Annals of University of Craiova, Math. Comp. Sci. Ser.
Volume 33, 2006, Pages 141–146
ISSN: 1223-6934

Linear Evolutionary Support Vector Machines for Separable Training Data

Ruxandra Stoean and Dumitru Dumitrescu

Abstract. Present paper addresses the famous machine learning paradigm, called support vector machines, from the viewpoint of evolutionary computation. Namely, the constrained optimization problem within support vector machines is solved through an evolutionary algorithm, for the sake of simplicity. The new approach has been so far applied solely to the case of linear support vector machines for separable data. Experiments are conducted on a fictitious 2-dimensional points data set and are very promising.

2000 Mathematics Subject Classification. 68T05, 68T20, 92D10.
Key words and phrases. support vector machines, separable data, evolutionary computation, 2-dimensional points data set.

1. Introduction

Recently developed and powerful tools, support vector machines (SVMs) have proven their success in various pattern recognition fields regarding classification and regression.

Although truly remarkable as learning machines, SVMs have a highly complicated mathematics acting as an engine. Concepts of convexity and extension of the method of Lagrange multipliers according to Karush-Kuhn-Tucker-Lagrange conditions are used in order to solve the constrained optimization problem within.

Proposed paper presents an alternative approach to this standard solving. The new approach is called evolutionary support vector machines (ESVMs) and accordingly relies on elements from the field of evolutionary optimization.

So far, ESVMs deal just with standard classification issues. Moreover, this first test of the newly developed technique has only been applied to the situation when training data are binary and linearly separable.

With respect to classification, standard SVMs are primarily concerned with binary labelled training data [2], [4]. They were initially built to handle linearly separable data; they were then extended to linear SVMs for nonseparable data and finally to nonlinear SVMs. Present paper is thus the beginning step towards the validation of ESVMs on all possible configurations of binary classification. Moreover, ESVMs for multi-class categorization are also to be developed in the near future.

Experiments for the assessment of ESVMs in the case of linear separability were conducted on a fictitious 2D points data set. Results reveal the suitability of the new approach and the guarantee for the success of extended ESVMs.

The structure of the paper is the following. Second section presents the basic concepts within linear SVMs for separable training data; next section introduces the linear ESVMs for data separability. Experimental results are outlined in the

Received: November 21st 2005.

fourth section; values for parameters of the evolutionary components are assigned and illustration of obtained results is given on a 2D points data set.

2. Linear support vector machines for separable data. An overview

Let it first be supposed that training data is of the following form:

$$\{(x_i, y_i)\}_{i=1,2,...,m} \tag{1}$$

where every $x_i \in R^n$ represents an input vector and each y_i an output (label).

As already mentioned, the two subsets of input vectors labelled with $+1$ and -1, respectively, are linearly separable. The positive and negative training vectors are then separated by the hyperplane:

$$\langle w, x \rangle - b = 0, \tag{2}$$

where $w \in R^n$ is the normal to the hyperplane, $b \in R$ and $\frac{|b|}{\|w\|}$ is the distance from the origin to the hyperplane.

Accordingly, two subsets of data are linearly separable iff there exist $w \in R^n$ and $b \in R$ such that:

$$\begin{cases} \langle w, x_i \rangle - b > 0, & y_i = 1, \\ \langle w, x_i \rangle - b < 0, & y_i = -1, i = 1, 2, ..., m. \end{cases} \tag{3}$$

According to [1], two subsets of data are linearly separable iff there exist $w \in R^n$ and $b \in R$ such that:

$$\begin{cases} \langle w, x_i \rangle - b > 1, & y_i = 1, \\ \langle w, x_i \rangle - b < -1, & y_i = -1, i = 1, 2, ..., m. \end{cases} \tag{4}$$

Basically, in the statement above, a scaling of the parameters for the separating hyperplane is performed. The separating hyperplane lies thus in the middle of the parallel supporting hyperplanes of the two classes.

Consider the data points (x_i, y_i) for which either the first line or the second of (4) holds with the equality sign. They are called **support vectors**. They are data points that lie closest to the decision surface. Their removal would change the solution found.

Now, following the structural risk minimization principle [5], [6], [7], the support vector machine has a high generalization ability if the separating hyperplane divides the training data with as few errors as possible and, at the same time, with a maximal margin of separation.

