Knowledge-Based Systems 237 (2022) 107705

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

The soft-margin Support Vector Machine with ordered weighted average

Alfredo Marín [a,1], Luisa I. Martínez-Merino [b,c,*,1], Justo Puerto [b,1], Antonio M. Rodríguez-Chía [c,1]

[a] Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Murcia, Murcia, Spain
[b] Instituto de Matemáticas de la Universidad de Sevilla (IMUS), Sevilla, Spain
[c] Departamento de Estadística e Investigación Operativa, Facultad de Ciencias, Universidad de Cádiz, Puerto Real (Cádiz), Spain

A R T I C L E   I N F O

Article history:
Received 4 June 2021
Received in revised form 21 September 2021
Accepted 5 November 2021
Available online 19 November 2021

Keywords:
Data science
Classification
Support Vector Machine
OWA operators
Mixed integer quadratic programming

A B S T R A C T

This paper deals with a cost sensitive extension of the standard Support Vector Machine (SVM) using an ordered weighted sum of the deviations of misclassified individuals with respect to their corresponding supporting hyperplanes. In contrast with previous heuristic approaches, an exact method that applies the ordered weighted average operator in the classical SVM model is proposed. Specifically, when weights are sorted in non-decreasing order, a quadratic continuous formulation is developed. For general weights, a mixed integer quadratic formulation is proposed. In addition, our results prove that nonlinear kernel functions can be also applied to these new models extending its applicability beyond the linear case. Extensive computational results reported in the paper show that the predictive performance provided by the proposed exact solution approaches are better than the ones provided by the classical models (linear and nonlinear kernel) and similar or better than the previous ones provided by the heuristic solution by Maldonado et al. (2018).

1. Introduction

Support Vector Machine (SVM) models have become one of the most used approaches of Mathematical Programming to address classification problems. SVM techniques have been applied in many different fields since the introduction of the classical soft margin SVM by [1,2]. Among them, image recognition, bioinformatics and face detection, see [3] and references therein.

In the literature, many models based on the classical soft margin SVM approach have been developed with the aim of improving its predictive performance. For instance, different norms have been used to measure the margin between classes, see [4] where the methodology for SVM and kernel functions is extended to the general case of ℓ_p-norms with $p > 1$. Other approaches take the presence of outliers or label noise into consideration, see [5–7] among others. The identification of outliers or label noise in these models has resulted especially effective in order to improve the predictive performance. Alternatively, other models also consider feature selection which provides a better interpretation of the resulting classifier, see [8–12]. Besides, cost sensitive SVM has recently become a useful approach to deal with class-imbalance datasets (see for instance [13,14]). In these cost sensitive SVM models, the penalizations of misclassified data are different depending whether the errors correspond to individuals in one class or the other. The model studied in this paper could be considered as a different perspective of the cost sensitive SVM approach.

In [15] a new approach to the classical soft-margin Support Vector Machine is proposed. This methodology proposes to apply the OWA operator to modify the hinge loss function of classical SVM. The idea is rather appealing in that it allows to tune the importance of deviations according to their size, that is to say, classification errors are differently accounted considering a preference ranking induced by deviations of misclassified data with respect to the corresponding supporting hyperplanes.

The penalization of the classification errors unevenly according to their sizes is related to the use of OWA operators that have become very popular in different areas of decision theory. Surprisingly, although very natural, this approach had been never tried in SVM until the paper [15] proposed a two-step heuristic method to solve their model: (1) the classical SVM is trained and its classification errors induce an order based on the deviations of misclassified data with respect to the corresponding supporting hyperplanes associated with their classes; (2) the SVM is re-trained using a weighted sum of classification errors with weights

* Corresponding author at: Universidad de Cádiz, Puerto Real (Cádiz), Spain.
E-mail addresses: amarin@um.es (A. Marín), luisa.martinez@uca.es (L.I. Martínez-Merino), puerto@us.es (J. Puerto), antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

[1] All the authors evenly contribute to all the aspects of the research.

https://doi.org/10.1016/j.knosys.2021.107705

induced from the order of the solution in the first step. In [16], the same authors propose an analogous method, but using the induced order of fuzzy density-based methods for outlier detection. This approach is very simple and has the same complexity as the classical SVM beyond of being applied twice. Moreover, as the authors show in their paper, its predictive performance is superior to the traditional SVM in a set of databases that are reported in the paper.

Actually, the approach by [15], denoted from now on as app-OWA-SVM, is a heuristic approximation to the exact application of OWA operators to classical SVM. As far as we know, the use of the ordered weighted average in the classical SVM has only been tackled from this heuristic perspective in the literature. Unlike previous heuristic studies, our exact method determines the optimal hyperplane penalizing the deviation of each individual by the weight associated with its position in the ordered vector of deviations. This operator allows to use very different ways of accounting the deviations in the objective function, among them, an alternative way of limiting the influence of outliers by assigning the smallest weights to the last positions of the vector. Particularly, Example 2.1 illustrates how the proposed exact approach is different from the previous heuristic approach introduced in [15]. Another novel aspect with respect to previous works is that we prove that nonlinear kernels can be used in this exact methodology. This is an important theoretical contribution which shows that the technique of using kernels to get nonlinear classifiers extends further to models beyond the standard one.

Our contributions in this paper are the following:

(i) We develop an exact methodology for the SVM considering the OWA operator to penalize the deviations for the first time in the literature.
(ii) We prove that nonlinear kernel functions can be accommodated in the formulations that are proposed.
(iii) Our analysis distinguishes between convex OWA operators (those induced by monotone non-decreasing weights) and non-convex ones. For the first family of methods the complexity of the exact OWA-SVM is similar to the classical SVM. However, the second family of methods, namely the non-convex ones, is more complex since it involves solving mixed-integer second-order cone programs.
(iv) We provide then two models that can be solved by using general MIP solvers as CPLEX, Gurobi or Xpress.
(v) We test the performance of OWA-SVM compared with classical SVM and with app-OWA-SVM. Our results confirm those already reported by [15]: OWA-SVM is superior to SVM, k nearest neighbors, naïve Bayes and the logistic regression and it performs similar or better than app-OWA-SVM.

The remainder of this paper is structured as follows. In Section 2, some notation and details about the problem are described. Moreover, an illustrative example is detailed. Section 3 is devoted to the development of an SVM model which includes the OWA when non-decreasing weights are considered. Besides, in Section 4 we introduce a general mixed integer quadratic model which allows the use of the OWA for general weights (not necessarily non-decreasing). Section 5 contains computational experiments carried out on several datasets. Finally, Section 6 includes conclusions and some future research lines.

2. Soft margin hinge loss SVM including OWA operators

In binary classification problems, we are given a training set of individuals, $N = \{1, \dots, n\}$, divided into two classes. Each individual, i, is represented by a pair $(\boldsymbol{x}_i, y_i) \in \mathbb{R}^d \times \{-1, 1\}$, where d is the number of considered features, $\boldsymbol{x}_i$ is a vector with features' values and y_i is the label associated with the class of the individual. The goal of SVM models is to determine a hyperplane $\boldsymbol{w}^T\boldsymbol{x} + b$ that optimally separates the training set and that allows the classification of new individuals.

The classical soft margin SVM model is a compromise between maximizing the distance (margin) between the two parallel class-supporting hyperplanes and minimizing the deviations of individuals. It is formulated as follows, see [17],

$$
\begin{aligned}
(\ell_2\text{-SVM}) \min_{\boldsymbol{w},b,\boldsymbol{\xi}} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{n}\xi_i, \\
\text{s.t.} \quad & y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 - \xi_i, \quad i \in N, \\
& \boldsymbol{w} \in \mathbb{R}^d, \\
& b \in \mathbb{R}, \\
& \xi_i \geq 0, \quad i \in N.
\end{aligned}
$$

In this formulation, $\boldsymbol{w}$- and b-variables are the coefficients of the optimal separating hyperplane and $\boldsymbol{\xi}$-variables represent the deviations of each individual with respect to the supporting hyperplane associated with its class. The margin between both supporting hyperplanes is given by $\frac{2}{\|\boldsymbol{w}\|_2}$. Consequently, as mentioned before, the objective function is a balance between the maximization of the margin and the minimization of the deviations. Observe that this balance is regulated by the constant parameter C.

