ESANN'2007 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2007, d-side publi., ISBN 2-930307-07-2.

Interval discriminant analysis using Support Vector Machines

Cecilio Angulo[1],Davide Anguita[2] and Luis González[3] *

1- Knowledge Engineering Research Group - Technical University of Catalonia
08800 Vilanova i la Geltrú - Spain.

2- Dept of Biophysical and Electronic Engineering - University of Genoa
Via All'Opera Pia 11a, 16145 Genoa - Italy

3- Qualitative Methods Research Group - University of Seville
Avda Ramón y Cajal, s/n, 41018 Seville - Spain

Abstract. Imprecision, incompleteness, prior knowledge or improved learning speed can motivate interval–represented data. Most approaches for SVM learning of interval data use local kernels based on interval distances. We present here a novel approach, suitable for linear SVMs, which allows to deal with interval data without resorting to interval distances. The experimental results confirms the validity of our proposal.

1 Introduction

Support Vector Machines (SVMs) are learning machines loosely implementing the Structural Risk Minimization inductive principle [1]. Our aim in this work is to extend the SVM to deal with information represented by intervals [2]. The use of intervals is motivated by several reasons: intervals allow to aggregate training data and therefore reduce the number of samples in the training set; they allow to deal with incomplete or imprecise data due to inaccurate or uncertain measurements, or information derived from linguistic assessments, or data transformed from continuous to discrete values.

The main drawback of Interval Arithmetic [2] is that the interval space does not inherit the Euclidean structure of the real line, therefore it is impossible to consider a direct extension of the most usual real norms. However, it is possible to define a distance: a SVM can build a regression model where coefficients are interval values [3]. For classification problems a distance can be defined in the interval space and combine it with a Gaussian kernel [4], or it is possible to define a nonlinear function being able to deal with intervals on a predefined feature space. All these methods make use of local kernels; our goal, instead, is to directly incorporate the interval concept into the SVM, without defining any particular distance on intervals.

Section 2 briefly recalls the SVM algorithm, while a new formulation, which give rise to an Interval-based SVM (I-SVM), is derived in Section 3. As the size of the resulting learning problem is large, a new approach is developed in Section 4, which drastically reduces the I-SVM size and can be interpreted as being directly based on interval arithmetics. Computational complexity issues and

*This research has been supported by the project EXODUS-ADA (DPI2006-15630-C02-01).

ESANN'2007 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2007, d-side publi., ISBN 2-930307-07-2.

an artificial example of the application of I-SVM are described in the following Section. Finally, conclusions are presented. Due to space limitations, several details and discussion about experimental results have been omitted but can be found in the technical report [5].

2 Support Vector Machine Learning

Let $\mathcal{Z} = \{z_1, \ldots, z_p\} \in (\mathcal{X} \times \mathcal{Y})^p$ be a training set, with $z_i = (x_i, y_i)$, $\mathcal{X} \subset \mathbb{R}^m$, $\mathcal{Y} = \{\pm 1\}$ and let $\phi : \mathcal{X} \to \mathcal{F}$ be a feature mapping: $\mathcal{F}$ is named *feature space* and is endowed with a dot product $\langle \cdot, \cdot \rangle$. Let $\phi(x) \in \mathcal{F}$ be the *representation* of $x \in \mathcal{X}$, then a linear classifier $f_{\mathbf{w}}(x) = \langle \phi(x), \mathbf{w} \rangle + b$ can be sought in the space $\mathcal{F}$, with $f_{\mathbf{w}} : \mathcal{X} \to \mathbb{R}$, $b \in \mathbb{R}$ and the class label is obtained by thresholding its output $h_{\mathbf{w}}(x) = \text{sign}(f_{\mathbf{w}}(x))$. Let us define $\beta = \min_{z_i \in \mathcal{Z}_+} \langle \phi(x_i), \mathbf{w} \rangle$, $\alpha = \max_{z_i \in \mathcal{Z}_-} \langle \phi(x_i), \mathbf{w} \rangle$, where $\mathcal{Z}_+$ and $\mathcal{Z}_-$ are, respectively, the patterns belonging to the classes labelled as $\{+1, -1\}$. The classifier $\mathbf{w}$ with the largest geometrical margin $\frac{\beta - \alpha}{\|\mathbf{w}\|}$ on a given training sample $\mathcal{Z}$ can be written as $\mathbf{w}_{SVM} = \arg\max_{\mathbf{w} \in \mathcal{F}} \frac{1}{\|\mathbf{w}\|} \cdot \min_{z_i \in \mathcal{Z}} y_i \langle \phi(x_i), \mathbf{w} \rangle$. The SVM original formulation for solving this problem suggests to minimize the norm $\|\mathbf{w}\|$ with $\beta - \alpha = 2$ and a bias term b is introduced by defining $\beta = b + 1$ and $\alpha = b - 1$, the solution resulting in the form $f_{\mathbf{w}_{SVM}}(x) = \sum_i \alpha_i \, y_i \, k(x_i, x) + b$, where $k(x, x') = \langle \phi(x), \phi(x') \rangle$ is the kernel function, and only some α_i have non zero values. In the following, the linear SVM formulation $k(x, x') = \langle x, x' \rangle$ will be considered.

