Pattern Recognition Letters 53 (2015) 93–99

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

New one-class classifiers based on the origin separation approach ☆

Mario Michael Krell [a,*], Hendrik Wöhrle [b]

[a] University of Bremen, Faculty 3 – Mathematics and Computer Science, Robotics Lab, Robert-Hooke-Str.1, 28359 Bremen, Germany
[b] German Research Center for Artificial Intelligence, DFKI Bremen, Robotics Innovation Center, Robert-Hooke-Str. 1, 28359 Bremen, Germany

ARTICLE INFO

Article history:
Received 18 October 2014
Available online 5 December 2014

Keywords:
Origin separation approach
Online learning
Unary classification
Support vector machines
Passive-aggressive algorithms
Balanced relative margin machine

ABSTRACT

The model of the one-class support vector machine (νoc-SVM) is based on the "origin separation approach," i.e., to add a sample at the origin to the training data for the second class and apply a maximum margin separation as known from the classical SVM (C-SVM). This has been proven only for hard margin separation but a clearly defined relation between the νoc-SVM and the C-SVM is not yet existing. In this work, the origin separation approach is analyzed in more detail. The approach reveals to be a more general concept to relate binary and unary (one-class) classifiers. We *prove* how its application to the ν-SVM, a variant of the C-SVM, directly results in the νoc-SVM. Furthermore, we apply this concept to the C-SVM and other related methods (balanced relative margin machine, regularized Fisher's discriminant analysis, online passive-aggressive algorithms) to derive entirely *new* classifiers. This includes variants that can be updated *online* which allows the application on large datasets or on systems with very limited resources.

1. Introduction

Focussing the classification on one class is a common approach if there are not enough examples for a second class (e.g., novelty and outlier detection [2]), or if the goal is to describe a single target class and its distribution [20]. Some unary (one-class) classifiers are modifications of binary ones like k-nearest-neighbors [2,15], decision trees [5], and SVMs [6,22,28]. This paper focusses on the connections between SVM variants, and their unary counterparts.

The C-SVM is based on the geometrical concept of maximizing the distance between two hyperplanes, which separate samples of two different classes [8,17,21,30]. The classical one-class SVM (νoc-SVM) was presented by Schölkopf et al. [20] as a model for "Estimating the support of a high-dimensional distribution" just 1 year after the publication of the ν-SVM [22]. In both cases, the algorithms were mainly motivated by their theoretical properties and a parameter ν has been introduced which is a lower bound in the fraction of support vectors. Support vectors are training samples that define the decision function of the classifier. For the ν-SVM, an equivalence to solutions of the C-SVM has been proven [4,22]. It has been shown [20] that the νoc-SVM is a generalization of the Parzen windows estimator [9]. Furthermore, in the motivation of the νoc-SVM [20] the authors stated that their "strategy is to map the data into the feature space corresponding to the kernel and to separate them from the origin with maximum margin." The important answer of how this strategy leads to the final model description and if there is a direct connection to the existing C-SVM or ν-SVM was not shown, despite similarities in the model formulations. Reference [14] published as a side remark a more concrete geometric motivation. They argue that "the objectives of 1-class SVMs are 2-fold:" "Develop a classifier or hyperplane in the feature space which returns a positive value for all samples that fall inside the normal cluster and a negative value for all values outside this cluster," and "Maximize the perpendicular distance of this hyperplane from the origin. This is because of the inherent assumption that the origin is a member of the faulty class." However, they did not provide a proof that the νoc-SVM fulfils these objectives and indicate that the C-SVM is the basis of this model, which is wrong. This concept is the basis of this paper and will be called *"origin separation approach."* We prove that, when applying a concrete implementation of this concept to the ν-SVM, solutions can be mapped one-to-one to the νoc-SVM (Section 2.1). We also show that this is a rather generic concept and use it to derive new unary classifiers from the C-SVM and its variants.