Nonlinear classifiers can also be obtained by using the classical SVM model. In order to determine a nonlinear separator, data of the training set N are mapped onto a higher dimension space by using a projection function $\phi(\cdot)$. By the use of duality theory and kernel functions, one can determine the optimal separator without explicitly knowing $\phi(\cdot)$. To clarify this aspect, it should be mentioned that kernel functions are those such that can be expressed as $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i) \cdot \phi(\boldsymbol{x}_j)$, where $\cdot$ denotes the scalar product. This, together with the fact that dual formulation of ℓ_2-SVM and the resulting optimal separating hyperplane only depend on the dot product of training samples, makes unnecessary the explicit use of $\phi(\cdot)$. For more details about this kernel-based method, see [18].

In the context of SVM models, OWA operators can be applied to the second term of the objective function of the classical SVM, i.e., considering the ordered weighted sum of deviations of individuals instead of the sum of them. The idea of OWA for a set of amounts is to consider the weighted sum of them but taking into account that the weights are assigned depending on the positions in the ordered sequence of these amounts. For instance, given a deviation vector $\boldsymbol{\xi}'$, the ordered weighted sum of the components of this vector is $\sum_{i=1}^{n}\lambda_i\xi'_{(i)}$, where $\boldsymbol{\xi}'_{()} = (\xi'_{(1)}, \xi'_{(2)}, \dots, \xi'_{(n)})$ is the vector $\boldsymbol{\xi}'$ with its elements sorted in non-decreasing order and $\lambda_i \geq 0$ represents the weight associated with the ith position of the ordered vector $\boldsymbol{\xi}'_{()}$. Observe that for the case $\lambda_i = 1, \forall i \in N$, we obtain the sum of these amounts. Hence, a new SVM model considering OWA operator can be expressed as follows,

$$
\begin{aligned}
\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{n}\lambda_i\xi_{(i)}, \\
\text{s.t.} \quad & y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 - \xi_i, \quad i \in N, \\
& \xi_{(i)} \leq \xi_{(i+1)}, \quad i = 1, \dots, n-1, \\
& \boldsymbol{w} \in \mathbb{R}^d, \\
& b \in \mathbb{R}, \\
& \xi_i \geq 0, \quad i \in N,
\end{aligned}
$$

where $\xi_{(i)}$ is a variable that represents the ith smallest deviation among the elements in the training set N. Note that elements of vector $\boldsymbol{\xi}_{()}$ are equal to the ones of $\boldsymbol{\xi}$ but sorted in non-decreasing order. Next sections focus on modeling the sorting of the deviations and provide valid exact formulations for this problem.

As mentioned in the introduction, a heuristic approach for SVM with OWA was introduced in [15]. The following example shows the difference between this approach and the exact method proposed in this paper.

Example 2.1. Suppose we are given the training sample detailed in Table 1 consisting of 18 individuals, 9 of them belonging to class -1 and the remaining to class 1. Particularly, the four columns of Table 1 report the label, class and data values of two features of each individual, respectively.

This sample can be represented in $\mathbb{R}^2$ as can be seen in Fig. 1. Observe that the coordinates of the individuals are given by the feature values, color blue represents the individuals of class -1 and color red represents the individuals belonging to class 1. It can be seen that individuals with labels 5 and 6 are outliers (individuals significantly different from the ones of their class) or label noise (wrongly classified individuals).

If the classical SVM, ℓ_2-SVM, is applied to this dataset, the resulting optimal separating hyperplane is the one appearing in Fig. 2. It can be seen that individuals 4, 9, 10, 14 and 15 are not correctly classified.

We would like to analyze whether the use of OWA in SVM, considering the λ-vector given in Table 2, could improve the performance of the classifier. If we apply the heuristic approach app-OWA-SVM, proposed in [15], the optimal solution $(\boldsymbol{w}^*, b^*, \boldsymbol{\xi}^*)$ of ℓ_2-SVM must be taken into consideration for the second step. Particularly, in the first step of the mentioned procedure, the ℓ_2-SVM is solved and then the order induced by $1 - y_i((\boldsymbol{w}^*)^T\boldsymbol{x}_i + b^*)$ for $i = 1, \ldots, n$ is used for the second step. Observe that $\xi_i^* = 0$ for the individuals correctly classified with respect to their supporting hyperplanes (in this instance: 18, 1, 13, 2, 16, 5 and 12) and $\xi_i^* = 1 - y_i((\boldsymbol{w}^*)^T\boldsymbol{x}_i + b^*)$ for the remaining elements. As can be seen in Fig. 2, the individuals sorted in a non-decreasing way are given in the following vector:

$$v = (18, 1, 13, 2, 16, 5, 12, 8, 6, 7, 11, 3, 17, 15, 10, 9, 14, 4).$$

This order satisfies that:

$$\xi_{18}^* \le \xi_1^* \le \xi_{13}^* \le \xi_2^* \le \xi_{16}^* \le \xi_5^* \le \xi_{12}^* \le \xi_8^* \le \xi_6^* \le \xi_7^* \le \xi_{11}^* \le \xi_3^* \le \xi_{17}^* \le \xi_{15}^* \le \xi_{10}^* \le \xi_9^* \le \xi_{14}^* \le \xi_4^*.$$

In the second step of app-OWA-SVM, the classical model is again solved but considering the weighted sum of the deviations where the weight of each deviation is given by the indices in v, i.e.,

$$\frac{1}{2}\|\boldsymbol{w}\|_2^2 + C(\lambda_1\xi_{18} + \lambda_2\xi_1 + \lambda_3\xi_{13} + \lambda_4\xi_2 + \lambda_5\xi_{16} + \lambda_6\xi_5 + \lambda_7\xi_{12} + \lambda_8\xi_8 + \lambda_9\xi_6 + \lambda_{10}\xi_7 + \lambda_{11}\xi_{11} + \lambda_{12}\xi_3 + \lambda_{13}\xi_{17} + \lambda_{14}\xi_{15} + \lambda_{15}\xi_{10} + \lambda_{16}\xi_9 + \lambda_{17}\xi_{14} + \lambda_{18}\xi_4).$$

By solving this second step, the obtained separating hyperplane is the one shown in Fig. 3. Observe that, with this heuristic approach, individuals 4, 9, 10, 14 and 15 are again wrongly classified. In this example, there is no improvement when using this two-step approach. Observe that the sorting of the second step solution does not entirely correspond to the order obtained in the first step. In fact, in the second step solution: $\xi_{11}^* \le \xi_7^* \le \xi_3^* \le \xi_6^*$ while in the first step solution $\xi_6^* \le \xi_7^* \le \xi_{11}^* \le \xi_3^*$. Consequently, the term of the deviations in the second step objective function is not an OWA operator, but a weighted sum based on the order of the first step.

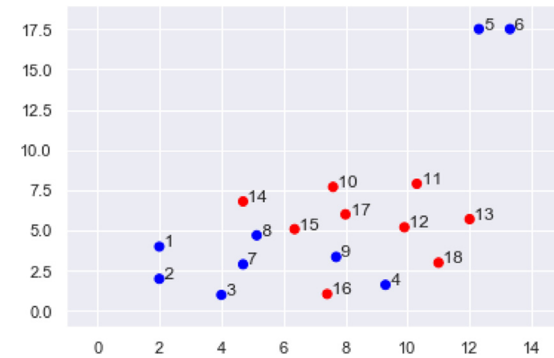

Fig. 1. Graphical representation of the training set.

Table 1
Data of the training sample.

Label	y	x_1	x_2
1	-1	2.00	4.00
2	-1	2.00	2.00
3	-1	4.00	1.00
4	-1	9.29	1.63
5	-1	12.3	17.5
6	-1	13.3	17.5
7	-1	4.70	2.90
8	-1	5.14	4.70
9	-1	7.70	3.36
10	1	7.60	7.70
11	1	10.30	7.90
12	1	9.90	5.20
13	1	12.00	5.70
14	1	4.70	6.80
15	1	6.36	5.08
16	1	7.41	1.06
17	1	8.00	6.00
18	1	11.00	3.00

Table 2
Weights used in the OWA operator.

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}
0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	1	1	1	1	1	0	0

Unlike this heuristic approach, the use of the formulations proposed in this paper gives the optimal hyperplane when the deviations are penalized by their corresponding weights depending on their positions in the deviations ordered vector. By using this exact single-step method, the optimal solution is the one presented in Fig. 4. It should be highlighted that, with this solution, only three individuals (5, 6 and 16) are not correctly classified. Note that the optimal solution provided by this exact approach is more robust against the presence of outliers (individuals 5 and 6).

