Information Sciences 430–431 (2018) 432–443

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Separability of set-valued data sets and existence of support hyperplanes in the support function machine☆

Jiqiang Chen[a,b], Xiaoping Xue[a,*], Litao Ma[b], Minghu Ha[b]

[a] Department of Mathematics, Harbin Institute of Technology, Harbin 150001, PR China
[b] School of Science, Hebei University of Engineering, Handan 056038, PR China

ARTICLE INFO

Article history:
Received 19 April 2017
Revised 27 November 2017
Accepted 29 November 2017

Keywords:
Support vector machine
Set-valued data
Separability
Support hyperplane
Support function

ABSTRACT

The support function machine (SFM) has been shown to be effective in separating set-valued data sets. However, in SFM, the separability of set-valued data and the existence of support hyperplanes, which can provide useful guidance for improving algorithms for use in applications, have not been discussed in theory. Therefore, in this paper, we firstly discuss the problem of whether the linearly separable set-valued data in $\mathbb{R}^d$ are still linearly separable after being mapped into the infinite-dimensional Banach space $C(S)$ by support functions. Secondly, we discuss the problem of whether the linearly inseparable set-valued data in $\mathbb{R}^d$ are linearly separable after being mapped into $C(S)$. If not, in which situations are they linearly separable? Thirdly, we discuss the existence of support hyperplanes in SFM. Finally, two experiments with set-valued data sets are provided to verify the reasoning in the above discussions and the correctness of their conclusions.

1. Introduction

In some practical problems, such as water quality evaluation [22] and gene expression experiments [35], multiple measurements or replicated experiments are often used to reduce the level of uncertainty, which leads to a new kind of learning task: set-based classification [2,13,33,37]. Currently, there are two methods of approaching set-based classification. One is to compute the statistics of the original data (such as mean and median) and describe the input set with a vector, such as CART [4], ID3 [24], C4.5 [24], and SVMs [7,9,14,20]. However, according to the law of large numbers, when the number of samples tends to infinity, the mean value converges to the real value, but in the actual problems, the number of measured data that make up a set-valued datum cannot reach infinity, so the sample mean cannot adequately represent the real value. Furthermore, when a set-valued datum is represented by a vector-valued datum, some other information (such as the variance) may be lost in the reprocessing [6]. The other method is to develop set-based classifiers directly [21,27,30,32]. However, these methods usually state some assumptions in advance (for example, they assume that the sample sets lie on some manifolds, whereas some classifications may lie on manifolds but others may not), and they do not work in some cases. Therefore, Chen et al. [6] presented the support function machine (SFM), which is a new learning machine for set-based classifications.

☆ Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author.
E-mail addresses: jiqiang516@163.com (J. Chen), xiaopingxue@hit.edu.cn (X. Xue).

https://doi.org/10.1016/j.ins.2017.11.057
0020-0255/

In SFM, the sets of feature vectors are mapped into an infinite-dimensional Banach space $C(S)$ (whose elements are the continuous functions defined on the unit ball S in $\mathbb{R}^d$) via support functions $\sigma(\mathbf{x})$ [12], and the set becomes a single point (namely a function) in this new Banach space $C(S)$. Then, the set-based tasks [16,26,28,38] in d-dimensional Euclid space $\mathbb{R}^d$ are converted into function-based tasks in Banach space $C(S)$. As $C(S)$ is not an inner space, the separating hyperplane in SFM is defined via a Radon measure μ whose theoretical basis is the Riesz representation theorem in Banach space [25], which is different from that in support vector machines (SVMs) [1,3,7,9,10,17,18,23,29,36]. Then, we construct the maximal margin algorithm in this new Banach space. Consequently, the SFM retains the classification information of the original set-valued data and can deal with the set-based classifications effectively.

Moreover, SFM is able to deal with function- (or distribution-) based classifications [6,31], learning tasks described with fuzzy sets, as we can map the fuzzy sets into Banach space $C(S)$ with membership functions [6,15]. After mapping, all of the above tasks are converted into function-based classifications. Then we can train classifiers in $C(S)$. In addition, as a vector $\mathbf{x}$ can be represented by a point set $\{\mathbf{x}\}$, vector-based classifications can also be handled with the proposed SFM. Therefore, the new algorithm is powerful for different data representations.

