An Evolutionary Support Vector Machines Approach to Regression

Ruxandra STOEAN

Faculty of Mathematics and Computer Science,
Department of Computer Science,
University of Craiova, Romania
ruxandra.stoean@inf.ucv.ro

Abstract. The paper proposes a first attempt to approach regression from the perspective of the novel technique of evolutionary support vector machines (ESVMs). ESVMs have been developed with the aim of providing a simpler architecture as compared to their canonical counterpart. As a consequence, they inherit the training mechanism of SVMs but approach the resulting optimization problem through evolutionary algorithms. In this paper, ESVMs for proposed task are built upon the classical learning engine of ϵ-SVMs for regression. The validation of the new approach on a set of 2-dimensional points once again demonstrates the advantages of ESVMs as well as their promise as concerns this particular pattern recognition case.
Keywords: evolutionary support vector machines, regression, ϵ-support vector machines, 2-dimensional points data set.
Math. Subjects Classification 2000: 68T05, 68T20, 92D10.

1 INTRODUCTION

Evolutionary support vector machines (ESVMs) [9], [10] are a novel learning paradigm based on the well-known and highly used technique of support vector machines (SVMs).

The reason for their development was that of bringing an easier alternative to the complicated mathematical procedure of solving the optimization problem that is reached by the learning mechanisms of SVMs. In order to determine the optimal learning function, ESVMs propose evolutionary algorithms (EAs) instead.

ESVMs have until now successfully addressed only classification tasks. It is the aim of present paper to prove that they are also competitive for regression issues.

ESVMs for regression inherit the engine of one classical technique in this respect, i.e. ϵ-SVMs, and estimate the regression coefficients through a canonical EA.

The paper is structured as follows. Next section presents the basic ideas behind SVMs. The third section briefly sketches the changes that are brought by ESVMs. Section 4 outlines the concepts specific to ϵ-SVMs for regression. Section 5 describes proposed application of ESVMs to the regression case. Experiments on a set of 2-dimensional points are illustrated in section 6. In the final section, conclusions are reached and implications for future work are presented.

2 SUPPORT VECTOR MACHINES. AN OVERVIEW

SVMs are a very powerful state-of-the-art learning technique that primarily targets classification and regression tasks.

A possible definition of SVMs can be stated as in [4]: "a system for efficiently training linear learning machines in kernel-induced feature spaces, while respecting the insights of generalization theory and exploiting optimization theory".

SVMs exhibit the mechanism that is typical to any machine learning technique. Given a training set $\{(x_i, y_i)\}_{i=1,2,...,m}$, where every $x_i \in R^n$ represents a data sample and each y_i a target, a training stage deals with the internal discovery and learning of the correspondence between every sample x_i and given target y_i.

That implies that, given a family of functions $\{f_{t \in T}|, f_t : R^n \rightarrow D\}$, the task consists in learning the optimal function f_{t^*} that minimizes the discrepancy between the given targets of data samples and the actual targets produced by the learning machine; the aim is then to find the optimal parameter t that corresponds to the appropriate learning function. According to what has been learnt, a test step is finally responsible for the prediction of targets for previously unknown samples.

Remark: If classification is envisaged, the domain D of the targets is discrete; conversely, if regression is considered, D is continuous.

The task of SVMs for classification is to achieve an optimal separation of given data into classes. SVMs regard learning in this situation from a geometrical point of view, i.e. they assume the existence of a separating surface between two classes labelled as -1 and 1, respectively. The aim of SVMs then becomes the discovery of the appropriate decision hyperplane, i.e. the detection of its optimal coefficients.

On the other hand, the standard assignment of SVMs for regression [11] is to find the optimal function to be fitted to the data such that it achieves at most ϵ deviation from the actual targets of samples; the aim becomes thus to estimate the optimal regression coefficients of such a function.

3 EVOLUTIONARY SUPPORT VECTOR MACHINES

Although a very competitive learning technique, SVMs have a highly complicated mathematics behind their engine. Concepts of convexity and an extension of the method of Lagrange multipliers according to Karush-Kuhn-Tucker-

Lagrange conditions are used in order to solve the constrained optimization problem that is reached by the learning mechanism.

