Deriving Weights in Multiple-Criteria Decision Making with Support Vector Machines

Emilio Carrizosa

Facultad de Matemáticas, Universidad de Sevilla
Avda Reina Mercedes s/n, 41012 Sevilla (Spain)
E-mail: ecarrizosa@us.es

Abstract

A key problem in Multiple-Criteria Decision Making is how to measure the importance of the different criteria when just a partial preference relation among actions is given. In this note we address the problem of constructing a linear score function (and thus how to associate weights of importance to the criteria) when a binary relation comparing actions and partial information (relative importance) on the criteria are given. It is shown that these tasks can be done via *Support Vector Machines*, an increasingly popular Data Mining technique, which reduces the search of the weights to the resolution of (a series of) nonlinear convex optimization problems with linear constraints. An interactive method is then presented and illustrated by solving a multiple-objective 0-1 knapsack problem. Extensions to the case in which data are imprecise (given by intervals) or intransitivities in strict preferences exist are outlined.

Key Words: Linear score functions, support vector machines, multiple-criteria decision making with partial information, data mining.

AMS subject classification: 90B50, 90C30, 62H30.

1 Introduction

Suppose we are given

- an integer N,

- a finite directed graph $(\mathcal{A}, P)$, whose set of edges P is non-empty and contains no loops,

- a function $\Psi : \mathcal{A} \longrightarrow \mathbb{R}^N$,

- a polyhedral cone Ω in $\mathbb{R}_+^N$, $\Omega \neq \{0\}$, represented in the form

$$\Omega = \left\{ \omega \in \mathbb{R}^N : q_j^\top \omega \geq 0,\, j \in J \right\},$$

  for a finite set $\{q_j : j \in J\} \subset \mathbb{R}^N$,

- a norm γ in $\mathbb{R}^N$,

and consider the optimization problem

$$\max_{\omega \in \Omega \setminus \{0\}} \min_{(a,a') \in P} \frac{\omega^\top \Psi(a) - \omega^\top \Psi(a')}{\gamma(\omega)}. \qquad (1.1)$$

Problem (1.1) has a clear interpretation in Multiple-Criteria Decision Making, as sketched below:

$\mathcal{A}$ represents a finite set of actions, where we have defined

- a vector-valued function $\Psi : \mathcal{A} \longrightarrow \mathbb{R}^N$, in such a way that $\Psi_j(a)$ represents the score of action a according to the j-th criterion, $j = 1, 2, \ldots, N$. All scores are assumed to be in a scale "the higher the better"

- a binary irreflexive relation $P \subset \mathcal{A} \times \mathcal{A}$. No further assumptions (e.g. transitivity or weak connectedness) are made on P. For instance, P might be the strict preferences obtained with Electre I, Roy (1968), or the strict preferences detected via a sample of pairwise comparisons among actions.

  By abuse of notation, we will write in what follows indifferently aPa' or $(a, a') \in P$.

Our aim is to extend the binary relation P to a total preorder on $\mathcal{A}$ via a linear function of the scores given by Ψ. In other words, we seek $\omega \in \Omega$ such that
$$\omega^\top \Psi(a) > \omega^\top \Psi(a') \qquad \forall a, a' \in \mathcal{A},\, aPa'. \qquad (1.2)$$

Any ω satisfying (1.2) induces a pair of binary relations (P_ω, I_ω) on $\mathcal{A}$,

$$\begin{array}{llll} aP_\omega a' & \text{iff} & \omega^\top \Psi(a) > \omega^\top \Psi(a') \\ aI_\omega a' & \text{iff} & \omega^\top \Psi(a) = \omega^\top \Psi(a'), \end{array} \qquad (1.3)$$

of strict preference and indifference.

We have that the strict preference P_ω is compatible with P, in the sense that the set P is included in the set P_ω, and thus no information contained in P is lost if P is replaced by P_ω. Moreover, P_ω enjoys properties usually considered to be desirable such as transitivity. By construction, each component of ω measures the importance of the corresponding criterion in P_ω, and hence we can also see ω_j as a measure of importance of Ψ_j in P.

We illustrate the model in the following Example.

Example 1.1. Consider the problem with 6 actions, $(\mathcal{A} = \{a_1, \ldots, a_6\})$, $N = 4$ criteria, $C_1, \ldots, C_4$, with rewards scored (in ordinal or cardinal scales) in Table 1.

	C_1 (max)	C_2 (max)	C_3 (max)	C_4 (max)
a_1	High	4	2	8
a_2	Low	2	2	7
a_3	High	3	4	3
a_4	Low	2	5	2
a_5	Average	8	1	1
a_6	Low	4	4	4

Table 1: *Decision table for Example 1.1*

Consider also the irreflexive binary relation P given on $\mathcal{A}$ represented in Figure 1.

The mechanism used to construct such P is simple: $a_i P a_j$ iff criterion C_1 in a_i is strictly better than C_1 in a_j, and, at the same time, the average of the three remaining scores in a_i is strictly greater than the average in a_j.

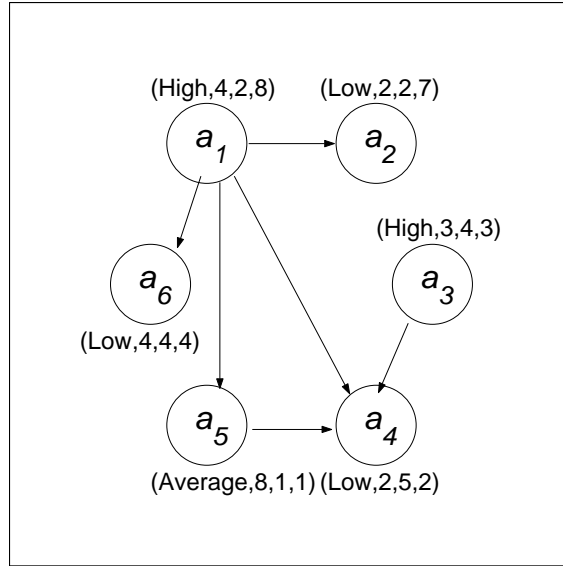In order to accommodate these data to the model considered, the cri-

Figure 1: *The graph $(\mathcal{A}, P)$ in Example 1.1*

teria must be measured in a cardinal scale. To do that, we define

$$\Psi_{11}(a) = \begin{cases} 1, & \text{if } a \text{ in } C_1 \text{ takes the value "High"} \\ 0, & \text{else} \end{cases}$$

$$\Psi_{12}(a) = \begin{cases} 1, & \text{if } a \text{ in } C_1 \text{ takes the value "Average"} \\ 0, & \text{else} \end{cases}$$

$$\Psi_{13}(a) = \begin{cases} 1, & \text{if } a \text{ in } C_1 \text{ takes the value "Low"} \\ 0, & \text{else} \end{cases}$$

$$\Psi_i(a) = \text{score of } a \text{ according to } C_i \text{ in Table 1}, \qquad i = 2, \ldots, 4,$$

and we obtain Table 2.

Under these assumptions, we seek a vector of weights ω, to be associated with the columns of Table 2, in such a way that ordinal values are replaced by ratio-scaled values, and the total ranking obtained by averaging the actions ω is compatible with P retaining all the information on the criteria: since the criteria are of the form "the higher the better," the weights should be non-negative; moreover, the weight associated with Ψ_{11} should not be smaller than the weight associated with Ψ_{12}, which should not be smaller than the weight associated with Ψ_{13}.