However, the hard margin SFM designed for linearly separable set-valued data is constructed directly via the maximal margin algorithm, and the soft margin SFM targeting linearly inseparable set-valued data is constructed by introducing slack variables. That is to say, we only care about establishing the SFMs themselves but do not consider the separability of set-valued data sets and the existence of support hyperplanes in the theoretical. The following three arguments indicate that it is meaningful to investigate the separability of set-valued data sets and the existence of support hyperplanes. First, linearly separating the finite function-valued data sets in $C(S)$ with the Hahn–Banach Theorem [8] is equivalent to separating their convex hulls linearly, and their convex hulls are infinite sets, so we have implicitly considered the linear separability of special infinite data sets when separating finite sets linearly [5]. Second, if we prove that the data sets are linearly separable in theory, but we cannot separate them completely via the hard margin SFM, we can infer that there is some mistake in the experiments or in the source code. Therefore, the discussions of separability can help us analyze the experimental results. Last, for a practical classification problem, certainly we wish to analyze the algorithm's complexity, including the number of support functions lying on the support hyperplanes, and this inspires us to discuss the existence of support hyperplanes. Thus, it is necessary to investigate the separability of set-valued data sets and the existence of support hyperplanes at least from the theoretical viewpoint, and such investigation can offer guidance for improving algorithms for use in practical problems [11,19,22,34].

Following the above considerations, this paper is organized as follows. Section 2 reviews some basic content related to SFM. Section 3 discusses the linear separability of two set-valued data sets. In Section 4, we discuss the problem of determining the situations in which linearly inseparable data sets are linearly separable after being mapped into $C(S)$. In Section 5, the existence of support hyperplanes is discussed. Section 6 provides two numerical experiments to verify the points made in the above discussions, and Section 7 draws conclusions and suggests future studies.

2. Some preliminaries about SFM

In order to make this paper self-contained, we provide some preliminaries about SFM in this section.

Definition 1 [8]**.** Let $\mathcal{K}$ be a normed linear space, let f be a continuous linear functional on $\mathcal{K}$, let E and F be two subsets of $\mathcal{K}$, and let $L = H_f^r \triangleq \{\mathbf{x} \in \mathcal{K} | f(\mathbf{x}) = r.\}$ $(r \in \mathbb{R})$ be a hyperplane. If for any $\mathbf{x} \in E$, we have $f(\mathbf{x}) \le r$ $(\text{or} \ge r)$, and for any $\mathbf{x} \in F$, we have $f(\mathbf{x}) \ge r$ $(\text{or} \le r)$, then we say that the hyperplane L separates sets E and F. If for any $\mathbf{x} \in E$, we have $f(\mathbf{x}) < r$ $(\text{or} > r)$, and for any $\mathbf{x} \in F$, we have $f(\mathbf{x}) > r$ $(\text{or} < r)$, then we say that the hyperplane L strongly separates sets E and F.

Theorem 1 [8] (Geometric Hahn–Banach Theorem)**.** *Let B^* be a normed linear space, let E_1 and E_2 be two non-empty disjoint convex sets in B^*, and let $\mathbf{x}_0$ be an interior point of E_1 and $\mathbf{x}_0 \notin E_2$. Then there exist $r \in \mathbb{R}$ and a nonzero continuous linear functional f such that hyperplane H_f^r separates E_1 and E_2. In other words, there exists a nonzero continuous linear functional f such that for any $\mathbf{x} \in E_1$, we have $f(\mathbf{x}) \le r$, and for any $\mathbf{x} \in E_2$, we have $f(\mathbf{x}) \ge r$.*

Remark 1. Theorem 1 is the key theorem discussing the separability of function-valued data sets; it shows that for any two non-empty disjoint convex sets in B^* there exists a hyperplane separating them completely (see Fig. 1).

Definition 2 [12]**.** The support function $\sigma_A : \mathbb{R}^d \to \mathbb{R}$ of a non-empty closed convex set A in $\mathbb{R}^d$ is given by $\sigma_A(\mathbf{x}) = \sup_{\mathbf{y} \in A}\{\langle \mathbf{x}, \mathbf{y} \rangle\}, \mathbf{x} \in \mathbb{R}^d$. For simplification, we also denote $\sigma_A(\mathbf{x})$ by σ_A.

As the Banach space $C(S)$ is not an inner space, the hyperplane in SFM is defined via the following Riesz representation theorem in Banach space.

Theorem 2 [25]**.** *Assume that X is a compact Hausdorff space. Then for any bounded linear functional Φ on $C(X)$, there is one and only one complex regular Borel measure μ such that*

$$\Phi(\sigma) = \int_X \sigma \, d\mu, \ \sigma \in C(X),$$

and

$$\|\Phi\| = |\mu|(X),$$