Applied Mathematical Modelling 39 (2015) 6351–6358

Contents lists available at ScienceDirect

Applied Mathematical Modelling

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

Semisupervised spherical separation

Annabella Astorino [a], Antonio Fuduli [b,*]

[a] Istituto di Calcolo e Reti ad Alte Prestazioni-C.N.R., c/o DIMES – Università della Calabria, 87036 Rende (CS), Italy
[b] Dipartimento di Matematica e Informatica, Università della Calabria, 87036 Rende (CS), Italy

ARTICLE INFO

Article history:
Received 25 July 2013
Received in revised form 7 October 2014
Accepted 8 January 2015
Available online 11 February 2015

Keywords:
Semisupervised classification
Spherical separation
Nonsmooth optimization

ABSTRACT

We embed the concept of spherical separation of two disjoint finite sets of points into the semisupervised framework. This approach improves efficiency in the solution of real-world classification problems in which the number of unlabeled points is very large and labeling data is in general expensive. We develop a model characterized by an error function which is nonconvex and nondifferentiable, that we minimize by means of a bundle method. Numerical results on some small/large datasets drawn from literature are reported.
© 2015 Elsevier Inc. All rights reserved.

1. Introduction

The objective of pattern classification is to categorize data into different classes on the basis of their similarities. The application field in this area is very vast: see for example text and web classifications, DNA and protein analysis, medical diagnosis, machine vision and many others.

The mathematical programming approaches for pattern classification are based on separation of sets, an interesting field which has become increasingly relevant in the last years. In particular, in binary classification, the separation problem consists of finding an appropriate surface separating two discrete point sets.

The objective of this paper is to focus on separating two finite disjoints sets of points by means of a sphere. Spherical classifiers have been used so far only in the supervised framework (see for example the recent works [1–4]), where the error function is calculated only on the basis of the labeled samples, while, in the applications, often happens that both labeled and unlabeled samples are available.

On the opposite, there exist also many approaches aimed at clustering the data by means of information coming only from the unlabeled points. This is the case of the unsupervised learning.

The semisupervised classification [5] is a relatively recent approach consisting in classifying the data by learning from both the labeled and unlabeled samples. It is something halfway between the supervised and unsupervised machine learning. The main advantage is that in practical cases most of the data are unlabeled and then it could be appealing to entirely exploit the available information. A possible drawback is that, in general, differently from the supervised case, the error function may become more difficult to minimize.

The literature in supervised machine learning is extremely rich. A central role is played by the well known Support Vector Machine (SVM) technique (see [6–10]), which revealed a very effective approach, as confirmed by the large number of papers in this field (see for example the recent works [11–14]). The main idea in the SVM technique is the introduction of the

* Corresponding author.
E-mail addresses: astorino@icar.cnr.it (A. Astorino), antonio.fuduli@unical.it (A. Fuduli).

http://dx.doi.org/10.1016/j.apm.2015.01.044
0307-904X/© 2015 Elsevier Inc. All rights reserved.

concept of "margin" in strict separation of two sets of points by means of a hyperplane. In fact the output of any SVM model is a hyperplane staying in the middle between two parallel hyperplanes supporting, respectively, the two sets. This is performed by maximizing the distance between the support hyperplanes and, at the same time, by minimizing a measure of the misclassification errors. A very relevant advantage of SVM is the possibility to use the "kernel trick", which allows to construct nonlinear separation surfaces in a higher dimensional space. We remark that a semisupervised version of the SVM approach is available too and it is known as TSVM (Transductive Support Vector Machine) [15]. A different recent approach providing a large margin classifier is introduced in [16], where each points set is approximated by a hyperdisk.

We note in passing that in the last years other papers [17–22] have been devoted to the separation of sets by means of nonlinear surfaces directly in the input space, leading in this way to classification rules which are closer to geometrical intuition.

In this paper we propose a classification model based on semisupervised spherical separation. In particular our error function, which is nonsmooth and nonconvex, is constructed by taking into account both label and unlabeled data. This may be appealing especially when the number of unlabeled data is much more big than that of the labeled ones and when to obtain the labels on the data is very costly: this is the case, for example, of webpage classification and of speech recognition.