Therefore, one is led to the following constrained optimization problem:

$$\begin{cases} \text{find } w \text{ and } b \text{ as to minimize } \frac{\|w\|^2}{2}, \\ \text{subject to } y_i(\langle w, x_i \rangle - b) \geq 1, i = 1, 2, ..., m. \end{cases} \tag{5}$$

Note. After basic calculations, one obtains for margin the value $\frac{2}{\|w\|}$.

3. Evolutionary linear support vector machines for separable data

The standard solving of (5) relies on constructing the Lagrangian function and afterwards applying the Karush-Kuhn-Tucker-Lagrange conditions. Issues of convexity are necessary and sufficient for the above.

The less complicated evolutionary algorithm to solve the constrained optimization problem in (5) is next presented.

Representation of chromosomes A chromosome is a vector of w and b of the form:

$$c = (w_1, ..., w_n, b) \tag{6}$$

where $w_i \in [-1, 1], i = 1, 2, ..., n$ and $b \in [-1, 1]$.

Initial population Each gene of a chromosome is randomly generated with a uniform distribution from the corresponding interval.

Fitness evaluation The fitness function has the subsequent expression:

$$f(c) = f(w_1, ..., w_n, b) = w_1^2 + ... + w_n^2 + \sum_{i=1}^{m}[t(y_i(\langle w, x_i\rangle - b) - 1)]^2, \tag{7}$$

where

$$t(a) = \begin{cases} a, & a < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

One is led to

$$minimize(f(c), c) \tag{9}$$

.

Selection operator Tournament selection is used [3].

Variation operators Intermediate crossover and mutation with normal perturbation are appointed [3].

Stop condition The algorithm stops after a fixed number of generations. At this point it provides the values for w and b. If equations for the supporting hyperplanes are also desired, an appropriate scaling of the decision function has to be performed, according to the proof of (4) from [1].

Once a support vector machine is trained and the equation of the separating hyperplane is found, the way to determine on which side of the hyperplane of decision a given test example x lies is to compute the value of the expression:

$$sgn(\langle w, x\rangle - b) \tag{10}$$

4. Experimental results. Application to a 2-dimensional points data set

A fictitious training data set of $m = 60$ points in a 2-dimensional environment of $[-50, 50] \times [-50, 50]$ was considered as to validate the evolutionary approach to solving the constrained optimization problem within linear support vector machines for the separable case. The data set was chosen as in Figure 1.

The values that were appointed for the evolutionary parameters are given in Table 1.

Illustration of obtained separating and supporting hyperplanes is depicted in Figure 2. It can be noticed how the decision hyperplane separates the squares from the circles. The circled points are the support vectors.

5. Conclusions and future work

The newly developed ESVMs prove to be very successful for binary classification of linearly separable data. The evolutionary approach is much easier for both the developer and the end user than that of the standard approach. The evolutionary solving of the optimization problem leads to the obtaining of w and b directly, while in the classical approach the equation of the optimal hyperplane is determined after Lagrange multipliers are found.

Future work will be held in extending ESVMs to handle nonseparable data, both linearly and nonlinearly. The design of multi-class ESVMs would also be of interest.

References

[1] R.A. Bosch, J.A. Smith, *Separating Hyperplanes and the Authorship of the Disputed Federalist Papers*, American Mathematical Monthly, Volume 105, Number 7, pp. 601-608, 1998

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

[3] D. Dumitrescu, B. Lazzerini, L.C. Jain, A. Dumitrescu, *Evolutionary Computation*, CRC Press, Boca Raton, Florida, 2000

[4] B. Scholkopf, *Support Vector Learning*, Dissertation, Berlin, 1997

[5] V. Vapnik, *Inductive Principles of Statistics and Learning Theory*, In Smolensky, Mozer and Rumelhart (Eds.), Mathematical Perspectives on Neural Networks, Lawrence Erlbaum, Mahwah, NJ, 1995

[6] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Verlag, New York, 1995

[7] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998

6. Tables and figures

ps	gen	pc	pm	ms
100	300	0.3	0.4	0.1

Table 1. Values for parameters of the evolutionary algorithm applied to linear SVMs for separable data; ps stands for population size, gen denotes the number of generations, while pc, pm and ms symbolize the probabilities for crossover, for mutation and the mutation strength, respectively

