mu_i^*) + \sum\limits_{i\in\Omega}(\lambda_i - \lambda_i^*)\tilde{u}_i + \sum\limits_{i\in\Omega}(\mu_i - \mu_i^*)\tilde{l}_i \\
s.t. & \sum\limits_{i\in\Omega}(\lambda_i - \lambda_i^* + \nu_i) = 0 \\
& \sum\limits_{i\in\Omega}(\mu_i - \mu_i^* - \nu_i) = 0 \\
& 0 \leq \lambda_i, \lambda_i^*, \mu_i, \mu_i^* \leq C, \ \ \forall i \in \Omega \\
& 0 \leq \nu_i, \quad \forall i \in \Omega.
\end{array} \tag{22}$$

Proof.
Since Problem (16) is a linearly-constrained convex quadratic program, one has that it admits a finite optimal solution iff its dual has a finite optimal solution and, in that case, the two values coincide (see Section 6.6 in [1]). Then, the only thing to prove is that formulation (22) is the dual program of Problem (16).
To build the dual of Problem (16), firstly, we introduce nonnegative Lagrange multipliers for every constraint and we compute the Lagrangean function L in the primal and dual variables,

$$
\begin{aligned}
L \;=\; & \frac{1}{2}\sum_{j=1}^{d}(\omega_{Lj}^2+\omega_{Uj}^2)+C\sum_{i\in\Omega}(\xi_i+\xi_i^*+\eta_i+\eta_i^*) \\
& +\sum_{i\in\Omega}\lambda_i(\tilde{u}_i-\sum_{j=1}^{d}\omega_{Uj}x_{ij}-\beta_U-\epsilon-\xi_i)+\sum_{i\in\Omega}\lambda_i^*(\sum_{j=1}^{d}\omega_{Uj}x_{ij}+\beta_U-\tilde{u}_i-\epsilon-\xi_i^*) \\
& +\sum_{i\in\Omega}\mu_i(\tilde{l}_i-\sum_{j=1}^{d}\omega_{Lj}x_{ij}-\beta_L-\epsilon-\eta_i)+\sum_{i\in\Omega}\mu_i^*(\sum_{j=1}^{d}\omega_{Lj}x_{ij}+\beta_L-\tilde{l}_i-\epsilon-\eta_i^*) \\
& +\sum_{i\in\Omega}\nu_i(\sum_{j=1}^{d}\omega_{Lj}x_{ij}+\beta_L-\sum_{j=1}^{d}\omega_{Uj}x_{ij}-\beta_U)-\sum_{i\in\Omega}(\gamma_i\xi_i+\gamma_i^*\xi_i^*+\delta_i\eta_i+\delta_i^*\eta_i^*).
\end{aligned}
$$

The partial derivatives of the Lagrangean function are used to build the dual problem. We obtain the following constraints,

$$\frac{\partial L}{\partial \omega_{Uj}} = \omega_{Uj}-\sum_{i\in\Omega}\lambda_i x_{ij}+\sum_{i\in\Omega}\lambda_i^* x_{ij}-\sum_{i\in\Omega}\nu_i x_{ij}=0,\ j=1,\ldots,d \tag{23}$$

$$\frac{\partial L}{\partial \omega_{Lj}} = \omega_{Lj}-\sum_{i\in\Omega}\mu_i x_{ij}+\sum_{i\in\Omega}\mu_i^* x_{ij}+\sum_{i\in\Omega}\nu_i x_{ij}=0,\ j=1,\ldots,d \tag{24}$$

$$\frac{\partial L}{\partial \beta_U} = -\sum_{i\in\Omega}\lambda_i+\sum_{i\in\Omega}\lambda_i^*-\sum_{i\in\Omega}\nu_i=0 \tag{25}$$

$$\frac{\partial L}{\partial \beta_L} = -\sum_{i\in\Omega}\mu_i+\sum_{i\in\Omega}\mu_i^*+\sum_{i\in\Omega}\nu_i=0 \tag{26}$$

$$\frac{\partial L}{\partial \xi_i} = C-\lambda_i-\gamma_i=0,\ \forall i\in\Omega \tag{27}$$

$$\frac{\partial L}{\partial \xi_i^*} = C-\lambda_i^*-\gamma_i^*=0,\ \forall i\in\Omega \tag{28}$$

$$\frac{\partial L}{\partial \eta_i} = C-\mu_i-\delta_i=0,\ \forall i\in\Omega \tag{29}$$

$$\frac{\partial L}{\partial \eta_i^*} = C-\mu_i^*-\delta_i^*=0,\ \forall i\in\Omega. \tag{30}$$

From constraints (23)-(24), we derive an expression for computing ω_U, ω_L as a function of the Lagrangean multipliers,

$$\omega_U = \sum_{i\in\Omega}(\lambda_i - \lambda_i^* + \nu_i)x_i \tag{31}$$

$$\omega_L = \sum_{i\in\Omega}(\mu_i - \mu_i^* - \nu_i)x_i. \tag{32}$$

From constraints (25)-(26), we obtain the following constraints for the dual problem,

$$\sum_{i\in\Omega}(\lambda_i - \lambda_i^* + \nu_i) = 0 \tag{33}$$

$$\sum_{i\in\Omega}(\mu_i - \mu_i^* - \nu_i) = 0. \tag{34}$$

Moreover, from constraints (27)-(30), the nonnegative multipliers λ_i, λ_i^*, μ_i and μ_i^* are bounded by the parameter C, that is,

$$0 \le \lambda_i, \lambda_i^*, \mu_i, \mu_i^* \le C, \ \forall i \in \Omega. \tag{35}$$

By replacing the values of ω_L and ω_U, obtained in (31)-(32), in the Lagrangean function L and by using expressions (33)-(34) and the sets of constraints (27)-(30), the objective function for the dual problem can be rewritten as follows,

$$\begin{aligned}
\tilde{L}(\lambda,\lambda^*,\mu,\mu^*,\nu) &= -\frac{1}{2}\sum_{j=1}^{d}\left(\sum_{i\in\Omega}(\lambda_i - \lambda_i^* + \nu_i)x_{ij}\right)\left(\sum_{l\in\Omega}(\lambda_l - \lambda_l^* + \nu_l)x_{lj}\right) \\
&\quad -\frac{1}{2}\sum_{j=1}^{d}\left(\sum_{i\in\Omega}(\mu_i - \mu_i^* - \nu_i)x_{ij}\right)\left(\sum_{l\in\Omega}(\mu_l - \mu_l^* - \nu_l)x_{lj}\right) \\
&\quad -\epsilon\sum_{i\in\Omega}(\lambda_i + \lambda_i^* + \mu_i + \mu_i^*) + \sum_{i\in\Omega}(\lambda_i - \lambda_i^*)\tilde{u}_i + \sum_{i\in\Omega}(\mu_i - \mu_i^*)\tilde{l}_i \\
&= -\frac{1}{2}\sum_{i,l\in\Omega}\left[(\lambda_i - \lambda_i^* + \nu_i)(\lambda_l - \lambda_l^* + \nu_l) + (\mu_i - \mu_i^* - \nu_i)(\mu_l - \mu_l^* - \nu_l)\right] x_i^\top x_l \\
&\quad -\epsilon\sum_{i\in\Omega}(\lambda_i + \lambda_i^* + \mu_i + \mu_i^*) + \sum_{i\in\Omega}(\lambda_i - \lambda_i^*)\tilde{u}_i + \sum_{i\in\Omega}(\mu_i - \mu_i^*)\tilde{l}_i.
\end{aligned} \tag{36}$$

Then, by maximizing the objective function (36) and by adding constraints (33)-(35), along with the non-negativity of the variables ν_i, we obtain the dual formulation (22), which is a concave quadratic maximization problem in the variables λ, λ^*, μ, μ^* and ν.

□

3.2 Reconstruction of an optimal solution for the primal problem

In the following, we show how to construct an optimal solution of the primal problem (16), given an optimal solution of the dual problem (22).

From expressions (31)-(32), we obtain the values for ω_U and ω_L. For the remaining variables (β_L, β_U, and the slack variables), we use that the following Karush-Kuhn-Tucker conditions must be satisfied (see Section 4.3 in [1]),

$$\lambda_i \cdot (\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \beta_U - \epsilon - \xi_i) = 0, \quad \forall i \in \Omega \tag{37}$$

$$\lambda_i^* \cdot (\sum_{j=1}^{d} \omega_{Uj} x_{ij} + \beta_U - \tilde{u}_i - \epsilon - \xi_i^*) = 0, \quad \forall i \in \Omega \tag{38}$$

$$\mu_i \cdot (\tilde{l}_i - \sum_{j=1}^{d} \omega_{Lj} x_{ij} - \beta_L - \epsilon - \eta_i) = 0, \quad \forall i \in \Omega \tag{39}$$

$$\mu_i^* \cdot (\sum_{j=1}^{d} \omega_{Lj} x_{ij} + \beta_L - \tilde{l}_i - \epsilon - \eta_i^*) = 0, \quad \forall i \in \Omega \tag{40}$$

$$\nu_i \cdot (\sum_{j=1}^{d} \omega_{Lj} x_{ij} + \beta_L - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \beta_U) = 0, \quad \forall i \in \Omega \tag{41}$$

$$\xi_i \cdot (C - \lambda_i) = 0, \quad \forall i \in \Omega \tag{42}$$

$$\xi_i^* \cdot (C - \lambda_i^*) = 0, \quad \forall i \in \Omega \tag{43}$$

$$\eta_i \cdot (C - \mu_i) = 0, \quad \forall i \in \Omega \tag{44}$$

$$\eta_i^* \cdot (C - \mu_i^*) = 0, \quad \forall i \in \Omega \tag{45}$$

$$0 \leq \lambda_i, \lambda_i^*, \mu_i, \mu_i^* \leq C, \quad \forall i \in \Omega. \tag{46}$$