The novel ESVMs [9] have therefore been proposed as a technique that, by means of EAs, provides a simpler alternative to this complex mathematical procedure. Subsequently, it has been observed that the method gains another advantage over canonical architecture, i.e. ESVMs determine the coefficients of the learning function in a direct fashion, which is generally not possible with SVMs.

ESVMs have been developed such as to inherit the training procedure of SVMs but to embed an EA [5] at the level of solving the optimization problem. So far, they have been designed and applied only for classification [9], [10].

4 SUPPORT VECTOR MACHINES FOR REGRESSION

Present paper proposes the first attempt of an ESVM approach for regression, which is chosen to be constructed through the hybridization between EAs and the standard ϵ-SVMs.

Recall the training environment $\{(x_i, y_i)\}_{i=1,2,\ldots,m}$, where every $x_i \in R^n$ represents a data sample and each $y_i \in R$ a target. Such a data set could represent exchange rates of a currency measured in subsequent days together with econometric attributes [8] or a medical indicator registered in multiple patients along with personal and medical information [1].

ϵ-SVMs for regression aim at determining the optimal coefficients of the function to be fitted to the training data while they allow for errors that are less than ϵ and, simultaneously, request high generalization ability, i.e. they require the function in question to be as flat as possible [7], [8].

The learning mechanism takes place as follows [8]. The first step is to observe if a linear regression model can fit the training samples. Accordingly, function f that suits data takes the form (1):

$$f(x) = \langle w, x \rangle - b, \tag{1}$$

where $w \in R^n$ is the slope of the regression hyperplane and $b \in R$ is the intercept.

The requirement that f approximates training data with ϵ precision is written as (2):

$$\begin{cases} y_i - \langle w, x_i \rangle + b \leq \epsilon \\ \langle w, x_i \rangle - b - y_i \leq \epsilon \end{cases} , i = 1, 2, ..., m \tag{2}$$

Conversely, the condition for a flat function is equivalent to demanding smallest slope, i.e. w, which leads to minimizing its squared norm $\|w\|^2$.

The task of linear ϵ-SVMs for regression is thus stated as the optimization problem (3):

$$\text{find } w \text{ and } b \text{ as to minimize } \|w\|^2$$
$$\text{subject to } \begin{cases} y_i - \langle w, x_i \rangle + b \leq \epsilon \\ \langle w, x_i \rangle - b - y_i \leq \epsilon. \end{cases} , i = 1, 2, ..., m \tag{3}$$

If a linear function f cannot match all training data, ϵ-SVMs will allow for some regression errors with the condition that they are in a minimal number [2], [6]. Therefore, the positive indicators for errors ξ_i and ξ_i^*, both attached to each sample, are introduced into the condition for approximation of training data which becomes (4):

$$\begin{cases} y_i - \langle w, x_i \rangle + b \leq \epsilon + \xi_i, \\ \langle w, x_i \rangle - b - y_i \leq \epsilon + \xi_i^*. \end{cases} , i = 1, 2, ..., m \tag{4}$$

Simultaneously, the sum of indicators for errors, $C \sum_{i=1}^{m} (\xi_i + \xi_i^*)$, is minimized. C symbolizes the penalty for errors.

The task of linear ϵ-SVMs for nonlinear regression is thus stated as the optimization problem (5):

$$\text{find } w \text{ and } b \text{ as to minimize } \|w\|^2 + C \sum_{i=1}^{m} (\xi_i + \xi_i^*)$$
$$\text{subject to } \begin{cases} y_i - \langle w, x_i \rangle + b \leq \epsilon + \xi_i \\ \langle w, x_i \rangle - b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} , i = 1, 2, ..., m \tag{5}$$

If a linear function still fails to satisfactorily fit training data, a nonlinear function has to be chosen. The procedure is as follows [3]. Data is mapped via a nonlinear function into a high enough dimensional space and linearly modelled there as previously. This corresponds to a nonlinear function in the initial space.