Another well-known approach is the support vector data description [28, SVDD]. The goal of the SVDD, which is also an SVM variant, is to find a hypersphere with minimal radius, which encloses all samples of one class. It is assumed that samples outside this hypersphere do not belong to the class. This geometric view is inherently different from the origin separation approach, which creates a separating hyperplane instead. Nevertheless, we will show and visualize a relation between SVM and SVDD.

Recently, it has been shown [11] that the balanced relative margin machine (BRMM) is a joint generalization of the regularized Fisher's discriminant analysis [16, RFDA] and the C-SVM, and, furthermore,

☆ This paper has been recommended for acceptance by Egon L. van den Broek.
* Corresponding author. Tel.: +49 421 178 45 6554; fax: +49 421 178 45 4150.
E-mail address: krell@uni-bremen.de (M.M. Krell).

http://dx.doi.org/10.1016/j.patrec.2014.11.008
0167-8655/

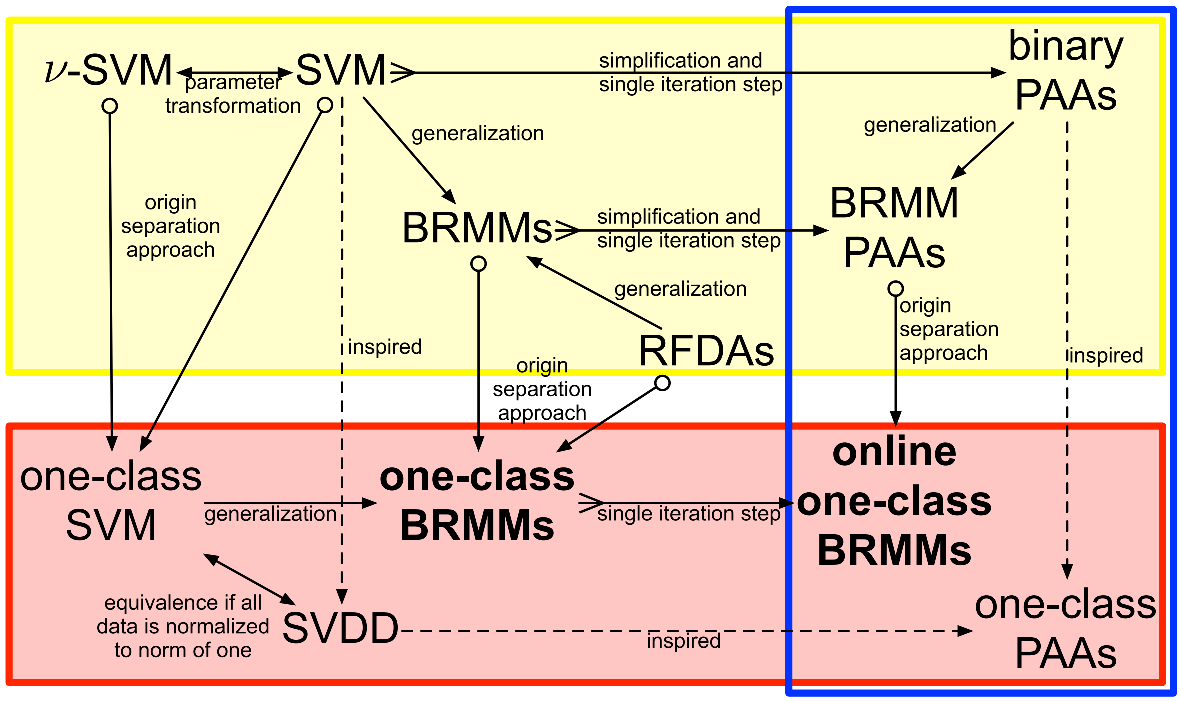

Fig. 1. Scheme of relations between binary classifiers (yellow) and their one-class (red) and online (blue) variants. The new variants introduced in this paper are in bold. The details are explained in Section 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that online passive-aggressive algorithms [6, PAAs] can be directly derived from these classifiers. We use these connections to the SVM and combine it with the origin separation approach to derive the respective new classifiers.