	Ψ_{11}	Ψ_{12}	Ψ_{13}	Ψ_2	Ψ_3	Ψ_4
a_1	1	0	0	4	2	8
a_2	0	0	1	2	2	7
a_3	1	0	0	3	4	3
a_4	0	0	1	2	5	2
a_5	0	1	0	8	1	1
a_6	0	0	1	4	4	4

Table 2: *Decision Table for Example 1.1. Cardinal scales*

Let Ω_P denote the set of solutions to (1.2),

$$\Omega_P = \left\{ \omega \in \Omega : \ \omega^\top \Psi(a) > \omega^\top \Psi(a') \qquad \forall (a, a') \in P \right\}.$$

In practice, of course, Ω_P can be empty. This is the case, for instance, when the graph $(\mathcal{A}, P)$ contains cycles. In such cases, no P_ω compatible with P exists. In what follows we assume that $\Omega_P \neq \emptyset$. How to address the case $\Omega_P = \emptyset$ will be outlined in Section 4.

In case $\omega \in \Omega_P$ exists, it is highly desirable that such ω makes maximal separation between actions, i.e., the slack $\omega^\top \Psi(a) - \omega^\top \Psi(a')$ should be not only positive but high in all pairs $a, a' \in \mathcal{A}, aPa'$. To do that, we can take as criterion the maximization of the lowest slack,

$$\min_{(a, a') \in P} \left\{ \omega^\top \Psi(a) - \omega^\top \Psi(a') \right\}. \tag{1.4}$$

Maximizing over Ω (or over Ω_P) the lowest slack is, as soon as $\Omega_P \neq \emptyset$, an optimization problem with unbounded solution: given $\omega \in \Omega_P$, $\vartheta\omega \in \Omega_P$ for any $\vartheta > 0$, and makes the smallest slack grow without limit by growing ϑ.

Hence, we have to introduce a normalization condition which enables us to identify ω and $\vartheta\omega$, $\vartheta > 0$. This can be done e.g., by setting $\gamma(\omega) = 1$ for a given norm, or, equivalently, as done in (1.1).

The inclusion of linear constraints on ω enables us to consider problems with *partial information* on the importance to be given to the different scalar functions Ψ_j, e.g., Carrizosa et al. (1995), and Carrizosa and Conde (2002). For instance, since, by assumption, Ψ_j scores a criterion of type

"the higher, the better", the ω sought should satisfy, together with (1.2), a sign constraint

$$\omega_j \geq 0. \tag{1.5}$$

Moreover, we may have information about the relative importance of criteria. For instance, from two criteria Ψ_i, Ψ_j of type "the higher, the better", we may impose that the weight associated with criterion i should not exceed K_{ij} times the weight associated with the j-th criterion. This way we generate the homogeneous linear constraint

$$\omega_i \leq K_{ij}\omega_j, \tag{1.6}$$

e.g., Carrizosa and Conde (2002), and Carrizosa et al. (1995).

More complex homogeneous linear constraints may appear, for instance, in decision problems under risk, with N scenarios, each ω_j representing the probability of the j-th scenario. Under such assumptions, since the weights represent probabilities, their sum must equal 1. This is modelled by taking as γ in (1.1) the ℓ_1 norm. Moreover, if we have interval information about the probability of the i-th scenario, in the form

$$L_i \leq \omega_i \leq U_i,$$

we can transform this condition into

$$\frac{L_i}{1 - L_i} \leq \frac{\omega_i}{1 - \omega_i} \leq \frac{U_i}{1 - U_i},$$

or into the pair of homogeneous linear constraints

$$\begin{array}{l} \frac{L_i}{1-L_i} \sum_{j \neq i} \omega_j \leq \omega_i \\ \frac{U_i}{1-U_i} \sum_{j \neq i} \omega_j \geq \omega_i. \end{array} \tag{1.7}$$

On the other hand, if two actions, a, a' are known to be indifferent, one would have

$$\begin{array}{rcl} (\Psi(a) - \Psi(a'))^\top \omega & \geq & 0 \\ (\Psi(a') - \Psi(a))^\top \omega & \geq & 0 \end{array} \tag{1.8}$$

As seen above, sign constraints and, among others, those of type (1.6), (1.7) and (1.8), can be modelled by imposing that ω belongs to a polyhedral cone Ω.

Example 1.2. With the data of Example 1.1, since the criteria in Table 1 are all of type "the higher the better", all weights must be non-negative. Since we have passed C_1 to a cardinal scale, see Table 2, we must impose also that $\omega_{11} \geq \omega_{12} \geq \omega_{13}$.

This information on the weights yields the polyhedral cone

$$
\begin{aligned}
\Omega \quad = \quad & \big\{\omega = (\omega_1, \omega_2, \dots, \omega_6) \in \mathbb{R}_+^6 : \\
& (1, -1, 0, 0, 0, 0)^\top \omega && \geq \quad 0 \\
& (0, 1, -1, 0, 0, 0)^\top \omega && \geq \quad 0\big\}
\end{aligned}
\tag{1.9}
$$

Solving numerically (1.1) taking as γ the Euclidean norm, the following weights (measures of importance of the criteria) are obtained after normalization to sum 1:

$$
\omega = (\omega_{11}, \omega_{12}, \omega_{13}, \omega_2, \omega_3, \omega_4) = (0.333, 0.000, 0.000, 0.333, 0.000, 0.333).
\tag{1.10}
$$

Hence, only the highest value of C_1, together with C_2 and C_4 are relevant, all being equally important.

With this information, the function $a \longmapsto \omega^\top \Psi(a)$ is given in Table 3, yielding the total strict order

$$
a_1 P_\omega a_2 P_\omega a_5 P_\omega a_6 P_\omega a_3 P_\omega a_4.
$$

a_i	$\omega^\top \Psi(a)$
a_1	4.334
a_2	3.001
a_3	2.333
a_4	1.333
a_5	3.000
a_6	2.667

Table 3: *Values of $\omega^\top \Psi(\cdot)$ in Example 1.2*

Suppose we also know that C_4 should not be weighted much stronger (say, 5 times) than C_3, modelled with the constraint
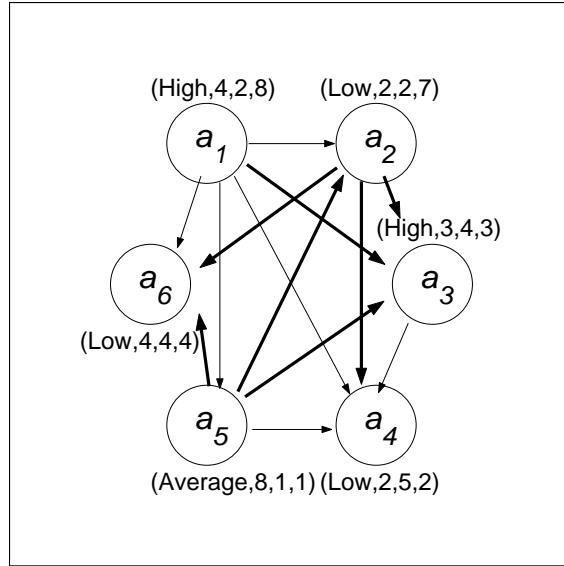
$$
\omega_4 \leq 5\omega_3.
\tag{1.11}
$$

Figure 2: *The graph $(\mathcal{A}, P_\omega)$ in Example 1.1*

This constraint must then be added to those given in (1.9).

