Support Vector Regression for imprecise data *

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

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

30th October 2007

Abstract

In this work, a regression problem is studied where the elements of the database are sets with certain geometrical properties. In particular, our model can be applied to handle data affected by some kind of noise or uncertainty and interval-valued data, and databases with missing values as well. The proposed formulation is based on the standard ϵ-Support Vector Regression approach. In the interval-data case, two different formulations will be obtained, according to the way of measuring the distance between the prediction and the actual intervals. Computational experiments with real databases are performed.

Keywords: Support Vector Regression, Interval Data, Missing Values, Quadratic Programming, Robustness, Gauges.

*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.

1 Introduction

Data cannot always be expressed by single feature vectors. In certain situations, interval data is a better way of representing the values of certain variables. For example, intervals are used for expressing ranges such as the range of temperature during a day, or age intervals for a group of individuals. Intervals can also be used when one measures several times a same variable for the same individual and this information must be summarized, like the fluctuations of blood pressure or pulse rate of a patient. Other examples of interval data appear in case of imprecise data, or when estimating a certain parameter by a confidence interval, and, in general, whenever uncertainty or vagueness arises in our problem.
Interval data also appear in the framework of Symbolic Data Analysis ([6, 8]). In certain problems, large databases cannot be handled in an efficient way and it is necessary to aggregate the data in such a way that the resulting dataset has a more manageable size and it retains enough knowledge from the original database. Different approaches exist to summarize the data by using classical variables (single values), multi-valued variables (categorical variables which can have several results), interval-valued variables (the data are aggregated into intervals and this is the case of our interest) or modal variables (a single, interval or nominal variable which can have different values with different probabilities associated).

From a Symbolic Data Analysis perspective, the first work published on this topic appeared in [4]. Consider the classical linear regression model (see e.g. [17]),

$$Y = X^\top \beta + \varepsilon,$$

with Y the dependent variable, $X = (1, X_1, \ldots, X_d)$ is the matrix of the predictor variables, $\beta = (\beta_0, \beta_1, \ldots, \beta_d)^\top$ is the vector of coefficients of the regression model and ε, the error. The approach in [4] consisted in fitting the classical linear regression model on the midpoint of the intervals of each variable of the dataset. The predicted lower and upper bounds for the dependent variables were computed on the obtained model. This model is improved in [14], where two linear regression models are used, one for predicting the midpoint of the output and the other one for predicting the range. The predicted lower and upper bounds for the dependent variable are recovered with the midpoint and range. In [29], a comparison between these two models is shown. Other extensions of these models can be found in [5, 30].

Related to our problem, we also find the concept of *interval regression analysis*, which is the simplest version of possibilistic regression analysis, as introduced by Tanaka et al. (see [28, 38, 39]). 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. To do 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 to minimize 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 [38]). Later, in [39], 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 [28], two regression models are built for each database by using quantile techniques, and two interval outputs are given for each observation, with the smallest one included in the biggest 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 such problem to build the two models (see [24]) and to the general interval regression analysis case. In [26], 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 [25], the quadratic formulation of [39] is integrated with the standard ϵ-SVR approach.

Concerning fuzzy regression analysis, several approaches have been developed. Roughly, they can be classified in two main groups: the possibilistic approach (initially proposed in [40]), where the objective function to be minimized is a measurement of the spread of the predicted output, and the least-squares approach (introduced in [11, 16]), which minimizes a distance, on fuzzy numbers, between the real and the predicted output.
SVMs have also been applied to fuzzy multiple linear regression models (see [22, 23]). 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 standard ϵ-SVR methodology is applied 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.

Another interesting situation related to our work is the case of data affected by some kind of noise or perturbation. A robust regressor must be constructed, insensitive to this noise in the data. One model of Robust Support Vector Regression has been studied in [41, 42], with noise in the input data (predictor variables). Although the data points are assumed to be uncertain or noisy, that perturbation is bounded by a given hypersphere of known radius. An optimization problem is formulated and solved via Second Order Cone Programming. It will be seen that our model generalizes the formulation proposed in [41, 42].

In this work, we study a regression problem with imprecise data, that is, the elements of the dataset are affected by uncertainty. We propose two formulations based on standard ϵ-Support Vector Regression (see [37]), by using two different distances (maximum and Hausdorff distances) for measuring the error between predicted and real intervals. The formulation is applied to the case of interval data, where our model has been tested on real databases. The case of data affected by some kind of noise is also handled, and it will be seen that our model generalizes the formulation proposed in [41, 42].

The technique described in our paper is also useful for modeling the case in which there exist missing values (see [32] for a complete study on statistical analysis in datasets with missing values), that is, when the database is formed by feature vectors but some of their coordinates do not appear in the dataset. Different techniques have been used in the literature to handle missing data in classification problems (for a survey on the topic, see [33, 34]). In our case, instead of imputing single values as usual for the missing coordinates, they will be replaced by intervals which will be built by using the non-missing values of the same class in the dataset. Different measurements will be taken into account to perform the construction of these intervals.

The structure of this paper is the following. In Section 2, after an introduction to the standard ϵ-Support Vector Regression for single-valued data, the extension to the case of non-single objects is described. A general optimization problem is given, from which two different formulations will be derived according to the distance used as a measurement of the error between the predicted interval and the observed one for each object of the dataset. The formulation with the maximum distance will be explained in depth in Section 3, whereas the formulation with the Hausdorff distance is given in Section 4. In each formulation, the general model is particularized to the case of interval data and perturbed data. In Section 5, a computational experiment with a cardiological database is performed. Section 6 includes a new methodology for imputation of missing values where the blanks are filled in by intervals built with the remaining values of the corresponding variable in the dataset. Finally, Section 7 contains some discussion and concluding remarks.

2 Modeling the problem

2.1 ϵ-Support Vector Regression with points

In the standard ϵ-Support Vector Regression, ϵ-SVR for short (see e.g. [13, 18, 19, 37, 44, 45]), 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 set 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 $\omega \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$ such that, for each instance $i \in \Omega$, the affine function $f(x) = \omega^\top x + \beta$ yields a small deviation(at most ϵ) between the observed value y_i and the predicted value $f(x_i)$.
Since the deviation between y_i and $f(x_i)$ must be at most ϵ, the following set of constraints is obtained

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

The optimization problem to solve, as stated in [37], is the following,

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

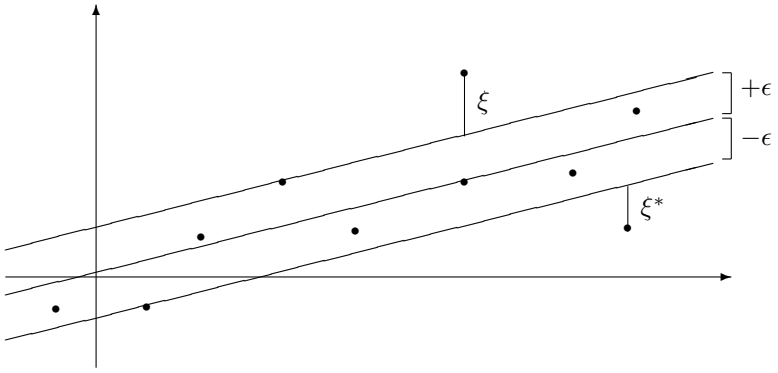

Figure 1: ϵ-Support Vector Regression

This optimization problem can be non-feasible. Hence, one must introduce some slack variables ξ, ξ^* in the constraints (as in the Soft-Margin case for Support Vector Machines, see e.g. [9, 12]) and a penalty term must be added to the objective function. The optimization problem has then the following form

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

with C and ϵ constants of the model, where ϵ is the maximum allowed deviation for the instances of the database and C represents the trade-off between the flatness of the prediction function and the sum of deviations larger than ϵ.
Figure 1 explains graphically the model. We seek a hyperplane to fit the points of the dataset, but only the points whose deviation from the predicted value (the corresponding point lying on the hyperplane) is bigger than ϵ will be penalized. That is, the points outside the band defined by the hyperplane and the parameter ϵ, the so-called *ϵ-insensitive tube*, will be penalized via the corresponding slack variable (variable ξ for points above the tube, and ξ^* for points below the tube).
Formulation (3), introduced by [44], corresponds to deal with the so-called *ϵ-insensitive loss function*, which is defined as

$$|\xi|_\epsilon = \begin{cases} 0 & \text{if } |\xi| \leq \epsilon, \\ |\xi| - \epsilon & \text{otherwise,} \end{cases} \tag{4}$$

see Figure 2.

2.2 ϵ-Support Vector Regression with objects

Whereas in the standard ϵ-SVR approach, each instance in the database is of the form $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$, now we consider a database $\Omega \subset \mathbb{R}^d \times \mathbb{R}$ formed by objects $i = (X_i, Y_i) \in$

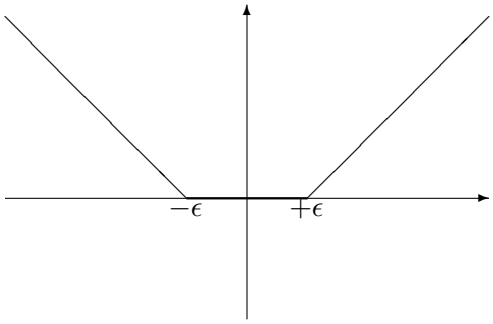

Figure 2: ϵ-insensitive loss function

Ω, where Y_i is an interval in $\mathbb{R}$, $Y_i = [\tilde{l}_i, \tilde{u}_i]$, with $\tilde{l}_i \leq \tilde{u}_i$, and X_i is of the form $X_i = x_i + B_i$, with $x_i \in \mathbb{R}^d$ and with B_i a subset of $\mathbb{R}^d$ with certain geometrical properties, namely, it is convex, symmetric with respect to the origin and contains the origin. In other words, B_i is the unit ball of a symmetric gauge γ_i (see [21]), that is,

$$B_i \quad = \quad \{s \in \mathbb{R}^d : \gamma_i(s) \leq 1\}. \tag{5}$$

We consider the following two particular cases of interest:

1. γ_i is given by

$$\gamma_i(s_1, \ldots, s_d) = \max_{j=1,\ldots,d} \frac{2|s_j|}{u_{ij} - l_{ij}}, \text{ for } l_{ij} < u_{ij}, \ j = 1, \ldots, d, \tag{6}$$

and then

$$\begin{aligned} B_i &= \{s \in \mathbb{R}^d : \gamma_i(s) \leq 1\} = \{s \in \mathbb{R}^d : \max_{j=1,\ldots,d} \frac{2|s_j|}{u_{ij} - l_{ij}} \leq 1\} \\ &= \{s \in \mathbb{R}^d : |s_j| \leq \frac{u_{ij} - l_{ij}}{2}, \ \forall j = 1, \ldots, d\}. \end{aligned} \tag{7}$$

2. γ_i is given by

$$\gamma_i(s_1, \ldots, s_d) = \frac{1}{r_i} \sum_{j=1}^{d} (|s_j|^p)^{\frac{1}{p}}, \text{ for some } p, \ 1 \leq p \leq \infty, \text{ for } r_i > 0, \tag{8}$$

and then

$$\begin{aligned} B_i &= \{s \in \mathbb{R}^d : \gamma_i(s) \leq 1\} = \{s \in \mathbb{R}^d : \frac{1}{r_i} \sum_{j=1}^{d} (|s_j|^p)^{\frac{1}{p}} \leq 1\} \\ &= \{s \in \mathbb{R}^d : \|s\|_p \leq r_i\}. \end{aligned} \tag{9}$$

When γ_i is of the form (6), taking x_i such that $x_{ij} = \frac{l_{ij} + u_{ij}}{2}$, $j = 1, \ldots, d$, one has that $X_i = x_i + B_i$, with B_i as in (7), is a Cartesian product of intervals, that is, $X_i = \prod_{j=1}^{d} [l_{ij}, u_{ij}]$.

Interval data can be used for expressing ranges, for grouping several measurements of the same variable from the same individual, in case of imprecise data or when the data have been aggregated or summarized in an interval (like in Symbolic Data Analysis, see [6, 8]). The second case, with γ_i of the form (8), can be used to model the case of noisy data. This problem was first tackled in [41, 42] using Support Vector Machines as well. The data, in that problem, have suffered some perturbations, which are supposed to be unknown, but a bound on them is known, for a chosen p-norm in the input space.
In that case, we can write $X_i = x_i + B_i$, with x_i the original value of the instance and B_i, as defined in (9), a ball representing the unknown perturbation and r_i being a positive constant which bounds the perturbation in p-norm, since $x \in X_i$ iff $x = x_i + s$, with $\gamma_i(s) \leq 1$, or equivalently, $\|s\|_p \leq r_i$, for each $i \in \Omega$.
In case of dealing with balls, the concept of ϵ-SVR must be modified and one has that our goal will be to compute the parameters ω and β of a hyperplane such that a given distance from (X_i, Y_i) to that hyperplane is at most ϵ, for every i in the database. Two distances will be considered: the maximum distance d_{max} and the Hausdorff distance d_H, defined on intervals $[\underline{a}, \overline{a}]$, $[\underline{b}, \overline{b}]$ as

$$\begin{aligned} d_{max}([\underline{a}, \overline{a}], [\underline{b}, \overline{b}]) &= \max\{|a - b| : \ a \in [\underline{a}, \overline{a}], \ b \in [\underline{b}, \overline{b}]\} \qquad (10) \\ &= \max\{|\underline{a} - \overline{b}|, |\overline{a} - \underline{b}|\} \\ d_H([\underline{a}, \overline{a}], [\underline{b}, \overline{b}]) &= \max\{\max_{a \in [\underline{a}, \overline{a}]} \min_{b \in [\underline{b}, \overline{b}]} |a - b|, \min_{a \in [\underline{a}, \overline{a}]} \max_{b \in [\underline{b}, \overline{b}]} |a - b|\} \qquad (11) \\ &= \max\{|\underline{a} - \underline{b}|, |\overline{a} - \overline{b}|\}. \end{aligned}$$

Then, our aim will be to seek ω and β such that

$$d([\min_{x \in X_i}(\omega^\top x + \beta), \max_{x \in X_i}(\omega^\top x + \beta)], [\tilde{l}_i, \tilde{u}_i]) \leq \epsilon, \ \forall i \in \Omega, \qquad (12)$$

where d is a distance (such as $d_{\max}$ or d_H) in the space of intervals.
Different solutions to the problem can be obtained. We are interested in finding the solution with minimum norm of ω, as done in the standard Support Vector Regression case (see [37, 45]).
The analog to Problem (2) is then

$$\begin{aligned} \min_{\omega, \beta} \quad & \frac{1}{2} \sum_{j=1}^{d} \omega_j^2 \\ \text{s.t.} \quad & d([\min_{x \in X_i}(\omega^\top x + \beta), \max_{x \in X_i}(\omega^\top x + \beta)], [\tilde{l}_i, \tilde{u}_i]) \leq \epsilon, \ \forall i \in \Omega. \end{aligned} \qquad (13)$$

In the next two sections, formulations for the problems with the maximum and the Hausdorff distances will be given.