3 I-SVM: A Convex Optimization Approach

As described in [6], prior knowledge in the form of multiple polyhedral sets can be introduced into a reformulation of a linear SVM. All points lying in a polyhedral set can be determined by a general set of linear inequalities, $\mathcal{B} = \{x | Bx \leq d\} \subset \mathbb{R}^m$ with $B \in \mathbb{R}^{s \times m}$ and $d \in \mathbb{R}^s$. By using the Farkas's theorem, it can be shown that, given a weight vector w, a nonempty polyhedron $\mathcal{B}$ lies in the half-space $\{x | w' \cdot x \geq 1\}$ if and only if there exists a vector u such that $B' \cdot u + w = 0$, $d' \cdot u + 1 \leq 0$ and $u \geq 0$. Restricted to intervals, like in our case, the above formulation results in

$$B = \begin{pmatrix} 1 & & \mathbf{0} \\ -1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & 1 \\ \mathbf{0} & & -1 \end{pmatrix}, d = \begin{pmatrix} x_1^U \\ -x_1^L \\ \vdots \\ x_m^U \\ -x_m^L \end{pmatrix} \tag{1}$$

with $B \in \mathbb{R}^{2m \times m}$ and $d \in \mathbb{R}^{2m}$.

We are interested in defining an interval approach for SVM, which we will call I-SVM, by directly inserting interval information into the SVM formulation. Let us suppose to add to the original training set $\mathcal{Z}$ a set of m-dimensional intervals,

$\mathcal{T} = \{(I_{p+1}, y_{p+1}), \ldots, (I_n, y_n)\} \in (\mathbb{I}^m \times \mathcal{Y})^r$, where each $I_i = (I_{i1}, \cdots, I_{im})$, with $I_{ij} = \left[x_{ij}^L, x_{ij}^U\right]$ and $x_{ij}^L \leq x_{ij}^U$, is an element of the space of m-dimensional closed intervals $\mathbb{I}^m$. The most extreme bounds of any interval will be quoted x^L and x^U, so $x^L \leq x^U$. By defining real values as degenerate intervals, where lower and upper bounds are the same $\{x_i\} = [x_i, x_i]$, a general training set $\mathcal{ZT} = \mathcal{Z} \cup \mathcal{T} = \{(I_1, y_1), \ldots, (I_n, y_n)\} = \{z_1, \ldots, z_p, z_{p+1}, \ldots, z_n\} \in (\mathbb{I}^m \times \mathcal{Y})^n$ is obtained.

Using (1) and defining $u^i = \left(u_{11}^i, u_{12}^i, \ldots, u_{m1}^i, u_{m2}^i\right)' \in \mathbb{R}^{2m}$, the QP problem in [6] can be written as $\min_{w \in \mathbb{R}^m, u^i \in \mathbb{R}^{2m}} \frac{1}{2}\|w\|^2$, s.t. $y_i w = u_2^i - u_1^i$, $y_i w' x_i^L \geq 1 + (\Delta x_i)' u_1^i$, $u^i \geq \mathbf{0}$, where $z_i \in \mathcal{ZT}$, $u_s^i = (u_{js}^i)_j$, $s = 1, 2$ and $\Delta x_i = \left(x_{i1}^U - x_{i1}^L, \ldots, x_{im}^U - x_{im}^L\right) = x_i^U - x_i^L \in \mathbb{R}^m$. It can be noted that this problem is over–parameterized, since w can be obtained in n distinct ways. This extra freedom can be exploited for robustness purposes by defining $w = \frac{1}{n}\sum_i y_i(u_2^i - u_1^i)$, which results in a simplified QP problem

$$\begin{array}{lrcl} \min\limits_{u_1^i, u_2^i \in \mathbb{R}^m} & \frac{1}{2n^2}\left\|\sum\limits_i y_i(u_2^i - u_1^i)\right\|^2 & & \\ \text{s.t.} & \left(u_2^i\right)' \cdot x_i^L - \left(u_1^i\right)' \cdot x_i^U & \geq & 1,\ z_i \in \mathcal{ZT} \\ & u_1^i, u_2^i & \geq & \mathbf{0},\ z_i \in \mathcal{ZT} \end{array} \quad (2)$$