Solving numerically (1.1) taking as γ the Euclidean norm, the following weights (measures of importance of the criteria) are obtained after normalization to sum 1:

$$\omega = (\omega_{11}, \omega_{12}, \omega_{13}, \omega_2, \omega_3, \omega_4) = (0.342, 0.000, 0.000, 0.340, 0.053, 0.265),$$

yielding the total strict order

$$a_1 P_\omega a_5 P_\omega a_2 P_\omega a_6 P_\omega a_3 P_\omega a_4.$$

Hence, the addition of the constraint (1.11) has produced slight changes in the optimal weight vector ω, but has led to a change in the ranking of the alternatives (please compare the position of a_3 and a_5 in both rankings.

The graph of P_ω is depicted in Figure 2. Preferences already present in P are plotted with a thick arrow, whereas those which did not exist in P and are added by P_ω are dotted.

In order to gain insight into Problem (1.1), and to extend the methodology to the case in which Ω_P is empty, we connect this with (a variant of)

Support Vector Machines. Other approaches to ordinal regression based on Support Vector Machines can be found e.g., in Freund et al. (2001), and Herbrich et al. (2000).

The remainder of the paper is organized as follows. In Section 2 we make a quick introduction to Support Vector Machines, and rephrase our problem within this framework. An interactive methodology is outlined in Section 3, where an illustrative example is also given. Possible extensions of this work are given in Section 4.

Throughout this paper, $conv(X)$ denotes the convex hull of a set X, and $cone(X)$ denotes the conic hull of X, i.e., $cone(X)$ is the set of points which can be expressed as a linear combination with non-negative coefficients of elements of X, Cramer and Singer (2001).

2 Support Vector Machines

Before particularizing to our problem, (see Section 2.2), we first briefly outline the main ideas underlying the Discriminant Analysis method known as Support Vector Machines, SVM. The reader is referred to e.g., Burges (1998), Cortes and Vapnik (1995), Cristianini and Shawe-Taylor (2000), Hastie et al. (2001), Suykens et al. (2002), Vapnik (1998), Vapnik (2000), and the references therein for further details.

2.1 General results

Let Ω be a polyhedral cone in $\mathbb{R}^N$. Let I be a finite non-empty set of individuals, partitioned as $I = I_+ \cup I_-$, with $I_+, I_- \neq \emptyset$. Each $i \in I$ has an associated vector $x_i \in \mathbb{R}^N$. One seeks an affine function *separating* I_+ and I_-, by strictly separating the sets $\{x_i : i \in I_+\}$ and $\{x_i : i \in I_-\}$: $(\omega, \beta) \in \Omega \times \mathbb{R}$ is sought such that

$$\begin{array}{lll} \omega^\top x_i + \beta & > & 0 \quad \forall i \in I_+ \\ \omega^\top x_i + \beta & < & 0 \quad \forall i \in I_-. \end{array} \tag{2.1}$$

If such (ω, β) exists, since, by assumption, $I_+, I_- \neq \emptyset$, one has that $\omega \neq 0$, and thus it defines a hyperplane $H(\omega, \beta) = \{x \in \mathbb{R}^N : \omega^\top x + \beta = 0\}$, which

bounds the closed halfspaces

$$
\begin{aligned}
H^-(\omega, \beta) &= \{x \in \mathbb{R}^N : \omega^\top x + \beta \le 0\} \\
H^+(\omega, \beta) &= \{x \in \mathbb{R}^N : \omega^\top x + \beta \ge 0\}.
\end{aligned}
$$

Define, for each $i \in I$, the label y_i as

$$
y_i = \begin{cases} 1, & \text{if } i \in I_+ \\ -1, & \text{if } i \in I_-. \end{cases}
$$

Condition (2.1) can then be rewritten as

$$
y_i \left(\omega^\top x_i + \beta \right) > 0 \quad \forall i \in I. \tag{2.2}
$$

The vector (ω, β) chosen will be used to classify, following (2.1), future entries $x \in \mathbb{R}^N$: if $\omega^\top x + \beta > 0$, x will be allocated to group I_+, whereas, if $\omega^\top x + \beta < 0$, then x will be allocated to I_-.

A key concept is (Ω-)separability.

Definition 2.1. Given a polyhedral cone Ω, the pair $(\{x_i : i \in I_+\}, \{x_i : i \in I_-\})$ is said to be Ω-*separable* if there exists $(\omega, \beta) \in \Omega \times \mathbb{R}$ satisfying (2.2). Such (ω, β) will be said to Ω-*separate* $(\{x_i : i \in I_+\}, \{x_i : i \in I_-\})$.

For $\Omega = \mathbb{R}^N$, separability tests, and, in the separable case, procedures for generating a separating hyperplane by means of Linear Programming date back to the 60s, e.g., Mangasarian (1965). The extension of these procedures to arbitrary polyhedral cones Ω is straightforward. One has in particular

Proposition 2.1. *Let Ω be a polyhedral cone in $\mathbb{R}^N$ of the form*

$$
\Omega = \{\omega \in \mathbb{R}^N : q_j^\top \omega \ge 0 \, \forall j \in J\}
$$

for a finite set J. The following statements are equivalent:

1. *$(\{x_i : i \in I_+\}, \{x_i : i \in I_-\})$ is Ω-separable.*

2. *For any norm $\|\cdot\|$ in $\mathbb{R}^N$, the linearly-constrained convex minimization problem*

$$
\begin{aligned}
\min \quad & \|\omega\| \\
\text{s.t.} \quad & y_i \left(\omega^\top x_i + \beta \right) \ge 1 \qquad \forall i \in I \\
& \omega \in \Omega
\end{aligned} \tag{2.3}
$$

   is feasible.

3. *The sets* $conv(\{x_i : i \in I_+\}) + cone(\{q_j : j \in J\})$ *and* $conv\big(\{x_i : i \in I_-\}\big)$ *are disjoint.*

Proof. Since Ω is a cone and I is finite, any (ω, β) satisfying (2.2) can be re-scaled so that the left hand sides in (2.2) are not only strictly positive, but also greater than or equal to 1. In other words, the pair $(\{x_i : i \in I_+\}, \{x_i : i \in I_-\})$ is Ω-separable iff there exist $\omega \in \Omega$, $\beta \in \mathbb{R}$ such that

$$y_i \left(\omega^\top x_i + \beta\right) \geq 1 \quad \forall i \in I,$$

which is equivalent to the feasibility of (2.3). Hence, separability of the pair $(\{x_i : i \in I_+\}, \{x_i : i \in I_-\})$ is equivalent to the feasibility of the linear problem

$$
\begin{aligned}
\min \quad & 0 \\
\text{s.t.} \quad & y_i \left(\omega^\top x_i + \beta\right) \geq 1 \quad \forall i \in I \\
& \omega^\top q_j \geq 0 \qquad\qquad \forall j \in J,
\end{aligned}
\tag{2.4}
$$

whose optimal value is, when feasible, zero. The dual of (2.4) is

$$
\begin{aligned}
\max \quad & \sum_{i \in I} \lambda_i \\
\text{s.t.} \quad & \sum_{i \in I} \lambda_i y_i x_i + \sum_{j \in J} \mu_j q_j = 0 \\
& \sum_{i \in I} \lambda_i y_i = 0 \\
& \lambda, \mu \geq 0.
\end{aligned}
\tag{2.5}
$$