3 Formulation based on the maximum distance

3.1 Formulation of the problem

Figure 3 gives a graphical explanation of the model for the maximum distance in the interval data case. A hyperplane is sought to fit the boxes, and penalties appear when the

maximum distance from the box to the corresponding vertical projection on the hyperplane is larger than ϵ, that is, when not every point of the box is inside the ϵ-insensitive tube. Variable ξ is used for the points above the tube and ξ^* for the points below the tube. Considering the maximum distance d_{max} between the predicted and the real interval, constraint (12) can be written as

$$\max_{(x,y)\in(X_i,Y_i)} |\omega^\top x + \beta - y| \leq \epsilon, \ \forall i \in \Omega. \tag{14}$$

Constraint (14) can be divided into the following pair of constraints,

$$\max_{x\in X_i} \max_{y\in Y_i} (y - \omega^\top x - \beta) \leq \epsilon, \quad \forall i \in \Omega \tag{15}$$

$$\max_{x\in X_i} \max_{y\in Y_i} (\omega^\top x + \beta - y) \leq \epsilon, \quad \forall i \in \Omega, \tag{16}$$

and by taking into account that Y_i is an interval $[\tilde{l}_i, \tilde{u}_i]$, we can rewrite them as

$$\max_{x\in X_i} (\tilde{u}_i - \omega^\top x - \beta) \leq \epsilon, \quad \forall i \in \Omega \tag{17}$$

$$\max_{x\in X_i} (\omega^\top x + \beta - \tilde{l}_i) \leq \epsilon, \quad \forall i \in \Omega. \tag{18}$$

Then, the hard-margin optimization problem (13) will be

$$\begin{array}{ll} \min\limits_{\omega,\beta} & \frac{1}{2}\sum\limits_{j=1}^{d} \omega_j^2 \\ \text{s.t.} & \max\limits_{x\in X_i} (\tilde{u}_i - \omega^\top x - \beta) \leq \epsilon, \ \forall i \in \Omega \\ & \max\limits_{x\in X_i} (\omega^\top x + \beta - \tilde{l}_i) \leq \epsilon, \ \forall i \in \Omega. \end{array} \tag{19}$$

In order to obtain a soft-margin version, one can introduce some slack variables ξ, ξ^* in the constraints (as done in the soft-margin case for Support Vector Machines, see [9, 12, 13]) and we must add a penalty term in the objective function, similar to (3),

$$\begin{array}{ll} \min\limits_{\omega,\beta,\xi,\xi^*} & \frac{1}{2}\sum\limits_{j=1}^{d} \omega_j^2 + C\sum\limits_{i\in\Omega}(\xi_i + \xi_i^*) \\ \text{s.t.} & \tilde{u}_i - \min\limits_{x\in X_i} \omega^\top x - \beta \leq \epsilon + \xi_i, \ \forall i \in \Omega \quad (20) \\ & \max\limits_{x\in X_i} \omega^\top x + \beta - \tilde{l}_i \leq \epsilon + \xi_i^*, \ \forall i \in \Omega \quad (21) \\ & \xi_i, \xi_i^* \geq 0, \ \forall i \in \Omega. \end{array}$$

3.2 An equivalent formulation

The following result gives an equivalent and more tractable formulation of our problem by using duality for the constraints (20)-(21). Recall that the dual gauge γ_i^0 of γ_i in ω is defined by $\gamma_i^0(\omega) = \max\limits_{\gamma_i(u)\leq 1} (\omega^\top u)$.

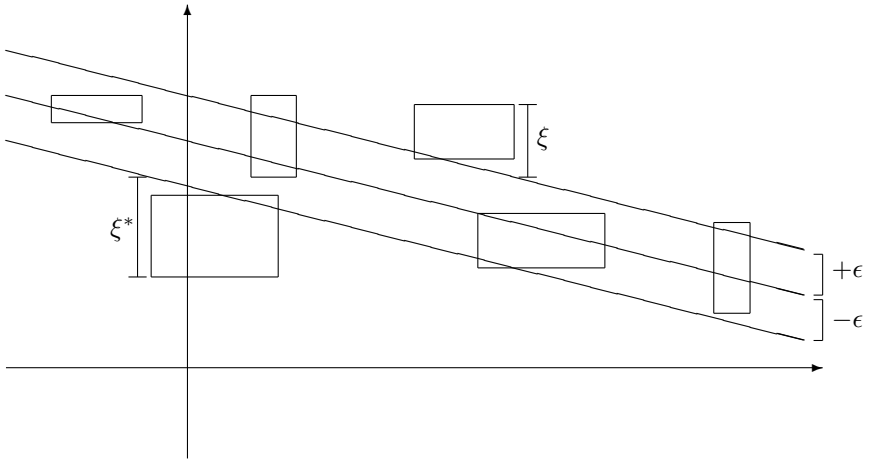

Figure 3: Formulation based on the maximum distance

Theorem 3.1. *Problem with constraints (20)-(21) admits the following equivalent formulation as a convex quadratic minimization problem with convex nonlinear constraints*

$$\begin{array}{ll} \min\limits_{\omega,\beta,\xi,\xi^*} & \frac{1}{2}\sum\limits_{j=1}^{d}\omega_j^2 + C\sum\limits_{i\in\Omega}(\xi_i+\xi_i^*) \\ s.t. & \tilde{u}_i - \omega^\top x_i + \gamma_i^0(\omega) - \beta \le \epsilon + \xi_i,\ \forall i\in\Omega \\ & \omega^\top x_i + \gamma_i^0(\omega) + \beta - \tilde{l}_i \le \epsilon + \xi_i^*,\ \forall i\in\Omega \\ & \xi_i,\xi_i^* \ge 0,\ \forall i\in\Omega, \end{array} \tag{22}$$

where γ_i is the gauge associated to the object $i\in I$ defining the ball B_i used in $X_i = x_i + B_i$ and γ_i^0 is its dual gauge.

Proof.
The proof is analogous to Theorem 2.1 in [10]. We change constraints (20)-(21) by using that $X_i = x_i + B_i$, B_i being the unit ball induced by the gauge γ_i for each X_i.
One has that,

$$\min_{x\in x_i+B_i}\omega^\top x = \min_{\gamma_i(u)\le 1}\omega^\top(x_i+u) = \omega^\top x_i + \min_{\gamma_i(u)\le 1}\omega^\top u = \omega^\top x_i - \max_{\gamma_i(u)\le 1}(-\omega^\top u).$$

By using that $\gamma_i^0(-\omega) = \max\limits_{\gamma_i(u)\le 1}(-\omega^\top u)$ (with γ_i^0 the dual gauge of γ_i) and since $\gamma_i^0(-\omega) = \gamma_i^0(\omega)$, one obtains that

$$\min_{x\in x_i+B_i}\omega^\top x = \omega^\top x_i - \gamma_i^0(-\omega) = \omega^\top x_i - \gamma_i^0(\omega). \tag{23}$$

Analogously, one has that

$$\max_{x\in x_i+B_i}\omega^\top x = \omega^\top x_i + \max_{\gamma_i(u)\le 1}(\omega^\top u) = \omega^\top x_i + \gamma_i^0(\omega). \tag{24}$$

Then, by using (23), the set of constraints (20) can be rewritten as

$$\tilde{u}_i - \min_{x\in x_i+B_i}\omega^\top x - \beta \le \epsilon + \xi_i \quad\leftrightarrow\quad \tilde{u}_i - \omega^\top x_i + \gamma_i^0(\omega) - \beta \le \epsilon + \xi_i,\ \forall i\in\Omega, \tag{25}$$

and by using (24), the set of constraints (21) remains as follows,

$$\max_{x\in x_i+B_i} \omega^\top x+\beta-\tilde{l}_i\leq \epsilon+\xi_i^* \quad\leftrightarrow\quad \omega^\top x_i+\gamma_i^0(\omega)+\beta-\tilde{l}_i\leq \epsilon+\xi_i^*,\ \forall i\in\Omega. \tag{26}$$

□

Below, we consider the two cases of interest for the definitions of γ_i given in (6) and (8). The first one is the case in which the elements of the database are boxes in dimension d, that is, $X_i=\prod_{j=1}^d[l_{ij},u_{ij}]$, for every $i\in\Omega$.

Corollary 3.1. *Let γ_i be the gauge defined in (6). Then, Problem (22) admits the following equivalent formulation as a convex quadratic problem with linear constraints*

$$\begin{array}{ll}\min\limits_{\sigma,\tau,\beta,\xi,\xi^*} & \frac{1}{2}\sum\limits_{j=1}^d(\sigma_j-\tau_j)^2+C\sum\limits_{i\in\Omega}(\xi_i+\xi_i^*)\\ s.t. & \tilde{u}_i+\sum\limits_{j=1}^d\tau_j u_{ij}-\sum\limits_{j=1}^d\sigma_j l_{ij}-\beta\leq\epsilon+\xi_i,\ \forall i\in\Omega\\ & \sum\limits_{j=1}^d\sigma_j u_{ij}-\sum\limits_{j=1}^d\tau_j l_{ij}+\beta-\tilde{l}_i\leq\epsilon+\xi_i^*,\ \forall i\in\Omega\\ & \xi_i,\xi_i^*,\sigma_j,\tau_j\geq 0,\ \forall i\in\Omega,\ j=1,\ldots,d.\end{array} \tag{27}$$

Proof.
For this proof, we need to observe that, for $s\in\mathbb{R}^d$, if $\gamma_i(s)=\max\limits_{j=1,\ldots,d}\ \dfrac{2|s_j|}{u_{ij}-l_{ij}}$ (for an object i of the database), then its dual gauge is

$$\gamma_i^0(s)=\sum_{j=1}^d\ \frac{u_{ij}-l_{ij}}{2}|s_j|. \tag{28}$$

Let us start with the set of constraints (25). If we replace $x_{ij}=\dfrac{l_{ij}+u_{ij}}{2}$, $j=1,\ldots,d$ and $\gamma_i^0(\omega)=\sum\limits_{j=1}^d|\omega_j|\dfrac{u_{ij}-l_{ij}}{2}$, one obtains the following constraints,

$$\tilde{u}_i-\sum_{j=1}^d\omega_j(\frac{l_{ij}+u_{ij}}{2})+\sum_{j=1}^d|\omega_j|(\frac{u_{ij}-l_{ij}}{2})-\beta\leq\epsilon+\xi_i,\ \forall i\in\Omega.$$

Let us define $\sigma_j=\max\{0,\omega_j\}$ and $\tau_j=\max\{0,-\omega_j\}$, for $j=1,\ldots,d$. One has that $\omega_j=\sigma_j-\tau_j$ and $|\omega_j|=\sigma_j+\tau_j$, and the constraints can be written as

$$\tilde{u}_i-\sum_{j=1}^d(\sigma_j-\tau_j)(\frac{l_{ij}+u_{ij}}{2})+\sum_{j=1}^d(\sigma_j+\tau_j)(\frac{u_{ij}-l_{ij}}{2})-\beta\leq\epsilon+\xi_i,\ \forall i\in\Omega.$$

which yields

$$\tilde{u}_i+\sum_{j=1}^d\tau_j u_{ij}-\sum_{j=1}^d\sigma_j l_{ij}-\beta\leq\epsilon+\xi_i,\ \forall i\in\Omega. \tag{29}$$

We proceed analogously with the set of constraints (26): we replace the values of x_i and $\gamma_i^0(\omega)$ and we obtain

$$\sum_{j=1}^{d} \omega_j (\frac{l_{ij}+u_{ij}}{2}) + \sum_{j=1}^{d} |\omega_j| (\frac{u_{ij}-l_{ij}}{2}) + \beta - \tilde{l}_i \leq \epsilon + \xi_i^*, \ \forall i \in \Omega.$$

Afterwards, we introduce the variables σ_j and τ_j, and after some computations, one obtains

$$\sum_{j=1}^{d} \sigma_j u_{ij} - \sum_{j=1}^{d} \tau_j l_{ij} + \beta - \tilde{l}_i \leq \epsilon + \xi_i^*, \ \forall i \in \Omega. \tag{30}$$

Joining constraints (29) and (30), we can rewrite our problem and we derive formulation (27). □

Remark 3.1. *When γ_i was defined in (6), we assumed that $l_{ij} < u_{ij}$, $\forall j = 1, \ldots, d$. In the case of degenerated boxes (that is, when $l_{ij} = u_{ij}$ for some coordinates), denote by J_F the set of indexes with $l_{ij} = u_{ij}$ and denote by J_V the set of indexes with $l_{ij} < u_{ij}$. Let us define γ_i as*

$$\gamma_i(s_1, \ldots, s_d) = \begin{cases} \max\limits_{j \in J_V} \dfrac{2|s_j|}{u_{ij} - l_{ij}}, & \text{if } s_j = 0, \ \forall j \in J_F \\ +\infty, & \text{otherwise.} \end{cases} \tag{31}$$

One has that $\gamma_i^0(s)$ has the same form as (28) and then, formulation (27) remains valid.

