Chapter XXVIII
Support Vector Machine:
Itself an Intelligent Systems

A. B. M. Shawkat Ali
Central Queensland University, Australia

ABSTRACT

From the beginning, machine learning methodology, which is the origin of artificial intelligence, has been rapidly spreading in the different research communities with successful outcomes. This chapter aims to introduce for system analysers and designers a comparatively new statistical supervised machine learning algorithm called support vector machine (SVM). We explain two useful areas of SVM, that is, classification and regression, with basic mathematical formulation and simple demonstration to make easy the understanding of SVM. Prospects and challenges of future research in this emerging area are also described. Future research of SVM will provide improved and quality access to the users. Therefore, developing an automated SVM system with state-of-the-art technologies is of paramount importance, and hence, this chapter will link up an important step in the system analysis and design perspective to this evolving research arena.

INTRODUCTION

Since the end of the last century, support vector machines (SVMs) have been introduced for classification and regression in the machine learning community. SVMs have a solid theoretical foundation rooted in statistical learning theory. SVMs work step by step. First, it maps the data into a high dimensional space via a nonlinear map, and in this high dimensional space it constructs an optimal separating hyperplane or linear regression function. This hyperplane or linear regression function obtained in the feature space couple with significant data points for prediction called support vectors (SVs). Therefore SVM do prediction based on SVs' information only. This process will involve a quadratic programming problem, and this will get a global optimal solution. This chapter formulates the statistical method of SVM based on classification and regression architecture. We explained both SVM classification methods: binary and muticlass classification. In each section we included a demonstration to easily understand SVM classification and regression methodology. Finally we give attention on how system analysers and designers can contribute to the construction of a full automated support vector learning algorithm.

BACKGROUND

The popular statistical learning algorithm, SVM, is an advanced version of the generalised portrait algorithm, which was developed in Russia in the late 60s (Smola & Schölkopf, 1998). After a large gap, Vapnik (1995) and his group simplified this theory and introduced SVM as an effective learning algorithm. Although SVM does not have a long history, it has already been successfully applied with significant outcomes in business, engineering, science, medicine, and many more.

SVM was introduced first to solve binary class pattern recognition problems, and then multiclass classification, regression estimation, and so many others. We can divide SVM literature into two parts: SVM classification and SVM regression. The following sections will cover both parts following the basic explanation with example.

SVM CLASSIFICATION

Let us consider a dataset D of l independently identically distributed (i.i.d) samples: $(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_l, y_l)$. Each sample is a set of feature vectors of length m, $\mathbf{x}_i = \langle x_1, \cdots, x_m \rangle$ and the target value $y_i \in \{1, -1\}$ that represents the binary class membership. Now, the pattern recognition problem or machine learning task is to learn the classes for each pattern by finding a classifier with decision functions $f(\mathrm{x}_i, \alpha_i)$, where, $f(\mathbf{x}_i, \alpha_i) = y_i, \alpha_i \in \Lambda, \forall \langle \mathbf{x}_i, y_i \rangle \in D$ and Λ is a set of abstract parameters.

Linear Hard Margin SVM

Now-a-days the general nonlinear SVM is quite popular to solve pattern recognition problems, but the root of this method is linear SVM. The original linear SVM was introduced to separate the binary class problem only.

Let us consider the above pattern recognition problem. Our aim is to find out the optimal hyperplane (OH) in the training phase with proper estimation of a weight vector w and the scalar bias factor b. All the training patterns are said to be linearly separable if there exists w and b such that the inequalities

$$(\omega \cdot \mathbf{x}_i) + b \geq 1 \qquad \text{if } y_i = 1 \text{ or } \bullet \tag{1}$$

$$(\omega \cdot \mathbf{x}_i) + b \leq -1 \qquad \text{if } y_i = -1 \text{ or } ☾ \tag{2}$$

These two sets of inequalities we can present into a single set such as

$$y_i = sign(\omega \cdot \mathbf{x}_i + b), \qquad i = 1, \ldots, \ell \tag{3}$$

After extracting the OH, the whole set of vectors $\{\mathrm{w}\}$ will satisfy the Equation (3) as described in Figure 1.

The margin could be found by measuring the distance d between the binary classes' data points as shown in Figure 1 as follows:

$$\begin{aligned} d = d_{+1} + d_{-1} &= \min_{i:(y_i=+1)} d(\omega, b; \mathbf{x}_i) - \min_{i:(y_i=-1)} d(\omega, b; \mathbf{x}_i) \\ \mathrm{m} &= \mathrm{in}_{i:(y_i=+1)} \frac{|\langle \omega, \mathbf{x}_i \rangle + b|}{\|\omega\|} - \min_{i:(y_i=-1)} \frac{|\langle \omega, \mathbf{x}_i \rangle + b|}{\|\omega\|} \\ \mathrm{m} &= \frac{1}{\|\omega\|} \left(\mathrm{in}_{i:(y_i=+1)} |\langle \omega, \mathbf{x}_i \rangle + b| - \min_{i:(y_i=-1)} |\langle \omega, \mathbf{x}_i \rangle + b| \right) \\ &= \frac{2}{\|\omega\|} \end{aligned} \tag{4}$$

where d_{+1} and d_{-1} are the distance of the closest positive and negative class data points from the OH.

Now the OH can be obtained through the minimisation:

$$\Phi(\omega) = \frac{1}{2} \|\omega\|^2 \tag{5}$$

subject to: $y_i[(\omega \cdot \mathbf{x}_i) + b] \geq 1$.

It is important to mention that the OH solution is independent with the scalar quantity bias b. The OH will shift up or down due to any changing of b, but the maximum margin will remain same.

Related Content

Characterizing PaaS Solutions Enabling Cloud Federations
Tamas Pflanzner, Roland Tornyai, Ákos Zoltán Goráczand Attila Kertesz (2018). *Application Development and Design: Concepts, Methodologies, Tools, and Applications (pp. 1095-1120).*
www.irma-international.org/chapter/characterizing-paas-solutions-enabling-cloud-federations/188247

Facilitating eLearning with Social Software: Attitudes and Usage from the Student's Point of View
Reinhard Bernsteiner, Herwig Ostermannand Roland Staudinger (2009). *Software Applications: Concepts, Methodologies, Tools, and Applications (pp. 1979-1997).*
www.irma-international.org/chapter/facilitating-elearning-social-software/29490

Ranking and Risk Factor Scheme for Malicious applications detection and Classifications
Kiran Khatterand Sapna Malik (2018). *International Journal of Information System Modeling and Design (pp. 67-84).*
www.irma-international.org/article/ranking-and-risk-factor-scheme-for-malicious-applications-detection-and-classifications/218172

Implications of Markup on the Description of Software Patterns
Pankaj Kamthan (2011). *Knowledge Engineering for Software Development Life Cycles: Support Technologies and Applications (pp. 136-160).*
www.irma-international.org/chapter/implications-markup-description-software-patterns/52881

Analysis of Existing Software Cognitive Complexity Measures
Sanjay Misra, Adewole Adewumi, Robertas Damaseviciusand Rytis Maskeliunas (2017). *International Journal of Secure Software Engineering (pp. 51-71).*
www.irma-international.org/article/analysis-of-existing-software-cognitive-complexity-measures/204524