The main advantage of the approach presented in this paper is that it generalizes the traditional way in which the deviations of misclassified individuals are measured. Instead of the classic sum of deviations, the use of the ordered weighted average operator allows to obtain a more flexible model. Consequently, the decisor can control how to penalize big and small deviations depending on the studied field or the specific dataset. For instance, in the above example we can observe that individuals 5 and 6 seem to be outliers in comparison with the remaining data. A small penalization of the individuals with a very big deviation is an advantage for the performance of the classifier as can be observed in Fig. 4. In Section 5, it will be observed that the use of the

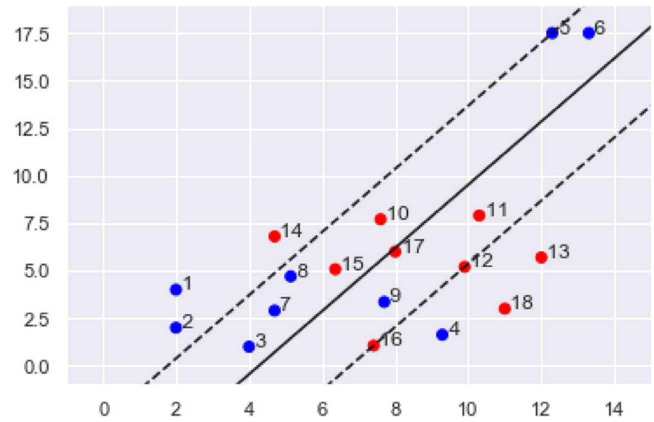

Fig. 2. Optimal hyperplane for ℓ_2-SVM.

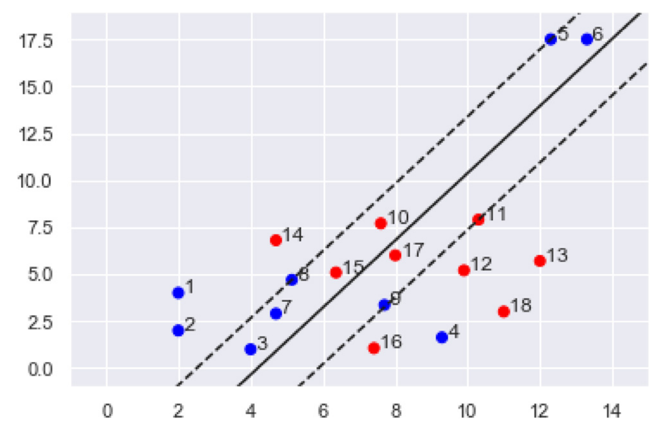

Fig. 3. Separating hyperplane obtained by the app-OWA-SVM proposed in [15].

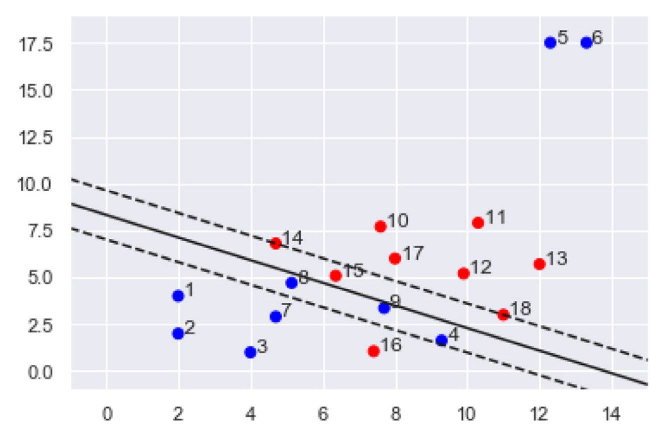

Fig. 4. Optimal hyperplane obtained by using the exact approach detailed in next sections.

OWA operator provides a better predictive performance in many datasets. Since this model is a generalization of the classical one, it also more complex. For this reason, bigger times are required to solve the models exactly.

In the next sections, we address the formulation of this problem. Recall that our objective is to provide an exact methodology for dealing with OWA operators with soft margin SVM. For this purpose, we distinguish between convex and non convex OWA operators. The reason of the aforementioned distinction is that, as we will detail in Section 3, the use of non-decreasing weights (convex case) allows to build a quadratic continuous formulation whose difficulty is similar to that of ℓ_2-SVM. In contrast, the use of non-convex OWA operators leads to the introduction of a mixed integer quadratic programming model which is computationally more complex. Section 4 deals with the use of these non-convex OWA operators. Besides, Sections 3 and 4 discuss whether nonlinear kernels can be accommodated in each model.

3. An SVM-model introducing convex OWA operators

In order to apply a correct OWA operator to the deviations of the SVM model, one has to multiply sorted deviation by the corresponding $\boldsymbol{\lambda}$-weight in the formulation. We begin analyzing the case of monotone non-decreasing $\boldsymbol{\lambda}$-weights since, as we will show, it induces simpler mathematical programming models.

With the aim of providing a formulation of this problem, together with the $\boldsymbol{w}$-, b- and $\boldsymbol{\xi}$-variables used in the classical ℓ_2-SVM, we need to include a new set of variables to model the order of the deviations of the individuals with respect to the supporting hyperplane associated with their classes. In particular, we define

$$z_{ij} = \begin{cases} 1, & \text{if deviation of observation } i \text{ is in the } j\text{th} \\ & \text{position of the sorted vector of deviations,} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

for $i, j \in N$. Given a vector of the deviation values related to each individual, $\boldsymbol{\xi}'$, the use of $\boldsymbol{z}$-variables allows us to express the ordered weighted average of these deviations with $\boldsymbol{\lambda}$-weights given in non-decreasing order as follows,

$$\sum_{i=1}^{n} \lambda_i \xi'_{(i)} = \max_{\boldsymbol{z}} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \xi'_i z_{ij},$$

$$\text{s.t.} \quad \sum_{i=1}^{n} z_{ij} = 1, \qquad j \in N, \tag{2}$$

$$\sum_{j=1}^{n} z_{ij} = 1, \qquad i \in N, \tag{3}$$

$$z_{ij} \geq 0, \qquad i, j \in N. \tag{4}$$

Constraints (2) and (3) ensure, respectively, that exactly one element of N is in each position and that each position is allocated to exactly one element of N. Besides, due to total unimodularity property, $\boldsymbol{z}$-variables can be relaxed as presented in (4). Hence, a formulation of the SVM with convex OWA operators is given by

$$\min_{\boldsymbol{w}, b, \boldsymbol{\xi}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \max_{\boldsymbol{z}} \sum_{i=1}^{n} \sum_{j=1}^{n} C \lambda_j \xi_i z_{ij},$$

$$\text{s.t.} \quad (2)\text{–}(4),$$

$$y_i(\boldsymbol{w}^T \boldsymbol{x}_i + b) \geq 1 - \xi_i, \qquad i \in N, \tag{5}$$

$$\boldsymbol{w} \in \mathbb{R}^d, \tag{6}$$

$$b \in \mathbb{R}, \tag{7}$$

$$\xi_i \geq 0, \qquad i \in N. \tag{8}$$

Like in the ℓ_2-SVM formulation, constraints (5) are the classical ones appearing in ℓ_2-SVM and the restrictions which determine the deviations of misclassified elements of N. Constraints (6)–(8) determine the domains of the corresponding variables.

Observe that this optimization model includes an inner maximization problem which intends to obtain the OWA of deviations taking advantage of the fact of using a non-decreasing weight vector. Considering the results of [19] in the context of facility location problems, we obtain the following quadratic continuous formulation dualizing the inner problem.

$$\text{(C-OWA-SVM)} \quad \min_{\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{u},\boldsymbol{v}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \sum_{i=1}^{n} u_i + \sum_{j=1}^{n} v_j,$$

$$\text{s.t.} \quad (5)\text{–}(8),$$

$$u_i + v_j \geq C\lambda_j \xi_i, \quad j \in N, \tag{9}$$

$$u_i \in \mathbb{R}, \quad i \in N, \tag{10}$$

$$v_j \in \mathbb{R}, \quad j \in N, \tag{11}$$

where $\boldsymbol{v}$- and $\boldsymbol{u}$-variables are dual variables associated with constraints (2) and (3), respectively. Note that C-OWA-SVM is a quadratic continuous model which determines a linear classifier considering ordered weighted average of individuals errors.