Remark 3.2. *When uncertainty only affects to the dependent variable Y_i, and then the predictive variables are single-valued, that is, $l_{ij} = u_{ij} = x_{ij}$, $\forall i \in \Omega$, $\forall j = 1, \ldots, d$, Problem (27) can be rewritten as*

$$\begin{aligned} \min_{\omega, \beta, \xi, \xi^*} \quad & \frac{1}{2} \sum_{j=1}^{d} \omega_j^2 + C \sum_{i \in \Omega} (\xi_i + \xi_i^*) \\ s.t. \quad & \tilde{u}_i - \sum_{j=1}^{d} \omega_j x_{ij} - \beta \leq \epsilon + \xi_i, \ \forall i \in \Omega \\ & \sum_{j=1}^{d} \omega_j x_{ij} + \beta - \tilde{l}_i \leq \epsilon + \xi_i^*, \ \forall i \in \Omega \\ & \xi_i, \xi_i^* \geq 0, \ \forall i \in \Omega, \ j = 1, \ldots, d. \end{aligned} \tag{32}$$

Likewise, if uncertainty only affects to the predictive variables and the dependent variable is single-valued, that is, $\tilde{l}_i = \tilde{u}_i = y_i$, $\forall i \in \Omega$, the problem to solve is

$$\begin{aligned} \min_{\sigma, \tau, \beta, \xi, \xi^*} \quad & \frac{1}{2} \sum_{j=1}^{d} (\sigma_j - \tau_j)^2 + C \sum_{i \in \Omega} (\xi_i + \xi_i^*) \\ s.t. \quad & y_i + \sum_{j=1}^{d} \tau_j u_{ij} - \sum_{j=1}^{d} \sigma_j l_{ij} - \beta \leq \epsilon + \xi_i, \ \forall i \in \Omega \\ & \sum_{j=1}^{d} \sigma_j u_{ij} - \sum_{j=1}^{d} \tau_j l_{ij} + \beta - y_i \leq \epsilon + \xi_i^*, \ \forall i \in \Omega \\ & \xi_i, \xi_i^*, \sigma_j, \tau_j \geq 0, \ \forall i \in \Omega, \ j = 1, \ldots, d. \end{aligned} \tag{33}$$

When we consider γ_i as defined in (8), we obtain our second case, which is of interest to model the case with data affected by some kind of perturbations. Then, by observing that the dual gauge of $\gamma_i = \frac{1}{r_i}\|.\|_p$ is $\gamma_i^0 = r_i\|.\|_q$ (with $\|.\|_q$ the dual norm of $\|.\|_p$, p and q satisfying that $\frac{1}{p}+\frac{1}{q}=1$), we obtain the following result, previously derived by [41, 42], as a straightforward consequence of our Theorem 3.1.

Corollary 3.2. *Let γ_i be the gauge defined in (8). Then, Problem (22) admits the following equivalent formulation,*

$$\begin{array}{ll} \min\limits_{\omega,\beta,\xi,\xi^*} & \frac{1}{2}\sum\limits_{j=1}^{d}\omega_j^2 + C\sum\limits_{i\in\Omega}(\xi_i+\xi_i^*) \\ s.t. & \tilde{u}_i - \omega^\top x_i + r_i\|\omega\|_q - \beta \le \epsilon + \xi_i,\ \forall i\in\Omega \\ & \omega^\top x_i + r_i\|\omega\|_q + \beta - \tilde{l}_i \le \epsilon + \xi_i^*,\ \forall i\in\Omega \\ & \xi_i,\xi_i^* \ge 0,\ \forall i\in\Omega, \end{array} \tag{34}$$

where $\|.\|_q$ is the dual norm of $\|.\|_p$, i.e., $\frac{1}{p}+\frac{1}{q}=1$.

This formulation (34), for $\tilde{l}_i = \tilde{u}_i = y_i$ (when the output is crisp) is equivalent to that given in [41, 42]. In these two papers, the authors formulate this problem by building the robust counterpart of the problem (by using robust optimization methods, [2, 3]) and they solve the problem in the Euclidean norm case, that is, for $p = q = 2$. Our formulation (22) for any kind of gauge γ_i is thus much more general than that obtained for Support Vector Regression with noisy data.

4 Formulation based on the Hausdorff distance

4.1 Formulation of the problem

Figure 4 explains graphically the model with the Hausdorff distance d_H. In this case, ξ and ξ^* penalize the case when the distance from $\tilde{u}_i$ to the highest value of the interval obtained projecting the box on the hyperplane is bigger than ϵ (ξ for points above the tube, ξ^* for points below the tube). Analogously, η and η^* are penalties for the distances between $\tilde{l}_i$ and the lowest value of the projection on the hyperplane.
If we use the distance d_H in (11) as a measurement between the predicted and the real interval-valued output, constraint (12) can be written as

$$\max\left\{|\tilde{u}_i - \max_{x\in X_i}(\omega^\top x+\beta)|, |\tilde{l}_i - \min_{x\in X_i}(\omega^\top x+\beta)|\right\} \le \epsilon,\ \forall i\in\Omega. \tag{35}$$

This is equivalent to say that

$$|\tilde{u}_i - \max_{x\in X_i}(\omega^\top x+\beta)| \le \epsilon, \quad \forall i\in\Omega \tag{36}$$

$$|\tilde{l}_i - \min_{x\in X_i}(\omega^\top x+\beta)| \le \epsilon, \quad \forall i\in\Omega. \tag{37}$$

Constraints (36)-(37) can be divided into

$$\tilde{u}_i - \max_{x \in X_i} \omega^\top x - \beta \leq \epsilon, \quad \forall i \in \Omega \tag{38}$$

$$\max_{x \in X_i} \omega^\top x + \beta - \tilde{u}_i \leq \epsilon, \quad \forall i \in \Omega \tag{39}$$

$$\tilde{l}_i - \min_{x \in X_i} \omega^\top x - \beta \leq \epsilon, \quad \forall i \in \Omega \tag{40}$$

$$\min_{x \in X_i} \omega^\top x + \beta - \tilde{l}_i \leq \epsilon, \quad \forall i \in \Omega. \tag{41}$$

Then, when using Hausdorff distance d_H in the constraints, Problem (13) can be written as follows,

$$\begin{aligned} \min_{\omega,\beta} \quad & \frac{1}{2} \sum_{j=1}^{d} \omega_j^2 \\ \text{s.t.} \quad & \tilde{u}_i - \max_{x \in X_i} \omega^\top x - \beta \leq \epsilon, \ \forall i \in \Omega \\ & \max_{x \in X_i} \omega^\top x + \beta - \tilde{u}_i \leq \epsilon, \ \forall i \in \Omega \\ & \tilde{l}_i - \min_{x \in X_i} \omega^\top x - \beta \leq \epsilon, \ \forall i \in \Omega \\ & \min_{x \in X_i} \omega^\top x + \beta - \tilde{l}_i \leq \epsilon, \ \forall i \in \Omega. \end{aligned} \tag{42}$$

As we did in Section 3, a soft-margin version, feasible even when (42) is unfeasible, is obtained here by adding slack variables ξ, ξ^*, η, η^* as follows,

$$\begin{aligned} \min_{\omega,\beta,\xi,\xi^*,\eta,\eta^*} \quad & \frac{1}{2} \sum_{j=1}^{d} \omega_j^2 + C \sum_{i \in \Omega} (\xi_i + \xi_i^* + \eta_i + \eta_i^*) \\ \text{s.t.} \quad & \tilde{u}_i - \max_{x \in X_i} \omega^\top x - \beta \leq \epsilon + \xi_i, \ \forall i \in \Omega && (43) \\ & \max_{x \in X_i} \omega^\top x + \beta - \tilde{u}_i \leq \epsilon + \xi_i^*, \ \forall i \in \Omega && (44) \\ & \tilde{l}_i - \min_{x \in X_i} \omega^\top x - \beta \leq \epsilon + \eta_i, \ \forall i \in \Omega && (45) \\ & \min_{x \in X_i} \omega^\top x + \beta - \tilde{l}_i \leq \epsilon + \eta_i^*, \ \forall i \in \Omega && (46) \\ & \xi_i, \xi_i^*, \eta_i, \eta_i^* \geq 0, \ \forall i \in \Omega. \end{aligned}$$

4.2 An equivalent formulation

By observing that $X_i = x_i + Bi$ (with B_i the unit ball induced by the gauge γ_i) and by using expressions (23)-(24) in constraints (43)-(46), then, following an analogous reasoning to that used in proof of Theorem 3.1, we obtain the following equivalent formulation.

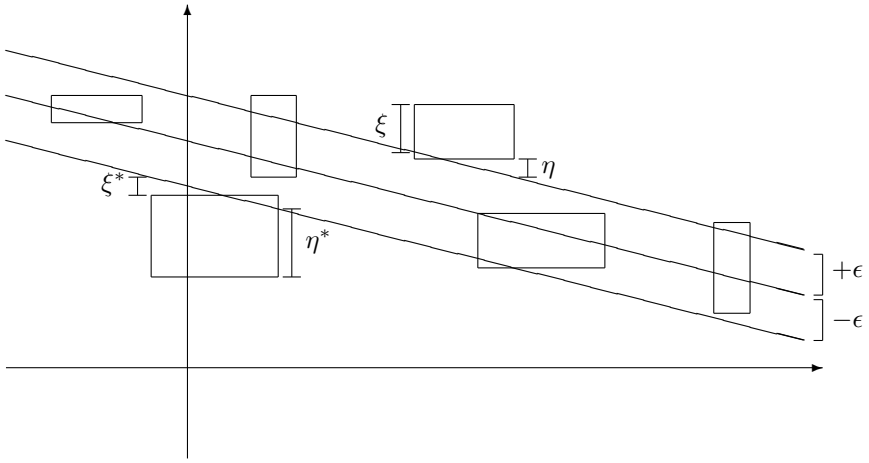

Figure 4: Formulation based on Hausdorff distance

Theorem 4.1. *Problem with constraints (43)-(44) admits the following equivalent formulation,*

$$\begin{array}{ll} \min\limits_{\omega,\beta,\xi,\xi^*,\eta,\eta^*} & \frac{1}{2}\sum\limits_{j=1}^{d}\omega_j^2 + C\sum\limits_{i\in\Omega}(\xi_i+\xi_i^*+\eta_i+\eta_i^*) \\ s.t. & \tilde{u}_i - \omega^\top x_i - \gamma_i^0(\omega) - \beta \le \epsilon + \xi_i,\ \forall i\in\Omega \\ & \omega^\top x_i + \gamma_i^0(\omega) + \beta - \tilde{u}_i \le \epsilon + \xi_i^*,\ \forall i\in\Omega \\ & \tilde{l}_i - \omega^\top x_i + \gamma_i^0(\omega) - \beta \le \epsilon + \eta_i,\ \forall i\in\Omega \\ & \omega^\top x_i - \gamma_i^0(\omega) + \beta - \tilde{l}_i \le \epsilon + \eta_i^*,\ \forall i\in\Omega \\ & \xi_i,\xi_i^*,\eta_i,\eta_i^* \ge 0,\ \forall i\in\Omega, \end{array} \tag{47}$$

where γ_i is the gauge associated to the object $i\in I$ defining the ball B_i used in $X_i = x_i + B_i$ and γ_i^0 is its dual gauge.

We consider now the cases for the definitions of γ_i given in (6) and (8). In the first case the objects are boxes in dimension d, that is, $X_i = \prod_{j=1}^{d}[l_{ij}, u_{ij}]$, for every $i\in\Omega$.

Corollary 4.1. *Let γ_i be the gauge defined in (6). Then, Problem (47) admits the following equivalent formulation as a convex quadratic problem with linear and equilibrium*

constraints,

$$
\begin{array}{ll}
\min\limits_{\sigma,\tau,\beta,\xi,\xi^*,\eta,\eta^*} & \dfrac{1}{2}\sum\limits_{j=1}^{d}(\sigma_j-\tau_j)^2 + C\sum\limits_{i\in\Omega}(\xi_i+\xi_i^*+\eta_i+\eta_i^*) \\
s.t. & \tilde{u}_i - \sum\limits_{j=1}^{d}\sigma_j u_{ij} + \sum\limits_{j=1}^{d}\tau_j l_{ij} - \beta \le \epsilon + \xi_i,\ \forall i\in\Omega \\
& \sum\limits_{j=1}^{d}\sigma_j u_{ij} - \sum\limits_{j=1}^{d}\tau_j l_{ij} + \beta - \tilde{u}_i \le \epsilon + \xi_i^*,\ \forall i\in\Omega \\
& \tilde{l}_i - \sum\limits_{j=1}^{d}\sigma_j l_{ij} + \sum\limits_{j=1}^{d}\tau_j u_{ij} - \beta \le \epsilon + \eta_i,\ \forall i\in\Omega \\
& \sum\limits_{j=1}^{d}\sigma_j l_{ij} - \sum\limits_{j=1}^{d}\tau_j u_{ij} + \beta - \tilde{l}_i \le \epsilon + \eta_i^*,\ \forall i\in\Omega \\
& \sigma_j\cdot\tau_j = 0,\ j=1,\dots,d \\
& \xi_i,\xi_i^*,\eta_i,\eta_i^*,\sigma_j,\tau_j \ge 0,\ \forall i\in\Omega,\ j=1,\dots,d.
\end{array}
\tag{48}
$$

Remark 4.1. *Since we define* $\sigma_j = \max\{0,\omega_j\}$ *and* $\tau_j = \max\{0,-\omega_j\}$*, for* $j=1,\dots,d$*, we have imposed the following constraint*

$$\sigma_j\cdot\tau_j = 0,\ j=1,\dots,d. \tag{49}$$

In principle, equilibrium constraints should have also been added to Problem (27). However, they are redundant due to the convexity of the problem.