An important concept related to the semisupervised classification is the "transductive" inference, where the classifier (prediction function) is derived from the information on all the available data, i.e. both the labeled and unlabeled points and it is not aimed at predicting the class label for newly incoming samples, but only at making a decision about the currently available unlabeled objects. Differently, in the "inductive learning" the classifier is constructed on the basis of the information concerning the labeled points with the aim at predicting the label of obviously not yet available newly incoming samples. For this reason the term "Transductive Support Vector Machine" (TSVM) is used to identify a kind of SVM technique, where the objective of the classifier, learned on labeled and unlabeled points, is to predict only the label of unlabeled points.

Different mathematical programming TSVM models exist in literature and they are basically of two types: mixed integer linear [15,23,24] and continuous optimization programs [25–29]. In all such works the basic idea is to obtain both the best support vector machine in terms of correct classification of the labeled data and as few as possible unlabeled points in the margin zone. An extension of the TSVM technique to the polyhedral separability is given in [30], while in [31] a data preprocessing technique for reducing the number of unlabeled samples has been proposed. Further references on optimization approaches in semisupervised learning can be found in [32].

Starting from the above observations, we propose a spherical classification model based on the same guidelines of the TSVM technique. In particular, we provide a separating sphere staying in the middle between two "support spheres" playing the same role of the support hyperplanes in SVM. More precisely, we embed in our approach the idea coming from TSVM taking into account the minimization of the number of unlabeled points in the margin zone, i.e. the area comprised between the two support spheres.

Throughout the paper we denote by $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^n$.

The paper is organized as follows. In the next section, we briefly describe the main properties of spherical separation, with particular emphasis to the concept of margin [3], which in the semisupervised learning plays a very important role. In Section 3 we present our semisupervised spherical model, that we treat by a nonsmooth bundle type method described in Section 4. Finally in Section 5 some numerical results are given.

2. The supervised spherical margin separation

In this section we recall the concept of margin spherical separation introduced in [3].

Given two nonempty and disjoint finite sets of sample points in the n-dimensional space $\mathbb{R}^n$

$$\mathcal{A} \triangleq \{a_1, \ldots, a_m\} \quad \text{and} \quad \mathcal{B} \triangleq \{b_1, \ldots, b_k\},$$

the objective of spherical separation is to find a separation sphere of the set $\mathcal{A}$ from the set $\mathcal{B}$, i.e. a sphere enclosing all points of $\mathcal{A}$ and no points of $\mathcal{B}$.

The sets $\mathcal{A}$ and $\mathcal{B}$ are defined to be spherically separated by $S(x_0, R)$, the sphere centered in $x_0 \in \mathbb{R}^n$ of radius $R \in \mathbb{R}$, if

$$\begin{cases} \|a_i - x_0\|^2 \leqslant R^2, & i = 1, \ldots, m \\ \|b_l - x_0\|^2 \geqslant R^2, & l = 1, \ldots, k. \end{cases} \tag{1}$$

We remark that the role of the two sets $\mathcal{A}$ and $\mathcal{B}$ is not symmetric, as it may happen that $\mathcal{A}$ is separable from $\mathcal{B}$ but the reverse is not true.

Moreover if inequalities (1) are strictly satisfied, then $\mathcal{A}$ and $\mathcal{B}$ are strictly spherically separated by $S(x_0, R)$, i.e. there exists M (the margin), with $0 < M \leqslant R$, such that

$$\begin{cases} \|a_i - x_0\|^2 \leqslant (R - M)^2, & i = 1, \ldots, m \\ \|b_l - x_0\|^2 \geqslant (R + M)^2, & l = 1, \ldots, k. \end{cases} \tag{2}$$

According to (2), the classification error function w associated to the sphere $S(x_0, R)$ is

Download English Version:

https://daneshyari.com/en/article/1703168

Download Persian Version:

https://daneshyari.com/article/1703168

Daneshyari.com