Lemma 3.1. *Given $(\lambda, \lambda^*, \mu, \mu^*, \nu)$ a solution of the Karush-Kuhn-Tucker system (37)-(46), one has that*

$$\lambda_i \cdot \lambda_i^* = 0, \quad \forall i \in \Omega \tag{47}$$

$$\mu_i \cdot \mu_i^* = 0, \quad \forall i \in \Omega. \tag{48}$$

Proof.
Suppose that there exists $i_0 \in \Omega$ such that $\lambda_{i_0}, \lambda_{i_0}^* > 0$. According to (37)-(38), the two corresponding constraints would become active and, for any $\epsilon > 0$, the following equalities would be satisfied simultaneously,

$$\begin{aligned} \tilde{u}_{i_0} - \sum_{j=1}^{d} \omega_{Uj} x_{i_0 j} - \beta_U &= \epsilon + \xi_{i_0} > 0 \\ \tilde{u}_{i_0} - \sum_{j=1}^{d} \omega_{Uj} x_{i_0 j} - \beta_U &= -\epsilon - \xi_{i_0}^* < 0, \end{aligned}$$

which is a contradiction. Hence, $\lambda_i \cdot \lambda_i^* = 0$, $\forall i \in \Omega$.
With a similar reasoning, one can show that $\mu_i \cdot \mu_i^* = 0$, $\forall i \in \Omega$. □

Remark 3.1. *The relative position of points with respect to the ϵ-insensitive tube around $H_U : \omega_U^\top x + \beta_U$ is obtained from the values of the dual variables λ, λ^*. Indeed, let $(\lambda, \lambda^*, \mu, \mu^*, \nu)$ be a solution of the Karush-Kuhn-Tucker system (37)-(46) and let $(\omega_L, \omega_U, \beta_L, \beta_U, \xi, \xi^*, \eta, \eta^*)$ be optimal for (16). By (47), there are five possible relative positions of points w.r.t. the ϵ-insensitive tube around the hyperplane H_U (see Figure 1), which are studied separately.*

1. *Above the tube: if $\lambda_i = C$, by (37) and since $\xi_i \geq 0$, one has that*

$$\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \beta_U = \epsilon + \xi_i \geq \epsilon. \tag{49}$$

2. *On the upper boundary of the tube: if $0 < \lambda_i < C$ then, by (42), $\xi_i = 0$, and by (37),*

$$\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \beta_U = \epsilon. \tag{50}$$

3. *Inside the tube: if $\lambda_i = \lambda_i^* = 0$, by (42)-(43), the slack variables are also zero, $\xi_i = \xi_i^* = 0$, and, since $(\omega_L, \omega_U, \beta_L, \beta_U, \xi, \xi^*, \eta, \eta^*)$ is feasible for (16),*

$$\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \beta_U \leq \epsilon \tag{51}$$

$$\sum_{j=1}^{d} \omega_{Uj} x_{ij} + \beta_U - \tilde{u}_i \leq \epsilon. \tag{52}$$

4. *On the lower boundary of the tube: if $0 < \lambda_i^* < C$ then, by (43), $\xi_i^* = 0$, and by (38),*

$$\sum_{j=1}^{d} \omega_{Uj} x_{ij} + \beta_U - \tilde{u}_i = \epsilon. \tag{53}$$

5. *Below the tube: if $\lambda_i^* = C$, by (38) and since $\xi_i^* \geq 0$,*

$$\sum_{j=1}^{d} \omega_{Uj} x_{ij} + \beta_U - \tilde{u}_i = \epsilon + \xi_i^* \geq \epsilon. \tag{54}$$

In the following remark, the same reasoning is repeated for the hyperplane $H_L : \omega_L^\top x + \beta_L$.

Remark 3.2. *Similarly, μ, μ^* determine the relative position of points w.r.t. the ϵ-insensititve tube. Let $(\lambda, \lambda^*, \mu, \mu^*, \nu)$ be a solution of the Karush-Kuhn-Tucker system (37)-(46) and let $(\omega_L, \omega_U, \beta_L, \beta_U, \xi, \xi^*, \eta, \eta^*)$ be optimal for (16). As done for $\tilde{u}_i$ in Remark 3.1, the position of the lower bounds $\tilde{l}_i$ of the elements of the database Ω with respect to the hyperplane $H_L : \omega_L^\top x + \beta_L$ can be derived from the values of the dual variables μ, μ^*. By (48), one has that μ_i or μ_i^* must be zero. Then, there are five types of points concerning their situation with respect to the ϵ-insensitive tube around H_L (see Figure 2):*

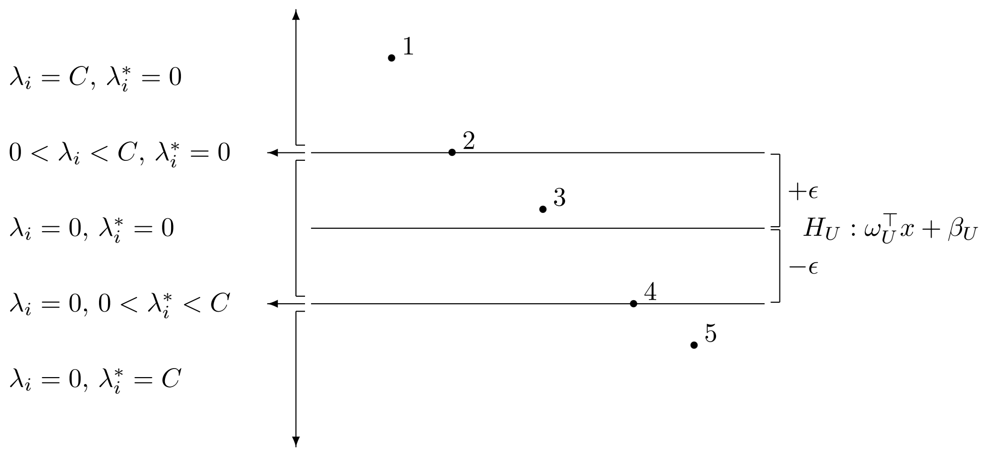

Figure 1: Geometrical idea of the position of $\tilde{u}_i$ according to the value of λ_i and λ_i^*

1. *Above the tube: if $\mu_i = C$, by (39) and since $\eta_i \geq 0$,*

$$\tilde{l}_i - \sum_{j=1}^{d} \omega_{Lj} x_{ij} - \beta_L = \epsilon + \eta_i \geq \epsilon. \tag{55}$$

2. *On the upper boundary of the tube: if $0 < \mu_i < C$ then, by (44), $\eta_i = 0$, and by (39),*

$$\tilde{l}_i - \sum_{j=1}^{d} \omega_{Lj} x_{ij} - \beta_L = \epsilon. \tag{56}$$

3. *Inside the tube: if $\mu_i = \mu_i^* = 0$, by (44)-(45), the slack variables are also zero, $\eta_i = \eta_i^* = 0$, and then*

$$\tilde{l}_i - \sum_{j=1}^{d} \omega_{Lj} x_{ij} - \beta_L \leq \epsilon \tag{57}$$

$$\sum_{j=1}^{d} \omega_{Lj} x_{ij} + \beta_L - \tilde{l}_i \leq \epsilon. \tag{58}$$

4. *On the lower boundary of the tube: if $0 < \mu_i^* < C$ then, by (45), $\eta_i^* = 0$, and by (40),*

$$\sum_{j=1}^{d} \omega_{Lj} x_{ij} + \beta_L - \tilde{l}_i = \epsilon. \tag{59}$$

5. *Below the tube: if $\mu_i^* = C$, by (40) and since $\eta_i^* \geq 0$,*

$$\sum_{j=1}^{d} \omega_{Lj} x_{ij} + \beta_L - \tilde{l}_i = \epsilon + \mu_i^* \geq \epsilon. \tag{60}$$

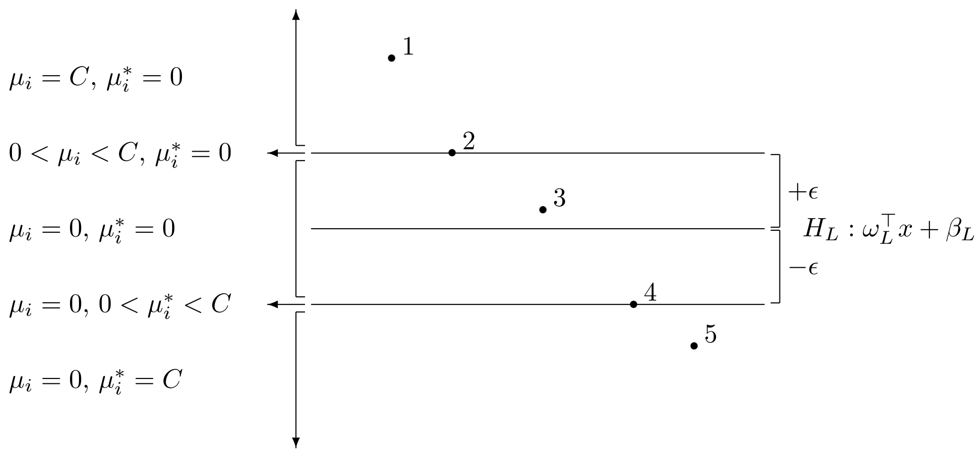

Figure 2: Geometrical idea of the position of $\tilde{l}_i$ according to the value of μ_i and μ_i^*

Theorem 3.2. *Let $(\lambda, \lambda^*, \mu, \mu^*, \nu)$ be a solution of the Karush-Kuhn-Tucker system (37)-(46) and let $(\omega_L, \omega_U, \beta_L, \beta_U, \xi, \xi^*, \eta, \eta^*)$ be optimal for (16). Then:*

1. *If there exists $i_0 \in \Omega$ such that $0 < \lambda_{i_0} < C$, then there exists a unique optimal value of β_U for Problem (16),*

$$\beta_U = \tilde{u}_{i_0} - \sum_{j=1}^{d} \omega_{Uj} x_{i_0 j} - \epsilon. \tag{61}$$

Analogously, if there exists $i_1 \in \Omega$ such that $0 < \lambda_{i_1}^* < C$, one has a unique optimal value of β_U for Problem (16),

$$\beta_U = \tilde{u}_{i_1} - \sum_{j=1}^{d} \omega_{Uj} x_{i_1 j} + \epsilon. \tag{62}$$