Remark 4.2. *When* γ_i *was defined in (6), we assumed that* $l_{ij} < u_{ij}$*,* $\forall j=1,\dots,d$*. In the case of degenerated boxes (that is, when* $l_{ij} = u_{ij}$ *for some coordinates),* γ_i *can be defined as in (31), and* $\gamma_i^0(s)$ *has the same form as (28). Then, formulation (48) remains valid.*

Remark 4.3. *If uncertainty only affects to* Y_i*, and* $l_{ij} = u_{ij} = x_{ij}$*,* $\forall i\in\Omega$*,* $\forall j=1,\dots,d$*, the problem to solve is the following convex quadratic problem with linear constraints*

$$
\begin{array}{ll}
\min\limits_{\omega,\beta,\xi,\xi^*,\eta,\eta^*} & \dfrac{1}{2}\sum\limits_{j=1}^{d}\omega_j^2 + C\sum\limits_{i\in\Omega}(\xi_i+\xi_i^*+\eta_i+\eta_i^*) \\
s.t. & \tilde{u}_i - \sum\limits_{j=1}^{d}\omega_j x_{ij} - \beta \le \epsilon + \xi_i,\ \forall i\in\Omega \\
& \sum\limits_{j=1}^{d}\omega_j x_{ij} + \beta - \tilde{u}_i \le \epsilon + \xi_i^*,\ \forall i\in\Omega \\
& \tilde{l}_i - \sum\limits_{j=1}^{d}\omega_j x_{ij} - \beta \le \epsilon + \eta_i,\ \forall i\in\Omega \\
& \sum\limits_{j=1}^{d}\omega_j x_{ij} + \beta - \tilde{l}_i \le \epsilon + \eta_i^*,\ \forall i\in\Omega \\
& \xi_i,\xi_i^*,\eta_i,\eta_i^* \ge 0,\ \forall i\in\Omega,\ j=1,\dots,d.
\end{array}
\tag{50}
$$

Likewise, if uncertainty only affects to the predictive variables and $\tilde{l}_i = \tilde{u}_i = y_i$, $\forall i \in \Omega$, *Problem (48) can be written as the following convex quadratic problem with linear and equilibrium constraints*

$$\begin{aligned}
\min_{\sigma,\tau,\beta,\xi,\xi^*,\eta,\eta^*} \quad & \frac{1}{2}\sum_{j=1}^{d}(\sigma_j - \tau_j)^2 + C\sum_{i\in\Omega}(\xi_i + \xi_i^* + \eta_i + \eta_i^*) \\
s.t. \quad & y_i - \sum_{j=1}^{d}\sigma_j u_{ij} + \sum_{j=1}^{d}\tau_j l_{ij} - \beta \leq \epsilon + \xi_i, \ \forall i \in \Omega \\
& \sum_{j=1}^{d}\sigma_j u_{ij} - \sum_{j=1}^{d}\tau_j l_{ij} + \beta - y_i \leq \epsilon + \xi_i^*, \ \forall i \in \Omega \\
& y_i - \sum_{j=1}^{d}\sigma_j l_{ij} + \sum_{j=1}^{d}\tau_j u_{ij} - \beta \leq \epsilon + \eta_i, \ \forall i \in \Omega \\
& \sum_{j=1}^{d}\sigma_j l_{ij} - \sum_{j=1}^{d}\tau_j u_{ij} + \beta - y_i \leq \epsilon + \eta_i^*, \ \forall i \in \Omega \\
& \sigma_j \cdot \tau_j = 0, \ j = 1, \ldots, d \\
& \xi_i, \xi_i^*, \eta_i, \eta_i^*, \sigma_j, \tau_j \geq 0, \ \forall i \in \Omega, \ j = 1, \ldots, d.
\end{aligned} \tag{51}$$

Corollary 4.2. *Let* γ_i *be the gauge defined in (8). Then, Problem (47) admits the following equivalent formulation,*

$$\begin{aligned}
\min_{\omega,\beta,\xi,\xi^*,\eta,\eta^*} \quad & \frac{1}{2}\sum_{j=1}^{d}\omega_j^2 + C\sum_{i\in\Omega}(\xi_i + \xi_i^* + \eta_i + \eta_i^*) \\
s.t. \quad & \tilde{u}_i - \omega^\top x_i - r_i\|\omega\|_q - \beta \leq \epsilon + \xi_i, \ \forall i \in \Omega \\
& \omega^\top x_i + r_i\|\omega\|_q + \beta - \tilde{u}_i \leq \epsilon + \xi_i^*, \ \forall i \in \Omega \\
& \tilde{l}_i - \omega^\top x_i + r_i\|\omega\|_q - \beta \leq \epsilon + \eta_i, \ \forall i \in \Omega \\
& \omega^\top x_i - r_i\|\omega\|_q + \beta - \tilde{l}_i \leq \epsilon + \eta_i^*, \ \forall i \in \Omega \\
& \xi_i, \xi_i^*, \eta_i, \eta_i^* \geq 0, \ \forall i \in \Omega,
\end{aligned} \tag{52}$$

where $\|.\|_q$ *is the dual norm of* $\|.\|_p$, *i.e.,* $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

5 Computational experiment with interval data

5.1 Error measures

For the numerical experiments, different measurements of the fitness of the model will be considered in each case (the standard measurements used in the literature of regression with interval data in the framework of Symbolic Data Analysis, [14, 30, 29]). They are obtained from the observed intervals $Y_i = [\tilde{l}_i, \tilde{u}_i]$ and the corresponding predicted intervals $\hat{Y}_i = [\hat{l}_i, \hat{u}_i]$, $i \in \Omega$. The measurements are the *lower bound root mean-squared error* ($RMSE_l$) and the *upper bound root mean-squared error* ($RMSE_u$), which are defined as

Pulse rate	Systolic blood pressure	Diastolic blood pressure
[44, 68]	[90, 100]	[50, 70]
[60, 72]	[90, 130]	[70, 90]
[56, 90]	[140, 180]	[90, 100]
[70, 112]	[110, 142]	[80, 108]
[54, 72]	[90, 100]	[50, 70]
[70, 100]	[130, 160]	[80, 110]
[72, 100]	[130, 160]	[76, 90]
[76, 98]	[110, 190]	[70, 110]
[86, 96]	[138, 180]	[90, 110]
[86, 100]	[110, 150]	[78, 100]
[63, 75]	[60, 100]	[140, 150]

Table 1: Cardiological interval-valued database

follows,

$$RMSE_l = \sqrt{\frac{1}{n}\sum_{i\in\Omega}(\tilde{l}_i - \hat{l}_i)^2} \tag{53}$$

$$RMSE_u = \sqrt{\frac{1}{n}\sum_{i\in\Omega}(\tilde{u}_i - \hat{u}_i)^2}, \tag{54}$$

where $\tilde{l} = (\tilde{l}_1, \ldots, \tilde{l}_n)^\top$, $\hat{l} = (\hat{l}_1, \ldots, \hat{l}_n)^\top$, $\tilde{u} = (\tilde{u}_1, \ldots, \tilde{u}_n)^\top$, $\hat{u} = (\hat{u}_1, \ldots, \hat{u}_n)^\top$, with n the cardinal of Ω.
Another measurement which we introduce to compute the fitness is the *mean Hausdorff distance* ($\overline{d}_H$), between the observed and predicted intervals, for the elements of the database, defined as

$$\overline{d}_H = \frac{1}{n}\sum_{i\in\Omega} d_H([\tilde{l}_i, \tilde{u}_i], [\hat{l}_i, \hat{u}_i]). \tag{55}$$

5.2 Results for resubstitution

We apply our methodology to solve the regression problem with interval data in a cardiological database. The first results for the regression analysis with this dataset were published in [4]. This dataset shows the records of the pulse rate, the systolic blood pressure and the diastolic blood pressure (these records being intervals) for eleven patients (see Table 1). The aim of the problem is to predict an interval for the 'pulse' variable, given the interval values of the 'systolic' and 'diastolic pressure' variables.
First of all, we compute the predicted interval for the 'pulse' variable via a resubstitution strategy (see [15]), that is, the complete set of instances will be our training sample, the regressor will be computed and we will assign the predicted interval to each patient of the training sample.
In Table 2 (left), we show the results obtained for different methods in the literature. CM stands for the center method explained in [4]. In that work, a linear regression model on the midpoint of the intervals was applied. MinMax [5] is a methodology where two models

	Resubstitution		Leave-one-out	
Method\ Measure	$RMSE_l$	$RMSE_u$	$RMSE_l$	$RMSE_u$
CM	11.09	10.41	24.78	28.41
MinMax	10.43	9.71	14.82	22.25
CRM	9.81	8.94	12.81	20.37
Interval ϵ-SVR	11.03	10.31	12.71	11.89

Table 2: Results via resubstitution (left) and leave-one-out (right) for the cardiological interval-valued database

Real value	CM	MinMax	CRM	Interval ϵ-SVR
[44, 68]	[59, 66]	[56, 72]	[52, 74]	[57, 65]
[60, 72]	[63, 79]	[60, 84]	[60, 82]	[62, 80]
[56, 90]	[83, 97]	[77, 100]	[80, 100]	[83, 99]
[70, 112]	[71, 86]	[67, 89]	[67, 90]	[71, 88]
[54, 72]	[59, 66]	[56, 72]	[52, 74]	[57, 65]
[70, 100]	[78, 92]	[73, 95]	[73, 97]	[78, 95]
[72, 100]	[77, 89]	[72, 93]	[73, 93]	[77, 90]
[76, 98]	[69, 102]	[65, 104]	[73, 99]	[69, 105]
[86, 96]	[82, 99]	[77, 101]	[79, 101]	[83, 102]
[86, 100]	[71, 87]	[67, 91]	[68, 90]	[70, 89]
[63, 75]	[65, 80]	[66, 81]	[62, 82]	[66, 82]

Table 3: Predicted values of 'pulse' variable

are fitted independently for the lower and the upper bounds. CRM stands for the center and range method in [14, 29], two linear independent models were used to predict the center and the range of the interval outputs and, this way, to build the predictions of the lower and upper bounds. We also present the best results obtained via our methodology (interval ϵ-SVR).

From these four methods, the best performance is obtained with CRM. Although our results are worse (for resubstitution) than those obtained with CRM, they are comparable in general with those obtained via the classical regression model. From this, one can think that a methodology based on ϵ-SVR can be competitive for this problem.

Table 3 shows the real interval values and the predicted outputs for the 'pulse' variable for these four methods.

For our methodology, the formulations based on the maximum distance and on the Hausdorff distance (in the interval case) were used, but the results corresponds to the best result (which was for Hausdorff-based formulation). Since the problems for the maximum distance were quadratic convex, they were solved by using AMPL+CPLEX. For the programs with the Hausdorff distance we had to use, however, AMPL+MINOS.

5.3 Results for leave-one-out

The next experiment shows the performance of the regressor built via our methodology when using a leave-one-out (LOO) strategy (see e.g. [20, 27]). It means that, in turns,

C	0.00001		0.0001		0.001		0.01		0.1	
$\epsilon \backslash RMSE$	l	u	l	u	l	u	l	u	l	u
0.0001	15.17	21.66	15.24	20.33	13.08	12.77	15.66	15.46	21.48	20.33
0.001	15.17	21.66	15.35	20.35	13.08	12.77	15.66	15.46	21.48	20.33
0.01	15.15	21.60	15.34	20.35	13.08	12.76	15.66	15.46	21.48	20.33
0.1	15.15	21.59	15.31	20.32	13.04	12.76	15.66	15.46	21.49	20.38
0.5	15.08	21.43	15.12	20.04	13.04	12.61	15.86	15.66	21.51	20.56
1	15.28	21.29	15.43	19.94	13.02	12.29	15.84	15.84	21.57	20.65
1.5	15.40	21.31	15.70	19.68	13.02	12.08	15.48	15.28	21.84	20.87
2	15.63	20.98	15.87	19.53	12.89	11.90	14.51	14.15	22.04	20.87
2.5	16.00	20.88	15.96	19.63	**12.71**	**11.89**	14.25	13.86	20.68	19.67
3	16.40	20.76	15.68	19.43	13.87	12.17	14.40	13.92	19.30	18.54
3.5	16.66	20.44	15.96	19.14	12.94	12.17	14.76	14.16	19.19	18.39
5	17.10	19.40	16.55	18.14	12.79	12.61	13.59	13.21	21.79	20.70
7	17.45	17.90	16.85	16.76	13.42	13.12	14.72	13.29	21.16	20.16
10	18.38	17.24	17.74	16.61	14.28	13.11	14.25	12.65	19.67	18.57

Table 4: $RMSE_l$ and $RMSE_u$ for the cardiological database via leave-one-out

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. We compute the error between the real output and the predicted output and we repeat the process for every element of the database. The fitness of the model will be studied via one of the measurements (53)-(54), and via the mean Hausdorff distance as well. The LOO strategy is more interesting than the resubstitution situation because it gives an idea of the behaviour of the regressor for new possible observations.