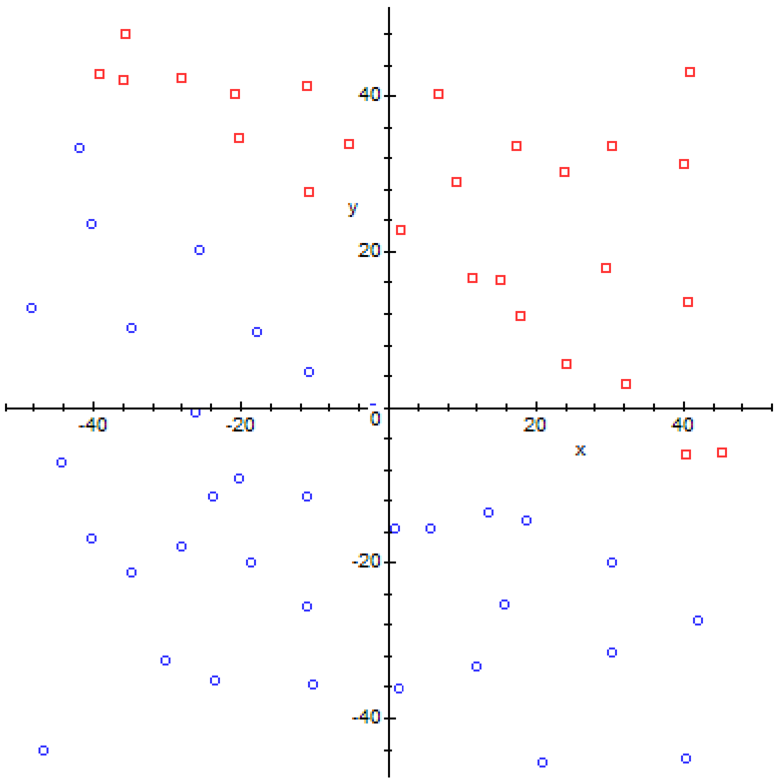

FIGURE 1. The 2D points set to validate linear ESVMs for separable data

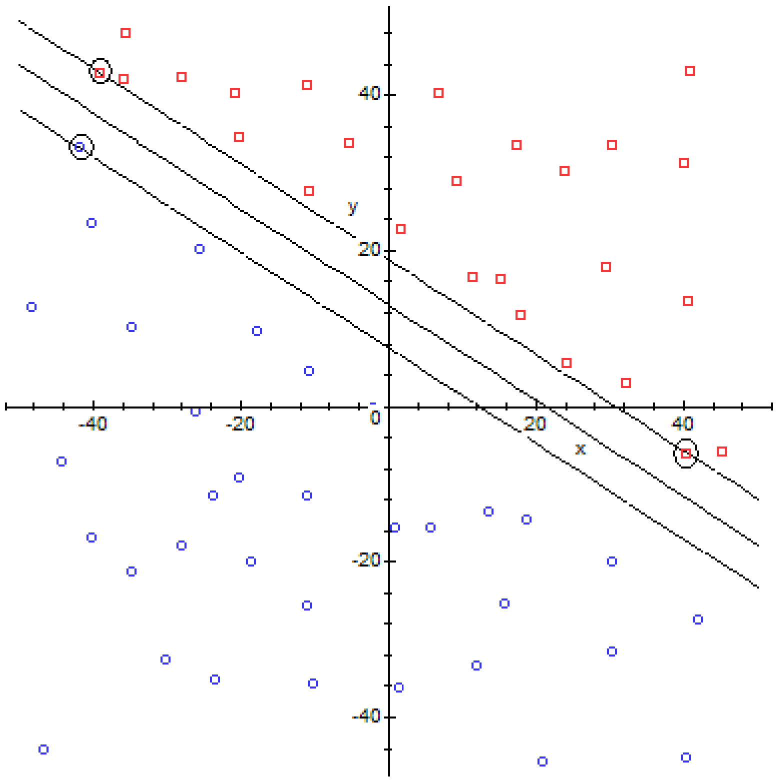

FIGURE 2. The separating hyperplane and the supporting hyperplanes in the case of linear ESVMs for separable data. The support vectors are circled

(Ruxandra Stoean) UNIVERSITY OF CRAIOVA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
DEPARTMENT OF COMPUTER SCIENCE
AL.I. CUZA STREET, NO. 13, CRAIOVA RO-200585, ROMANIA
E-mail address: ruxandra.stoean@inf.ucv.ro

(Dumitru Dumitrescu) "BABES- BOLYAI" UNIVERSITY
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
DEPARTMENT OF COMPUTER SCIENCE
CLUJ-NAPOCA
ROMANIA
E-mail address: ddumitr@cs.ubbcluj.ro