As in the usual SVM framework, the dual formulation can be derived by using the Lagrangian $L\left(u_1^i, u_2^i\right) = \frac{1}{2n^2}\left\|\sum_i y_i(u_2^i - u_1^i)\right\|^2 - \sum_i u_2^i \cdot \left(\alpha_i \cdot x^L + \nu_i\right) + \sum_i u_1^i \cdot \left(\alpha_i \cdot x^U - \mu_i\right) + \sum_i \alpha_i$ with $\alpha_i \geq 0$, $\mu_i, \nu_i \in \mathbb{R}^m$ and $\mu_i, \nu_i \geq \mathbf{0}$. The Karush-Kuhn-Tucker (KKT) conditions for optimality

$$\frac{1}{n^2} y_i \sum_j y_j \left(u_2^j - u_1^j\right) = \alpha_i \cdot x_i^U - \mu_i \ , \ \frac{1}{n^2} y_i \sum_j y_j \left(u_2^j - u_1^j\right) = \alpha_i \cdot x_i^L + \nu_i \quad (3)$$

can be used to compute the weight vector. Here, without losing generality, the second condition is used to obtain the dual QP problem,

$$\begin{array}{ll} \min\limits_{\alpha_i, \nu_{ij}} & \frac{1}{2}\gamma' \cdot \begin{pmatrix} \mathbf{q_L}'\mathbf{q_L} & \mathbf{q_L}'\left(\mathbf{I_m}, \ldots, \mathbf{I_m}\right) \\ \left(\mathbf{I_m}, \ldots, \mathbf{I_m}\right)' \mathbf{q_L} & \mathbf{I_{nm}} \end{pmatrix} \cdot \gamma - (\mathbf{1}', \mathbf{0}') \cdot \gamma \\ \text{s.t.} & \alpha_i \geq 0,\ z_i \in \mathcal{ZT} \\ & \nu_i \geq \mathbf{0},\ z_i \in \mathcal{ZT} \end{array} \quad (4)$$

given that $\mathbf{q_L} = \left(y_i \cdot x_i^L\right)_i$ and $\gamma' = (\alpha', \nu')$. The vector solution can be written, $w = \mathbf{q_L} \cdot \alpha + \sum_i y_i \cdot \nu_i$ and the bias b is obtained a–posteriori in the usual way.

4 I-SVM: an Interval Arithmetic Approach

Despite the simplifications introduced in the previous section, the size of the QP problem (2), i.e. the number of parameters and constraints, is still very large. A new alternative formulation is developed, leading to a SVM that is directly based on interval arithmetic: let us define $(u^i_{j1})_j = w^-, (u^i_{j2})_j = w^+$ for $y_i = 1$, $(u^i_{j1})_j = w^+, (u^i_{j2})_j = w^-$ for $y_i = -1$, and

$$f_i^L \overset{def}{=} \sum_{j=1}^{m} \left(w_j^+\right)' \cdot x_{ij}^L - \sum_{j=1}^{m} \left(w_j^-\right)' \cdot x_{ij}^U \;,\; f_i^U \overset{def}{=} \sum_{j=1}^{m} \left(w_j^+\right)' \cdot x_{ij}^U - \sum_{j=1}^{m} \left(w_j^-\right)' \cdot x_{ij}^L \quad (5)$$

then the primal QP problem can be rewritten as

$$\begin{array}{cl} \min\limits_{w^+,w^-\in\mathbb{R}^m} & \frac{1}{2}\|w^+ - w^-\|^2 \\ & y_i \cdot f_i^L \geq 1, \quad y_i = 1 \\ \text{s.t.} & y_i \cdot f_i^U \geq 1, \quad y_i = -1 \\ & w^+, w^- \geq \mathbf{0} \end{array} \quad (6)$$

where $w = w^+ - w^-$, a much simpler formulation than those obtained in the previous Section and clearly showing the connection with the conventional SVM.

