

Support Vector Machines

Cecilio Angulo

Technical University of Catalonia, Spain

Luis Gonzalez-Abril

Technical University of Catalonia, Spain

INTRODUCTION

Support Vector Machines -- SVMs -- are learning machines, originally designed for bi-classification problems, implementing the well-known **Structural Risk Minimization (SRM) inductive principle** to obtain good **generalization** on a limited number of learning patterns (Vapnik, 1998). The optimization criterion for these machines is maximizing the margin between two classes, i.e. the distance between two parallel hyperplanes that split the vectors of each one of the two classes, since larger is the margin separating classes, smaller is the **VC dimension** of the learning machine, which theoretically ensures a good generalization performance (Vapnik, 1998), as it has been demonstrated in a number of real applications (Cristianini, 2000). In its formulation is applicable the **kernel trick**, which improves the capacity of these algorithms, learning not being directly performed in the original space of data but in a new space called feature space; for this reason this algorithm is one of the most representative of the called **Kernel Machines (KMs)**.

Main theory was originally developed on the sixties and seventies by V. Vapnik and A. Chervonenkis (Vapnik et al., 1963, Vapnik et al., 1971, Vapnik, 1995, Vapnik, 1998), on the basis of a separable binary classification problem, however generalization in the use of these learning algorithms did not take place until the nineties (Boser et al., 1992). SVMs has been used thoroughly in any kind of learning problems, mainly in classification problems, although also in other problems like regression (Schölkopf et al., 2004) or clustering (Ben-Hur et al., 2001).

The fields of Optic Character Recognition (Cortes et al., 1995) and Text Categorization (Sebastiani, 2002) were the most important initial applications where SVMs were used. With the extended application of new kernels, novel applications have taken place in the field of Bioinformatics, concretely many works

are related with the classification of data in Genetic Expression (Microarray Gene Expression) (Brown et al., 1997) and detecting structures between proteins and their relationship with the chains of DNA (Jaakkola et al., 2000). Other applications include image identification, voice recognition, prediction in time series, etc. A more extensive list of applications can be found in (Guyon, 2006).

BACKGROUND

Regularization Networks (RNs), obtained from the penalization inductive principle, are algorithms based on a deep theoretical background, but their purely asymptotic approximation properties and the expansion of the solution function on a large number of vectors convert them in a no practical choice in its original definition. Looking for a more reduced expansion of the solution some researchers observe the good behaviour of the SVM, being able to consider a finite training set as hypothesis in its theoretical discourse as well as building the final solution by considering nested approximation spaces.

As well regularization inductive principles as structural risk minimization establish inserting 'a priori' information on the shape of the solution without considering any assumption about the unknown probability density function relating working spaces. The regularization principle considers a regularizer or regularization operator ensuring find a good solution in asymptotic form on nested function spaces when the number of elements in the training set tends to be infinite. Besides, the SRM principle also is based on nested spaces but the solution is found by ensuring an upper bound for the risk functional considering only a finite set of empirical data.

Both inference processes are obviously not equivalents, but their similarities have been projected on the

learning methods having their theoretical background on these principles, in such a form that a number of researchers approaching the learning problem from different perspectives are implied in establishing a common framework allowing to deal SVMs and RNs like particular cases of a more general learning methodology, let us call it Kernel Methods (Campbell, 2000), when it is emphasized the key rule played by the kernel function generating the feature space, or Large Margin Classifiers (Cristianini, 2000), when the measure to be optimized to ensure maximal generalization is emphasized.

Both, results obtained and theoretical framework from SRM seem offer better theoretical warranties than other previous approaches when looking for solutions with good generalization for problems based on a finite empirical set. Hence, the integration of machine learning methods on mixed models is a state of the art research field. Besides, the use of Bayesian inference in these mixed models is being avoided because the user must beforehand define a probability density function.