Otherwise, if $\lambda_i, \lambda_i^* \in \{0, C\}$, $\forall i \in \Omega$, the set of solutions for β_U is the following interval

$$\beta_U \in \Big[\max\{\max_{\{i:\ \lambda_i+\lambda_i^*=0\}} (\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij}) - \epsilon, \max_{\{i:\ \lambda_i^*=C\}} (\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij}) + \epsilon\},$$
$$\min\{\min_{\{i:\ \lambda_i+\lambda_i^*=0\}} (\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij}) + \epsilon, \min_{\{i:\ \lambda_i=C\}} (\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij}) - \epsilon\}\Big]. \tag{63}$$

2. *If there exists $i_0 \in \Omega$ with $0 < \mu_{i_0} < C$, then there exists a unique optimal solution for β_L in Problem (16),*

$$\beta_L = \tilde{l}_{i_0} - \sum_{j=1}^{d} \omega_{Lj} x_{i_0 j} - \epsilon. \tag{64}$$

Analogously, if there exists $i_1 \in \Omega$ with $0 < \mu^*_{i_1} < C$, one has a unique optimal solution for β_L in Problem (16),

$$\beta_L = \tilde{l}_{i_1} - \sum_{j=1}^{d} \omega_{Lj} x_{i_1 j} + \epsilon. \tag{65}$$

Otherwise, if $\mu_i, \mu^*_i \in \{0, C\}$, $\forall i \in \Omega$, the set of solutions for β_L is the following interval

$$\beta_L \in \big[\max\{\max_{\{i:\ \mu_i+\mu_i^*=0\}} (\tilde{l}_i - \sum_{j=1}^{d} \omega_{Lj} x_{ij}) - \epsilon, \max_{\{i:\ \mu_i^*=C\}} (\tilde{l}_i - \sum_{j=1}^{d} \omega_{Lj} x_{ij}) + \epsilon\},$$
$$\min\{\min_{\{i:\ \mu_i+\mu_i^*=0\}} (\tilde{l}_i - \sum_{j=1}^{d} \omega_{Lj} x_{ij}) + \epsilon, \min_{\{i:\ \mu_i=C\}} (\tilde{l}_i - \sum_{j=1}^{d} \omega_{Lj} x_{ij}) - \epsilon\}\big]. \tag{66}$$

Proof.
By Lemma 3.1, for any $i \in \Omega$, either λ_i or λ^*_i are zero (and the same for μ_i and μ^*_i).

1. The cases $0 < \lambda_{i_0} < C$ or $0 < \lambda_{i_1} < C$ for some i_0, i_1, are analyzed in Remark 3.1 (cases 2 and 4).

   Now, we analyze the case in which $\lambda_i, \lambda^*_i \in \{0, C\}$, $\forall i \in \Omega$. For every i such that $\lambda_i = 0$, we have that either $\lambda^*_i = C$, and in such a case, by Remark 3.1, case 5,

$$\beta_U \ \geq \ \tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} + \epsilon, \tag{67}$$

   or $\lambda^*_i = 0$, and thus, by Remark 3.1, case 3,

$$\beta_U \ \geq \ \tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \epsilon \tag{68}$$
$$\beta_U \ \leq \ \tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} + \epsilon. \tag{69}$$

   Moreover, for every i such that $\lambda_i = C$, we have by Remark 3.1, case 1, that

$$\beta_U \ \leq \ \tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \epsilon. \tag{70}$$

   Hence,

$$\beta_U \ \geq \ \max\{\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \epsilon :\ \lambda_i = 0, \lambda^*_i = 0\}$$

and

$$\beta_U \quad \geq \quad \max\{\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} + \epsilon : \ \lambda_i = 0, \lambda_i^* = C\}.$$

This leads to

$$\beta_U \geq \max\{ \max_{\{i:\ \lambda_i+\lambda_i^*=0\}} (\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij}) - \epsilon, \max_{\{i:\ \lambda_i^*=C\}} (\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij}) + \epsilon\}. \quad (71)$$

Analogously, we obtain the upper bound for the interval of solutions for β_U as

$$\beta_U \leq \min\{ \min_{\{i:\ \lambda_i+\lambda_i^*=0\}} (\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij}) + \epsilon, \min_{\{i:\ \lambda_i=C\}} (\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij}) - \epsilon\}. \quad (72)$$

2. With an analogous reasoning on the sets of variables μ_i, μ_i^* to that used with variables λ_i and λ_i^*, we obtain expressions (64)-(66).

□

Lemma 3.2. *Let $(\lambda, \lambda^*, \mu, \mu^*, \nu)$ be optimal for the dual problem (22). Then, for any k such that $\nu_k > 0$, one has that*

$$\sum_{i\in\Omega} (\mu_i - \mu_i^* - \lambda_i + \lambda_i^* - 2\nu_i) x_i^\top x_k = \delta, \quad (73)$$

where δ is a constant.

Proof.
If there exists k such that $\nu_k > 0$, then, by (41), one has that

$$\sum_{j=1}^{d} \omega_{Lj} x_{kj} + \beta_L = \sum_{j=1}^{d} \omega_{Uj} x_{kj} + \beta_U. \quad (74)$$

By using the expressions for ω_U and ω_L, given in (31)-(32), we obtain that

$$\sum_{i\in\Omega} (\mu_i - \mu_i^* - \nu_i) x_i^\top x_k + \beta_L = \sum_{i\in\Omega} (\lambda_i - \lambda_i^* + \nu_i) x_i^\top x_k + \beta_U, \quad (75)$$

and this leads to

$$\sum_{i\in\Omega} (\mu_i - \mu_i^* - \lambda_i + \lambda_i^* - 2\nu_i) x_i^\top x_k = \beta_U - \beta_L = \delta. \quad (76)$$

□

Theorem 3.3. *Let $(\lambda, \lambda^*, \mu, \mu^*, \nu)$ be optimal for the dual problem (22), such that λ_i, λ_i^*, μ_i, $\mu_i^* \in \{0, C\}$. Let ω_U, ω_L be defined by (31)-(32). One has that:*

1. *If $\nu_i = 0$, for all $i \in \Omega$, then any β_U, β_L satisfying (63) and (66), respectively, are such that an optimal solution exists for Problem (16) of the form $(\omega_L, \omega_U, \beta_L, \beta_U, \xi, \xi^*, \eta, \eta^*)$.*

2. *If there exists $k \in \Omega$ such that $\nu_k > 0$, then any β_U, β_L satisfying (63) and (66), respectively, and also satisfying*

$$\beta_U - \beta_L = \delta, \tag{77}$$

with δ the constant defined in (73), are such that an optimal solution exists for Problem (16) of the form $(\omega_L, \omega_U, \beta_L, \beta_U, \xi, \xi^*, \eta, \eta^*)$.

Proof.
We have to show that for ω_U, ω_L defined in (31)-(32) and for any β_U, β_L in the intervals (63) and (66), there exist ξ, ξ^*, η, η^* such that the pairs $(\omega_L, \omega_U, \beta_L, \beta_U, \xi, \xi^*, \eta, \eta^*)$ and $(\lambda, \lambda^*, \mu, \mu^*, \nu)$ jointly satisfy the KKT system (37)-(46).
If $\nu_i = 0$, for all $i \in \Omega$, then (41) is automatically satisfied, else, by construction of β_U and β_L, they satisfy (77), and thus (41) also holds.
Hence, we only need to show that for any β_U and β_L in the intervals (63) and (66), ξ, ξ^*, η, η^* exist satisfying (37)-(40) and (42)-(45).
Consider β_U in the interval (63), this means that

$$\tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \epsilon \leq \beta_U \leq \tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} + \epsilon, \quad \forall i \in \Omega: \ \lambda_i = \lambda_i^* = 0 \tag{78}$$

$$\beta_U \geq \tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} + \epsilon, \qquad \forall i \in \Omega: \ \lambda_i = 0, \lambda_i^* = C \tag{79}$$

$$\beta_U \leq \tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \epsilon, \qquad \forall i \in \Omega: \ \lambda_i = C, \lambda_i^* = 0. \tag{80}$$

We check that constraints (37)-(38) and (42)-(43) are satisfied, by considering the possible values for λ_i and λ_i^* (by taking into account (47)).

- For every $i \in \Omega$ such that $\lambda_i = \lambda_i^* = 0$, (37)-(38) are trivially satisfied and (42)-(43) are satisfied for $\xi_i = \xi_i^* = 0$.

- For every $i \in \Omega$ such that $\lambda_i = 0$ and $\lambda_i^* = C$, (37) is trivially satisfied, and (42) is satisfied with $\xi_i = 0$.

  By (38), we compute the value of ξ_i^*,

$$\xi_i^* = \sum_{j=1}^{d} \omega_{Uj} x_{ij} + \beta_U - \tilde{u}_i - \epsilon, \tag{81}$$

  which is non-negative by expression (79).

- For every $i \in \Omega$ such that $\lambda_i = C$ and $\lambda_i^* = 0$, (38) is trivially satisfied, and (43) is satisfied with $\xi_i^* = 0$.

  By (37), we obtain the value ξ_i as

$$\xi_i = \tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \beta_U - \epsilon, \tag{82}$$

  which is bigger than or equal to zero by expression (80).

A similar reasoning can be done with β_L to check that (39)-(40) and (44)-(45) are satisfied. □

Corollary 3.1. *Let $(\lambda, \lambda^*, \mu, \mu^*, \nu)$ be a solution of the Karush-Kuhn-Tucker system (37)-(46) and let $(\omega_L, \omega_U, \beta_L, \beta_U, \xi, \xi^*, \eta, \eta^*)$ be optimal for (16). Then:*

1. *If $\lambda_i = C$, then*

$$\xi_i = \tilde{u}_i - \sum_{j=1}^{d} \omega_{Uj} x_{ij} - \beta_U - \epsilon. \tag{83}$$

2. *If $\lambda_i^* = C$, then*

$$\xi_i^* = \sum_{j=1}^{d} \omega_{Uj} x_{ij} + \beta_U - \tilde{u}_i - \epsilon. \tag{84}$$

3. *If $\mu_i = C$, then*

$$\eta_i = \tilde{l}_i - \sum_{j=1}^{d} \omega_{Lj} x_{ij} - \beta_L - \epsilon. \tag{85}$$

4. *If $\mu_i^* = C$, then*

$$\eta_i^* = \sum_{j=1}^{d} \omega_{Lj} x_{ij} + \beta_L - \tilde{l}_i - \epsilon. \tag{86}$$

In the following result, the uniqueness of solution for Problem (16), under conditions of Proposition 2.1, is studied.