4.1 Interval Arithmetic Interpretation

Given two m-dimensional intervals, I^1,$I^2 \in \mathbb{I}^m$, with $I^j_i = [x^L_{ji}, x^U_{ji}]$, then the inequality relation is defined as $I^1 \succeq I^2 \Leftrightarrow x^L_1 \geq x^U_2$, and $I^1 \succeq q \Leftrightarrow x^L_1 \geq q$, where $q \in \mathbb{R}^m$. From (5), since $w^+, w^- \geq \mathbf{0}$, it can be shown that $\min_{x\in I_i} w'\cdot x = f_i^L$, $\max_{x\in I_i} w'\cdot x = f_i^U$ and $w' \cdot I_i = \{w' \cdot x \,|\, x \in I_i\} = [f_i^L, f_i^U]$. Therefore, the QP problem (6) can be considered, from an interval arithmetic perspective, as

$$\begin{array}{cl} \min\limits_{w^+,w^-\in\mathbb{R}^m} & \frac{1}{2}\|w^+ - w^-\|^2 \\ \text{s.t.} & y_i \cdot \left[f_i^L, f_i^U\right] \succeq 1, \quad z_i \in \mathcal{ZT} \\ & w^+, w^- \geq \mathbf{0} \end{array} \quad (7)$$

Let us define $\beta = \min\limits_{z_i\in\mathcal{Z}_+} y_i f_i^L$, $\alpha = \max\limits_{z_i\in\mathcal{Z}_-} y_i f_i^U$, then the QP problem (7) corresponds to finding the classifier with the largest geometrical margin $\frac{\beta-\alpha}{w}$ on a given training set of intervals, $w_{I-SVM} \overset{def}{=} \arg\max\limits_{w\in\mathbb{R}^m} \frac{1}{\|w\|} \cdot \min\limits_{z_i\in\mathcal{ZT}} \left(\min\limits_{x\in I_i} y_i \cdot w' \cdot x\right)$. Thus, using an interval arithmetic framework, (7) can be rewritten as

$$\begin{array}{cl} \min\limits_{w\in\mathbb{R}^m} & \frac{1}{2}\|w\|^2 \\ & y_i \cdot w' \cdot I_i \succeq 1, \quad z_i \in \mathcal{ZT} \end{array}$$

which is a direct generalization of the standard SVM learning problem, where a function from $\mathbb{I}^m$ to $\mathbb{I}^1$ is defined as $f_w(I) \overset{def}{=} \left[f_w^L(I), f_w^U(I)\right] = w' \cdot I$, given that $I = [x_i^L, x_i^U]$, $w = [w_i^L, w_i^U]$ with $w_i^L = w_i^U = w_i$.

4.2 I-SVM: QP Dual by Interval Analysis

From the primal formulation, the Lagrangian function $L\left(w^+, w^-\right)$ can be written as $\frac{1}{2}\left\|w^+ - w^-\right\|^2 - (w^+)'\left(\sum_{\mathcal{Z}_+} \gamma_i x_i^L - \sum_{\mathcal{Z}_-} \beta_i x_i^U + \mu\right) + \sum_{\mathcal{Z}_+} \gamma_i + \sum_{\mathcal{Z}_-} \beta_i -$

Approach	Param.	Const.	Approach	Param.	Const.
Primal Convex	$2nm$	$(2m+1)n$	Dual (Eq. 4)	$(m+1)n$	$(m+1)n$
Interval	$2m$	$n+2m$	(Eq. 9)	$n+m$	$n+m$
Standard	m	$n2^m$		$n2^m$	$n2^m$

Table 1: Complexity comparison between I-SVM approaches.

$(w^-)'\left(\sum_{\mathcal{Z}_+}\gamma_i x_i^U + \sum_{\mathcal{Z}_-}\beta_i x_i^L + \nu\right)$, with $\gamma_i, \beta_i \geq 0$, $\mu, \nu \geq \mathbf{0}$. As in the previous case, two expressions are derived from the KKT conditions

$$w^+ - w^- = \sum_{\mathcal{Z}_+}\gamma_i x_i^L - \sum_{\mathcal{Z}_-}\beta_i x_i^U + \mu \ , \ w^+ - w^- = \sum_{\mathcal{Z}_+}\gamma_i x_i^U - \sum_{\mathcal{Z}_-}\beta_i x_i^L - \nu \quad (8)$$

and both can be employed to calculate the weight vector $w = w^+ - w^-$. Without any loss in generality, the first expression is used and it is defined $\alpha_i \stackrel{def}{=} \gamma_i$ for the positive labels (assumed as the first patterns) and $\alpha_i \stackrel{def}{=} \beta_i$ for the negative ones (the latter). Dual QP problem can be derived,