Remark 3.1. By considering the model proposed in [20] for minimizing the sum of k largest functions, an alternative formulation to C-OWA-SVM is

$$\text{(OT-C-OWA-SVM)} \quad \min_{\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{t},\boldsymbol{d}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \sum_{k=1}^{n} (\lambda_{n-k+1} - \lambda_{n-k}) \left(kt_k + \sum_{i=1}^{n} d_{ik} \right)$$

$$\text{s.t.} \quad (5)\text{–}(8),$$

$$d_{ik} \geq C\xi_i - t_k, \quad i,k \in N,$$

$$d_{ik} \geq 0, \quad i,k \in N,$$

$$t_k \in \mathbb{R}, \quad k \in N.$$

Some preliminary computational results show that formulation C-OWA-SVM outperforms, in terms of computational times, formulation OT-C-OWA-SVM.

As in classical SVM, it would be interesting to check whether it is possible to develop a methodology for obtaining nonlinear separators by applying the kernel trick. For this reason, once we have a primal formulation of C-OWA-SVM, we present its dual version that will be very useful to build nonlinear classifiers. The following results give a formulation of the dual problem.

Proposition 3.1. *The dual form of C-OWA-SVM is given by:*

$$\max_{\boldsymbol{\alpha},\boldsymbol{\eta}} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j,$$

$$s.t. \quad \sum_{i=1}^{n} \alpha_i y_i = 0, \tag{12}$$

$$\alpha_i \leq \sum_{j=1}^{n} \eta_{ij} C\lambda_j, \quad i \in N, \tag{13}$$

$$\sum_{i=1}^{n} \eta_{ij} = 1, \quad j \in N, \tag{14}$$

$$\sum_{j=1}^{n} \eta_{ij} = 1, \quad i \in N, \tag{15}$$

$$0 \leq \alpha_i, \quad i \in N, \tag{16}$$

$$0 \leq \eta_{ij}, \quad i,j \in N. \tag{17}$$

Proof. The Lagrangian function associated with model C-OWA-SVM is

$$L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{u}, \boldsymbol{v}) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \sum_{i=1}^{n} u_i + \sum_{j=1}^{n} v_j + \sum_{i=1}^{n} \alpha_i [1 - \xi_i - y_i(\boldsymbol{w}^T \boldsymbol{x}_i + b)] - \sum_{i=1}^{n} \mu_i \xi_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_{ij}(C\lambda_j \xi_i - u_i - v_j),$$

where $\boldsymbol{\alpha} \geq 0$, $\boldsymbol{\mu} \geq 0$ and $\boldsymbol{\eta} \geq 0$ are positive Lagrangian multipliers. The necessary and sufficient optimality conditions for C-OWA-SVM result in:

$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{u}, \boldsymbol{v})}{\partial w_j} = w_j - \sum_{i=1}^{n} \alpha_i y_i x_{ij} = 0, \ j \in N, \tag{18}$$

$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{u}, \boldsymbol{v})}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i = 0, \tag{19}$$

$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{u}, \boldsymbol{v})}{\partial \xi_i} = -\alpha_i - \mu_i + \sum_{j=1}^{n} \eta_{ij} C\lambda_j = 0, i \in N, \tag{20}$$

$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{u}, \boldsymbol{v})}{\partial u_i} = 1 - \sum_{j=1}^{n} \eta_{ij} = 0, \quad i \in N, \tag{21}$$

$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{u}, \boldsymbol{v})}{\partial v_j} = 1 - \sum_{i=1}^{n} \eta_{ij} = 0, \quad j \in N, \tag{22}$$

$$\alpha_i[1 - \xi_i - y_i(\boldsymbol{w}^T \boldsymbol{x}_i + b)] = 0, \quad i \in N, \tag{23}$$

$$\mu_i \xi_i = 0, \quad i \in N, \tag{24}$$

$$\eta_{ij}(C\lambda_j \xi_i - u_i - v_j) = 0, \quad i,j \in N, \tag{25}$$

$$\alpha_i, \mu_i, \eta_{ij} \geq 0, \quad i,j \in N. \tag{26}$$

From (18), it can be shown that $w_j = \sum_{i=1}^{n} \alpha_i y_i x_{ij}$ for $j \in N$. Besides, as in the classic SVM, condition (19) results in constraint (12). In addition, conditions (20) can be replaced by inequalities (13). Finally, constraints (14) and (15) can be deduced from conditions (21) and (22), respectively.

By using the complementary slackness conditions and replacing $w_j = \sum_{i=1}^{n} \alpha_i y_i x_{ij}$ in the Lagrangian function, we obtain

$$L(\boldsymbol{w}, b, \xi, u, v) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j. \quad \square$$

Observe that the formulation of the dual problem depends on the observed data through the scalar product of two observations. Hence, replacing in C-OWA-SVM the scalar products of training data by a kernel function $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i) \cdot \phi(\boldsymbol{x}_j)$, where ϕ is a map of the observations in a higher dimension space, the resulting formulation is

$$\text{(C-OWA-SVM}_K\text{)} \quad \max_{\boldsymbol{\alpha},\boldsymbol{\eta}} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j),$$

$$\text{s.t.} \quad (12)\text{–}(17).$$

Thus, kernel trick can be applied. Recall that this trick consists in using kernel functions in such a way that it is not necessary to explicitly know the transformation $\phi(\cdot)$ and this formulation does not depend on the dimension of feature space.

Given a sample, $\boldsymbol{x}$, belonging to an unknown class, the separating function of a nonlinear SVM is given by

$$\boldsymbol{w}_\phi^T \phi(\boldsymbol{x}) + b = \sum_{i=1}^{n} \alpha_i^* y_i \phi(\boldsymbol{x}_i) \cdot \phi(\boldsymbol{x}) + b$$

$$= \sum_{i=1}^{n} \alpha_i^* y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b. \quad (27)$$

When using C-OWA-SVM$_K$ to obtain a nonlinear separator, $\boldsymbol{\alpha}^*$ values are given by the optimal solution of C-OWA-SVM$_K$. The following result states how to determine the value of b coefficient.

Proposition 3.2. *b-coefficient of the separating function associated with C-OWA-SVM$_K$ is given by*

$$b = \frac{1 - y_k(\sum_{i=1}^{n} \alpha_i^* y_i K(\boldsymbol{x}_i, \boldsymbol{x}_k))}{y_k},$$

where $k \in N$ verifies that $0 < \alpha_k^* < C \sum_{j=1}^{n} \eta_{kj}^* \lambda_j$, and $(\boldsymbol{\alpha}^*, \boldsymbol{\eta}^*)$ are the optimal values of C-OWA-SVM$_K$.

Proof. Let $k \in N$ be such that $0 < \alpha_k^* < C \sum_{j=1}^{n} \eta_{kj}^* \lambda_j$. Then, due to conditions (20), $\mu_k^* > 0$. Since (24) holds, $\xi_k^* = 0$. Considering (23), we obtain

$$1 - y_k \left(\sum_{i=1}^{n} \alpha_i^* y_i K(\boldsymbol{x}_i, \boldsymbol{x}_k) + b \right) = 0.$$

Then,

$$b = \frac{1 - y_k(\sum_{i=1}^{n} \alpha_i^* y_i K(\boldsymbol{x}_i, \boldsymbol{x}_k))}{y_k}. \quad \square$$

The discussion above proves that kernel trick can be used in OWA-SVM provided that $\boldsymbol{\lambda}$-weights are given in non-decreasing order and that C-OWA-SVM$_K$ formulation is used. Furthermore, the nonlinear separator can be easily obtained by using (27) and Proposition 3.2.

Regarding formulation C-OWA-SVM$_K$, we find some differences with respect to the classical kernel extension of ℓ_2-SVM. Specifically, it is necessary to include some variables ($\boldsymbol{\eta}$) and constraints ((13)–(15),(17)) which do not appear in the classical dual model. Despite this, the resulting formulation C-OWA-SVM$_K$ is a convex quadratic continuous formulation that can be solved in reasonable small times comparable to the times of classical kernel-version SVM model as we will see in Section 5. Then, we have obtained an exact OWA-SVM approach for non-decreasing $\boldsymbol{\lambda}$-weights that can be efficiently solved.

4. An SVM-model introducing non convex OWA operators

The main goal of this Section is to introduce OWA in the SVM model when general $\boldsymbol{\lambda}$-weights are considered, not necessarily given in non decreasing order. In contrast with the formulation addressed in the previous section, the use of general $\boldsymbol{\lambda}$-weights forces the introduction of binary variables in the formulation in order to model the sorting of the SVM related deviations. As a consequence, the resulting model is a quadratic mixed integer formulation which is computationally more complex than C-OWA-SVM. Besides, in spite of using a MIQP to model OWA-SVM with general $\boldsymbol{\lambda}$-weights, we are able to deal with a kernel extension in a different way.