Proposition 3.1. *Let $\epsilon > 0$ be a parameter in Problem (16) satisfying*

$$\epsilon \geq \frac{1}{2} \max \left\{ \max_{i\in\Omega}\{\tilde{u}_i\} - \min_{i\in\Omega}\{\tilde{u}_i\}, \max_{i\in\Omega}\{\tilde{l}_i\} - \min_{i\in\Omega}\{\tilde{l}_i\} \right\}. \tag{87}$$

Then, Problem (16) has a unique optimal solution iff

$$\epsilon = \frac{1}{2}(\max_{i\in\Omega}\{\tilde{u}_i\} - \min_{i\in\Omega}\{\tilde{u}_i\}) = \frac{1}{2}(\max_{i\in\Omega}\{\tilde{l}_i\} - \min_{i\in\Omega}\{\tilde{l}_i\}). \tag{88}$$

Proof.
When $\epsilon > 0$ satisfies (87), by Proposition 2.1 one has that the optimal solution of Problem (16) satisfies that $\omega_{Uj} = \omega_{Lj} = 0$, $j = 1, \ldots, d$, and the slack variables ξ_i, ξ_i^*, η_i, η_i^* are also zero. The constraints of Problem (16) are reduced to

$$\max_{i\in\Omega}\{\tilde{u}_i\} - \epsilon \quad \leq \beta_U \leq \quad \min_{i\in\Omega}\{\tilde{u}_i\} + \epsilon \tag{89}$$

$$\max_{i\in\Omega}\{\tilde{l}_i\} - \epsilon \quad \leq \beta_L \leq \quad \min_{i\in\Omega}\{\tilde{l}_i\} + \epsilon \tag{90}$$

$$\beta_L \leq \beta_U. \tag{91}$$

Then, the uniqueness must be studied only in the variables β_L and β_U.

In one direction, if expression (88) is satisfied, constraints (89)-(91) are transformed into the following expressions,

$$\begin{aligned} \frac{1}{2}(\min_{i\in\Omega}\{\tilde{u}_i\} + \max_{i\in\Omega}\{\tilde{u}_i\}) \quad &\leq \beta_U \leq \quad \frac{1}{2}(\min_{i\in\Omega}\{\tilde{u}_i\} + \max_{i\in\Omega}\{\tilde{u}_i\}) \\ \frac{1}{2}(\min_{i\in\Omega}\{\tilde{l}_i\} + \max_{i\in\Omega}\{\tilde{l}_i\}) \quad &\leq \beta_L \leq \quad \frac{1}{2}(\min_{i\in\Omega}\{\tilde{l}_i\} + \max_{i\in\Omega}\{\tilde{l}_i\}) \\ &\beta_L \leq \beta_U. \end{aligned}$$

Denoting by $u = \frac{1}{2}(\min_{i\in\Omega}\{\tilde{u}_i\} + \max_{i\in\Omega}\{\tilde{u}_i\})$ and by $l = \frac{1}{2}(\min_{i\in\Omega}\{\tilde{l}_i\} + \max_{i\in\Omega}\{\tilde{l}_i\})$, the bounds of these degenerate intervals, one has that the unique possible solution is $\beta_U = u$ and $\beta_L = l$, since it is clear that $l \leq u$. Then, Problem (16) admits a unique optimal solution.

In the other direction, we denote by $\epsilon_1 = \frac{1}{2}(\max_{i\in\Omega}\{\tilde{u}_i\} - \min_{i\in\Omega}\{\tilde{u}_i\})$ and by $\epsilon_2 = \frac{1}{2}(\max_{i\in\Omega}\{\tilde{l}_i\} - \min_{i\in\Omega}\{\tilde{l}_i\})$. We prove that if ϵ is strictly bigger than ϵ_1 or ϵ_2, then the solution is not unique. We consider two different cases, when $\epsilon_1 = \epsilon_2$ and when $\epsilon_1 \neq \epsilon_2$.

- Case 1: $\epsilon_1 = \epsilon_2$.

  Suppose that $\epsilon > \epsilon_1$, that is, there exists an amount $\delta > 0$, such that $\epsilon = \epsilon_1 + \delta = \epsilon_2 + \delta$. Then, replacing ϵ in constraints (89)-(91), we obtain

  $$\begin{aligned} u - \delta \quad &\leq \beta_U \leq \quad u + \delta \\ l - \delta \quad &\leq \beta_L \leq \quad l + \delta \\ &\beta_L \leq \beta_U. \end{aligned}$$

  Since $\delta > 0$, non-degenerate intervals can be found for β_L and β_U satisfying this set of constraints, for example, all the solutions such $\beta_L \in [l - \delta, l]$ and $\beta_U \in [u, u + \delta]$.

- Case 2: $\epsilon_1 \neq \epsilon_2$.

  We study, for instance, the case $\epsilon_1 < \epsilon_2$ (the opposite is analogous), that is, there exists an amount $\delta_1 > 0$ such that $\epsilon_2 = \epsilon_1 + \delta_1$. Suppose that $\epsilon \geq \epsilon_2$, that is, there exists an amount $\delta_2 \geq 0$ such that $\epsilon = \epsilon_2 + \delta_2$, or analogously, $\epsilon = \epsilon_1 + \delta_1 + \delta_2$.

Then, replacing ϵ in constraints (89)-(91), one obtains that

$$\begin{array}{rcl} u-\delta_1-\delta_2 & \leq \beta_U \leq & u+\delta_1+\delta_2 \\ l-\delta_2 & \leq \beta_L \leq & l+\delta_2 \\ & \beta_L \leq \beta_U. & \end{array}$$

Then, all the solutions in the following intervals, $\beta_L \in [l-\delta_2, l]$ and $\beta_U \in [u, u+\delta_1+\delta_2]$, satisfy this set of constraints, and, since $\delta_1 > 0$, at least the interval for β_U is non-degenerate.

□

4 Kernel-based dual formulation

Kernels are used to project the data from the input space $\mathcal{X} \subseteq \mathbb{R}^d$ into a high dimensional feature space, where more abstract features of the data can be exploited. This way, since the data are projected to the feature space via a usually non-linear mapping, non-linear relations between the data can be extracted by means of a linear regressor (see [6, 12, 20]).

Definition 4.1. *Let $\phi : \mathcal{X} \subseteq \mathbb{R}^d \to \mathcal{F}$ be a (usually non-linear) mapping from the input space $\mathcal{X}$ into the feature space $\mathcal{F}$. A kernel K is a function $K : \mathcal{X} \times \mathcal{X} \subseteq \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, such that for every $x, y \in \mathcal{X}$,*

$$K(x,y) = \phi(x)^\top \phi(y).$$

From Problem (22) and by replacing $x_i^\top x_l$ by another general kernel structure $K(x_i, x_l)$, we can rewrite the dual of Problem (16) as the following concave quadratic maximization problem,

$$\begin{array}{ll} \max\limits_{\lambda,\lambda^*,\mu,\mu^*,\nu} & -\dfrac{1}{2}\sum\limits_{i,l\in\Omega} \left[(\lambda_i-\lambda_i^*+\nu_i)(\lambda_l-\lambda_l^*+\nu_l)+(\mu_i-\mu_i^*-\nu_i)(\mu_l-\mu_l^*-\nu_l)\right] K(x_i,x_l) \\ & -\epsilon\sum\limits_{i\in\Omega}(\lambda_i+\lambda_i^*+\mu_i+\mu_i^*)+\sum\limits_{i\in\Omega}(\lambda_i-\lambda_i^*)\tilde{u}_i+\sum\limits_{i\in\Omega}(\mu_i-\mu_i^*)\tilde{l}_i \\ \text{s.t.} & \sum\limits_{i\in\Omega}(\lambda_i-\lambda_i^*+\nu_i)=0 \\ & \sum\limits_{i\in\Omega}(\mu_i-\mu_i^*-\nu_i)=0 \\ & 0\leq \lambda_i,\lambda_i^*,\mu_i,\mu_i^*\leq C,\ \forall i\in\Omega \\ & 0\leq \nu_i, \quad \forall i\in\Omega. \end{array} \tag{92}$$

When a general kernel structure has been introduced in the problem, explicit expressions of ω_U and ω_L cannot be computed, because one has that

$$\omega_U = \sum_{i\in\Omega}(\lambda_i-\lambda_i^*+\nu_i)\phi(x_i) \tag{93}$$

$$\omega_L = \sum_{i\in\Omega}(\mu_i-\mu_i^*-\nu_i)\phi(x_i) \tag{94}$$

and the mapping ϕ is unknown in general. However, given a new element x, we can always obtain the predicted interval output, since the kernel is known.
Expressions for β_U and β_L are obtained from expressions (61)-(66) by only replacing $x_l^\top x_i$ by the corresponding kernel value. Thus, when $\exists i_0 \in \Omega : 0 < \lambda_{i_0} < C$, β_U can be built by using expression (61), as follows,

$$\begin{aligned} \beta_U &= \tilde{u}_{i_0} - \sum_{l\in\Omega}(\lambda_l - \lambda_l^* + \nu_l)\phi(x_l)^\top\phi(x_{i_0}) - \epsilon \\ &= \tilde{u}_{i_0} - \sum_{l\in\Omega}(\lambda_l - \lambda_l^* + \nu_l)K(x_l, x_{i_0}) - \epsilon. \end{aligned} \quad (95)$$