Problem (2.5) is always feasible (take $\lambda = 0 \in \mathbb{R}^I$ and $\mu = 0 \in \mathbb{R}^J$). Hence, the pair $(\{x_i : i \in I_+\}, \{x_i : i \in I_-\})$ is Ω-separable iff (2.5) has 0 as optimal value, which happens iff the polyhedral cone $conv(\{x_i : i \in I_+\}) + cone(\{q_j : j \in J\})$ and the polytope $conv(\{x_i : i \in I_-\})$ are disjoint sets. Indeed, if there exists $u \in (conv(\{x_i : i \in I_+\}) + cone(\{q_j : j \in J\})) \cap conv(\{x_i : i \in I_-\})$, then there exist $(\lambda_i)_{i \in I_+} \in \mathbb{R}_+^{I_+}$, $(\lambda_i)_{i \in I_-} \in \mathbb{R}_+^{I_-}$, and $\mu \in R^J$ such that

$$
\begin{aligned}
u \;&=\; \sum_{i \in I_+} \lambda_i x_i + \sum_{j \in J} \mu_j q_j \\
&=\; \sum_{i \in I_-} \lambda_i x_i \\
\sum_{i \in I_+} \lambda_i \;&=\; \sum_{i \in I_-} \lambda_i = 1,
\end{aligned}
$$

i.e.,

$$\sum_{i\in I}\lambda_i y_i x_i + \sum_{j\in J}\mu_j q_j = 0$$
$$\sum_{i\in I}\lambda_i y_i = 0$$
$$\sum_{i\in I}\lambda_i = 1,$$

thus the optimal value of (2.5) cannot be zero.

Conversely, if the optimal value of (2.5) is not zero, then, taking $\lambda \in R_+^I$, $\mu \in \mathbb{R}_+^J$, feasible for (2.5), with $\sum_{i\in I}\lambda_i > 0$, define $\widehat{\lambda}, \widehat{\mu}$ as

$$\widehat{\lambda} = \frac{2}{\sum_{i\in I}\lambda_i}\lambda$$
$$\widehat{\mu} = \frac{2}{\sum_{i\in I}\lambda_i}\mu.$$

Then, one has

$$\sum_{i\in I_+}\widehat{\lambda}_i = 2\frac{\sum_{i\in I_+}\lambda_i}{\sum_{i\in I}\lambda_i} = 1$$
$$\sum_{i\in I_-}\widehat{\lambda}_i = 1.$$

Moreover,

$$\sum_{i\in I_+}\lambda_i x_i + \sum_{j\in J}\mu_j q_j = \sum_{i\in I_-}\lambda_i x_i,$$

thus

$$u = \sum_{i\in I_+}\widehat{\lambda}_i x_i + \sum_{j\in J}\widehat{\mu}_j q_j = \sum_{i\in I_-}\widehat{\lambda}_i x_i,$$

implying $u \in (conv(\{x_i : i \in I_+\}) + cone(\{q_j : j \in J\})) \cap conv(\{x_i : i \in I_-\})$. $\square$

Following (2.1), for $(\omega, \beta) \in (\Omega \setminus \{0\}) \times \mathbb{R}$ given, the half-space misclassifying those $i \in I_+$ (respectively $i \in I_-$) is $H(\omega, \beta)^- = \{x \in \mathbb{R}^N : \omega^\top x + \beta \leq 0\}$ (respectively $H(\omega, \beta)^+ = \{x \in \mathbb{R}^N : \omega^\top x + \beta \geq 0\}$).

Let $\|\cdot\|$ be a norm in $\mathbb{R}^N$. Denote by $f_i(\omega, \beta)$ the distance (according to the metric induced by $\|\cdot\|$) from x_i to its half-space of misclassification (either $H(\omega, \beta)^-$ or $H(\omega, \beta)^+$).

For $(\omega, \beta) \in \mathbb{R}^N \times \mathbb{R}$, $\omega \neq 0$, the *margin* is defined as the minimum of the distances of the points to their corresponding half-space of misclassification. Hence, the problem of finding the Ω-separating hyperplane maximizing the margin can be written as the following optimization problem with non-closed feasible region,

$$
\begin{array}{ll}
\max & (\min_{i \in I} f_i(\omega, \beta)) \\
\text{s.t.} & \omega \in \Omega \setminus \{0\} \\
& y_i\left(\omega^\top x_i + \beta\right) > 0 \qquad \forall i \in I
\end{array}
\tag{2.6}
$$

Problem (2.6) has been extensively studied in the particular case $\Omega = \mathbb{R}^N$ and the Euclidean norm as $\|\cdot\|$, e.g., Burges (1998), Cortes and Vapnik (1995), Cristianini and Shawe-Taylor (2000), Smola and Schölkopf (2002), Vapnik (1998), under the name of Support Vector Machines. The consideration of an arbitrary norm, as done here, is relatively recent, Mangasarian (2000), although no constraints were included there.

From the formula of the distance from a point to a hyperplane for an arbitrary norm $\|\cdot\|$, Carrizosa and Fliege (2002), Plastria and Carrizosa (2001), one has for any $(\omega, \beta) \in \mathbb{R}^N \times \mathbb{R}$, $\omega \neq 0$, that

$$
f_i(\omega, \beta) = \max\{\frac{y_i(\omega^\top x_i + \beta)}{\|\omega\|^\circ}, 0\},
\tag{2.7}
$$

where $\|\cdot\|^\circ$ denotes the norm dual to $\|\cdot\|$,

$$
\|\omega\|^\circ = \max_{\|u\|=1} u^\top \omega.
$$

When the pair $(\{x_i : i \in I_+\}, \{x_i : i \in I_-\})$ is Ω-separable, the constraints in the form

$$
y_i(\omega^\top x_i + \beta) > 0 \quad \forall i \in I
\tag{2.8}
$$

are redundant: any $(\omega, \beta) \in (\mathbb{R}^N \setminus \{0\}) \times \mathbb{R}$ not satisfying (2.8) will have 0 as objective value, which will never be the optimum in the Ω-separable case. We can thus drop under Ω-separability assumptions the constraints (2.8), yielding the equivalent optimization problem

$$
\max_{\omega \in \Omega \setminus \{0\}} \quad \min_{i \in I} \quad \frac{y_i(\omega^\top x_i + \beta)}{\|\omega\|^\circ}
\tag{2.9}
$$

Moreover, since the objective function in (2.9) is homogeneous, and, by assumption, Ω is a cone, we obtain the equivalent formulation

$$
\begin{aligned}
\min \quad & \|\omega\|^{\circ} \\
\text{s.t.} \quad & y_i(\omega^{\top} x_i + \beta) \geq 1 \; \forall i \in I \\
& \omega \in \Omega.
\end{aligned}
\tag{2.10}
$$

Observe that the case $\omega = 0$ is naturally excluded by this formulation.

This will be our canonical formulation for the SVM.