As previously mentioned, OWA operators with general $\boldsymbol{\lambda}$-weights have been applied in many combinatorial optimization problems. Particularly, in [21], OWA problems are analyzed from a modeling point of view and several formulations are compared. Based on this analysis, we present a quadratic mixed integer formulation for the SVM model that we are studying.

To this purpose, it is necessary to use the $\boldsymbol{z}$-variables described in (1) and to introduce a new family of continuous variables, for $k \in N$, defined as:

$\theta_k =$ deviation associated with the individual which is in the kth position of the sorted vector of deviations,

The resulting formulation is the following:

$$\text{(NC-OWA-SVM)} \quad \min_{\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{z},\boldsymbol{\theta}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{k=1}^{n} \lambda_k \theta_k,$$

$$\text{s.t.} \quad (2), (5)\text{–}(8),$$

$$\theta_k \geq \xi_i - M(1 - \sum_{\substack{j=1 \\ j \leq k}}^{n} z_{ij}), \qquad i, k \in N, \quad (28)$$

$$z_{ik} \in \{0, 1\}, \qquad i, k \in N, \quad (29)$$

$$\theta_k \geq 0, \qquad k \in N. \quad (30)$$

Constraints (28) ensure that deviation value in position k is at least the deviation of element i, if i is in a position smaller than or equal to k, for $i, k \in N$. Constraints (28) use a big M parameter to establish this link between θ_k- and ξ_i-variables. Note that the maximum distance between two points of the training data is a valid value of M.

Observe that, in contrast with formulation C-OWA-SVM, it is necessary to include binary variables to correctly model the order. As a consequence, completely different techniques must be applied to extend kernel trick to this formulation. In what follows, we develop a model to accommodate nonlinear kernel functions in NC-OWA-SVM, see [22].

Remark 4.1. In model NC-OWA-SVM, assume that $\boldsymbol{z}$-variables are fixed to $\hat{\boldsymbol{z}}$ (feasible assignment). Then the following formulation can be stated,

$$\text{(NC-OWA-SVM}(\hat{\boldsymbol{z}})) \quad \min_{\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\theta}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{k=1}^{n} \lambda_k \theta_k,$$

$$\text{s.t.} \quad (5)\text{–}(8), (30),$$

$$\theta_k \geq \xi_i - M\left(1 - \sum_{\substack{j=1 \\ j \leq k}}^{n} \hat{z}_{ij}\right), \quad i, k \in N. \quad (31)$$

The dual formulation of NC-OWA-SVM($\hat{\boldsymbol{z}}$) is

$$\text{(NC-OWA-SVM}_D(\hat{\boldsymbol{z}})) \quad \max_{\boldsymbol{\alpha},\boldsymbol{\mu}} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j - M\sum_{i=1}^{n}\sum_{k=1}^{n} \mu_{ik}\left(1 - \sum_{j \leq k} \hat{z}_{ij}\right),$$

$$\text{s.t.} \quad (12), (16),$$

$$\alpha_i \leq \sum_{k=1}^{n} \mu_{ik}, \qquad i \in N, \quad (32)$$

$$\sum_{i=1}^{n} \mu_{ik} \leq C\lambda_k, \qquad i, k \in N, \quad (33)$$

$$\mu_{ik} \geq 0, \qquad i, k \in N. \quad (34)$$

Besides, from necessary and sufficient optimality conditions, $w_j = \sum_{i=1}^{n} \alpha_i y_i x_{ij}$ for $j \in N$.

Based on the link between NC-OWA-SVM($\hat{\boldsymbol{z}}$) and NC-OWA-SVM$_D$($\hat{\boldsymbol{z}}$), we propose an alternative formulation of OWA-SVM for general weights. This formulation allows the use of kernel functions and, consequently, the use of nonlinear separators.

$$
\text{(NC-OWA-SVM}_D) \quad \min_{\alpha,b,\xi,\theta,z} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_iy_j\alpha_i\alpha_j\boldsymbol{x}_i\cdot\boldsymbol{x}_j + C\sum_{k=1}^{n}\lambda_k\theta_k,
$$

$$
\text{s.t.} \quad (2),(7),(8),(28)\text{–}(30),
$$

$$
y_i(\sum_{j=1}^{n} y_j\alpha_j\boldsymbol{x}_i\cdot\boldsymbol{x}_j + b) \geq 1-\xi_i,\ i\in N. \tag{35}
$$

Proposition 4.1. *Given an optimal solution of NC-OWA-SVM, $(\boldsymbol{w}^*, b^*, \boldsymbol{\xi}^*, \boldsymbol{\theta}^*, \boldsymbol{z}^*)$, it can be built a feasible solution of NC-OWA-SVM$_D$, $(\boldsymbol{\alpha}^*, b^*, \boldsymbol{\xi}^*, \boldsymbol{\theta}^*, \boldsymbol{z}^*)$, with the same objective value.*

Proof. Given a solution of NC-OWA-SVM, $(\boldsymbol{w}^*, b^*, \boldsymbol{\xi}^*, \boldsymbol{\theta}^*, \boldsymbol{z}^*)$, then $(\boldsymbol{w}^*, b^*, \boldsymbol{\xi}^*, \boldsymbol{\theta}^*)$ is an optimal solution of NC-OWA-SVM($\boldsymbol{z}^*$). From necessary and sufficient optimality conditions, $\boldsymbol{w}^* = \sum_{i=1}^n \alpha'_i y_i x_{ij}$, where $\boldsymbol{\alpha}'$ are the optimal values of $\boldsymbol{\alpha}$ variables appearing in NC-OWA-SVM$_D$($\boldsymbol{z}^*$). By defining $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}'$, $(\boldsymbol{\alpha}^*, b^*, \boldsymbol{\xi}^*, \boldsymbol{\theta}^*, \boldsymbol{z}^*)$ is feasible for NC-OWA-SVM$_D$ and

$$
\frac{1}{2}\|\boldsymbol{w}^*\|_2^2 + C\sum_{k=1}^{n}\lambda_k\theta_k^* = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_iy_j\alpha_i^*\alpha_j^*\boldsymbol{x}_i\cdot\boldsymbol{x}_j + C\sum_{k=1}^{n}\lambda_k\theta_k^*. \quad \square
$$

Proposition 4.2. *Given an optimal solution of NC-OWA-SVM$_D$, $(\boldsymbol{\alpha}^*, b^*, \boldsymbol{\xi}^*, \boldsymbol{\theta}^*, \boldsymbol{z}^*)$, a feasible solution of NC-OWA-SVM with the same objective value, $(\boldsymbol{w}^*, b^*, \boldsymbol{\xi}^*, \boldsymbol{\theta}^*, \boldsymbol{z}^*)$, can be built.*

Proof. Given an optimal solution of NC-OWA-SVM$_D$, $(\boldsymbol{\alpha}^*, b^*, \boldsymbol{\xi}^*, \boldsymbol{\theta}^*, \boldsymbol{z}^*)$, we define $w_j^* = \sum_{i=1}^n \alpha_i^* y_i x_{ij}$. This solution is feasible for NC-OWA-SVM since NC-OWA-SVM$_D$ has been built from NC-OWA-SVM, by replacing w_j by $\sum_{i=1}^n \alpha_i y_i x_{ij}$. $\square$

Propositions 4.1 and 4.2 show that formulations NC-OWA-SVM and NC-OWA-SVM$_D$ are equivalent in the sense that optimal solutions to NC-OWA-SVM can be built from optimal solutions of NC-OWA-SVM$_D$ with the same objective value, and viceversa. Note that formulation NC-OWA-SVM$_D$ allows us to accommodate nonlinear kernel functions. Replacing scalar products by a general kernel function, the resulting formulation is

$$
\text{(NC-OWA-SVM}_K) \quad \min_{\alpha,b,\xi,\theta,z} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_iy_j\alpha_i\alpha_jK(\boldsymbol{x}_i,\boldsymbol{x}_j) + C\sum_{k=1}^{n}\lambda_k\theta_k,
$$

$$
\text{s.t.} \quad (2),(7),(8),(28)\text{–}(30),
$$

$$
y_i(\sum_{j=1}^{n} y_j\alpha_jK(\boldsymbol{x}_i,\boldsymbol{x}_j) + b) \geq 1-\xi_i,\ i\in N. \tag{36}
$$