The connection to PAAs is of special interest. The original unary PAAs [6] were motivated from the SVDD and not from the C-SVM which was the original motivation of the two-class PAAs. Based on the connection of the PAAs to the C-SVM and thus the BRMM, we apply the origin separation approach to derive new unary classifiers from C-SVM, BRMM, and RFDA for *online learning* which can be used to apply the algorithms when resources are limited. This completes the picture on PAAs given in [6].

The paper is structured as follows. First, the origin separation approach will be applied to the SVM variants in Section 2, including different losses and kernels. Connections between these variants will be given, too. This is followed by an application on the classification of handwritten digits in Section 3. Finally, a conclusion is given in Section 4. The supplementary material [26] provides additional formulas and proofs.

2. The origin separation approach

In the origin separation approach, the origin is added as a negative training example to a unary classification problem with only positive training samples. With this modified data, classical binary classifiers are trained.[1] In this chapter, this approach will be first applied to the ν-SVM and it will be proven that the resulting classifier is equivalent to the νoc-SVM. In the next steps, *new* classifiers will be derived by applying the same approach to the C-SVM and related algorithms like the BRMMs, RFDA (Section 2.2), and PAAs (Section 2.3). A summary of existing and new model relations is depicted in Fig. 1 and will be explained in this chapter step-by-step. In Section 2.2 the effect of using different kernels and loss functions is highlighted including the special case where one-class SVMs and SVDD are equivalent.

[1] A strict (hard margin) separation at the origin is required to avoid a degeneration of the classifier.

2.1. Proving the connection between ν-SVM and νoc-SVM

Reference [20, Proposition 1] proved under the assumption of separability and hard margin separation that the νoc-SVM defines the hyperplane with maximum distance for separating the data from the origin. This concept is similar to the well known maximum margin principle in binary classification. In the following, we will generalize this proposition to arbitrary data and maximum margin separation with a so called soft margin, e.g., as specified for the ν-SVM. The model of the ν-SVM [22] is defined by the optimization problem

$$\begin{aligned} \min_{w',t',\rho',b'} \quad & \frac{1}{2}\|w'\|_2^2 - \nu\rho' + \frac{1}{l'}\sum t_i' \\ \text{s.t.} \quad & y_i(\langle w', x_i'\rangle + b') \geq \rho' - t_i' \quad \text{and} \quad t_i' \geq 0\ \forall i\,. \end{aligned} \tag{1}$$

$x_i \in \mathbb{R}^n$ are the training data with labels $y_i \in \{-1, +1\}$ and l' is the number of training samples. w' and b' define the classification function $f(x) = \operatorname{sgn}(\langle w', x\rangle + b')$. The slack variables t_i' are used to handle outliers which do not fit the model of linear separation. ν is a hyperparameter to regulate the maximum number of these samples. The original restriction $\rho' \geq 0$ is omitted for simplicity as suggested in [7].

In the origin separation approach, only the origin (zero) is taken as the negative class ($y_0 = -1$). In this case, the origin must not be an outlier ($t_0 = 0$), because it is the only sample of the negative class.[2] Consequently, the respective inequality becomes: $-(\langle w', 0\rangle + b') = \rho'$. Accordingly, b' can be automatically set to $-\rho'$. To achieve class balance, as many samples, as we have original samples of the positive class, can be added to the origin for the negative class. This step only affects the total number of samples which is doubled ($l' = 2l$), such that l only represents the number of real positive training samples and not the artificially added ones. Putting everything together ($y_i = 1$, $b' = -\rho'$, $l' = 2l$), results in

$$\begin{aligned} \min_{w',t',\rho'} \quad & \frac{1}{2}\|w'\|_2^2 - \nu\rho' + \frac{1}{2l}\sum t_i' \\ \text{s.t.} \quad & \langle w', x_i\rangle \geq 2\rho' - t_i' \quad \text{and} \quad t_i' \geq 0\ \forall i\,. \end{aligned} \tag{2}$$

[2] Also known as hard margin separation.