Example 2.1. As an illustration, consider the set of points (triangles or discs) in Figure 3. If distances are measured by the Euclidean norm, the optimal hyperplane is depicted in Figure 4. However, if distances are measured by the ℓ_{∞}, and thus $\| \cdot \|^{\circ} = \ell_1$, the optimal solution is depicted in Figure 5.
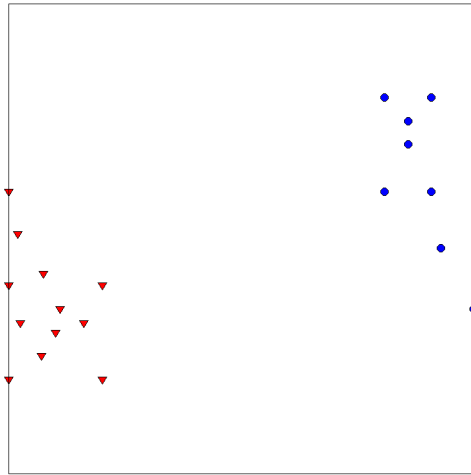

Figure 3: *The data points of Example 2.1*

In general, (2.10) is an optimization problem with convex objective and linear constraints, solvable by general-purpose methods such as those described in Hiriart-Urruty and Lemaréchal (1996), and Carrizosa and Fliege (2002).
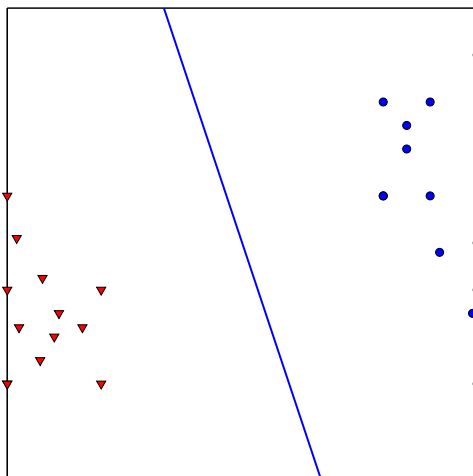
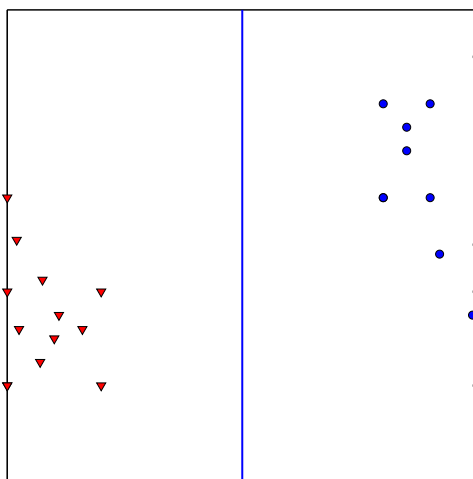**Figure 4**: *Large margin solution* $(\|\cdot\| : \ell_2)$

Figure 5: *Large margin solution* $(\|\cdot\| : \ell_\infty)$

Particular choices, however, lead to more structured problems for which more efficient algorithms can be used. For instance, if $\|\cdot\|$ is the Euclidean distance, (2.10) leads to a linearly-constrained convex quadratic problem, whereas if $\|\cdot\|$ is a polyhedral norm (i.e., a norm whose unit ball is a polyhedron), then (2.10) can easily be transformed into an equivalent linear program.

2.2   Linear scores and SVM

Once we have described our Multiple-Criteria decision problem, and after deriving the canonical formulation (2.10) for the SVM, we show how to interpret our problem, as described in Section 1, as a particular instance of (2.10). To do this, in what follows we construct a classification problem with the elements given in Section 1.

Given the directed graph $(\mathcal{A}, P)$, define the index sets I_+, I_- as

$$
\begin{aligned}
I_+ &= \{(a, a') \in \mathcal{A} \times \mathcal{A} : aPa'\} \\
I_- &= \{(a', a) \in \mathcal{A} \times \mathcal{A} : aPa'\}.
\end{aligned}
\tag{2.11}
$$

Each pair $(a, a') \in I_+ \cup I_-$ has associated a vector $x_{aa'} \in \mathbb{R}^N$,

$$
x_{aa'} = \Psi(a) - \Psi(a').
\tag{2.12}
$$

Ω_P-separation and Ω-separation are related as follows:

Proposition 2.2. *One has*

$$
\begin{aligned}
\Omega_P &= \big\{\omega \in \Omega : (\omega, 0)\ \Omega\text{-}separates\ \{x_{aa'} : aPa'\}, \{x_{a'a} : aPa'\}\big\} \\
&= \big\{\omega \in \Omega : (\omega, \beta)\ \Omega\text{-}separates\ \{x_{aa'} : aPa'\}, \{x_{a'a} : aPa'\} \\
&\qquad for\ some\ \beta \in \mathbb{R}\big\}
\end{aligned}
$$

Proof. Given $\omega \in \Omega_P$, one has by definition of Ω_P that

$$
\omega^\top(\Psi(a) - \Psi(a')) > 0 \quad \forall a, a',\, aPa',
$$

which amounts to saying that $(\omega, 0)$ Ω-separates the sets $\{x_{aa'} : aPa'\}, \{x_{a'a} : aPa'\}$. To finish the proof, let (ω, β) Ω-separate $\{x_{aa'} : aPa'\}, \{x_{a'a} : aPa'\}$, and let us conclude that $\omega \in \Omega_P$. By definition, one has that

$$
\begin{aligned}
\omega^\top\left(\Psi(a) - \Psi(a')\right) + \beta &> 0 \quad \forall a, a', aPa' \\
\omega^\top\left(\Psi(a') - \Psi(a)\right) + \beta &< 0 \quad \forall a, a', aPa',
\end{aligned}
$$

thus

$$\omega^\top \left(\Psi(a) - \Psi(a')\right) > |\beta| \geq 0 \quad \forall a, a', aPa'.$$

Hence, $\omega \in \Omega_P$. $\qquad\square$

Proposition 2.3. *For $I_+, I_-, x = (x_{aa'})_{aPa'}$, as defined in (2.11)-(2.12), if (ω, β) is an optimal solution to (2.10), then $\beta = 0$.*

Proof. The constraints in (2.10) are written as

$$
\begin{array}{rcll}
\omega^\top \left(\Psi(a) - \Psi(a')\right) + \beta & \geq & 1 & \forall a, a' \in \mathcal{A}, \, aPa' \\
\omega^\top \left(\Psi(a') - \Psi(a)\right) + \beta & \leq & -1 & \forall a, a' \in \mathcal{A}, \, aPa' \\
\omega & \in & \Omega,
\end{array}
$$

or, in condensed form,

$$
\begin{array}{rcll}
\omega^\top \left(\Psi(a) - \Psi(a')\right) & \geq & 1 + |\beta| & \forall a, a' \in \mathcal{A}, \, aPa' \\
\omega & \in & \Omega.
\end{array}
\tag{2.13}
$$

Given (ω_0, β_0), feasible for (2.10), with $\beta_0 \neq 0$, one has that the solution $(\frac{1}{1+|\beta_0|}\omega_0, 0)$ is also feasible, with objective value $\frac{1}{1+|\beta_0|}\|\omega_0\|^\circ < \|\omega_0\|^\circ$. Hence, (ω_0, β_0) cannot be optimal. $\qquad\square$

Hence, we can impose in (2.10) that $\beta = 0$, and then (2.13) yields a finite set of linear constraints. This way we rewrite (2.10) as

$$
\begin{array}{ll}
\min & \|\omega\|^\circ \\
\text{s.t.} & \omega^\top \left(\Psi(a) - \Psi(a')\right) \geq 1 \, \forall a, a' \in \mathcal{A}, \, aPa' \\
& \omega \in \Omega,
\end{array}
\tag{2.14}
$$

which is equivalent to (1.1), taking as norm $\gamma = \|\cdot\|^\circ$.