Observe that, in NC-OWA-SVM$_K$ formulation, a valid value for big M parameter should be determined and, the tighter the formulation, the better the performance. In order to obtain a good estimate for M, one can solve the following auxiliary problem:

$$
(\text{UB}_\text{M}) \quad \max_{\alpha,b,\xi,\theta} \quad \theta,
$$

$$
\text{s.t.} \quad (7),(8),(12),(36),
$$

$$
\theta \geq \xi_i, \qquad i\in N \tag{37}
$$

$$
\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_iy_j\alpha_i\alpha_jK(\boldsymbol{x}_i,\boldsymbol{x}_j) + C\lambda_n\theta \leq \text{UB}, \tag{38}
$$

$$
0\leq\alpha_i\leq C\sum_{k}\lambda_k, \qquad i\in N, \tag{39}
$$

$$
\theta\geq 0, \tag{40}
$$

where UB is an upper bound on the optimal value of NC-OWA-SVM$_K$ and θ is a variable that represents the deviation of the individual that is in the nth position of the sorted vector of deviations, the largest one. In UB$_\text{M}$ formulation, constraint (12) must be satisfied since it appears in NC-OWA-SVM$_D$($\boldsymbol{z}'$) for each feasible assignment $\boldsymbol{z}'$. In addition, constraints (36) ensure that $\boldsymbol{\alpha}$ solutions satisfy the constraints of NC-OWA-SVM$_K$ and constraints (37) establish that θ is the largest deviation. Besides, we include constraint (38) which restrict the objective value of the original problem to be smaller than or equal to a certain upper bound. Finally, the remaining constraints determine the bounds of the problem variables. Note that constraints (39) result from the combination of constraints (32) and (33) appearing in formulation NC-OWA-SVM$_D$($\hat{\boldsymbol{z}}$).

In a similar way, upper and lower bounds on b-variable could be obtained. Particularly,

$$
(\text{UB}_\text{b}) \quad \max_{\alpha,b,\xi,\theta} \quad b,
$$

$$
\text{s.t.} \quad (7),(8),(12),(36)\text{–}(40)
$$

provides an upper bound on b-variable. Analogously,

$$
(\text{LB}_\text{b}) \quad \min_{\alpha,b,\xi,\theta} \quad b,
$$

$$
\text{s.t.} \quad (7),(8),\ (12),\ (36)\ \text{–}(40)
$$

allows us to obtain a lower bound on b-variable.

We can conclude that, by using an initial upper bound on C-OWA-SVM$_K$ (UB) and auxiliary problems (UB$_\text{M}$, UB$_\text{b}$, LB$_\text{b}$), a valid big M value and bounds on the b-variable can be determined. Then, NC-OWA-SVM$_K$ can be solved more efficiently. In Algorithm 1, the method for solving NC-OWA-SVM$_K$ is outlined.

Algorithm 1: Method for solving NC-OWA-SVM$_K$.

Data: Training sample composed by a set of n individuals with d features.

Result: OWA-SVM classifier using non convex weights and a certain kernel function $K(\cdot,\cdot)$.

1. Solve the dual form of problem ℓ_2-SVM with kernel function $K(\cdot,\cdot)$ obtaining a solution $(\boldsymbol{\alpha}', b')$.
2. Consider the deviations associated with the optimal solution $(\boldsymbol{\alpha}', b')$ and sort them in non-decreasing order, obtaining a sorted vector of deviations θ'.
3. Build a feasible solution for NC-OWA-SVM$_K$:

$$
\text{UB}^* := \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_iy_j\alpha'_i\alpha'_jK(\boldsymbol{x}_i,\boldsymbol{x}_j) + C\sum_{k=1}^{n}\lambda_k\theta'_k
$$

4. Solve the problems LB$_\text{b}$ and UB$_\text{b}$ establishing UB=UB* in constraints (38). Obtain optimal objective values: l_b from LB$_\text{b}$ and u_b from UB$_\text{b}$.
5. Solve the problem UB$_\text{M}$ establishing UB=UB* in constraints (38) and adding constraint:

$$
l_b\leq b\leq u_b. \tag{41}
$$

The optimal solution of UB$_\text{M}$ is denoted as ub_M.

6. Solve NC-OWA-SVM$_K$ including constraints (41) and using $M = ub_M$. The optimal solutions of NC-OWA-SVM$_K$ are denoted by $(\boldsymbol{\alpha}^*, b^*, \boldsymbol{\xi}^*, \boldsymbol{\theta}^*, \boldsymbol{z}^*)$.

Next section will be devoted to some computational studies on the different OWA-SVM models.

Table 3
Analyzed datasets.

Label	Complete name	n	d	Class(%)
DUKE	Duke breast-cancer	44	7129	52.3/47.7
COLON	Colon cancer	62	2000	35.5/64.5
SONAR	Sonar	208	60	46.6/53.4
IONO	Ionosphere	351	34	64.1/35.9
WBC	Wisconsin Breast Cancer	569	30	62.7/37.3
AUS	Australian Credit	690	14	55.5/44.5
DIA	Pima Indians Diabetes	768	8	65.1/34.9
GC	German Credit	1000	24	70.0/30.0
SPL	Splice	1000	60	51.7/48.3
SVMG3	svmguide3	1243	21	23.8/76.2

5. Computational experiments

As mentioned in Section 1, [15], which is our benchmark, presents a heuristic approach to the use of OWA in the soft-margin SVM, that for the sake of presentation we denote by app-OWA-SVM. The app-OWA-SVM approach is a two-step method. In the first step, the classical soft margin SVM is solved and in the second step, the order induced by this optimal solution on the deviations is used to assign fixed weights to a new soft-margin SVM model.

In this section, we analyze the performance of our exact approach to OWA-SVM in comparison with app-OWA-SVM, the classical soft-margin SVM, k nearest neighbors, naïve Bayes and the logistic regression. In our computational studies, we implemented our exact OWA-SVM, app-OWA-SVM and the classical SVM. The reader is referred to results in [15], where it is shown that even the classical soft margin SVM outperforms k nearest neighbors, naïve Bayes and the logistic regression. We report results based on ten datasets described in Table 3, taken from the publicly available repository [23], which include those used in our benchmark [15]. Observe that Table 3 details the complete names of the datasets, the sample size, the number of features and the proportion of each class in the sample.

Our comparison does not only include the previously mentioned models using linear kernel, but also their corresponding Gaussian kernel versions. Recall that the Gaussian kernel function can be expressed as

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{2\sigma^2}\right),$$

where $\sigma > 0$ is known as the width parameter, see [24]. For that reason, in the reported results, we distinguish between the results of the model with and without the Gaussian kernel.

In order to define the parameters, we follow the same strategy as in [15]. Specifically, we performed a ten fold cross validation for C and σ in $\{2^{-7}, 2^{-6}, \ldots, 2^{6}, 2^{7}\}$.

Moreover, we analyze four OWA weights based on linguistic quantifiers (see [25,26]): basic, quadratic, exponential and trigonometric. For the sake of completeness, we recall the expressions of these quantifiers:

- Basic quantifier: $Q_b(r) = r^{\widetilde{\alpha}}$, $\widetilde{\alpha} \geq 0$.
- Quadratic quantifier: $Q_q(r) = \left(\frac{1}{1-\widetilde{\alpha}(r)^{0.5}}\right)$, $\widetilde{\alpha} \geq 0$.
- Exponential quantifier: $Q_e(r) = e^{-\widetilde{\alpha} r}$, $\widetilde{\alpha} \geq 0$.
- Trigonometric quantifier: $Q_t(r) = \arcsin(r\widetilde{\alpha})$, $\widetilde{\alpha} \geq 0$.

Considering these quantifiers, the associated weights can be determined by calculating

$$\lambda_i' = Q\left(1 - \frac{i-1}{n}\right) - Q\left(1 - \frac{i}{n}\right), \text{ for } i \in N.$$

Hence, the final weights are given by

$$\lambda_i = \frac{\lambda_i'}{\bar{\lambda}'},$$

where $\bar{\lambda}'$ is the average of $\boldsymbol{\lambda}'$-vector.

The specific choice of these weights is motivated since they are the ones reported in [15]. Note that the quantifiers related to the weights include a new parameter $\widetilde{\alpha}$ which is also validated in $\widetilde{\alpha} \in \{0.2, 0.4, 0.6, 0.8\}$.