The regression problem has been solved for several combinations of the parameters C and ϵ, namely, for every pair (C, ϵ), with $C = 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$, and $\epsilon = 0.0001, 0.001, 0.01, 0.1, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 5, 7, 10$. We have considered the two choices of d, but the results that we present belong to d_H, because they are systematically better than those obtained with d_{max}.

Table 4 displays the results for the measurements $RMSE_l$ and $RMSE_u$ for the different combinations of the parameters. One can observe that the results obtained in the case of $C = 0.001$ and $C = 0.01$ are better than the rest (especially, the former). The best results for these measurements have been marked in bold in the table.

Table 5 shows the results obtained when we use the mean Hausdorff distance (55) to measure the error between the predicted interval and the real one to study the fitness of our model to the data. Good values can be found again for $C = 0.001, 0.01$ and the lowest value of the distance is in bold.

Finally, in Table 2 (right), a comparison for the measurements (53)-(54) obtained for the cardiological dataset via different methods is given. In particular, we present the results obtained for CM ([4]), MinMax ([5]), CRM ([14, 29]) and our methodology.

One can observe that the results obtained with our method are better than with the other models. In fact, attending to the $RMSE_l$ and $RMSE_u$ measurements, any result for $C = 0.001$ or $C = 0.01$ would be better than those obtained so far in the literature. The other methods were good in general with the training sample, but the error is bigger in

$\epsilon \backslash C$	0.00001	0.0001	0.001	0.01	0.1
0.0001	23.74	23.03	14.92	16.11	18.47
0.001	23.74	23.08	14.92	16.11	18.47
0.01	23.69	23.08	14.91	16.11	18.47
0.1	23.71	23.04	14.89	16.10	18.50
0.5	23.68	22.76	14.84	16.15	18.60
1	23.82	22.89	14.65	16.10	18.57
1.5	23.93	22.91	14.49	15.82	18.62
2	23.89	22.94	14.31	15.24	18.52
2.5	24.04	23.07	14.26	14.94	17.76
3	24.18	22.71	14.98	15.01	17.11
3.5	24.13	22.69	14.68	15.13	16.94
5	23.70	22.40	15.05	**14.15**	17.58
7	22.95	21.67	15.90	14.78	17.47
10	22.84	22.09	16.07	14.69	17.03

Table 5: Mean Hausdorff distance ($\overline{d}_H$) between the predicted interval and the real interval (leave-one-out)

the test sample, due to a problem of overfitting. The improvement with respect to CRM, which was the best result so far, is quite remarkable in $RMSE_l$ and $RMSE_u$, and thus, one can conclude that our method is competitive to deal with regression with interval data.

6 Computational experiment with missing data

6.1 Imputation for missing values via intervals

The term 'missing data' is used, when editing survey data, to denote invalid blanks in an entry of any field of the survey (invalid in the sense that this value should appear).
Several strategies can be adopted when handling missing data, such as the imputation for given records, which means to replace the missing values of a dataset by other plausible values in such a way that the data must remain consistent.
In the literature, different methodologies for imputation can be found (see [31, 32, 34] for a list of them), like the use of the mean (for quantitative variables) or the mode (for qualitative variables) of the non-missing values of the database (see e.g. [1]). One of the drawbacks of this method is that the variability within the sample is ignored and it can contain relevant information which should be taken into account during the imputation process (see [36]).
The methodology that we propose for imputation consists in replacing each blank by an interval (instead of a single value) constructed with the non-missing values of the dataset. That is, if a blank appears in the j-th variable of an observation, the non-missing values in the j-th variable of the rest of observations are used to build the interval. Two different strategies will be followed to construct these intervals.
The first one is to build an interval based on the mean and deviation of the remaining values. This way, the standard deviation will have an important role in the imputation

of missing values. For a missing value in the j-th variable of an instance of the training sample, we fill it in by computing the mean $\overline{x}_j$ and the standard deviation σ_{x_j} for the values in this column of the remaining observations, and afterwards, we replace the blank by the interval $[\overline{x}_j - k\sigma_{x_j}, \overline{x}_j + k\sigma_{x_j}]$, where k is a parameter to be tuned.
The second strategy is based on the quantiles. We consider the interval which is defined as $[Q_a, Q_{1-a}]$, where Q_a represents the a-th quantile, and thus the interval contains all but a fraction $2a$ of all non-missing values.

6.2 Description of the experiment

Our formulation for regression with interval data has been applied to a real database, obtained from the UCI Machine Learning Repository [7], for dealing with missing values. The 'automobile' database contains 205 observations. Each record describes different characteristics of a determined automobile, with nominal and numerical variables. However, the nominal variables have been discarded for our experiment and thus, each observation is represented by 16 numerical variables: 15 of them will be the predictor variables and the last one, which is the price of the car, will be the dependent variable to be approximated via regression. There are several missing values, in some predictor variables (variables 2, 9, 10, 12 and 13) and in the dependent variable as well, which will be imputed via an interval.
The regression problem has been solved through 10-fold cross-validation (see [27]), that is, the elements of the database are grouped in 10 sets, which form a partition, and each one has been used in turn as test set against all 9 others taken together as training set (that is, the process is repeated ten times).
We have used the formulation using d_H. Before solving the corresponding optimization problem (48), the two different strategies explained before have been used for imputing the missing values. In the first strategy, we replace the blank by the interval $[\overline{x}_j - k\sigma_{x_j}, \overline{x}_j + k\sigma_{x_j}]$, with $\overline{x}_j$ and σ_{x_j} computed with the non-missing values in the j-th column of the elements of the training sample. The values studied for k are 0, 0.01, 0.05, 0.1, 0.2, 0.5, 0.75 and 1. Observe that the case $k = 0$ corresponds to considering the imputation to the mean.
In the second strategy, the blank is replaced by the interval $[Q_a, Q_{1-a}]$, Q_a being the a-th quantile. The values chosen for $2a$ are 0, 0.01, 0.05, 0.1, 0.2, 0.5 and 1. Observe that, when $2a = 0$, we obtain the range of the variable, when $2a = 0.5$, we obtain the interquartile range, and when $2a = 1$, the interval is reduced to a singleton, which is the median of the variable.
For each record of the database, we obtain a predicted interval $\hat{Y}_i = [\hat{l}_i, \hat{u}_i]$. Since the values of the dependent variable (the 'price' of the car) are punctual, we compute $\hat{y}_i = \frac{\hat{l}_i + \hat{u}_i}{2}$, the midpoint of the bounds of the interval, and we use it to compare the predicted and the real values for the dependent variable.
Two measurements have been chosen to compute the fitness of our model in this database:

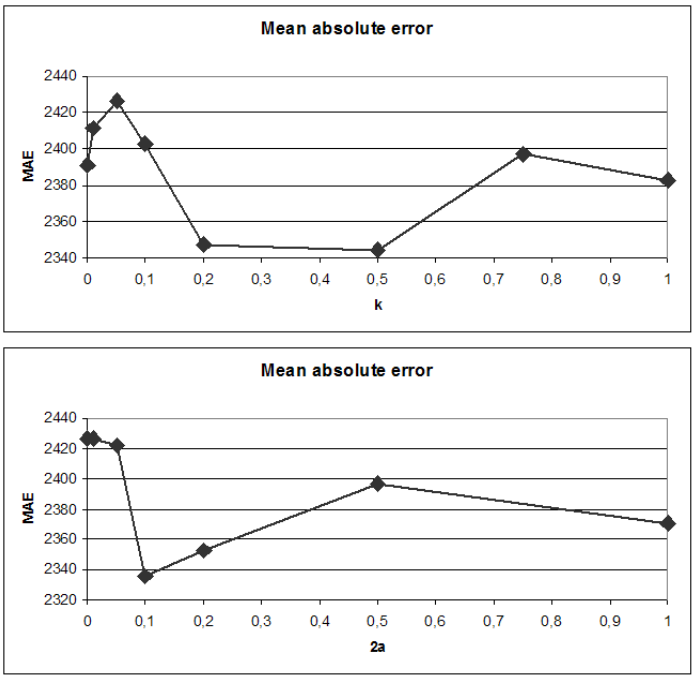

Figure 5: Best results for the mean absolute error. Up: interval $[\overline{x}_j - k\sigma_{x_j}, \overline{x}_j + k\sigma_{x_j}]$. Down: interval $[Q_a, Q_{1-a}]$

the *mean absolute error* (MAE) and the *root mean-squared error* (RMSE), defined as

$$MAE = \frac{1}{n}\sum_{i\in\Omega}|y_i - \hat{y}_i| \tag{56}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i\in\Omega}(y_i - \hat{y}_i)^2}, \tag{57}$$

between the predicted value $\hat{y}_i$ and the real value y_i of the variable 'price' in the database. There are four records in the database with a missing value in this variable. These missing values have been transformed into an interval for the experiment to build the hyperplanes, but they are not taken into account to measure the fitness.
The imputation process and all the modifications of the database have been performed with Matlab 6.5. The optimization problems have been implemented with AMPL and solved with LOQO [43] (by using the NEOS server, [35]). Different combinations of the parameters C and ϵ have been considered.

6.3 Numerical results

The results for MAE for the different intervals are displayed in Tables 6-9 and for RMSE in Tables 10-13. The best results for each k and a are shown in bold and are depicted in Figures 5-6.
One can observe that the best results are obtained, in both imputation strategies, for non-degenerate intervals with medium-size intervals. When imputing by $[\overline{x}_j - k\sigma_{x_j}, \overline{x}_j + k\sigma_{x_j}]$,

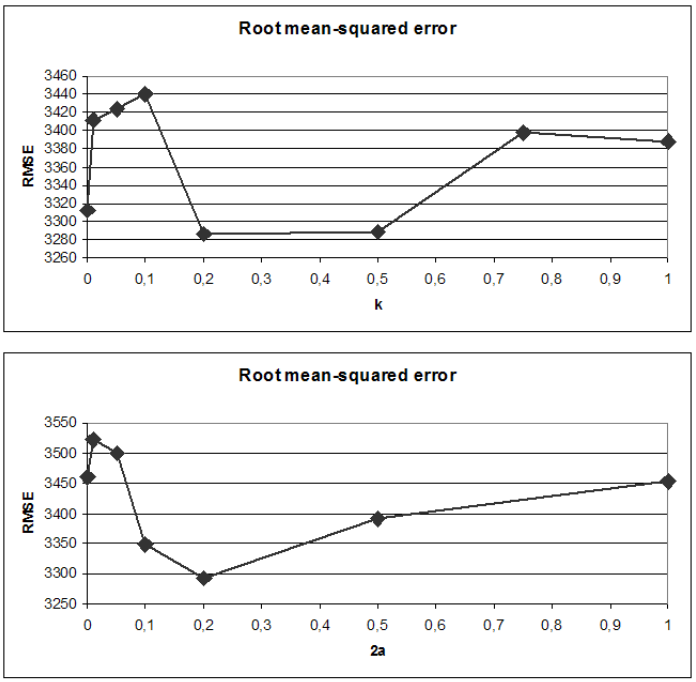

Figure 6: Best results for the root mean-squared error. Up: interval $[\overline{x}_j - k\sigma_{x_j}, \overline{x}_j + k\sigma_{x_j}]$. Down: interval $[Q_a, Q_{1-a}]$

although the best results for the two measurements (MAE and RMSE) are obtained for $k = 0.2$ and $k = 0.5$, and we improve the results obtained when imputing to the mean (case $k = 0$). It means that, in this case, it is better to use the value of the standard deviation for imputing the missing value than only using the mean.
The situation is quite similar when imputing by $[Q_a, Q_{1-a}]$. Better results are obtained for $2a = 0.1$ and $2a = 0.2$ than for $2a = 1$ (imputation to the median).
Concerning the values of the parameters C and ϵ, one can assert that big values of C (around 1000 or 10000) seem to be more suitable for this dataset. However, the variation is bigger in the case of the parameter ϵ.
Then, we conclude that imputation via intervals seems to be a good strategy when dealing with missing values in regression problems.