In other words, finding the linear score function in Ω maximizing the minimum slack (normalization done via γ) is reduced to the problem of finding the hyperplane of largest margin, distances being measured through $\|\cdot\| = \gamma^\circ$.

For more structured problems, further results can be obtained. This is the case, for instance, when the norm has some monotonicity properties. We recall that $\|\cdot\|^\circ$ is said to be monotonic in $\mathbb{R}^N_+$ if

$$(0 \leq \omega \leq \overline{\omega}) \Rightarrow \|\omega\|^\circ \leq \|\overline{\omega}\|^\circ,$$

whereas $\| \cdot \|^\circ$ is said to be strictly monotonic in $\mathbb{R}_+^N$ if

$$(0 \leq \omega \leq \overline{\omega}, \omega \neq \overline{\omega}) \Rightarrow \|\omega\|^\circ < \|\overline{\omega}\|^\circ.$$

For monotonic norms we have the following

Proposition 2.4. *Let* $u \in \mathbb{R}_+^N$, $u \neq 0$, *such that*

$$
\begin{aligned}
u^\top \left(\Psi(a) - \Psi(a')\right) &\leq 0 &&\forall a, a' \in \mathcal{A}, \, aPa' \\
q_j^\top u &\leq 0 &&\forall j \in J.
\end{aligned}
$$

1. *If* $\| \cdot \|^\circ$ *is monotonic in* $\mathbb{R}_+^N$, *then there exists* ω^*, *optimal for (2.14), such that* $\min\{\omega_i^* : u_i > 0\} = 0$.

2. *If* $\| \cdot \|^\circ$ *is strictly monotonic in* $\mathbb{R}_+^N$, *then any* ω^* *optimal for (2.14) satisfies* $\min\{\omega_i^* : u_i > 0\} = 0$.

Proof. Let $\|\cdot\|$ be monotonic in $\mathbb{R}_+^N$, and let ω^* be optimal for (2.14). Define Δ as

$$\Delta = \min\left\{\frac{\omega_i^*}{u_i} : u_i > 0\right\}.$$

If $\Delta = 0$, the result follows, so suppose $\Delta > 0$. By construction of Δ and the assumptions,

$$
\begin{aligned}
\omega_i^* - \Delta u_i &\geq 0 &&\forall i = 1, 2, \ldots, N \\
q_j^\top(\omega^* - \Delta u) &\geq 0 &&\forall j \in J.
\end{aligned}
$$

Hence, $\omega^* - \Delta u \in \Omega$. Moreover, for any $a, a' \in \mathcal{A}, aPa'$, one has that

$$(\omega^* - \Delta u)^\top \left(\Psi(a) - \Psi(a')\right) \geq \omega^{*\top} \left(\Psi(a) - \Psi(a')\right) \geq 1,$$

thus $\omega^* - \Delta u$ is feasible for (2.14).

Hence,

$$
\begin{aligned}
0 &\leq \omega^* - \Delta u \leq \omega^* \\
\omega^* - \Delta u &\neq \omega^*
\end{aligned}
$$

If $\| \cdot \|^\circ$ is monotonic in $\mathbb{R}_+^N$, we have that

$$\|\omega^* - \Delta u\|^\circ \leq \|\omega^*\|^\circ,$$

and thus $\omega := \omega^* - \Delta u$ is also optimal and satisfies, by construction, $\min\{\omega_i : u_i > 0\} = 0$.

If $\|\cdot\|^\circ$ is strictly monotonic in $\mathbb{R}^N_+$, we have that

$$\|\omega^* - \Delta u\|^\circ < \|\omega^*\|^\circ,$$

which is a contradiction, and thus, in this case, ω^*, as any optimal solution, must satisfy $\min\{\omega_i^* : u_i > 0\} = 0$. $\qquad\square$

As an application of this result, consider a decision problem in which criterion C_k is measured in an ordinal scale, with n_k different ranked levels, $L_1, \ldots, L_{n_k}$, L_1 being the best and L_{n_k} the worst.

Following the strategy of introducing dummy variables, as described in Example 1.1, we define

$$\Psi_{k1}(a) \;=\; \left\{ \begin{array}{ll} 1, & \text{if } a \text{ takes the value } L_1 \text{ in criterion } C_k \\ 0, & \text{else} \end{array} \right.$$

$$\vdots \qquad \vdots$$

$$\Psi_{kn_k}(a) \;=\; \left\{ \begin{array}{ll} 1, & \text{if } a \text{ takes the value } L_{n_k} \text{ in criterion } C_k \\ 0, & \text{else} \end{array} \right.$$

No further constraints are imposed on the weights associated with C_k, thus the only constraints involving the weights $\omega_{k,1} \ldots, \omega_{k,n_k}$ should be, as in Example 1.2,

$$\omega_{k,1} \geq \omega_{k,2} \ldots \geq \omega_{k,n_k}$$

If $\|\cdot\|^\circ$ is monotonic in $\mathbb{R}^N_+$, then there exists an optimal ω with $\omega_{kn_k} = 0$. Moreover, if $\|\cdot\|^\circ$ is strictly monotonic in $\mathbb{R}^N_+$, then we can assert that any optimal ω satisfies $\omega_{kn_k} = 0$. This is a direct consequence of Proposition 2.4. Indeed, take as u the vector with 0 in all its components excepting those corresponding to the columns of the form Ψ_{kj}, which are set to 1. Such a vector u fulfills the assumptions of Proposition 2.4, and thus the result follows.

3 An interactive method

SVM are known to enjoy excellent *generalization properties* both in theory and in practice, i.e., the rule constructed via (2.10) tends to be useful to

classify future instances in the classification problem, see Cristianini and Shawe-Taylor (2000), Vapnik (1998), and Vapnik (2000).

In the context of Multiple-Criteria Decision Making, this property can be extremely useful for designing a method of progressive articulation of preferences, in which, unlike other classical interactive methods such as STEM, Benayoun et al. (1971), or the Tchebycheff method of Steuer and Choo (1983), the system at each step asks the decision-maker to find *somehow* an action better than the one proposed by the system, or to increase the amount of partial information incorporated in the polyhedral cone Ω.