SUPPORT VECTOR MACHINE

Let us consider a bi-classification problem (another kind of problems are analyzed in the cited references). Thus, let $Z = \{z_i = (x_i, y_i), i = 1, 2, \dots, n\}$ be a training set with $x_i \in X \subset \mathbb{R}^d$ as the input space and $y_i \in \{\theta_1, \theta_2\}$ (the output

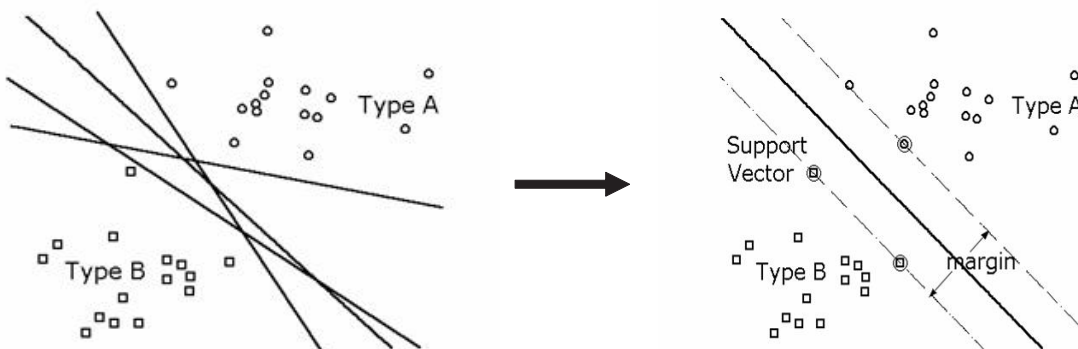
space) ($\theta_1 \neq \theta_2$). Let us initially suppose that classes are linearly separable (two sets are linearly separable in n -dimensional space if they can be separated by an $d-1$ dimensional hyperplane) then a hyperplane, denoted by $\pi : w x - b = 0$ (where b is called bias), is sought which separates the two classes, that is $w x_i - b > 0$ if $y_i = \theta_1$ and $w x_i - b < 0$ if $y_i = \theta_2$. Nevertheless, there are many hyperplanes with this condition (see **Figure 1**), so a new condition is imposed that is the distance between the optimal hyperplane and the nearest training pattern (margin) is maximal. Let us see detailed this condition: In the first place without loss of generality let us suppose that $\theta_1 = 1$ and $\theta_2 = -1$. Hence, let β and α be the minimum (class $+1$) and the maximum (class -1) absolute values of the unbiased hyperplane effectively attained for some patterns $z_1 \in Z_1$ and $z_2 \in Z_2$ i.e.

$$\alpha = \max_{z_i \in Z_2} w x_i \text{ and } \beta = \min_{z_i \in Z_1} w x_i,$$

where Z_1 and Z_2 are the patterns belonging to the classes labelled as $\{+1, -1\}$ respectively. It is considered that $\alpha \leq \beta$, otherwise vector $-w$ is chosen. Thus, given a vector w , the margin is defined as the distance between parallel hyperplanes $\pi_\alpha : w x - \alpha = 0$ and $\pi_\beta : w x - \beta = 0$, that is

$$\text{margin} = d(\pi_\alpha, \pi_\beta) = \frac{\beta - \alpha}{\|w\|}$$

Figure 1. Type A denotes the class $+1$ (θ_1) and Type B denotes the class -1 (θ_2).



4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/support-vector-machines/10439

Related Content

Reassessing the Foundations of Semiotics: Preliminaries

Mihai Nadin (2012). *International Journal of Signs and Semiotic Systems* (pp. 1-31).

www.irma-international.org/article/reassessing-foundations-semiotics/64637

Synergizing AI and Blockchain: Transforming Aerospace Engineering Operations

R. Elakya, R. Thanga Selvi, T. Manoranjitham and S. Shanthana (2024). *AI and Blockchain Optimization Techniques in Aerospace Engineering* (pp. 193-209).

www.irma-international.org/chapter/synergizing-ai-and-blockchain/341334

Hybrid Approach Using Deep Autoencoder and Machine Learning Techniques for Cyber-Attack Detection

Vikash Kumar and Ditipriya Sinha (2022). *International Journal of Ambient Computing and Intelligence* (pp. 1-21).

www.irma-international.org/article/hybrid-approach-using-deep-autoencoder-and-machine-learning-techniques-for-cyber-attack-detection/293098

Extension of the TOPSIS for Multi-Attribute Group Decision Making under Atanassov IFS Environments

Deng-Feng Li and Jiang-Xia Nan (2011). *International Journal of Fuzzy System Applications* (pp. 47-61).

www.irma-international.org/article/extension-topsis-multi-attribute-group/60380

Personalized Decision Support Systems

Neal Shambaugh (2009). *Encyclopedia of Artificial Intelligence* (pp. 1310-1315).

www.irma-international.org/chapter/personalized-decision-support-systems/10409