7 Concluding remarks and extensions

In this work, a regression problem based on Support Vector Regression has been studied, where the elements of the database are, instead of points, sets with certain geometrical properties. Two different formulations have been proposed, depending on the distance used to measure the error between the predicted interval and the real one: the maximum distance and the Hausdorff distance.
The obtained models generalize the standard ϵ-Support Vector Regression approach to the case of having interval-valued data. In particular, the model for the maximum distance generalizes the formulation given in [41, 42] for data with some kind of noise or

k	$\epsilon \backslash C$	0.1	1	10	100	1000	10000	100000
	0.001	7084.05	2597.63	2870.03	2437.70	2508.48	2551.69	3245.35
	0.01	2619.93	2563.47	3574.43	2445.85	2487.04	2481.80	3199.14
	0.1	2960.54	2831.10	2636.85	2394.53	2523.59	2433.77	3249.50
	1	2953.87	2639.04	2483.76	2544.29	2457.99	2445.48	3388.20
	2	4223.66	2643.36	2454.52	2476.50	2676.40	2433.98	3355.55
0	5	3082.13	3682.27	2504.87	2466.14	2774.77	3684.02	3377.33
	10	2890.56	2726.73	2624.86	2636.46	2932.34	2503.47	3259.52
	50	2960.48	3092.09	3159.90	4052.34	**2390.83**	2502.70	3385.84
	100	4636.50	2992.28	2533.54	3198.47	2580.09	2424.05	3389.89
	500	2718.73	2512.29	2463.92	2697.66	2572.34	2681.71	3371.90
	1000	2765.21	2846.93	3305.12	2499.69	2988.98	2646.16	3353.96
	1500	3305.86	2929.59	2497.11	2954.70	2767.32	2951.69	3320.76
	0.001	3607.65	6784.85	6821.07	6805.25	6998.63	2487.02	4237.73
	0.01	5540.41	6808.26	6832.51	6804.68	6647.92	2598.80	4375.45
	0.1	7261.81	6807.43	6827.32	6805.62	6630.29	4863.78	4297.54
	1	7233.58	6804.80	6821.61	7400.09	6600.58	2723.24	4383.17
	2	7130.19	6791.17	6826.25	6784.95	6592.02	2519.69	4399.95
0.01	5	6561.17	6800.90	6826.71	7341.40	6548.18	2561.43	4358.44
	10	6716.50	6798.11	6807.13	6777.38	6519.67	2501.01	4386.30
	50	3564.61	2556.67	2689.38	2561.91	2481.90	2451.19	4442.67
	100	2455.32	2716.08	2641.32	2555.88	2510.63	**2411.16**	4377.38
	500	2883.64	2823.38	4171.47	2522.09	2497.71	2642.20	4471.80
	1000	3173.87	2746.49	10541.40	6774.40	2474.22	2959.28	4421.30
	1500	2634.10	2676.56	6912.14	2467.96	2445.17	2466.92	4373.35
	0.001	2640.62	2580.88	2775.28	2485.06	2472.95	2517.13	3597.68
	0.01	3079.12	2574.54	2517.89	2785.32	2493.22	2553.24	3601.02
	0.1	2726.56	4329.75	2570.95	2426.76	2604.81	3608.90	3593.12
	1	2817.53	4008.66	2828.27	4628.60	2473.74	2548.19	3633.81
	2	2772.31	2571.82	2745.44	2637.02	2995.04	2503.41	3620.94
0.05	5	3851.89	2575.74	2541.39	2457.55	2467.97	2464.02	3035.87
	10	2615.64	2499.67	2718.85	2439.87	2685.50	2859.88	2481.40
	50	2482.93	4104.34	3117.36	3068.55	3072.74	3498.52	2689.93
	100	2741.33	2922.98	2947.70	4000.05	**2426.08**	3586.14	6873.31
	500	3942.21	2693.75	2604.93	3303.11	2458.49	3563.21	4754.05
	1000	3138.34	2493.04	2481.85	2433.61	2502.43	3538.68	4437.42
	1500	3281.92	2637.69	2900.00	2464.44	2479.78	3507.60	4627.46
	0.001	2487.27	4287.86	4272.01	4140.04	2512.39	2614.55	3886.54
	0.01	2708.81	4243.74	4285.99	2609.96	2468.04	2418.56	3863.26
	0.1	4724.23	4200.64	4264.98	2518.12	2668.89	2556.64	3972.21
	1	4107.99	4148.51	4315.94	2482.56	2607.79	2470.73	4002.59
	2	4056.07	4166.55	4212.34	2465.01	2800.15	2525.05	3862.22
0.1	5	4051.51	4191.97	4061.04	2648.82	2453.27	3981.42	4070.27
	10	4054.58	4254.53	4133.81	2643.40	2480.42	4025.49	4185.52
	50	2500.49	2564.10	2580.00	3119.76	2470.50	4041.87	3990.38
	100	2544.58	2562.04	2519.60	2495.86	2513.04	3816.27	3852.54
	500	2839.62	3011.66	2458.90	2465.17	2413.05	4136.40	4111.86
	1000	3327.81	2799.12	2643.84	**2402.48**	2669.99	4321.36	4092.51
	1500	2516.18	2751.96	3136.47	2527.16	2580.17	4231.19	3902.97

Table 6: MAE. Interval $[\overline{x}_j - k\sigma_{x_j}, \overline{x}_j + k\sigma_{x_j}]$

k	$\epsilon \backslash C$	0.1	1	10	100	1000	10000	100000
	0.001	3266.65	2487.20	3445.20	2515.65	2865.73	2536.32	4856.38
	0.01	2494.97	2491.87	2511.66	2684.68	3242.01	2522.41	4813.76
	0.1	2717.68	2408.15	2720.75	3131.61	2660.02	2547.16	4788.83
	1	2593.12	2451.42	2410.59	2452.80	**2347.55**	2908.88	4812.79
	2	3277.46	2691.19	2400.75	2439.05	3822.34	2993.63	4813.51
0.2	5	3147.73	2642.14	2494.03	2535.89	3469.24	3369.44	4796.80
	10	2723.01	3594.18	2722.38	2825.97	2515.16	2898.94	4791.76
	50	4836.95	2974.49	3491.70	2611.63	2538.02	4914.68	4808.73
	100	3177.15	2486.29	2666.38	2493.73	2491.49	4944.50	4805.36
	500	2767.91	4651.08	2347.60	2451.43	2411.56	4909.12	4828.31
	1000	3137.61	2479.42	2664.66	2459.33	2467.53	4838.63	4598.14
	1500	2597.40	2750.55	2553.94	2725.11	2454.62	4731.55	4534.28
	0.001	2687.74	2857.61	2927.70	2443.09	2520.80	2530.71	**2344.20**
	0.01	2457.83	2576.90	2578.53	2579.90	2438.63	2546.31	3353.26
	0.1	2525.85	2704.67	2423.09	2440.30	2483.32	2614.82	3272.27
	1	2698.14	2475.53	2461.57	2515.08	3074.58	2454.19	2918.04
	2	2936.45	3152.54	2460.78	2770.73	2857.64	2810.02	3490.04
0.5	5	3045.15	2863.01	2990.63	3180.89	2377.73	2775.91	3684.54
	10	2526.67	2780.05	2924.33	2391.30	2475.90	2714.20	3593.76
	50	3027.80	2405.48	3456.79	2582.61	2483.41	3453.92	3867.74
	100	2789.75	2423.10	2528.53	2548.84	2478.88	2737.95	3921.96
	500	2515.75	2504.36	2568.48	2644.91	2527.35	2989.54	3694.38
	1000	2477.65	2954.94	3576.56	2729.56	2470.91	2521.73	3763.21
	1500	3261.32	2469.13	2771.71	2486.26	2568.64	3522.15	3658.83
	0.001	4053.53	2604.81	4455.91	2436.62	2435.64	2520.19	3475.13
	0.01	3186.41	2628.14	3047.64	2440.83	2441.36	2501.16	3658.09
	0.1	3632.12	2899.63	2596.82	2463.25	2552.02	2542.94	3231.59
	1	2825.60	3755.32	2527.77	2553.73	4088.33	2873.78	2825.54
	2	2995.87	2786.43	2584.52	2759.69	3916.15	2523.66	2532.38
0.75	5	2862.27	2538.60	2431.11	2640.30	3369.97	2441.96	2471.89
	10	2707.33	2524.26	2467.86	3150.26	**2397.07**	2694.33	3133.98
	50	3717.33	2778.60	2452.70	2505.23	2438.64	2559.73	3150.64
	100	2577.34	2679.39	5498.55	2611.74	2437.13	2963.45	2585.39
	500	3187.86	3222.31	4806.54	2486.58	2403.41	2401.88	3702.54
	1000	2765.24	2659.92	2726.10	2499.20	2461.56	2769.43	2451.55
	1500	3850.97	3079.70	2691.85	2687.26	2767.25	2540.09	3004.25
	0.001	3048.78	2968.94	3617.10	2437.22	2460.91	2537.11	3231.23
	0.01	2873.22	2597.23	2441.06	**2382.82**	2447.18	2460.87	3350.50
	0.1	3156.44	2754.81	2450.10	2403.00	2431.00	2510.49	3376.16
	1	2847.69	2445.55	2783.80	2408.15	4545.61	2536.21	3343.22
	2	2602.67	2816.98	2554.30	2547.31	4510.02	2481.35	3358.79
1	5	3725.82	2868.65	2610.87	2484.54	4556.55	2617.10	3371.79
	10	3602.39	2717.06	2462.77	2454.40	4581.36	3755.46	3381.21
	50	4891.44	5154.38	5123.45	5112.75	5151.09	2991.89	3425.14
	100	4886.65	5123.78	5109.16	5132.04	5160.88	3686.20	3278.78
	500	4849.02	5218.67	5180.17	5081.83	5092.26	3280.42	3420.74
	1000	4713.15	5030.80	5125.62	5174.21	5238.59	3446.51	3372.58
	1500	4683.77	5236.12	5122.57	5341.90	5147.82	3600.50	3347.99

Table 7: MAE. Interval $[\overline{x}_j - k\sigma_{x_j}, \overline{x}_j + k\sigma_{x_j}]$

$2a$	$\epsilon \backslash C$	0.1	1	10	100	1000	10000	100000
	0.001	4291.69	2737.43	2505.15	2672.89	**2426.22**	2801.65	4485.76
	0.01	4275.44	3355.66	2540.66	2498.26	2943.94	2456.37	4431.00
	0.1	4309.66	6200.28	3959.74	2481.90	2471.10	4225.71	4478.33
	1	4166.17	3400.28	2566.21	4247.22	3736.12	2521.28	4507.87
	2	3721.11	2474.54	2463.66	3120.38	4889.16	2533.68	4514.86
0	5	3781.38	5091.83	2530.84	2780.67	9073.02	3810.04	4430.49
	10	3700.08	2831.72	2569.97	2596.62	6988.85	4448.89	4467.91
	50	4275.74	3948.21	4910.17	8588.59	7703.47	4822.49	4956.24
	100	4036.68	4129.18	6783.78	7997.96	7713.47	5217.57	4521.87
	500	4033.36	4599.32	7028.35	7781.45	8128.09	4496.62	4502.25
	1000	4292.28	4463.94	7019.17	6792.90	7292.11	4436.09	4508.77
	1500	4169.41	4396.28	7842.06	7304.79	8621.70	5043.53	4402.42
	0.001	5311.40	5079.97	6540.28	6424.83	6263.68	2646.48	5071.51
	0.01	5889.63	5172.37	5673.55	6495.11	6329.34	2651.91	5087.50
	0.1	5570.80	5420.45	6062.06	6695.22	5623.35	**2426.38**	5141.43
	1	4815.58	5990.27	6534.49	6153.96	6697.46	2548.35	5135.84
	2	5204.30	5526.95	6440.05	6742.79	7087.13	2621.72	5101.77
0.01	5	5126.71	5639.25	5784.98	7383.08	6789.23	4135.82	5128.00
	10	5495.88	7119.94	6168.64	6170.35	6275.96	5344.65	5153.16
	50	4275.74	3948.21	4910.17	8588.59	7703.47	5307.97	5100.92
	100	4036.68	4129.18	6783.78	7997.96	7713.47	5196.32	5318.96
	500	4033.36	4599.32	7028.35	7781.45	8128.09	5092.20	5336.05
	1000	4292.28	4463.94	7019.17	6792.90	7292.11	5043.82	5215.09
	1500	4169.41	4396.28	7842.06	7304.79	8621.70	5046.74	5114.58
	0.001	4394.94	4283.89	4204.72	4173.29	4180.29	2822.52	7113.29
	0.01	4379.66	4242.87	4197.99	4193.68	4021.62	2442.97	6726.00
	0.1	4355.82	4242.08	4188.42	4166.00	4006.45	**2422.46**	6614.03
	1	4337.90	4238.24	4186.07	4156.91	4019.94	4061.94	6656.69
	2	4301.24	4233.77	4173.84	4172.52	4101.51	2959.08	6561.90
0.05	5	4339.47	4243.06	4177.87	4199.46	4122.69	5070.31	6561.98
	10	4324.31	4235.61	4176.33	4152.86	4246.53	10942.30	6480.23
	50	2872.30	3826.83	2982.67	2485.20	2520.76	7959.29	6650.94
	100	2740.72	2587.04	2784.63	2823.22	2485.34	7401.63	6958.09
	500	2954.76	2847.77	2764.50	2534.77	3702.37	7357.32	6901.62
	1000	2909.84	3022.14	2640.74	2457.48	5057.82	7264.25	6886.16
	1500	4304.01	3020.11	3295.84	2834.52	5720.23	7175.78	6835.20
	0.001	2730.15	2808.35	2640.26	2643.20	5949.74	2686.89	5168.97
	0.01	2602.98	2535.97	2477.90	3687.32	5985.68	2459.33	5120.17
	0.1	2526.36	2676.62	2498.03	6466.95	5956.40	3002.50	5129.39
	1	2614.56	3001.23	2432.44	6228.66	5967.36	2688.10	5475.37
	2	2861.44	2476.09	3082.08	5892.53	6069.97	2427.51	5125.53
0.1	5	3007.66	2460.56	2504.33	5973.97	5961.31	5196.28	5262.28
	10	2717.68	2484.63	2501.98	5959.15	5969.95	5302.96	5065.44
	50	4582.01	2665.71	2556.45	2697.51	**2335.63**	5435.45	3310.36
	100	5229.03	4246.36	5500.82	3367.34	2850.65	5377.14	2878.69
	500	6055.45	2910.80	10261.70	2570.64	2577.44	5409.49	2970.82
	1000	4149.11	3249.08	3504.15	3092.69	2830.00	5539.88	3327.95
	1500	4374.19	2606.48	3885.12	3922.38	2545.78	5277.26	2569.53