If $\exists i_1 \in \Omega : 0 < \lambda_{i_1}^* < C$, the value of β_U is obtained from (62),

$$\beta_U = \tilde{u}_{i_1} - \sum_{l\in\Omega}(\lambda_l - \lambda_l^* + \nu_l)K(x_l, x_{i_1}) + \epsilon, \quad (96)$$

and, if $\lambda_i, \lambda_i^* \in \{0, C\}$, $\forall i \in \Omega$, from (63), β_U belongs to

$$\begin{aligned} &\max\{\max_{\{i:\lambda_i+\lambda_i^*=0\}} \tilde{u}_i - \sum_{l\in\Omega}(\lambda_l - \lambda_l^* + \nu_l)K(x_l, x_i) - \epsilon, \max_{\{i:\lambda_i^*=C\}} \tilde{u}_i - \sum_{l\in\Omega}(\lambda_l - \lambda_l^* + \nu_l)K(x_l, x_i) + \epsilon\} \leq \beta_U \\ &\leq \min\{\min_{\{i:\lambda_i+\lambda_i^*=0\}} \tilde{u}_i - \sum_{l\in\Omega}(\lambda_l - \lambda_l^* + \nu_l)K(x_l, x_i) + \epsilon, \min_{\{i:\lambda_i=C\}} \tilde{u}_i - \sum_{l\in\Omega}(\lambda_l - \lambda_l^* + \nu_l)K(x_l, x_i) - \epsilon\}. \end{aligned}$$

Analogous expressions are derived for β_L from expressions (64)-(66). If $\exists i_0 \in \Omega : 0 < \mu_{i_0} < C$ or $\exists i_1 \in \Omega : 0 < \mu_{i_1}^* < C$, we obtain, from expressions (64) and (65), respectively, the following expressions,

$$\beta_L = \tilde{l}_{i_0} - \sum_{l\in\Omega}(\mu_l - \mu_l^* - \nu_l)K(x_l, x_{i_0}) - \epsilon \quad (97)$$

$$\beta_L = \tilde{l}_{i_1} - \sum_{l\in\Omega}(\mu_l - \mu_l^* - \nu_l)K(x_l, x_{i_1}) + \epsilon, \quad (98)$$

(99)

and if $\mu_i, \mu_i^* \in \{0, C\}$, $\forall i \in \Omega$, one has that, from (66), β_L satisfies

$$\begin{aligned} &\max\{\max_{\{i:\mu_i+\mu_i^*=0\}} \tilde{l}_i - \sum_{l\in\Omega}(\mu_l - \mu_l^* - \nu_l)K(x_l, x_i) - \epsilon, \max_{\{i:\mu_i^*=C\}} \tilde{l}_i - \sum_{l\in\Omega}(\mu_l - \mu_l^* - \nu_l)K(x_l, x_i) + \epsilon\} \leq \beta_L \\ &\leq \min\{\min_{\{i:\mu_i+\mu_i^*=0\}} \tilde{l}_i - \sum_{l\in\Omega}(\mu_l - \mu_l^* - \nu_l)K(x_l, x_i) + \epsilon, \min_{\{i:\mu_i=C\}} \tilde{l}_i - \sum_{l\in\Omega}(\mu_l - \mu_l^* - \nu_l)K(x_l, x_i) - \epsilon\}. \end{aligned}$$

Finally, given a new element x, the predicted interval is the following,

$$\begin{aligned} x \hookrightarrow f(x) &:= [\omega_L^\top x + \beta_L, \omega_U^\top x + \beta_U] \\ &= [\sum_{i\in\Omega}(\mu_i - \mu_i^* - \nu_i)K(x_i, x) + \beta_L, \sum_{i\in\Omega}(\lambda_i - \lambda_i^* + \nu_i)K(x_i, x) + \beta_U] \end{aligned} \quad (100)$$

where the only thing that changes with respect to the primal problem is that the scalar products $\omega_L^\top x$ and $\omega_U^\top x$ have been replaced by the corresponding kernel values to avoid using explicit expressions of ω_L and ω_U, which could not be computed, according to (93)-(94).

5 Computational experiments

5.1 Error measures

For the numerical experiments, several measures have been used to study the fitness of the model. These measurements are the *lower* and *upper bound root mean-squared error*, $RMSE_l$ and $RMSE_u$, and the *mean Hausdorff distance*, $\overline{d}_H$, between the predicted and the real interval outputs, defined as follows,

$$RMSE_l = \sqrt{\frac{1}{n}\sum_{i\in\Omega}(\tilde{l}_i - \hat{l}_i)^2} \quad (101)$$

$$RMSE_u = \sqrt{\frac{1}{n}\sum_{i\in\Omega}(\tilde{u}_i - \hat{u}_i)^2} \quad (102)$$

$$\overline{d}_H = \frac{1}{n}\sum_{i\in\Omega}\max\{|\tilde{l}_i - \hat{l}_i|, |\tilde{u}_i - \hat{u}_i|\}, \quad (103)$$

where $\tilde{l} = (\tilde{l}_1, \dots, \tilde{l}_n)^\top$, $\hat{l} = (\hat{l}_1, \dots, \hat{l}_n)^\top$, $\tilde{u} = (\tilde{u}_1, \dots, \tilde{u}_n)^\top$, $\hat{u} = (\hat{u}_1, \dots, \hat{u}_n)^\top$, with n the cardinal of Ω.

5.2 Results for resubstitution

In this section, the formulation (16) is applied to solve the problem with single-valued input and interval-valued output in a database of cars. This dataset, which has been used previously in [9], shows the relationship between a set of (single-valued) characteristics of a determined car (predictor variables) and an interval score given by a set of experts (dependent variable). Observe that in the original dataset (see [9]), the scores are given in a fuzzy way (but these fuzzy spreads are always built under the same assumption: one to the left and to the right when possible, but taking into account that the final interval must be contained in [0,10]).

The database with the interval output for 50 different cars is shown in the Table 1. The characteristics of each car studied are: X_1 = *price,* X_2 = *displacement,* X_3 = *potential,* X_4 = *speed,* X_5 = *acceleration,* X_6 = *urban fuel consumption,* X_7 = *extra urban fuel consumption* and X_8 = *cost/Km.* The variable Y shows the interval score given by the experts according to the characteristics.

First, we compute the predicted interval for the score of each car via resubstitution (see [7]), i.e., when the training sample contains all the elements of the dataset.

The primal and dual formulations have been tested by using this database. All these programs have been solved by using AMPL+MINOS. Concerning the kernel-based dual formulation, the two following kernels ([20, 21]) have been studied:

- Polynomial:

$$K(x, y) = (x^\top y + b)^p, \ p \in \mathbb{N}, \ b \geq 0 \quad (104)$$

- Radial Basis Function (RBF):

$$K(x, y) = \exp\{-\frac{\|x - y\|^2}{2\sigma^2}\}, \ \sigma > 0. \quad (105)$$

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Y
1	21330	1598	120	200	10.5	8.8	15.6	0.41	[5, 8]
2	29864	1781	150	222	8.9	8.8	15.6	0.5	[7, 10]
3	26830	1895	118	206	10.4	8.8	16.7	0.49	[5, 8]
4	26004	1997	110	191	12.5	13.7	22.2	0.29	[4, 6]
5	17613	1998	133	195	9	8.2	15.6	0.42	[1, 4]
6	18120	1596	103	187	10.7	8.9	15.9	0.38	[4, 6]
7	13170	1396	75	165	15.6	10.5	15.6	0.36	[3, 6]
8	19290	1997	136	200	9.6	7.9	14.9	0.44	[1, 4]
9	26100	1998	150	210	9.6	7.2	13.3	0.48	[4, 6]
10	29128	1988	155	215	9.5	7.5	12.8	0.52	[7, 10]
11	28715	1998	129	210	11	7.3	14.9	0.62	[6, 9]
12	26494	1995	140	203	11.3	8.1	13.7	0.5	[3, 6]
13	22931	1997	136	208	10.8	8.7	15.4	0.51	[5, 8]
14	24248	1985	152	215	8.5	7.9	14.9	0.52	[5, 8]
15	19898	1595	105	192	11.3	8.6	15.5	0.41	[5, 8]
16	22200	1948	136	205	9.7	8.5	15.9	0.46	[4, 6]
17	30320	1970	150	215	8.5	7.5	14.7	0.53	[0, 3]
18	19095	1390	75	173	12	16.7	12.2	0.43	[5, 8]
19	37390	2393	165	222	9.2	7.3	13.5	0.65	[7, 10]
20	63812	2771	193	232	10.1	5.7	11.9	0.88	[6, 9]
21	32656	1781	180	228	7.4	9.2	15.9	0.55	[4, 6]
22	54021	2793	193	228	8.6	7.3	12.7	0.84	[6, 9]
23	20199	1997	90	175	14.5	14.3	21.7	0.3	[6, 9]
24	15250	1596	103	180	11.5	9.6	16.9	0.37	[3, 6]
25	28379	1997	147	208	10	8.2	14.1	0.15	[4, 6]
26	60942	3996	280	240	7.3	5.8	11.2	0.86	[1, 4]
27	83666	3996	280	240	7.3	5.8	11.2	1.13	[4, 6]
28	10750	1242	80	174	11.2	13.5	20	0.37	[5, 8]
29	66623	4293	281	250	6.7	5.7	11.2	1.01	[7, 10]
30	36772	1998	163	223	9.1	7.4	14.3	0.65	[6, 9]
31	23235	1796	120	193	9	9.8	17.5	0.54	[4, 6]
32	22176	1781	125	202	9.7	9.4	16.1	0.45	[5, 8]
33	40852	2171	170	226	9.1	8.2	14.1	0.63	[4, 6]
34	37701	2446	129	201	12.1	9.2	16.9	0.39	[1, 4]
35	22125	1998	133	206	10.2	7.7	14.1	0.46	[1, 4]
36	12100	1242	60	155	14.3	13.7	20.8	0.3	[5, 8]
37	14530	1242	80	170	12.5	10.6	18.9	0.32	[1, 4]
38	11078	1242	75	167	13.1	11.5	17.2	0.3	[5, 8]
39	15597	1596	101	185	11	11	18.5	0.37	[6, 9]
40	72562	3996	281	250	6.7	5.8	11.6	1	[5, 8]
41	32030	1998	220	243	7.3	6.8	12.3	0.61	[5, 8]
42	32660	1796	118	202	5.9	10.4	17.5	0.52	[5, 8]
43	20193	1598	102	182	10.8	10.4	18.9	0.4	[5, 8]
44	40619	2597	170	219	9.5	6.1	11.8	0.71	[7, 10]
45	64764	3199	224	240	8.2	5.8	12.2	0.92	[6, 9]
46	93117	4966	306	250	6.5	5.3	11.4	1.23	[7, 10]
47	11104	1360	75	170	13.2	11.2	18.9	0.3	[7, 10]
48	76132	3387	300	280	5.2	5.8	11.8	1.03	[7, 10]
49	11336	1149	60	160	15	12.7	19.2	0.45	[7, 10]
50	15423	1390	75	171	13.5	11.8	18.9	0.34	[7, 10]