$$\begin{array}{ll} \min\limits_{\alpha_i,\mu_j} & \frac{1}{2}\gamma' \cdot \left(\begin{array}{cc|c} \mathbf{q_1}'\mathbf{q_1} & \mathbf{q_1}'\mathbf{q_2} & \mathbf{q_1}' \\ \mathbf{q_2}'\mathbf{q_1} & \mathbf{q_2}'\mathbf{q_2} & -\mathbf{q_2}' \\ \hline \mathbf{q_1} & -\mathbf{q_2} & \mathbf{I_m} \end{array}\right) \cdot \gamma - (\mathbf{1}', \mathbf{0}') \cdot \gamma \\ \text{s.t.} & \begin{array}{l} \alpha_i \geq 0 \ , \quad z_i \in \mathcal{ZT} \\ \mu_j \geq 0 \ , \quad j = 1, \ldots, m \end{array} \end{array} \quad (9)$$

with $\gamma' = (\alpha', \mu')$ and $\mathbf{q} = (\mathbf{q_1}, \mathbf{q_2}) = \left((y_i \cdot x_i^L)_{\mathcal{Z}_+}, (y_i \cdot x_i^U)_{\mathcal{Z}_-}\right)$. Vector w is computed as $w = \mathbf{q} \cdot \alpha + \mu$ and bias term b is obtained as usual.

It is possible to show a connection between the Convex Optimization approach developed in previous sections and these results obtained through Interval Arithmetic. By observing that $\sum_i y_i\left(\alpha_i x_i^L + \nu_i\right) = \sum_{\mathcal{Z}_+}\alpha_i x_i^L - \sum_{\mathcal{Z}_-}\alpha_i x_i^U + \sum_{\mathcal{Z}_+}\nu_i + \sum_{\mathcal{Z}_-}\mu_i$, it can be defined $\mu = \sum_{\mathcal{Z}_+}\nu_i + \sum_{\mathcal{Z}_-}\mu_i \geq 0$ and the dual QP problem (9) is obtained.

5 Computational Complexity and Experimental Result

In order to completely analyze the computational complexity of the two presented approaches, the translation of each multidimensional interval contained in the training set to an equivalent set of points is also considered. In this case, the QP problem of the conventional SVM becomes $\min_{w\in\mathbb{R}^m} \frac{1}{2}\|w\|^2$ s.t. $y_i \cdot w' \cdot x_i^j \geq 1$, $z_i \in \mathcal{ZT}$, $j = 1, \ldots, 2^m$, where x_i^j is any of the 2^m vertices of the interval $I_i \in \mathcal{ZT}$. Table 1 summarizes the computational complexity of the different approaches: the Interval Arithmetic approach results in the more balanced QP problem, showing both a small number of parameters to be optimized and a small number of constraints.

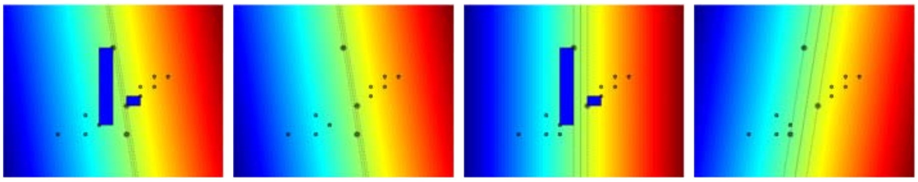

Fig. 1: I-SVM versus standard SVM.

Fig. 1 shows an example of this approach on two artificial datasets. Interval information is not critical for the first training set (left pictures), so I-SVM recovers the conventional SVM. The interval information (parameter $\mu \neq 0$), instead, is used in the second training set, improving the pointwise information used for the conventional SVM (right pictures).

6 Conclusion

Imprecision in the input information, incompleteness on the patterns, discretization procedures, prior knowledge insertion or learning speed-up can motivate interval represented data. Differently from existing SVM approaches working on interval data, a new formulation for a linear SVM classifier has been derived, called I-SVM, by inserting directly interval information in the SVM algorithm. The new approach drastically reduces the original convex-based approach complexity and increases its robustness, whereas it can be interpreted as based on Interval Arithmetic. Future research will address the extension to the nonlinear case by introducing the interval information in nonlinear kernels.

References

[1] V. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, New York, 1998.

[2] R. E. Moore. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, 1966.

[3] 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(8):1114–1125, 2006.

[4] Y. Zhao, Q. Chen, and Q. He. An interval set classification based on support vector machines. In *Proc. of ICAS/ICNS 2005*, pages 81–86, 2005.

[5] C. Angulo, D. Anguita, and L. González. Interval discriminant analysis using support vector machines. Technical report, Technical Univ. Catalonia, June 2006. http://www.epsevg.upc.edu/cecilio/articles/esaiirr0610.pdf.

[6] G.M. Fung, O.L. Mangasarian, and J.W. Shavlik. Knowledge-based support vector machine classifiers. In *Advances in Neural Information Processing Systems 15*, pages 521–528. MIT Press, Cambridge, MA, 2003.