The procedure might work as follows

Step 0: `Initialize:`

  - `Choose a norm` $\| \cdot \|$ `in` $\mathbb{R}^N$.
  - `Take` Ω_1, `polyhedral cone in` $\mathbb{R}^N_+$, `modelling the relative importance of criteria.`
  - `Construct` P_1 `by pairwise comparison among some elements of` $\mathcal{A}$, P_1.
  - `Set` $k = 1$ `and go to Step 1.`

Step k : `Find` ω^k, `optimal solution to`

$$
\begin{array}{ll}
\min & \|\omega\|^\circ \\
\text{s.t.} & \omega^\top (\Psi(a) - \Psi(a')) \geq 1 \, \forall a, a' \in \mathcal{A}, \, a P_k a' \qquad (3.1) \\
& \omega \in \Omega_k,
\end{array}
$$

`and find` $a^k \in \mathcal{A}$, `optimal to`

$$
\begin{array}{ll}
\max & \Psi(a)^\top \omega^k \\
\text{s.t.} & a \in \mathcal{A}
\end{array}
$$

`If` (a^k, ω^k) `is considered to be satisfactory then`
`STOP with` a^k `as optimal solution and` ω^k `as vector of weights.`
`Else, enlarge` P_k `(e.g., by showing some` a `preferred to` a^k,
`and setting` $P_{k+1} = P_k \cup \{(a, a^k)\}$) `or reduce` Ω_k `(by adding a homogeneous constraint , and thus setting` $\Omega_{k+1} = \Omega_k \cap \{\omega : q^\top \omega \leq 0\}$). `GoTo Step` $k + 1$.

A toy example is given below.

Example 3.1. Consider the 4-objective knapsack problem

$$\begin{array}{ll} \max & \left(r_1^\top x, \ldots, r_4^\top x\right) \\ \text{s.t.} & d^\top x \leq 30 \\ & x_i \in \{0,1\}, \ i = 1, 2, \ldots, 10, \end{array} \tag{3.2}$$

with coefficients r_j and d given in Table 4. Since this is a synthetic problem, we can assume that the weight vector ω^* of the decision-maker is at hand, and will check the output of the algorithm against such ω^*. In fact, the unknown ω^* will be given by

$$\omega^* = (0.1, 0.3, 0.2, 0.4), \tag{3.3}$$

although no information on the weights (apart from non-negativity) will be provided to the system. Hence, $\Omega = \mathbb{R}_+^N$.

It is clear that the high cardinality of the feasible set $\mathcal{A}$ advices against the use of any methodology which starts with a complete enumeration of $\mathcal{A}$. Instead, the system samples some feasible solutions and defines strict preferences among them, leading to a strict preference P, from which a weight vector ω and an action a are obtained.

r_1	7	6	1	9	8	2	7	7	10	4
r_2	2	1	-2	3	4	-8	2	6	9	-6
r_3	10	2	4	9	1	5	3	6	3	9
r_4	-1	4	-15	4	1	1	-6	7	-7	-1
d	7	1	6	4	7	3	6	6	10	4

Table 4: *Data for Example 3.1.*

We solve the single-objective knapsack problems

$$\begin{array}{ll} \max & r_i^\top x \\ \text{s.t.} & d^\top x \leq 30 \\ & x_j \in \{0,1\} \quad \forall j, \end{array}$$

and denote by y_i the corresponding optimal solutions, as given in Table 5.

The decision-maker is asked to sort the actions in $\{y_1, y_2, y_3, y_4\}$, thus defining $(\mathcal{A}, P_1)$. The decision-maker (in fact, we, following (3.3)) gives the following sorting

$$y_2 P_1 y_4 P_1 y_1 P_1 y_3$$

y_i											$r_i^\top y_i$
y_1	0	1	0	1	0	1	1	1	1	0	41
y_2	0	1	0	1	1	0	0	1	1	0	23
y_3	1	0	1	1	0	1	0	1	0	1	43
y_4	0	1	0	1	1	1	0	1	0	0	17

Table 5: *Optimal solutions x_i for single-objective knapsack problems.*

As customary in SVM, we have chosen as $\| \cdot \|$ the Euclidean norm. Solving the convex quadratic problem with linear constraints (3.1) for $\Omega_1 = \mathbb{R}_+^N$ and P_1, one obtains

$$
\begin{aligned}
\omega^1 &= (0, 0.1209, 0, 0.1319) \\
x^1 &= (0, 1, 0, 1, 1, 0, 0, 1, 1, 0) = y_2
\end{aligned}
$$

Such solution is not considered to be satisfactory, and P_1 is enriched, yielding P_2. In particular, the decision maker provides some feasible action, y_5,

$$
y_5 = (1, 1, 0, 1, 0, 0, 0, 1, 1, 0),
$$

and P_2 is defined as

$$
P_2 = P_1 \cup \{(y_5, x^1)\}.
$$

Problem (3.1) is solved for P_2, yielding ω^2, x^2,

$$
\begin{aligned}
\omega^2 &= (0, 0.1994, 0.2101, 0.2462) \\
x^2 &= (1, 1, 0, 1, 1, 0, 0, 1, 0, 1).
\end{aligned}
$$

The importance weights ω^2 obtained are not considered to be acceptable: ω^2 incorporates the fact that the least important criterion is the first one, and the most important is the last one, but fails to fit the rank of the second and third criteria, which are reversed with respect to ω^*.

If we were concerned not only with finding the most preferred action, but also a full ranking on $\mathcal{A}$, we could go on, e.g., by imposing the weight constraint $\omega_2 \geq \omega_3$. Adding this constraint to the problem, and solving again, we obtain

$$
\begin{aligned}
\omega^3 &= (0, 0.2162, 0.2162, 0.2568) \\
x^3 = x^2 &= (1, 1, 0, 1, 1, 0, 0, 1, 0, 1).
\end{aligned}
$$

Adding now the cut $\omega_4 \geq 1.5\omega_3$ leads to

$$\begin{aligned} \omega^4 &= (0.0286, 0.2571, 0.2571, 0.3857) \\ x^4 = x^3 = x^2 &= (1, 1, 0, 1, 1, 0, 0, 1, 0, 1). \end{aligned}$$

At this moment, it seems the process is stabilized in the solution x^2 and we stop, yielding x^2 as optimal action, and ω^4 as vector of importance of weights. In fact, x^2 can be shown to be optimal for the knapsack problem aggregating the four criteria with weights ω^*.

Hence, we have quickly obtained a most preferred action.

4 Concluding remarks and extensions

In this note we have addressed the problem of extending an irreflexive binary relation (modelling strict preferences in a Multiple-Objective Decision-Making problem) to a total preorder induced by a linear function of the criteria $\Psi_1, \ldots, \Psi_N$. Posed as a non-parametric classification problem, we have shown that (hard-margin) SVM can be used to come up with a procedure of progressive elicitation of preferences.