Table 1: 'Car scores' database (single-valued input and interval-valued output)

	$RMSE_l$	$RMSE_u$	$\overline{d}_H$
Primal	1.6104	1.6834	1.3856
Polynomial, $b = 1$	1.8051	1.8999	1.5370
Polynomial, $b = 100$	1.7880	1.9022	1.5456
Polynomial, $b = 10000$	1.7844	1.8912	1.5388
RBF, $\sigma = 1000$	1.6054	1.6663	1.2025
RBF, $\sigma = 2000$	**1.5730**	**1.6266**	**1.0395**
RBF, $\sigma = 5000$	1.6125	1.6925	1.0878
RBF, $\sigma = 7500$	1.6349	1.7202	1.2098
RBF, $\sigma = 10000$	1.6515	1.7203	1.3276

Table 2: Best results for $RMSE_l$, $RMSE_u$ and $\overline{d}_H$ for different methods via resubstitution

Different experiments have been performed with several values of the parameters C and ϵ, namely, for every pair (C, ϵ), with $C = 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1$, and $\epsilon = 0.0001, 0.001, 0.01, 0.05, 0.1, 0.5, 1, 1.5, 2, 2.5$. Several kernels have been used by changing the parameters b ($p = 1$ for all the experiments), in the polynomial kernel, and σ, in the RBF kernel. The best results (those with the smallest value for the sum $RMSE_l + RMSE_u$ or with the smallest value of the mean Hausdorff distance) for the primal and several kernel-based dual formulations are shown in Table 2. The best performance in the primal formulation was obtained with $\epsilon = 0.5$ and with $C = 1000$. Concerning the dual formulation, the best results were obtained for the RBF-kernel with $\sigma = 2000$ with for $C = 10$. With $\epsilon = 0.0001$ the smallest value for the mean Hausdorff distance was obtained, and with $\epsilon = 1$ we obtained the smallest sum of the root mean-squared errors. One can observe that the best results (especially, when considering the mean Hausdorff distance as the fitness measurement) are obtained with RBF-kernels, which seem to work well to fit this dataset.
The predicted intervals for the minimum mean Hausdorff distances in the primal and dual formulations are presented in Table 3 ($C = 1000$, $\epsilon = 0.5$ for the primal, $C = 10$, $\epsilon = 0.001$, $\sigma = 2000$ for the dual). One can observe that the non-linear regressor approximates better most of the intervals than the linear regressor.

5.3 Results for leave-one-out

Below, the performance of the regressors built via our methodology when applying the leave-one-out (LOO) strategy (see e.g. [11, 18]) to the cars dataset is shown. Leave-one-out means that, in turns, we consider only one element in the test sample, we train the model with the remaining elements and we test this model with the unitary test sample. The LOO strategy is more interesting than the resubstitution because this way we can study the behaviour of the regressor when a new element arrives.
Several combinations of the parameters C and ϵ and for the parameters of the kernels b and σ have been considered. The performance with the primal and dual formulations has been studied.
Tables 4-5 display the results for the root mean-squared errors, for the different combinations of the parameters C and ϵ with the primal formulation and with the RBF-kernel formulation. The best results for the sum of the two errors have been marked in bold.

	Real	Primal	Dual		Real	Primal	Dual
1	[5 , 8]	[4,4 , 7,3]	[5,0 , 8,0]	26	[1 , 4]	[2,6 , 5,5]	[1,0 , 4,0]
2	[7 , 10]	[5,5 , 8,5]	[6,8 , 9,9]	27	[4 , 6]	[4,7 , 7,7]	[4,0 , 6,0]
3	[5 , 8]	[5,7 , 8,7]	[4,0 , 6,2]	28	[5 , 8]	[5,4 , 8,4]	[5,9 , 9,0]
4	[4 , 6]	[5,4 , 8,4]	[4,0 , 6,0]	29	[7 , 10]	[4,6 , 7,5]	[7,0 , 10,0]
5	[1 , 4]	[2,8 , 5,7]	[3,3 , 5,2]	30	[6 , 9]	[5,9 , 8,9]	[6,0 , 9,0]
6	[4 , 6]	[3,6 , 6,5]	[4,0 , 6,0]	31	[4 , 6]	[4,9 , 7,9]	[4,9 , 7,4]
7	[3 , 6]	[4,4 , 7,3]	[3,0 , 6,0]	32	[5 , 8]	[4,7 , 7,6]	[4,7 , 7,5]
8	[1 , 4]	[3,4 , 6,2]	[4,0 , 6,7]	33	[4 , 6]	[5,7 , 8,7]	[4,1 , 7,0]
9	[4 , 6]	[3,6 , 6,5]	[4,0 , 6,0]	34	[1 , 4]	[4,0 , 6,9]	[5,6 , 8,6]
10	[7 , 10]	[4,2 , 7,1]	[6,4 , 9,5]	35	[1 , 4]	[4,2 , 7,1]	[4,6 , 7,4]
11	[6 , 9]	[6,5 , 9,4]	[6,0 , 9,0]	36	[5 , 8]	[4,7 , 7,7]	[5,0 , 8,0]
12	[3 , 6]	[4,3 , 7,2]	[3,9 , 6,0]	37	[1 , 4]	[3,9 , 6,9]	[2,4 , 5,5]
13	[5 , 8]	[5,5 , 8,4]	[4,7 , 7,4]	38	[5 , 8]	[3,8 , 6,8]	[5,9 , 9,0]
14	[5 , 8]	[4,6 , 7,5]	[5,0 , 7,3]	39	[6 , 9]	[4,5 , 7,5]	[3,4 , 6,2]
15	[5 , 8]	[4,3 , 7,2]	[5,0 , 8,0]	40	[5 , 8]	[4,5 , 7,5]	[5,0 , 8,0]
16	[4 , 6]	[4,4 , 7,3]	[4,6 , 7,4]	41	[5 , 8]	[4,5 , 7,5]	[5,0 , 8,0]
17	[0 , 3]	[4,4 , 7,4]	[6,5 , 9,6]	42	[5 , 8]	[4,5 , 7,5]	[4,8 , 7,8]
18	[5 , 8]	[5,5 , 8,5]	[4,9 , 7,4]	43	[5 , 8]	[4,2 , 7,2]	[5,1 , 8,2]
19	[7 , 10]	[5,5 , 8,4]	[5,8 , 8,7]	44	[7 , 10]	[4,9 , 7,8]	[4,6 , 7,5]
20	[6 , 9]	[6,5 , 9,5]	[6,0 , 9,0]	45	[6 , 9]	[5,9 , 8,8]	[6,0 , 9,0]
21	[4 , 6]	[5,1 , 8,2]	[4,8 , 7,8]	46	[7 , 10]	[4,8 , 7,8]	[7,0 , 10,0]
22	[6 , 9]	[6,0 , 9,0]	[6,0 , 9,0]	47	[7 , 10]	[4,3 , 7,3]	[5,9 , 9,0]
23	[6 , 9]	[5,5 , 8,5]	[4,6 , 7,7]	48	[7 , 10]	[6,8 , 9,9]	[7,0 , 10,0]
24	[3 , 6]	[3,5 , 6,4]	[3,0 , 6,0]	49	[7 , 10]	[6,5 , 9,5]	[5,9 , 9,0]
25	[4 , 6]	[0,6 , 3,5]	[5,6 , 8,5]	50	[7 , 10]	[4,9 , 7,9]	[3,3 , 6,2]

Table 3: Predicted interval output for the primal and dual formulations with the smallest values of mean Hausdorff distance