Hence, data samples are mapped into some Euclidean space H through a function $\Phi : R^n \mapsto H$. Therefore, the optimization problem in H has the formulation (6):

$$\text{find } w \text{ and } b \text{ as to minimize } \langle \Phi(w), \Phi(w) \rangle + C \sum_{i=1}^{m} (\xi_i + \xi_i^*)$$
$$\text{subject to } \begin{cases} y_i - \langle \Phi(w), \Phi(x_i) \rangle + b \leq \epsilon + \xi_i \\ \langle \Phi(w), \Phi(x_i) \rangle - b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} , i = 1, 2, ..., m \tag{6}$$

Nevertheless, the choice of a function Φ with required properties is not an easy task. However, as in the training algorithm vectors appear only as part of dot products, if there were a kernel function K such that $K(x,y) = \langle \Phi(x), \Phi(y)$, where $x, y \in R^n$, one would use K in the training algorithm and would never need to explicitly even know what Φ is.

The kernel functions that meet this condition are given by Mercer's theorem from functional analysis [2]. Still, it may not be easy to check whether the condition is satisfied for every new kernel. There are, however, a couple of classical kernels that had been demonstrated to meet Mercer's condition [2]. The most commonly used ones are the polynomial kernel of degree p, $K(x,y) = \langle x, y\rangle^p$, and the radial kernel of parameter σ, $K(x,y) = e^{\frac{\|x-y\|^2}{\sigma}}$.

The task of nonlinear ϵ-SVMs for regression is thus stated as the optimization problem (7):

$$\text{find } w \text{ and } b \text{ as to minimize } K(w,w) + C\sum_{i=1}^{m}(\xi_i + \xi_i^*)$$
$$\text{subject to } \begin{cases} y_i - K(w,x_i) + b \leq \epsilon + \xi_i \\ K(w,x_i) - b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}, i = 1, 2, ..., m \tag{7}$$

Once training of ϵ-SVMs is completed, the predicted target for test samples is computed following obtained learning function that may either refer the found coefficients or, as usually w and b cannot be directly determined, some other variables derived from mathematical artifices.

5 EVOLUTIONARY SUPPORT VECTOR MACHINES FOR REGRESSION

Training in ϵ-ESVMs for regression follows the same steps as in the canonical technique. The components of the EA to solve the inherent optimization problem were experimentally chosen as follows.

REPRESENTATION OF INDIVIDUALS. An individual encodes the regression coefficients together with the indicators for errors of regression (included for reasons of reference in the EA formulation of the optimization problem), i.e. is a vector of w, b, ξ and ξ^* (8):

$$c = (w_1, ..., w_n, b, \xi_1, ..., \xi_m, \xi_1^*, ..., \xi_m^*) \tag{8}$$

INITIAL POPULATION. Individuals are randomly generated following a uniform distribution, such that $w_i \in [-1,1], i = 1, 2, ..., n$, $b \in [-1,1]$ and ξ_j and $\xi_j^* \in [0,1], j = 1, 2, ..., m$.

FITNESS EVALUATION. The fitness function is derived from the objective function of the optimization problem and has to be minimized. Constraints are handled through penalizing the infeasible individuals by appointing $t : R \to R$ which returns the value of the argument if negative, while otherwise 0. The expression of the fitness function is considered as follows (9):

$$f(w_1, ..., w_n, b, \xi_1, ..., \xi_m, \xi_1^*, ..., \xi_m^*) = K(w, w) + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) +$$
$$\sum_{i=1}^{m}[t(\epsilon + \xi_i - y_i + K(w, x_i) - b)]^2 + \sum_{i=1}^{m}[t(\epsilon + \xi_i^* + y_i - K(w, x_i) + b)]^2, \quad (9)$$

One is led to:

$$\text{minimize } (f(c), c). \quad (10)$$

GENETIC OPERATORS. Tournament selection is used. Intermediate crossover and mutation with normal perturbation are considered. Mutation is restricted only for ξ and ξ^*, preventing the indicators for errors from taking negative values.

STOP CONDITION. The algorithm stops after a predefined number of generations. In the end, one obtains the optimal estimated regression coefficients, i.e. w and b, which are subsequently applied to the test data.