The strategy proposed in this note can also be used for problems with uncertainty or vagueness in the data. Indeed, suppose the criteria are not scored in a precise way, and, instead, each score function Ψ_j does not take values in $\mathbb{R}$ but in the set of compact intervals of $\mathbb{R}$,

$$\Psi_j(a) = \left[\underline{\Psi_j(a)}, \overline{\Psi_j(a)}\right] \subset \mathbb{R}, \qquad j = 1, 2, \ldots, N. \tag{4.1}$$

We could still use (3.1) taking into account that now, each constraint of type

$$\omega^\top \left(\Psi(a) - \Psi(a')\right) \geq 1$$

has an interval as left-hand-side. Since the ω is non-negative (recall that, by assumption, $\Omega \subset \mathbb{R}_+^N$), we can avoid interval-type constraints and re-write (3.1) as

$$\begin{aligned} &\min \quad \|\omega\|^\circ \\ &\text{s.t.} \quad \textstyle\sum_{j=1}^N \omega_j \left(\underline{\Psi_j(a)} - \overline{\Psi_j(a')}\right) \geq 1 \, \forall a, a' \in \mathcal{A}, \, aP_k a' \\ &\quad\quad \omega \in \Omega_k. \end{aligned} \tag{4.2}$$

SVM are used not only to linearly separate separable sets, but also to construct linear rules which misclassify "few points" in the non-separable case, i.e., for the case in which Problem (2.10), or (4.2) if, as in (4.1), intervals are given for Ψ_j. is unfeasible. Several strategies have been proposed to do so. In particular, in the so-called *soft-margin approach*, Cortes and Vapnik (1995), (2.10) is replaced by a problem in the form

$$
\begin{aligned}
\min \quad & \|\omega\|_2^2 + C \sum_{i \in I} \eta_i^p \\
\text{s.a.} \quad & y_i \left(\omega^\top x_i + \beta \right) \geq 1 - \eta_i \, \forall i \in I \\
& \eta \geq 0 \\
& \omega \in \Omega,
\end{aligned}
\tag{4.3}
$$

where $p \geq 1$ (typically $p = 1$) and C is a strictly positive tuning parameter. See, e.g., Cristianini and Shawe-Taylor (2000), Suykens and Vandewalle (1999), Suykens et al. (2002), and Zhu et al. (2003), for related proposals, all seen as Goal-Programming strategies for solving unfeasible optimization problems, Carrizosa and Fliege (2002).

In our Multiple-Criteria Decision-Making problem, it may be the case that Ω_P is empty, since, e.g., P contains cycles of strict preference. Our aim of seeking P_ω enriching P is then replaced by the less ambitious aim of seeking P_ω somehow similar to P. The interactive procedure described in Section 3 can be used by replacing the SVM problem (3.1) by

$$
\begin{aligned}
\min \quad & \|\omega\|_2^2 + C \sum_{aP_k a'} \eta_{aa'}^p \\
\text{s.t.} \quad & \omega^\top \left(\Psi(a) - \Psi(a') \right) \geq 1 - \eta_{aa'} \, \forall a, a' \in \mathcal{A}, \, aP_k a' \\
& \eta_{aa'} \geq 0 \, \forall a, a' \in \mathcal{A}, \, aP_k a' \\
& \omega \in \Omega_k.
\end{aligned}
\tag{4.4}
$$

With this, one obtains a vector ω which hopefully satisfies most of the constraints defining Ω_P. It should be observed, however, that the solution obtained this way is not necessarily a vector minimizing the number of constraints in Ω_P, and can only be seen as a heuristic approach.

Whereas the aim of this note was to show that SVM can be used in Multiple-Criteria Decision Making, it is also possible to see the problem the other way round. Indeed, the problem does not ask for the use of the standard SVM; on the contrary, certain issues are relevant here, but mostly ignored in the SVM literature. Among others,

1. the choice of the norm $\| \cdot \|$. Since we are using γ as a normalization, there is no natural choice for it. This asks for a deep study of SVM for arbitrary norms, extending the results of Mangasarian (2000).

2. the partial information on the criteria has been modelled via homogeneous linear constraints. Whereas SVM problems with constraints have already been addressed in the literature, e.g., Fung et al. (2001), this seems to be more the exception than the rule. Further results, both at analytical and algorithmic levels, are needed.

These issues deserve further study and will be the subject of future research.

References

Benayoun R., Montgolfier J. de, Tergny J. and Laritchev O. (1971). Linear Programming and Multiple Objective Functions: STEP Method (STEM). *Mathematical Programming* 1, 366–375.

Burges C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition". *Data Mining and Knowledge Discovery* 2, 121–167.

Carrizosa E. and Conde E. (2002). A Fractional Model for Locating Semi-Desirable Facilities on Networks. *European Journal of Operational Research* 136, 67–80.

Carrizosa E., Conde E., Fernández F.R. and Puerto J. (1995). Multi-Criteria Analysis with Partial Information about the Weighting Coefficients. *European Journal of Operational Research* 81, 291–301.

Carrizosa E. and Fliege J. (2002). Generalized Goal Programming: Polynomial methods and applications. *Mathematical Programming* 93, 281–303.

Cortes C. and Vapnik V. (1995). Support Vector Networks. *Machine Learning* 20, 273–297.

Crammer K. and Singer Y. (2001). Pranking with Ranking. Proceedings of the Fourteenth Annual Conference on Neural Information Processing Systems, 641–647.

Cristianini N. and Shawe-Taylor J. (2000). *An Introduction to Support Vector Machines.* Cambridge University Press.

Freund Y., Iyer R., Schapire R.E. and Singer Y. (1998). An Efficient Boosting Algorithm for Combining Preferences. Proceedings of the Fifteenth International Conference on Machine Learning, Madison, Wisconsin, USA, 170–178.

Fung G., Mangasarian O.L. and Shavlik J. (2001). Knowledge-Based Support

Vector Machine Classifiers. Technical Report 01-09. Data Mining Institute, University of Winsconsin.

Hastie T., Tibshirani R. and Friedman J. (2001). *The Elements of Statistical Learning.* Springer.

Herbrich R., Graepel T. and Obermayer K. (2000). Large Margin Rank Boundaries for Ordinal Regression. In: Smola A., Barlett P., Schölkopf B. and Schuurmans D. (eds.), *Advances in Large Margin Classifiers.* MIT Press, 115-132.

Hiriart-Urruty J.B. and Lemaréchal C. (1996). *Convex Analysis and Minimization Algorithms*, vol. 2. Springer.

Mangasarian O.L. (1965). Linear and Nonlinear Separation of Patterns by Linear Programming. *Operations Research* 13, 444–452.

Mangasarian O.L. (2000). Generalized Support Vector Machines. In: Smola A., Barlett P., Schölkopf B. and Schuurmans D. (eds.), *Advances in Large Margin Classifiers.* MIT Press, 135–146.

Plastria F. and Carrizosa E. (2001). Gauge Distances and Median Hyperplanes. *Journal of Optimization Theory and Applications* 110, 173–182.

Roy B. (1968). Classement et Choix en Présence de Points de Vue Multiples (la Méthode Electre). *Revue Française d'Informatique et de Recherche Opérationnelle* 2, 57–75.

Smola A. and Schölkopf B. (2002). *Learning with kernels.* MIT Press.

Steuer R.E. and Choo E.U. (1983). An Interactive Weighted Tchebycheff Procedure for Multiple Objective Programming. *Mathematical Programming* 26, 326–344.

Suykens J.A.K. and Vandewalle J. (1999). Least Squares Support Vector Machine Classifiers. *Neural Processing Letters* 9, 293–300.

Suykens J.A.K., Van Gestel T., Brabanter J. De, De Moor B. and Vandewalle J. (2002). *Least Squares Support Vector Machines.* World Scientific Publishing.

Vapnik V.N. (1998). *Statistical Learning Theory.* Wiley.

Vapnik V.N. (2000). *The Nature of Statistical Learning.* Springer.

Zhu J., Rosset S., Hastie T. and Tibshirani R. (2003). L_1 Norm Support Vector Machines. Technical report, Standford University.

Top

Volume 14, Number 2
December 2006

CONTENTS   Page