Table 8: MAE. Interval $[Q_a, Q_{1-a}]$

$2a$	$\epsilon \backslash C$	0.1	1	10	100	1000	10000	100000
	0.001	3713.64	4507.69	2739.16	2631.08	2485.23	2503.60	5096.85
	0.01	3470.88	3140.55	2490.55	2620.89	2451.64	2541.10	5466.23
	0.1	2720.40	3511.33	2582.64	2558.39	2413.21	2468.90	5120.19
	1	3297.89	2517.90	3283.58	2475.44	2542.20	2449.83	5387.51
	2	3023.56	2685.01	2787.95	2469.61	3650.91	2512.33	5283.59
0.2	5	4213.48	2598.65	3504.22	2441.16	3784.45	3690.87	5259.67
	10	7344.67	3162.90	3073.63	2477.49	2776.21	2504.30	5258.30
	50	3742.55	3620.95	2536.11	2539.38	2565.72	2881.43	5281.44
	100	3137.27	4717.39	2985.77	2410.96	2459.75	3722.22	5096.50
	500	2845.11	3410.62	3147.78	2490.84	2464.36	2536.91	5114.92
	1000	3016.49	2582.90	2479.67	3174.91	**2352.88**	4414.38	5179.80
	1500	3112.88	2580.87	2569.94	4780.88	2560.05	4336.05	5201.37
	0.001	5582.25	2686.18	2523.61	2443.64	2455.10	2551.12	4652.50
	0.01	2858.91	2704.00	2700.13	2517.27	2434.09	2526.01	4667.21
	0.1	2462.98	2566.49	3997.70	2440.53	2459.68	2499.97	4664.88
	1	2922.46	2602.72	3114.66	2458.81	2451.84	2477.96	4664.77
	2	3196.45	3023.06	2611.94	2516.18	2547.76	**2396.60**	4620.23
0.5	5	2827.81	3079.18	2635.87	2433.38	2415.74	2518.03	4690.50
	10	2482.14	2555.84	2482.86	2442.57	2817.84	4960.84	4671.36
	50	2813.10	2592.58	2446.82	2493.37	2469.03	4710.72	4690.14
	100	3124.95	2413.04	2462.09	2628.51	2455.13	4764.77	4709.35
	500	3641.64	2581.96	2455.13	2670.04	2430.39	4714.70	4706.35
	1000	2960.57	2681.66	2469.85	2412.67	2838.71	4707.30	4692.92
	1500	3278.53	7794.27	2425.08	2526.81	2544.17	4624.87	4686.43
	0.001	2936.66	2513.41	3493.27	2542.79	2583.91	2473.06	5549.70
	0.01	3498.00	2591.50	2887.49	2506.64	4155.10	2446.67	3180.51
	0.1	3206.63	2551.58	3224.95	2500.39	2547.80	3768.62	**2370.66**
	1	3124.48	2616.88	2539.67	3672.65	2526.31	2517.42	2814.52
	2	2651.65	2485.84	2517.61	3136.46	2533.92	2395.90	2539.67
1	5	2618.35	2460.57	2676.35	3332.55	2454.69	2791.44	2580.10
	10	2747.55	4577.51	2478.98	2462.48	2505.84	3470.90	2580.50
	50	2615.08	2471.55	2476.49	2511.89	2431.25	3437.53	4084.30
	100	2566.75	3063.95	2765.69	2577.83	3672.04	2776.26	3994.31
	500	2928.37	2803.62	2639.36	2507.08	4427.28	4404.41	3922.63
	1000	3021.27	2587.06	2877.91	2467.84	4482.99	4602.07	4091.80
	1500	2700.60	2590.89	2849.27	2482.05	4508.11	4594.18	3809.55

Table 9: MAE. Interval $[Q_a, Q_{1-a}]$

k	$\epsilon \backslash C$	0.1	1	10	100	1000	10000	100000
	0.001	11319.10	3575.72	4071.37	3626.59	3671.97	3905.55	5612.24
	0.01	3637.03	3731.72	5344.16	3534.49	3692.20	3769.12	5528.25
	0.1	4128.08	3769.99	3822.51	3538.38	3768.07	3556.32	5576.15
	1	4114.21	3934.61	3679.58	3639.90	3546.70	3663.06	5645.37
	2	5336.41	3684.13	3633.02	3636.32	3881.87	**3312.15**	5631.39
0	5	4214.51	6327.43	3636.38	3689.72	3961.85	5003.38	5656.34
	10	3952.54	3809.05	3769.64	3972.95	4126.21	3773.16	5610.92
	50	4341.77	4308.66	4296.98	5289.17	3442.81	3684.07	5685.06
	100	5864.19	4244.98	3764.20	4512.33	3561.44	3610.47	5668.92
	500	3835.73	3619.88	3481.19	3866.38	3696.39	3859.00	5647.23
	1000	3905.33	3826.13	4247.28	3669.99	3764.37	3611.11	5618.51
	1500	4632.51	3923.86	3553.37	3757.30	3755.07	3997.66	5583.04
	0.001	5133.28	8218.76	8236.87	8215.47	8700.49	3735.69	7480.06
	0.01	6883.62	8225.75	8245.97	8213.51	8023.19	3806.07	7594.46
	0.1	9130.52	8224.52	8242.91	8214.26	8005.52	6725.50	7533.80
	1	9033.00	8222.46	8236.50	9147.00	7974.81	4219.05	7587.40
	2	8958.34	8207.06	8240.85	8191.92	7965.77	3744.07	7594.77
0.01	5	8248.49	8223.48	8239.37	9117.14	7914.36	3731.60	7591.21
	10	8137.41	8216.37	8218.28	8170.04	7884.24	3851.65	7613.09
	50	6197.62	3628.40	3622.79	3770.23	3630.50	3659.00	7623.50
	100	3821.61	3677.69	3750.67	3764.96	3711.95	3510.43	7613.53
	500	4037.90	3957.69	5499.51	3677.59	3615.02	3571.24	7615.88
	1000	3991.06	4191.06	12322.40	8997.97	3568.99	4565.58	7577.69
	1500	3935.83	3694.13	8844.46	3475.48	3446.13	**3411.15**	7472.52
	0.001	3789.05	3739.20	4005.49	3660.79	3613.42	3861.29	4826.76
	0.01	4208.50	3696.61	3571.41	3806.09	3752.19	3742.04	4831.36
	0.1	3920.85	5134.08	3612.19	3513.57	3656.49	4820.96	4819.56
	1	3973.86	5657.70	4046.10	8168.62	**3424.11**	3736.03	4862.35
	2	4186.57	3811.67	3800.23	3778.87	4378.85	3606.65	4838.58
0.05	5	6708.62	3835.02	3609.89	3557.34	3514.21	3636.19	4031.60
	10	3594.31	3642.04	3779.38	3494.61	3550.80	3877.10	3510.46
	50	3760.81	5612.43	4165.71	4479.93	4941.81	4460.55	3959.37
	100	3869.34	4207.07	4013.66	6083.34	3526.17	4794.56	10046.00
	500	4884.67	3891.77	3659.79	4935.17	3515.85	4765.23	7755.52
	1000	4252.70	3548.48	3558.99	3518.94	3603.99	4730.03	7747.10
	1500	4802.01	3884.58	4042.52	3610.62	3546.17	4692.98	7909.41
	0.001	3703.82	6397.87	6283.50	5953.76	3657.51	3871.00	6935.82
	0.01	3948.21	6304.60	6239.19	3962.57	3597.24	3573.01	7003.83
	0.1	7673.07	6201.15	6324.74	3648.30	3499.53	3835.95	6967.70
	1	6072.82	6215.08	6350.76	3575.25	3756.95	3686.94	6971.93
	2	5940.69	6351.52	6228.91	3556.87	4595.88	3592.04	7021.61
0.1	5	5990.23	6408.07	6062.28	3724.65	3615.24	6974.15	7027.77
	10	6008.50	6333.60	6095.35	3721.05	3771.07	6851.05	7201.36
	50	3775.48	3710.31	3695.59	5370.07	3562.66	6857.61	7076.19
	100	3808.47	3702.25	3664.39	3680.84	3667.41	6819.46	6921.27
	500	4056.45	4293.35	3589.10	3528.88	**3439.63**	6992.88	7093.70
	1000	4455.30	4180.02	3733.45	3495.54	4074.33	7097.27	7077.80
	1500	3729.98	3785.89	4258.38	3558.07	3949.87	7039.57	6991.32

Table 10: RMSE. Interval $[\overline{x}_j - k\sigma_{x_j}, \overline{x}_j + k\sigma_{x_j}]$

k	$\epsilon \backslash C$	0.1	1	10	100	1000	10000	100000
	0.001	4574.58	3708.67	5225.38	3673.44	3970.93	3870.86	6993.52
	0.01	3748.62	3627.15	3550.45	3827.51	4776.32	3813.32	6912.75
	0.1	3677.97	3483.01	3714.66	4179.88	3764.01	3813.42	6806.42
	1	3811.98	3593.63	3540.67	3648.11	**3286.57**	3746.24	6947.60
	2	4414.92	3905.95	3500.10	3540.76	6296.83	4417.05	6898.57
0.2	5	4862.08	3732.55	3655.27	3620.85	5739.01	5458.80	6895.83
	10	3894.01	4470.63	3762.47	3890.03	3631.28	3733.86	6894.30
	50	6714.45	4262.35	5102.91	3636.12	3782.98	7000.88	6878.72
	100	4222.14	3518.63	3869.43	3585.53	3678.72	7072.95	6873.88
	500	4054.90	6667.62	3460.92	3610.74	3570.69	7024.06	6774.66
	1000	4148.46	3552.59	3790.67	3572.28	3609.00	6933.84	6648.84
	1500	3660.92	3918.37	3635.78	3605.39	3521.59	6804.78	6444.16
	0.001	3827.30	3756.42	4182.10	3626.63	3671.05	3896.95	**3288.79**
	0.01	3638.59	3751.82	3608.76	3632.17	3601.46	3798.62	5649.47
	0.1	3678.91	3638.62	3595.67	3619.23	3470.27	3821.41	5321.87
	1	3879.33	3641.26	3514.20	3600.65	4752.06	3617.68	3994.55
	2	4753.87	4504.39	3623.02	3749.10	3907.21	3800.86	4542.81
0.5	5	4892.93	4145.48	4070.11	4216.28	3457.10	4084.12	4701.40
	10	3682.97	4299.43	3968.99	3441.36	3305.31	3813.45	4664.51
	50	4272.42	3583.43	4787.04	3735.99	3645.94	4513.71	4897.98
	100	4546.60	3574.94	3819.47	3664.57	3664.18	3785.50	4911.87
	500	3718.34	3573.21	3770.37	3833.47	3579.05	3946.12	4757.06
	1000	3619.25	3941.42	4793.98	3947.26	3618.14	3637.59	4791.72
	1500	4807.29	3554.54	4039.48	3551.92	3582.28	4733.42	4699.28
	0.001	5692.42	3646.05	5367.72	3469.55	3601.36	3907.02	4788.49
	0.01	4401.95	3693.44	4365.82	3508.17	3582.36	3793.01	4685.99
	0.1	5847.80	4064.32	3665.05	3574.77	3611.85	3626.17	4184.07
	1	4123.32	4724.75	3575.64	3459.81	5234.26	3821.28	3824.70
	2	4426.78	3748.97	3733.31	3797.69	5367.48	3810.22	3772.94
0.75	5	4316.04	3626.03	3575.34	3641.89	4666.53	3577.78	3636.48
	10	3783.91	3594.02	3701.41	4700.05	**3398.10**	3640.88	4242.38
	50	4922.31	3983.20	3415.49	3553.63	3592.78	3488.82	4222.12
	100	3689.64	3913.53	7551.17	3654.46	3584.48	4313.92	3772.08
	500	3983.94	4661.12	6806.15	3675.50	3492.38	3494.97	6035.49
	1000	4029.09	3995.88	3737.45	3639.03	3587.36	3723.70	3658.42
	1500	5999.36	4100.38	3754.91	3690.52	3844.67	3796.85	3926.40
	0.001	4128.16	4183.07	5183.64	3604.22	3594.73	3857.57	5583.93
	0.01	4565.07	3665.57	3553.93	3478.67	3643.82	3756.18	5598.08
	0.1	4468.78	3753.63	3620.01	3543.12	**3388.10**	3600.83	5603.10
	1	4235.98	3714.19	3868.05	3517.73	7611.00	3718.95	5583.40
	2	3969.07	3937.75	3669.67	3668.33	7745.81	3664.83	5583.25
1	5	5150.88	3812.56	3626.69	3639.56	7839.28	3573.74	5582.38
	10	5377.14	3962.76	3552.08	3559.00	7905.63	5283.59	5583.96
	50	7027.88	7320.04	7338.28	7346.38	7325.76	4755.49	5623.92
	100	7022.92	7314.54	7352.96	7336.78	7345.32	5881.55	5563.69
	500	6963.32	7445.11	7373.33	7347.38	7271.79	5684.45	5605.52
	1000	6780.99	7225.78	7385.82	7270.21	7428.40	5604.50	5471.64
	1500	6765.66	7431.36	7231.79	7529.79	7351.45	5699.92	5422.52

Table 11: RMSE. Interval $[\overline{x}_j - k\sigma_{x_j}, \overline{x}_j + k\sigma_{x_j}]$