	C	0.0001		0.001		0.01		0.1		1	
	$\epsilon \backslash RMSE$	l	u	l	u	l	u	l	u	l	u
Primal	0.0001	2.09	2.24	2.15	2.30	2.10	2.25	2.06	2.49	2.09	2.71
	0.001	2.09	2.24	2.15	2.30	2.10	2.25	2.06	2.49	2.09	2.71
	0.01	2.08	2.23	2.15	2.29	2.10	2.25	2.06	2.48	2.09	2.70
	0.05	2.10	2.24	2.12	2.26	2.10	2.25	2.05	2.46	2.07	2.67
	0.1	2.08	2.23	2.07	2.21	2.11	2.26	2.02	2.39	2.04	2.62
	0.5	1.96	2.19	1.95	2.27	1.96	2.20	1.99	2.23	2.26	2.47
	1	2.16	2.11	2.05	2.08	2.12	2.13	2.22	2.23	2.58	2.65
	1.5	2.07	2.17	2.02	2.15	2.04	2.27	2.11	2.32	2.38	2.70
	2	2.00	2.21	1.98	2.18	1.92	2.12	1.95	1.99	2.64	2.64
	2.5	**1.89**	**1.99**	1.91	1.99	1.98	2.01	2.06	2.07	2.34	2.33
RBF $\sigma = 1000$	0.0001	1.96	2.10	1.99	2.15	1.95	2.10	1.96	2.10	2.06	2.20
	0.001	1.96	2.10	1.94	2.08	1.99	2.15	1.96	2.10	2.06	2.20
	0.01	1.96	2.09	1.94	2.08	1.94	2.08	2.00	2.16	2.06	2.20
	0.05	1.96	2.08	1.94	2.07	1.94	2.07	1.95	2.08	2.06	2.20
	0.1	1.95	2.07	1.93	2.06	1.93	2.06	1.94	2.06	2.05	2.20
	0.5	2.05	2.01	1.97	2.01	1.97	2.01	1.97	2.01	2.05	2.11
	1	2.71	2.07	2.52	2.14	2.52	2.14	2.50	2.12	2.06	2.06
	1.5	2.01	1.99	2.01	1.99	2.01	1.99	1.99	2.00	1.94	2.09
	2	1.97	2.11	1.97	2.11	1.97	2.10	1.96	2.08	**1.94**	**1.98**
	2.5	1.96	2.05	1.96	2.05	1.96	2.05	1.97	2.05	2.06	2.09
RBF $\sigma = 2000$	0.0001	1.94	2.15	1.93	2.10	1.95	2.10	1.97	2.11	2.14	2.29
	0.001	1.99	2.15	1.99	2.15	1.96	2.12	1.97	2.11	2.14	2.29
	0.01	1.99	2.06	1.99	2.15	1.99	2.15	1.97	2.14	2.13	2.29
	0.05	**1.93**	**2.06**	1.98	2.14	1.98	2.15	2.00	2.17	2.17	2.35
	0.1	1.97	2.13	1.97	2.13	1.98	2.13	2.00	2.16	2.16	2.29
	0.5	2.14	1.99	2.14	2.09	2.14	2.09	2.14	2.10	2.30	2.32
	1	2.33	2.07	2.51	2.18	2.51	2.18	2.51	2.17	2.45	2.30
	1.5	1.94	2.15	2.29	2.01	2.29	2.01	2.29	2.01	1.99	2.08
	2	1.97	2.11	1.97	2.11	1.97	2.10	1.96	2.08	2.00	2.74
	2.5	1.96	2.05	1.96	2.05	1.96	2.05	1.97	2.05	2.16	2.12

Table 4: $RMSE_l$ and $RMSE_u$ via LOO (primal and RBF-kernel, $\sigma = 1000, 2000$)

	C	0.0001		0.001		0.01		0.1		1	
	$\epsilon\backslash RMSE$	l	u	l	u	l	u	l	u	l	u
RBF $\sigma = 5000$	0.0001	2.07	2.28	1.98	2.08	2.03	2.14	1.99	2.18	2.07	2.25
	0.001	2.01	2.23	2.14	2.39	1.99	2.08	2.03	2.15	2.07	2.25
	0.01	1.99	2.21	1.96	2.11	2.15	2.32	1.99	2.09	2.10	2.25
	0.05	2.01	2.21	1.96	2.10	1.95	2.09	2.03	2.16	2.22	2.32
	0.1	2.00	2.44	1.95	2.09	1.94	2.08	2.14	2.31	2.18	2.32
	0.5	2.18	2.17	1.99	2.04	1.99	2.04	1.98	2.04	2.28	2.19
	1	2.46	2.29	2.65	2.14	2.69	2.18	2.70	2.14	2.48	2.06
	1.5	1.95	2.16	1.98	2.15	2.34	2.01	2.40	2.02	2.05	2.44
	2	2.00	2.35	1.97	2.32	1.97	2.33	1.94	2.37	1.99	2.13
	2.5	2.01	2.09	1.96	2.05	1.96	2.05	**1.96**	**2.04**	2.01	2.06
RBF $\sigma = 7500$	0.0001	2.26	2.19	2.20	2.21	2.20	2.25	2.08	2.24	1.96	2.09
	0.001	1.96	2.17	2.01	2.28	2.08	2.21	2.09	2.26	2.05	2.22
	0.01	1.96	2.17	1.97	2.14	1.99	2.23	2.09	2.22	2.08	2.32
	0.05	1.96	2.17	1.96	2.13	1.96	2.13	1.99	2.20	2.11	2.35
	0.1	1.96	2.17	1.96	2.12	1.96	2.12	1.99	2.23	2.09	2.14
	0.5	1.92	2.20	1.95	2.09	1.98	2.09	1.95	2.09	2.21	2.27
	1	2.32	2.16	2.32	2.34	2.32	2.34	2.32	2.28	2.22	2.10
	1.5	2.05	2.39	2.05	2.60	2.05	2.77	2.01	2.59	1.86	2.70
	2	1.90	1.99	1.90	1.99	1.90	1.99	1.90	1.99	1.94	2.55
	2.5	2.03	2.05	1.97	2.02	1.98	2.02	2.00	2.04	**1.91**	**1.95**
RBF $\sigma = 10000$	0.0001	2.11	2.23	1.97	2.19	1.95	2.12	2.04	2.17	2.10	2.24
	0.001	1.94	2.12	1.88	2.04	1.98	2.19	2.02	2.17	2.08	2.29
	0.01	1.94	2.12	1.94	2.04	**1.92**	**1.98**	1.99	2.20	2.29	2.47
	0.05	2.03	2.12	1.93	2.18	1.96	2.07	2.12	2.23	2.06	2.18
	0.1	1.93	2.07	1.92	2.04	1.88	2.03	1.88	2.05	2.04	2.27
	0.5	1.95	2.03	1.93	2.01	2.02	1.97	1.93	2.13	2.09	2.37
	1	2.54	2.07	2.67	2.09	2.67	2.06	2.69	2.04	2.89	2.31
	1.5	2.05	2.25	2.48	2.59	2.47	2.77	2.15	2.67	1.93	2.74
	2	1.91	1.99	1.91	1.99	1.91	1.99	1.92	1.99	2.04	2.02
	2.5	2.03	2.05	2.03	2.05	2.04	2.05	2.07	2.08	2.10	2.08

Table 5: $RMSE_l$ and $RMSE_u$ via LOO (RBF-kernel, $\sigma = 5000, 7500, 10000$)

$\epsilon\backslash C$		0.0001	0.001	0.01	0.1	1		0.0001	0.001	0.01	0.1	1
0.0001		1.78	1.87	1.80	2.10	2.26		1.80	**1.62**	1.72	1.73	1.77
0.001		1.78	1.87	1.80	2.10	2.26		1.72	1.82	1.62	1.73	1.77
0.01		1.77	1.87	1.80	2.09	2.25		1.70	1.70	1.78	1.63	1.84
0.05		1.78	1.84	1.81	2.07	2.21		1.72	1.71	1.67	1.69	1.90
0.1	Primal	1.79	1.80	1.83	2.01	2.15	RBF	1.96	1.71	1.68	1.79	1.84
0.5		1.85	1.91	1.88	1.88	2.00	$\sigma = 5000$	1.96	1.78	1.78	1.74	2.04
1		1.94	1.85	1.92	2.00	2.12		2.52	2.58	2.57	2.51	2.18
1.5		1.85	1.84	1.90	1.95	2.06		1.95	1.84	2.16	2.14	2.26
2		1.83	1.84	1.81	1.76	2.07		2.04	2.00	2.01	2.05	1.83
2.5		**1.69**	1.71	1.70	1.79	1.93		1.82	1.76	1.76	1.77	1.81
0.0001		1.64	1.68	1.64	1.65	1.77		1.86	1.78	1.80	1.77	1.70
0.001		1.64	**1.60**	1.68	1.65	1.77		1.72	1.86	1.72	1.71	1.78
0.01		1.64	1.60	1.60	1.69	1.77		1.72	1.70	1.82	1.73	1.88
0.05		1.65	1.61	1.61	1.62	1.77		1.74	1.71	1.71	1.77	1.82
0.1	RBF	1.66	1.62	1.62	1.62	1.78	RBF	1.76	1.72	1.72	1.83	1.73
0.5	$\sigma = 1000$	1.84	1.78	1.78	1.77	1.80	$\sigma = 7500$	1.90	1.84	1.86	1.82	2.11
1		2.54	2.46	2.45	2.41	1.89		2.46	2.50	2.49	2.39	2.06
1.5		1.84	1.84	1.84	1.84	1.92		2.17	2.33	2.42	2.30	2.33
2		1.64	1.64	1.64	1.66	1.71		**1.62**	1.62	1.62	1.63	2.23
2.5		1.76	1.76	1.76	1.78	1.84		1.74	1.74	1.74	1.79	1.71
0.0001		1.68	1.64	1.64	1.66	1.81		1.96	1.76	1.72	1.71	1.69
0.001		1.68	1.68	1.68	1.66	1.81		1.86	1.64	1.74	1.71	1.81
0.01		1.66	1.68	1.68	1.70	1.81		1.86	1.62	**1.62**	1.75	2.01
0.05		**1.61**	1.69	1.69	1.70	1.92		1.93	1.73	1.69	1.75	1.67
0.1	RBF	1.70	1.70	1.70	1.70	1.85	RBF	1.68	1.66	1.66	1.66	1.75
0.5	$\sigma = 2000$	1.88	1.88	1.88	1.86	2.00	$\sigma = 10000$	1.75	1.86	1.90	1.85	2.05
1		2.30	2.40	2.40	2.36	2.24		2.60	2.63	2.61	2.56	2.70
1.5		1.81	2.06	2.06	2.07	1.81		2.03	2.69	2.79	2.50	2.38
2		1.64	1.64	1.64	1.66	2.24		1.72	1.72	1.72	1.73	1.84
2.5		1.76	1.76	1.76	1.78	1.91		1.74	1.74	1.74	1.79	1.86

Table 6: $\overline{d}_H$ via LOO (primal and RBF-kernel, σ =1000, 2000, 5000, 7500, 10000)

Likewise, Table 6 present the results obtained when we use the Hausdorff distance to measure the error between the predicted and the observed interval and we compute the mean Hausdorff distance ($\overline{d}_H$) for all the elements of the database, for the primal and the RBF-kernel formulation. The lowest value of the mean distance is in bold in each table.
One can observe that, in Tables 4-6, the behaviour of the measurements $RMSE_l$, $RMSE_u$ and $\overline{d}_H$ is not very stable with respect to the parameter ϵ, especially for those values $\epsilon \geq 0.5$.
In Table 7, the best results for the two measurements, for the different formulations are shown. Although all the results are quite similar, one can observe that the polynomial kernel seems to be better suited when using $RMSE_l$ and $RMSE_u$ as fitness measurements, whereas the RBF-kernel has a better behaviour when studying the fitness via the mean Hausdorff distance.