PREDICTION OF TEST SAMPLES. Given a test data sample x, its predicted target is computed following (11):

$$f(x) = K(w, x) - b \quad (11)$$

6 EXPERIMENTAL RESULTS. APPLICATION TO A 2-DIMENSIONAL POINTS DATA SET

A fictitious training data set of points in a 2-dimensional environment was considered in order to validate the novel ϵ-ESVMs for regression. The data set was chosen as in Figure 1.

A polynomial kernel was chosen as it achieved the best results in preliminary experiments. The values that were appointed for the specific parameters of support vector machines and evolutionary algorithms are given in Table 1.

Illustration of obtained regression model is depicted in Figure 1. In the plot, the squares are the actual data points and the line denotes the function that was fitted to the data.

Table 1. Values for parameters of ϵ-ESVMs for regression

Parameter	Manually picked value
C	1
ϵ	0
p	2
Population size	100
Number of generations	100
Crossover probability	0.3
Mutation probability	0.4
Mutation probability of indicators for errors	0.5
Mutation strength	0.1
Mutation strength of indicators for errors	0.1

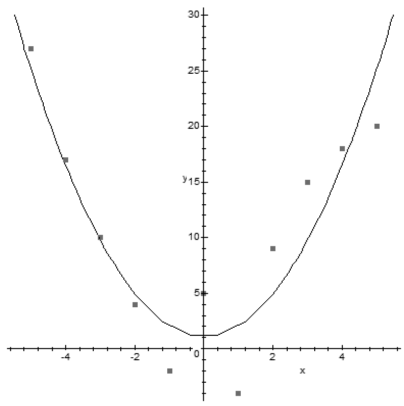

Fig. 1. Training samples (*squares*) and function that was fitted to data by ϵ-ESVMs

7 CONCLUSIONS AND FUTURE WORK

Present paper constitutes the first approach of the novel technique of ESVMs to the regression case, which is achieved through the hybridization between EAs and the classical ϵ-SVMs engine. Validation of the new approach is performed on a data set of 2-dimensional points.

Obtained regression model proved to perform well for the given task, in addition to the fact that, in comparison to SVMs, ESVMs have a simpler nature and a direct handling of the learning function through its coefficients.

As concerns future work, it would be interesting to attempt a second approach of ESVMs to regression, through inheriting the training engine of another classical technique, i.e. ν-SVMs.

References

[1] **Altman D. G.**, Practical Statistics for Medical Research, Chapman and Hall, 1991
[2] **Boser B. E., Guyon I. M., Vapnik V.**, A Training Algorithm for Optimal Margin Classifiers, *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, Pittsburgh, PA, ACM Press, 11-152, 1992
[3] **Cover T. M.**, Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition, *IEEE Transactions on Electronic Computers*, vol. EC-14, 326-334, 1965
[4] **Cristianini N., Shawe-Taylor J.**, An Introduction to Support Vector Machines, Cambridge University Press, 2000
[5] **Eiben A. E., Smith J. E.**, Introduction to Evolutionary Computing, Springer - Verlag, 2003
[6] **Haykin, S.**, Neural Networks, A Comprehensive Foundation, Prentice Hall, New Jersey, 1999
[7] **Rosipal R.**, Kernel-based Regression and Objective Nonlinear Measures to Access Brain Functioning, PhD thesis, Applied Computational Intelligence Research Unit School of Information and Communications Technology University of Paisley, Scotland, September 2001
[8] **Smola A. J., Scholkopf B.**, A Tutorial on Support Vector Regression, Technical report, NeuroCOLT2 Technical Report Series, NC2-TR-1998-030, October 1998
[9] **Stoean R., Stoean C., Preuss M., El-Darzi E., Dumitrescu D.**, Evolutionary Support Vector Machines for Diabetes Mellitus Diagnosis, *Proceedings of IEEE IS'06*, London, UK, in press, 2006
[10] **Stoean R., Dumitrescu D., Preuss M., Stoean C.**, Different Techniques of Multi-class Evolutionary Support Vector Machines, BIC-TA, China, submitted, 2006
[11] **Vapnik V.**, The Nature of Statistical Learning Theory, Springer Verlag, New York, 1995