$2a$	$\epsilon \backslash C$	0.1	1	10	100	1000	10000	100000
	0.001	7968.61	3910.66	3562.75	3675.89	**3460.60**	4308.48	8193.99
	0.01	7938.79	4683.53	3574.54	3654.03	3979.48	3574.31	8205.21
	0.1	7955.27	8417.59	5318.17	3579.03	3561.28	5535.62	8172.67
	1	6734.64	5352.00	3542.58	5981.18	4841.99	3727.49	8189.25
	2	4852.48	3734.37	3664.87	4251.67	6572.50	3506.82	8207.09
0	5	4973.24	6335.80	3678.04	3820.11	10625.30	5510.75	8165.20
	10	5207.01	3965.71	3710.32	3713.61	8744.16	8116.04	8174.45
	50	7706.58	6846.75	8156.00	11650.40	10358.50	8354.07	8524.69
	100	7596.27	7426.39	9532.59	10707.50	10315.40	8681.55	8197.47
	500	7516.51	7657.07	9775.03	10414.20	10861.30	8099.91	8161.31
	1000	7718.19	7538.32	9791.77	9451.85	9987.97	8087.81	8025.75
	1500	7339.38	7478.43	10475.40	9907.25	11624.50	8524.43	7936.27
	0.001	8892.04	8581.60	10067.90	9243.79	9104.63	3942.26	7681.38
	0.01	9928.79	8651.02	8772.08	9317.38	9116.48	3685.38	7695.02
	0.1	9057.31	8770.58	8892.17	9485.06	7927.92	**3521.19**	7800.90
	1	8379.20	9357.95	9399.96	8909.44	9393.96	3679.78	7786.19
	2	8715.52	8769.67	9246.86	9575.23	9775.81	3741.27	7735.61
0.01	5	8673.46	8845.89	8430.49	11249.00	9380.46	5552.93	7752.27
	10	9031.78	11147.50	8864.21	8989.99	8995.75	8131.85	7759.13
	50	7706.58	6846.75	8156.00	11650.40	10358.50	8069.66	7729.19
	100	7596.27	7426.39	9532.59	10707.50	10315.40	7894.22	8117.64
	500	7516.51	7657.07	9775.03	10414.20	10861.30	7705.56	8145.81
	1000	7718.19	7538.32	9791.77	9451.85	9987.97	7579.69	7953.88
	1500	7339.38	7478.43	10475.40	9907.25	11624.50	7514.13	7820.83
	0.001	7886.85	7747.20	7695.03	7660.69	7622.62	4153.65	9860.66
	0.01	7869.29	7726.26	7689.52	7659.47	7540.20	3549.52	9370.97
	0.1	7827.95	7723.30	7680.84	7646.88	7491.78	3579.58	9308.24
	1	7805.06	7717.44	7675.62	7645.76	7474.06	5039.17	9433.61
	2	7781.42	7715.29	7669.08	7646.78	7485.76	4049.39	9329.04
0.05	5	7778.51	7717.78	7668.35	7654.05	7511.77	7453.97	9425.49
	10	7770.10	7710.68	7665.13	7640.24	7594.50	14022.10	9475.32
	50	4397.67	5021.15	4101.25	3620.44	3680.86	10855.90	9791.83
	100	3979.29	3616.86	3817.97	3748.67	**3498.77**	10093.80	10269.40
	500	4012.51	3888.82	3647.72	3727.22	5062.00	10026.20	10140.20
	1000	4210.01	3872.18	3689.93	3587.36	6670.04	9955.74	10017.80
	1500	5464.91	4054.66	4302.52	3826.73	8160.03	9855.79	9980.20
	0.001	3992.11	3830.82	3775.93	3617.94	7715.03	4042.37	8536.53
	0.01	3636.74	3657.39	3431.14	4961.07	7718.45	3614.14	8557.16
	0.1	3785.67	3847.89	3673.14	8370.87	7714.75	4111.38	8581.80
	1	3780.40	3930.47	3574.96	8022.96	7716.50	3772.44	8746.28
	2	3934.12	3671.28	4199.21	7681.82	7753.24	**3348.94**	8471.32
0.1	5	4032.47	3660.44	3618.07	7719.23	7713.52	8674.74	8609.38
	10	3627.16	3677.70	3628.80	7723.33	7716.01	8712.29	8324.23
	50	6429.00	3895.79	3729.50	3818.91	3378.95	8799.26	5145.93
	100	6009.14	5435.54	6937.80	4621.81	3992.93	8754.91	3903.91
	500	7584.26	4069.28	17300.90	3559.37	3729.92	8728.74	4158.34
	1000	5973.41	4198.71	5428.70	4827.98	3777.36	8792.98	4301.99
	1500	6663.31	3693.41	5278.47	5465.87	3612.07	8649.68	3583.70

Table 12: RMSE. Interval $[Q_a, Q_{1-a}]$

$2a$	$\epsilon \backslash C$	0.1	1	10	100	1000	10000	100000
	0.001	4904.20	5977.53	3779.70	3780.04	3713.63	3899.80	8984.28
	0.01	5605.16	4370.99	3640.12	3686.47	3673.52	3709.78	9073.41
	0.1	3972.21	4818.86	3545.38	3746.84	3569.74	3667.19	9035.25
	1	4991.24	3677.55	4312.87	3652.74	3479.27	3599.70	9164.58
	2	4142.54	3760.06	4126.02	3631.37	5494.16	3684.83	9110.87
0.2	5	4941.05	3740.66	5289.66	3522.49	5953.85	4675.27	9101.67
	10	10211.20	4497.64	4159.79	3691.60	4056.16	3507.80	9099.23
	50	5234.92	4479.16	3707.86	3655.00	3711.63	3744.11	9093.04
	100	4211.10	6210.85	4078.55	3566.51	3662.80	4908.79	8938.82
	500	4011.09	4884.89	4351.68	3494.58	3641.45	3498.33	8996.82
	1000	3993.23	3670.69	3627.86	4506.54	**3292.85**	7525.17	8986.34
	1500	4302.15	3674.50	3526.23	8138.59	3727.43	7432.20	8991.76
	0.001	7163.69	3970.14	3715.58	3599.01	3635.16	3925.24	8408.00
	0.01	4157.72	3770.84	3777.42	3586.69	3521.96	3605.14	8406.13
	0.1	3699.77	3707.69	5791.96	3553.77	3537.00	3439.75	8407.97
	1	4125.27	4101.13	4794.39	3638.59	3580.44	3759.09	8413.68
	2	3943.05	3963.96	3663.00	3546.00	3605.33	3507.68	8392.49
0.5	5	4007.52	4347.11	3681.40	3569.72	**3391.72**	3559.15	8405.38
	10	3708.23	3660.54	3646.33	3577.59	3979.59	8747.95	8395.11
	50	3936.52	3602.18	3642.20	3553.99	3646.75	8410.29	8419.92
	100	4592.71	3558.81	3636.78	3632.43	3583.06	8456.23	8424.25
	500	5224.26	3737.88	3635.14	3929.26	3517.29	8416.79	8420.44
	1000	4390.37	3827.70	3512.69	3441.22	3984.74	8400.83	8405.19
	1500	4736.09	13161.50	3563.73	3519.90	3610.17	8357.03	8385.13
	0.001	4561.81	3653.66	5075.52	3733.18	3796.50	3824.43	7271.65
	0.01	4604.51	3653.84	4994.71	3689.30	5342.39	3588.68	4003.83
	0.1	5056.89	3797.15	5038.06	3590.37	3705.12	4931.95	**3453.53**
	1	4258.65	3673.15	3599.40	5071.63	3650.23	3729.52	3885.88
	2	3824.71	3657.81	3635.45	4393.82	3624.19	3537.44	3643.37
1	5	3854.98	3619.83	3838.18	4231.54	3572.68	3964.56	3607.47
	10	4064.15	7036.31	3598.74	3598.12	3705.30	4540.52	3757.86
	50	3753.95	3602.27	3614.52	3620.03	3624.17	4451.16	7087.42
	100	3779.20	4301.07	3627.77	3587.51	4821.67	3912.93	6824.13
	500	4085.39	3940.26	3743.10	3634.84	5704.04	5521.75	6674.31
	1000	4261.96	3580.81	4019.48	3585.85	5771.53	5886.58	6757.16
	1500	3893.08	3707.88	4017.28	3587.48	5802.82	5898.73	6337.29

Table 13: RMSE. Interval $[Q_a, Q_{1-a}]$

perturbations, which are supposed to be unknown but bounded for a given norm.
Several computational experiments with real datasets have been performed. In particular, our formulation allows to improve the results obtained in [29] for a cardiological example. Our tool has also been tested when imputation for missing data is done via intervals based on the mean and deviation of the non-missing values or on quantiles, obtaining good results for regression when using non-degenerate intervals. All these experiments display our tool as a competitive model for regression with uncertainty on the data.
As possible extensions, we propose the study of other formulations for our model. One can observe that, in formulation (13), the optimization problem has been posed when using the Euclidean norm for the objective function. However, the use of the l_1-norm or the l_∞-norm gives us similar expressions which are linear programs in case of considering the maximum distance for the constraints of the problem. Furthermore, other different distances (apart from maximum and Hausdorff distances) can be introduced in the constraints of Problem (13). Experiments with all these formulations can be done in the future to try to select the most suitable formulation for our problem.
Likewise, the introduction of kernels in the model is another topic which deserves further studies.

References

[1] A. A. Afifi and R. M. Elashoff. Missing Observations in Multivariate Statistics: II. Point Estimation in Simple Linear Regression. *Journal of the American Statistical Association*, 62(317):10–29, (1967).

[2] A. Ben-Tal and A. Nemirovski. Robust Convex Optimization. *Mathematics of Operations Research*, 23(4):769–805, (1996).

[3] A. Ben-Tal and A. Nemirovski. Robust Solutions of Uncertain Linear Programs. *Operations Research Letters*, 25:1–13, (1998).

[4] L. Billard and E. Diday. Regression Analysis for Interval-valued Data. In H. A. L. Kiers, J-P. Rasson, P. J. F. Groenen, and M. Schader, editors, *Data Analysis, Classification and Related Methods*, pages 369–374. Springer-Verlag, Berlin, (2000).

[5] L. Billard and E. Diday. Symbolic Regression Analysis. In K. Jajuga, A. Sokolowski, and H-H. Bock, editors, *Classification, Clustering and Data Analysis*, pages 281–288. Springer-Verlag, Berlin, (2002).

[6] L. Billard and E. Diday. From the Statistics of Data to the Statistics of Knowledge: Symbolic Data Analysis. *Journal of the American Statistical Association*, 98(462):470–487, (2003).

[7] C. L. Blake and C. J. Merz. UCI Repository of Machine Learning Databases. http://www.ics.uci.edu/∼mlearn/MLRepository.html, (1998).

[8] H-H. Bock and E. Diday, editors. *Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data*. Springer-Verlag, Berlin, (2000).

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

[10] E. Carrizosa, J. Gordillo, and F. Plastria. Classification Problems with Imprecise Data through Separating Hyperplanes. Technical Report MOSI/33, MOSI Department, Vrije Universiteit Brussel, http://www.vub.ac.be/MOSI/papers/CarrizosaGordilloPlastria.pdf, (2007).

[11] A. Celminš. Multidimensional Least-squares Fitting of Fuzzy Models. *Mathematical Modelling*, 9(9):669–690, (1987).

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

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

[14] F. A. T. De Carvalho, E. A. Lima-Neto, and C. P. Tenorio. A New Method to Fit a Linear Regression Model for Interval-valued Data. In *Proc. of the 27th Annual German Conference on Artificial Intelligence*, pages 295–306, (2004).

[15] 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).

[16] P. Diamond. Fuzzy Least Squares. *Information Sciences*, 46:141–157, (1988).

[17] N. R. Draper and H. Smith. *Applied Regression Analysis*. John Wiley, New York, (1998).

[18] 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).

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

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

[21] J-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Springer-Verlag, Berlin, (1993).

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

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

[24] 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).

[25] 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).

[26] 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).

[27] 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).

[28] 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).

[29] E. A. Lima-Neto and F. A. T. De Carvalho. Centre and Range Method for Fitting a Linear Regression Model to Symbolic Interval Data. *Computational Statistics and Data Analysis*, To appear, (2007).

[30] E. A. Lima-Neto, F. A. T. De Carvalho, and E. S. Freire. Applying Constrained Linear Regression Models to Predict Interval-valued Data. In *Proc. of the 28th Annual German Conference on Artificial Intelligence*, pages 92–106, (2005).

[31] R. J. A. Little. Regression with Missing X's. *Journal of the American Statistical Association*, 87(420):1227–1237, (1992).

[32] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. John Wiley & Sons, New Jersey, (2002).

[33] W. Z. Liu, A. P. White, S. G. White, S. G. Thompson, and M. A. Bramer. Techniques for Dealing with Missing Values in Classification. In *Proc. of the Second International Symposium on Intelligent Data Analysis*, pages 527–536, (1997).

[34] M. Magnani. Techniques for Dealing with Missing Data in Knowledge Discovery Tasks. http://magnanim.web.cs.unibo.it/data/pdf/missingdata.pdf, (2004).

[35] NEOS. Server for Optimization. http://www-neos.mcs.anl.gov/.

[36] J. Scheffer. Dealing with Missing Data. *Research Letters in the Information and Mathematical Sciences*, 3:153–160, (2002).

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

[38] 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).

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

[40] H. Tanaka, S. Uejima, and K. Asai. Linear Regression Analysis with Fuzzy Model. *IEEE Transactions on Systems, Man and Cybernetics*, 12:903–907, (1982).

[41] T. B. Trafalis and S. A. Alwazzi. Support Vector Regression with Noisy Data: a Second Order Cone Programming Approach. *International Journal of General Systems*, 36(2):237–250, (2007).

[42] T. B. Trafalis and R. C. Gilbert. Robust Classification and Regression Using Support Vector Machines. *European Journal of Operational Research*, 173(3):893–909, (2006).

[43] R. J. Vanderbei. LOQO User's Manual - Version 4.05. Technical Report ORFE-99, Operations Research and Financial Engineering, Princeton University, New Jersey, (2000).

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

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