	$RMSE_l$	$RMSE_u$	$\overline{d}_H$
Primal	1.8852	1.9924	1.6944
Polynomial, $b = 1$	1.8259	1.9722	1.6229
Polynomial, $b = 100$	**1.8256**	**1.9720**	1.6225
Polynomial, $b = 10000$	1.8260	1.9720	1.6227
RBF, $\sigma = 1000$	1.9415	1.9758	**1.6002**
RBF, $\sigma = 2000$	1.9341	2.0646	1.6100
RBF, $\sigma = 5000$	1.9610	2.0388	1.6201
RBF, $\sigma = 7500$	1.9062	1.9464	1.6200
RBF, $\sigma = 10000$	1.9179	1.9798	1.6219

Table 7: Best results for $RMSE_l$, $RMSE_u$ and $\overline{d}_H$ for different methods via leave-one-out

	Resubstitution			Leave-one-out		
Formulation	$RMSE_l$	$RMSE_u$	$\overline{d}_H$	$RMSE_l$	$RMSE_u$	$\overline{d}_H$
Maximum distance	2.1907	2.1403	2.6676	2.5564	2.3673	3.0196
Hausdorff distance	2.0617	2.2531	2.6556	2.3519	2.4633	2.9861
Primal	1.6104	1.6834	1.3856	1.8852	1.9924	1.6944
Dual (kernel)	1.5730	1.6266	1.0395	1.8256	1.9720	1.6002

Table 8: Comparison between the best results obtained for formulations with one hyperplane (top) and for formulations with two hyperplanes (bottom)

5.4 Comparison with point estimation

Table 8 compares the performance of the two possible models described in [3], formulations (6)-(7), to solve the problem with single-valued input and interval-valued output, with the model proposed in this paper.

The table shows the best results for the two fitness measurements when performing resubstitution and leave-one-out. The first two rows display the best results when using formulations (6)-(7) for the maximum and Hausdorff distances (with only one hyperplane to compute), respectively. The drawback of these formulations is that the predicted interval for each record of the database is degenerate (point estimation instead of interval estimation). In the last two rows, we present the results with formulation (16) (the new one introduced for the single-input and interval-output situation) and its kernel-based dual formulation (92) (with two hyperplanes to compute), respectively. All the kernels used in the computational experiment have been taken into account to select the best results.

One can observe that the results obtained with formulations with two hyperplanes are much better on this dataset than those obtained with the formulations with only one hyperplane, that is, one can say that interval estimates are much better than only point estimates. This way, the introduction of formulation (16) becomes justified.

Finally, with the results obtained in these experiments, we have performed the following measurement. We have considered the midpoint of the interval outputs in the original dataset, and we have also obtained the midpoint of the predicted intervals via formulation

	Resubstitution	Leave-one-out
Hausdorff distance formulation (7)	1.2456	1.5761
Primal formulation (16)	1.2755	1.5666
Kernel-based dual formulation (92)	0.9349	1.5102

Table 9: $\overline{d}_1$ for formulations with one hyperplane (top) and for formulations with two hyperplanes (bottom)

(16) and the dual formulation (92). We consider the mean l_1-distance, defined as

$$\overline{d}_1 = \frac{1}{n} \sum_{i \in \Omega} |\tilde{y}_i - \hat{y}_i|, \tag{106}$$

where $\tilde{y}_i$ is the center of the interval output of the element $i \in \Omega$, $\hat{y}_i$ is the center of the predicted interval for $i \in \Omega$, and n is the cardinal of Ω.
We measure the mean l_1-distance $\overline{d}_1$ between the midpoints of the real interval outputs and the prediction given by formulation (7) (the model described in [3]) for the Hausdorff distance (only one hyperplane). We also measure $\overline{d}_1$ between the center of the real intervals and the centers of the predicted intervals obtained with formulations (16) and (92). We have selected, in each model, the combination of parameters (C, ϵ) which gave the minimum $\overline{d}_H$ in Table 8. The values of this measure, for resubstitution and leave-one-out, are displayed in Table 9.
One can observe that the best results, for resubstitution and for leave-one-out, are obtained for the kernel-based dual formulation. Hence, formulation (92) is still better in this dataset when only a point-prediction (instead of an interval-prediction) is asked for, since the midpoints of the intervals obtained with this formulation are closer to the midpoints of the real intervals than the point estimations obtained with formulation (7).
This indicates that, in a regression problem with imprecise output, when one wants a single-valued estimate, one obtains quite similar result-quality when one does a single-point estimation directly (using the Hausdorff distance formulation (7)) than rather an interval estimation (primal formulation (16)) and then taking its midpoint (that is, eliminating the additional uncertainty information given by the interval output). In fact, for leave-one-out, the results are slightly better for the interval estimation methodology (formulation (16)) and clearly better (for resubstitution and leave-one-out) when kernels are introduced (formulation (92)), since they allow to extend the model to study more abstract relations between data.

6 Conclusions

In this work, a regression problem for data with single-valued predictive variables and interval-valued dependent variable has been analyzed.
The proposed model is based on the standard ϵ-Support Vector Regression approach, and two hyperplanes must be computed to approximate the lower and upper bounds of the dependent variable. The dual formulation of this optimization problem has been obtained, allowing the introduction of a kernel structure to use non-linear regressors in the data.

The computational experiments show that the introduction of these kernel structures allows to improve the results when measuring the error between the predicted and observed intervals.
The results for the kernel-based formulation via leave-one-out are quite unstable with respect to the parameter ϵ. Thus, the correct choice of the parameter ϵ remains as a critical issue for future work, which can be approached by using some of the strategies proposed in the literature. Thus, in [4], for example, values of C and ϵ are given, based on the outputs of the training sample.
The formulation has been obtained by using the Hausdorff distance in the constraints and by minimizing the sum of the Euclidean norms of the hyperplanes in the objective function. The use of other distances and norms to define the optimization problem is another topic which deserves further study.

References

[1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley and Sons, New York, (1993).

[2] C. J. C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, (1998).

[3] E. Carrizosa, J. Gordillo, and F. Plastria. Support Vector Regression for Imprecise Data. Technical Report MOSI/35, MOSI Department, Vrije Universiteit Brussel, http://www.vub.ac.be/MOSI/papers/CarrizosaPlastriaGordillo_interval_SVR.pdf, (2007).

[4] V. Cherkassky and Y. Ma. Selection of Meta-Parameters for Support Vector Regression. In *Proc. of the International Conference on Artificial Neural Networks*, pages 687–693, (2002).

[5] C. Cortes and V. Vapnik. Support Vector Networks. *Machine Learning*, 20:273–297, (1995).

[6] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and other Kernel-based Learning Methods*. Cambridge University Press, Cambridge, (2000).

[7] L. Devroye and T. J. Wagner. Distribution-free Performance Bounds with the Resubstitution Error Estimate. *IEEE Transactions on Information Theory*, 25(2):208–210, (1979).

[8] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik. Support Vector Regression Machines. In *Proc. Advances in Neural Information Processing Systems*, volume 9, pages 155–161, (1997).

[9] P. D'Urso. Linear Regression Analysis for Fuzzy/Crisp Input and Fuzzy/Crisp Output Data. *Computational Statistics and Data Analysis*, 42:47–72, (2003).

[10] S. Gunn. Support Vector Machines for Classification and Regression. Technical Report ISIS-1-98, Department of Electronics and Computer Science, University of Southampton, (1998).

[11] D. J. Hand, H. Mannila, and P. Smyth. *Principles of Data Mining.* MIT Press, Cambridge, (2001).

[12] R. Herbrich. *Learning Kernel Classifiers: Theory and Algorithms.* MIT Press, Cambridge, (2002).

[13] D. H. Hong and C. Hwang. Support Vector Fuzzy Regression Machines. *Fuzzy Sets and Systems*, 138(2):271–281, (2003).

[14] D. H. Hong and C. Hwang. Extended Fuzzy Regression Models Using Regularization Method. *Information Sciences*, 164:31–36, (2004).

[15] C. Hwang, D. H. Hong, E. Na, H. Park, and J. Shim. Interval Regression Analysis Using Support Vector Machine and Quantile Regression. In L. Wang and Y. Jin, editors, *Fuzzy Systems and Knowledge Discovery*, pages 100–109. Springer-Verlag, Berlin, (2005).

[16] C. Hwang, D. H. Hong, and K. H. Seok. Support Vector Interval Regression Machine for Crisp Input and Output Data. *Fuzzy Sets and Systems*, 157:1114–1125, (2006).

[17] J-T. Jeng, C-C. Chuang, and S-F. Su. Support Vector Interval Regression Networks for Interval Regression Analysis. *Fuzzy Sets and Systems*, 138:283–300, (2003).

[18] R. Kohavi. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1137–1143, (1995).

[19] H. Lee and H. Tanaka. Upper and Lower Approximation Models in Interval Regression Using Regression Quantile Techniques. *European Journal of Operational Research*, 116:653–666, (1999).

[20] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis.* Cambridge University Press, Cambridge, (2004).

[21] A. J. Smola and B. Schölkopf. A Tutorial on Support Vector Regression. *Statistics and Computing*, 14:199–222, (2004).

[22] H. Tanaka and H. Ishibuchi. Possibilistic Regression Analysis Based on Linear Programming. In J. Kacprzyk and M. Fedrizzi, editors, *Fuzzy Regression Analysis*, pages 47–60. Omnitech Press, Warsow, (1992).

[23] H. Tanaka and H. Lee. Interval Regression Analysis by Quadratic Programming Approach. *IEEE Transactions on Fuzzy Systems*, 6:473–481, (1998).

[24] V. N. Vapnik. *The Nature of Statistical Learning.* Springer-Verlag, New York, (1995).

[25] V. N. Vapnik. *Statistical Learning Theory.* John Wiley and Sons, New York, (1998).