To compare the results of the models, two classification performance metrics are presented: the accuracy (ACC) and the area under the curve (AUC). The accuracy is calculated as

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP are true positives, TN are true negatives, FP false positives and FN false negatives. The area under the curve is given by

$$\text{AUC} = \frac{\frac{TP}{TP+FN} + \frac{TN}{TN+FP}}{2}.$$

Regarding the solution methods used for solving the models, classical soft-margin ℓ_2-SVM and the ℓ_2-SVM model using Gaussian kernel are solved with *SVC* function of Scikit Learn module in Python, see [27]. Moreover, app-OWA-SVM and app-OWA-SVM$_K$, the models with linear and Gaussian kernel (respectively) appearing in [15] are solved by using the two-step method proposed by them.

The exact OWA-SVM models that we propose can be solved depending on the weights with different approaches. Specifically, C-OWA-SVM$_K$ is used for the weights based on basic and exponential quantifiers since they are monotone non-decreasing. Besides, the weight based on the quadratic quantifier is also monotone non-decreasing for $\widetilde{\alpha} = 0.2$. For this reason, C-OWA-SVM$_K$ is also used with these weights. Note that C-OWA-SVM$_K$ formulation is also the one used in the linear kernel case, i.e., $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i \cdot \boldsymbol{x}_j$. For the remaining weights, NC-OWA-SVM and NC-OWA-SVM$_K$ (following Algorithm 1) are applied to obtain the classifiers. It should be noted that all our computational studies were performed using CPLEX 20.1.0 in Python on an Intel(R) Xeon(R) W-2245 CPU 256 GB RAM computer. For further information, the implemented code can be found in https://github.com/LuisammGh/OWA_SVMs_and_related_models.

Before comparing the predictive performance of the models, we would like to emphasize the level of adequacy provided by app-OWA-SVM with respect to the correct final ranking of the deviations. Table 4 reports an illustrative example of how the order of deviations behave in app-OWA-SVM methods compared with the correct final ranking. Specifically, Table 4 reports, for the considered datasets, the average percentage of coincidences between the positions that occupy the individuals in the sorted deviations vector of step one and their positions in step two of the app-OWA-SVM method. In addition, in the third column of Table 4, we report the average percentage of coincidences between the final order in the app-OWA-SVM method and the order provided by C-OWA-SVM$_K$. It should be highlighted that these results are obtained when applying the ten fold cross validation to the models with $C = 1$, $\sigma = 1$, basic quantifier weight and $\widetilde{\alpha} = 0.6$.

Results of Table 4 show how the induced order of classical SVM, in step 1 in [15], is not the same as the one resulting in the second step of app-OWA-SVM. This order is neither the same as the sorting obtained by applying C-OWA-SVM$_K$. In contrast with the approaches in [15], the exact OWA-SVM models, proposed in this paper, set the order of the deviations while solving the model

Table 4
Average percentage of coincidence in the positions of the sorted deviations vector.

Data	Step 1 - Step 2 (%)	Step 2 - C-OWA-SVM$_K$ (%)
DUKE	1.02%	0.77%
COLON	1.80%	2.16%
SONAR	1.44%	3.05%
IONO	1.68%	2.06%
WBC	1.17%	0.18%
AUS	0.86%	1.71%
DIA	0.56%	1.62%
GC	0.40%	0.36%
SPL	0.03%	0.00%
SVMG3	0.87%	1.52%

itself and therefore they actually apply exact OWA operators to SVM. This shows that the method in [15] is a OWA-like approach but actually, it is not an exact application of OWA operators.

Focusing on the performance classification metrics, Table 5 reports the best results in terms of ACC provided by the different models. Note that ℓ_2-SVM and ℓ_2-SVM$_K$ show the ACC of the classic model with linear and Gaussian kernels, respectively. The column corresponding to app-OWA-SVM presents the results of the model proposed in [15] and column app-OWA-SVM$_K$ shows the results for the model in [15] using the Gaussian kernel. Finally, ex-OWA-SVM and ex-OWA-SVM$_K$ report the best results of our proposed exact OWA-SVM methods using the formulations in Sections 3 and 4 . Observe that, for all datasets, the best ACC are either the one provided by the method presented in [15] using the Gaussian kernel, or the ACC of the exact OWA-SVM model using the Gaussian kernel.

Particularly, ex-OWA-SVM$_K$ provides the best ACC results for SONAR, WBC, AUS and SVMG3 datasets; app-OWA-SVM$_K$ and ex-OWA-SVM$_K$ seem to yield the same results in the DUKE, COLON, IONO and GC cases; whereas the best results for DIA and SPL are obtained by app-OWA-SVM$_K$. In general, we can observe that the accuracy (ACC) of app-OWA-SVM$_K$ and ex-OWA-SVM$_K$ are similar. This indicates that both approaches are worthy in the sense that they improve this classification performance metric with respect to the classical SVM, which in turns outperforms k nearest neighbors, naïve Bayes and the logistic regression (see [15]), achieving almost the same value.

Table 6 reports the AUC for the best combination of parameter values in each case. As for the ACC measure, the best AUC results are always provided by app-OWA-SVM$_K$ and ex-OWA-SVM$_K$. Regarding the results, we observe that the AUC of both approaches are quite similar. For the DUKE, COLON, IONO, DIA, SPL and SVMG3 datasets, the same AUC is achieved by app-OWA-SVM$_K$ and ex-OWA-SVM$_K$; ex-OWA-SVM$_K$ provides the best results for the SONAR, WBC and AUS datasets; and finally, the app-OWA-SVM$_K$ reports the best values of AUC for GC dataset. Both approaches, app-OWA-SVM$_K$ and ex-OWA-SVM$_K$, improve the AUC of the classical ℓ_2-SVM and ℓ_2-SVM$_K$.

To conclude the analysis, we wish to include some information on the CPU times needed to solve these problems. Table 7 reports the average solving time (in seconds) per fold of the models for the parameter values that provide the best AUC. For the approach in [15], we report the time required by the two steps that are involved in the method.

Concerning the times of ex-OWA-SVM and ex-OWA-SVM$_K$, we note in passing that they correspond to formulation C-OWA-SVM$_K$ since, in all cases tested, the best results in terms of accuracy and AUC are obtained using monotone non-decreasing weights. Table 7 shows that the exact approach for OWA-SVM requires more time than the classical SVM and also more than the methods presented in [15]. This is due to the fact that models in [15] have essentially the same complexity as the classical SVM.

The aforementioned results show that the exact OWA-SVM models introduced in this paper improve the classification performance metrics of the classical SVM. Furthermore, these measures are similar to the ones reported in previous approaches to OWA-SVM models in terms of ACC and AUC, although they have the advantage of actually capturing the essence of OWA in its application to SVM.

6. Conclusions

OWA operators have been applied to different problems of decision theory. This paper proposes an exact approach that allows the introduction of OWA operators for the deviation errors appearing in the soft-margin SVM. For this aim, we have distinguished between OWA-SVM with non-decreasing $\boldsymbol{\lambda}$-weights and OWA-SVM with general $\boldsymbol{\lambda}$-weights (not necessarily non-decreasing).

The use of non-decreasing weights allowed to formulate the model using only continuous variables. As consequence, a quadratic continuous formulation was developed, C-OWA-SVM. In addition, it was shown that nonlinear kernels could be accommodated in this model by the use of the dual formulation. The required solving time of this formulation is similar to the classical one. In contrast, the use of non-monotone weights in the OWA-SVM leads us to the use of binary variables to model the order of the deviations vector. Then, a mixed integer quadratic formulation, NC-OWA-SVM, is necessary to solve this problem. Hence, NC-OWA-SVM is more complex than C-OWA-SVM. Despite this, it is also possible to apply nonlinear kernel by the use of an alternative formulation.

We have compared the proposed models with the heuristic approach to OWA-SVM appearing in [15]. Regarding the results, we can conclude that app-OWA-SVM provides solutions very different from the optimal solutions of the resulting model of applying actual OWA operators to SVM. However, both methodologies show similar predictive performance improving the ones obtained with the classical SVM (linear and nonlinear), k nearest neighbors, naïve Bayes and the logistic regression.

As future research, it could be interesting to analyze the OWA-SVM model using other ℓ_p-norms. For instance, it could be evaluated whether the use of ℓ_1-norm to measure the margin associated with the separating hyperplane together with OWA operator for deviations provides a better predictive performance. It could also be studied if nonlinear kernels could be applied when using ℓ_1-norm. Another line of research could be the development of a more complete exact model including other aspects such as feature selection or outlier detection. The development of this new model would imply the use of extra constraints and/or new solution techniques. The above mentioned subjects will be the basis of a follow up paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Alfredo Marín has been supported by Spanish Ministry of Science and Innovation under project PID2019-110886RB-I00. Part of this research was conducted while he was on sabbatical at Universidad de Sevilla, Spain. Luisa I. Martínez-Merino, Antonio M.

Table 5
Best ACC results of each model.

ACC (%)						
Data	ℓ_2-SVM	app-OWA-SVM	ex-OWA-SVM	ℓ_2-SVM$_K$	app-OWA-SVM$_K$	ex-OWA-SVM$_K$
DUKE	88.50%	88.50%	88.50%	86.00%	**90.50%**	**90.50%**
COLON	85.48%	85.48%	85.48%	88.57%	**90.48%**	**90.48%**
SONAR	80.36%	80.79%	78.38%	70.26%	90.38%	**91.33%**
IONO	90.60%	90.89%	90.89%	95.44%	**95.72%**	**95.72%**
WBC	98.07%	98.07%	97.71%	98.24%	98.42%	**98.77%**
AUS	85.51%	86.09%	85.51%	86.38%	87.25%	**87.39%**
DIA	77.60%	77.73%	77.73%	77.34%	**78.38%**	78.12%
GC	76.90%	77.50%	77.30%	77.30%	**77.50%**	**77.50%**
SPL	81.30%	82.00%	81.70%	88.40%	**89.80%**	89.40%
SVMG3	82.46%	82.62%	82.46%	83.27%	85.03%	**85.27%**

Table 6
Best AUC results for each model.

AUC (%)						
Data	ℓ_2-SVM	app-OWA-SVM	ex-OWA-SVM	ℓ_2-SVM$_K$	app-OWA-SVM$_K$	ex-OWA-SVM$_K$
DUKE	87.50%	87.50%	87.50%	85.00%	**90.00%**	**90.00%**
COLON	86.67%	86.67%	86.67%	87.50%	**90.42%**	**90.42%**
SONAR	80.36%	80.97%	78.65%	70.12%	90.27%	**91.52%**
IONO	88.67%	88.89%	88.89%	94.55%	**95.15%**	**95.15%**
WBC	97.70%	97.70%	97.14%	97.84%	98.08%	**98.44%**
AUS	86.20%	86.67%	86.20%	86.54%	87.46%	**87.71%**
DIA	72.32%	72.96%	73.00%	72.19%	**73.30%**	**73.30%**
GC	68.98%	71.00%	69.19%	68.74%	**71.67%**	69.90%
SPL	81.41%	82.02%	81.78%	88.41%	**89.86%**	**89.86%**
SVMG3	65.60%	65.94%	65.48%	70.40%	**74.85%**	**74.85%**

Table 7
Times per fold of each model for the best parameter values.

Time (s)								
Data	ℓ_2-SVM	app-OWA-SVM		ex-OWA-SVM	ℓ_2-SVM$_K$	app-OWA-SVM$_K$		ex-OWA-SVM$_K$
		Step 1	Step 2			Step 1	Step 2	
DUKE	0.006	0.008	1.179	0.364	0.003	0.002	0.043	0.056
COLON	0.002	0.002	0.472	0.258	0.002	0.001	0.077	0.114
SONAR	0.002	0.002	0.105	1.685	0.003	0.003	0.734	1.178
IONO	0.023	0.014	0.212	4.725	0.006	0.003	1.900	4.967
WBC	0.003	0.003	0.494	12.628	0.003	0.002	4.915	9.345
AUS	0.011	0.425	0.297	19.900	0.014	0.014	8.105	26.609
DIA	0.013	0.009	0.286	25.133	0.013	0.008	9.944	21.536
GC	0.033	0.016	0.560	41.788	0.024	0.025	17.090	31.707
SPL	0.022	0.522	0.815	40.561	0.057	0.038	16.387	30.866
SVMG3	0.181	0.088	1.900	71.205	0.072	0.031	24.438	61.576

Rodríguez-Chía and Justo Puerto thank the Agencia Estatal de Investigación (AEI) and the European Regional Development's funds (ERDF): project PID2020-114594GB-C21/C22; Regional Government of Andalusia: projects FEDER-UCA18-106895, FEDER-US-1256951 and P18-FR-1422; and Fundación BBVA: project NetmeetData (Ayudas Fundación BBVA a equipos de investigación científica 2019). Luisa I. Martínez-Merino has been supported by Subprograma Juan de la Cierva Formación 2019 (FJC2019-039023-I) and PAIDI 2020 postdoctoral fellowship financed by European Social Fund and Junta de Andalucía.

References

[1] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.

[2] V. Vapnik, Statistical Learning Theory, first ed., Wiley, 1998.

[3] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, A. Lopez, A comprehensive survey on support vector machine classification: Applications, challenges and trends, Neurocomputing 408 (2020) 189–215.

[4] V. Blanco, J. Puerto, A.M. Rodríguez-Chía, On ℓ_p-support vector machines and multidimensional kernels, J. Mach. Learn. Res. 21 (14) (2020) 1–29.

[5] M. Baldomero-Naranjo, L.I. Martínez-Merino, A.M. Rodríguez-Chía, Tightening big Ms in integer programming formulations for support vector machines with ramp loss, European J. Oper. Res. 286 (1) (2020) 84–100.

[6] M. Baldomero-Naranjo, L.I. Martínez-Merino, A.M. Rodríguez-Chía, A robust SVM-based approach with feature selection and outliers detection for classification problems, Expert Syst. Appl. 178 (2021) 15017.

[7] V. Blanco, A. Japón, J. Puerto, A mathematical programming approach to binary supervised classification with label noise, 2020, Preprint, arXiv:2004.10170v1.

[8] M. Gaudioso, E. Gorgone, M. Labbé, A.M. Rodríguez-Chía, Lagrangian relaxation for SVM feature selection, Comput. Oper. Res. 87 (2017) 137–145.

[9] A. Jiménez-Cordero, J.M. Morales, S. Pineda, A novel embedded min–max approach for feature selection in nonlinear support vector machine classification, European J. Oper. Res. 293 (1) (2021) 24–35.

[10] M. Labbé, L.I. Martínez-Merino, A.M. Rodríguez-Chía, Mixed integer linear programming for feature selection in support vector machine, Discrete Appl. Math. 261 (2019) 276–304.

[11] I.G. Lee, Q. Zhang, S.W. Yoon, D. Won, A mixed integer linear programming support vector machine for cost-effective feature selection, Knowl.-Based Syst. 203 (2020) 106145.

[12] S. Maldonado, J. Pérez, R. Weber, M. Labbé, Feature selection for support vector machines via mixed integer linear programming, Inform. Sci. 279 (2014) 163–175.

[13] S. Maldonado, J. López, J, Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for SVM classification, Appl. Soft Comput. 67 (2018) 94–105.

[14] B.X. Wang, N. Japkowicz, Boosting support vector machines for imbalanced data sets, J. Knowl. Inf. Syst. 4994 (2008) 38–47.

[15] S. Maldonado, J. Merigó, J. Miranda, Redefining support vector machines with the ordered weighted average, Knowl.-Based Syst. 148 (2018) 41–46.

[16] S. Maldonado, J. Merigó, J. Miranda, IOWA-SVM: A density-based weighting strategy for svm classification via owa operators, IEEE Trans. Fuzzy Syst. 28 (9) (2020) 2143–2150.
[17] P.S. Bradley, O.L. Mangasarian, Feature selection via concave minimization and support vector machines, ICML 98 (1998) 82–90.
[18] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov. 2 (2) (1998) 121–167.
[19] V. Blanco, J. Puerto, S. El Haj Ben Ali, Revisiting several problems and algorithms in continuous location with ℓ_τ norms, Comput. Optim. Appl. 58 (3) (2014) 563–595.
[20] W. Ogryczak, A. Tamir, Minimizing the sum of the k largest functions in linear time, Inform. Process. Lett. 85 (3) (2003) 117–122.
[21] E. Fernández, M.A. Pozo, J. Puerto, Ordered weighted average combinatorial optimization: Formulations and their properties, Discrete Appl. Math. 169 (2014) 97–118.
[22] J.P. Brooks, Support vector machines with the ramp loss and the hard margin loss, Oper. Res. 59 (2) (2011) 467–479.
[23] C.C. Chang, C.J. Lin, LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (3) (2011) 1–27, URL http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[24] B. Schölkopf, A. Smola, Learning with Kernels, MIT Press, 2002.
[25] P. Luukka, O. Kurama, Similarity classifier with ordered weighted averaging operators, Expert Syst. Appl. 40 (4) (2013) 995–1002.
[26] R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decisionmaking, IEEE transactions on systems, Man Cybern. 18 (1) (1988).
[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.