Maximum Margin Separations in Finite Closure Systems

Florian Seiffarth[1], Tamás Horváth[1,2,3]✉, and Stefan Wrobel[1,2,3]

[1]Dept. of Computer Science, University of Bonn, Bonn, Germany
[2]Fraunhofer IAIS, Schloss Birlinghoven, Sankt Augustin, Germany
[3]Fraunhofer Center for Machine Learning, Sankt Augustin, Germany
{seiffarth,horvath,wrobel}@cs.uni-bonn.de

Abstract. Monotone linkage functions provide a measure for proximities between elements and subsets of a ground set. Combining this notion with Vapnik's idea of support vector machines, we extend the concepts of maximal closed set and half-space separation in finite closure systems to those with maximum margin. In particular, we define the notion of margin for finite closure systems by means of monotone linkage functions and give a greedy algorithm computing a maximum margin closed set separation for two sets efficiently. The output closed sets are maximum margin half-spaces, i.e., form a partitioning of the ground set if the closure system is Kakutani. We have empirically evaluated our approach on different synthetic datasets. In addition to binary classification of finite subsets of the Euclidean space, we considered also the problem of vertex classification in graphs. Our experimental results provide clear evidence that maximal closed set separation with maximum margin results in a much better predictive performance than that with arbitrary maximal closed sets.

Keywords: closure systems · maximum margin separations · monotone linkages · binary classification

1 Introduction

Motivated by different applications of *finite closure systems*, including e.g. closed itemset mining [12], inductive logic programming [11], and formal concept analysis [5], in [14] we studied the algorithmic properties of *half-space* and *maximal closed set* separation in this kind of set systems. One of our results in [14] is a greedy algorithm, which takes as input two sets and returns two *disjoint* maximal closed sets containing them if their closures are disjoint. It is shown in [14] that this greedy algorithm provides an algorithmic characterization of the special class of *Kakutani* closure systems [2,9]. That is, for any two sets it returns two complementary *half-spaces* containing them if and only if the closures of the input sets are disjoint, where a half-space is a closed set such that its complement is also closed. For the case that the separating maximal closed sets or half-spaces are not unique, the greedy algorithm returns one of them selected arbitrarily.

This is similar to Rosenblatt's perceptron algorithm [13], which fulfills also the minimum requirement the output hyperplane to separate the input point sets. A major drawback of such unconstrained solutions is that they provide no control of *overfitting*. This problem has been addressed by Vapnik and his co-authors' work on *support vectors machines* (SVM) [1], which have become a well-established tool within machine learning for its well-founded theory and excellent predictive performance on a broad range of real-world problems. In particular, SVM resolve the problem of overfitting by separating the data points in an inner product feature space by the hyperplane maximizing the minimum of the distances to the sets of positive and negative examples.

Motivated by the same problem as SVM, in this work we adapt the idea of maximum margin hyperplanes to the binary separation problems studied in [14] for *finite closure systems*. We stress that our adaptation does *not* generalize SVM to finite closure systems. While in case of SVM the inner product induces a distance, in case of finite closure systems the ground set is typically *not* a metric space. To overcome this problem, we assume that the closure systems are provided by some *weak* measure of proximity defined by means of *monotone linkage functions* [10]. While this kind of functions strongly generalize distance functions (e.g., they are not required to fulfill symmetry or the triangle inequality), they preserve the *anti-monotonicity* of distances. That is, the linkage from a point to a set is anti-monotonic for set inclusion. Similarly to SVM, this feature is essential for the separation problems considered in this work. A second issue is how to define margins for closed set and half-space separations in finite closure systems. While there are different equivalent characterizations of maximum margins for SVM, it turns out that their equivalence does not hold when adapting them to abstract closure systems equipped with monotone linkage functions. In particular, in contrast to SVM, the linkage of the set of positive examples to a half-space can be different from that of the negative examples to the complementary half-spaces for *all* half-spaces. We therefore define the *margin* by the smallest linkage from the closures of the input sets to the complementary half-spaces. Furthermore, we generalize this concept to arbitrary closed set separations as well.

Using these notions, we formulate the computational problems of finding closed set and half-space separations *maximizing* the margin in finite closure systems equipped with monotone linkage functions. This problem preserves several key features of SVM for abstract closure systems. For the above problem we give another *greedy* algorithm and prove that it is correct and requires a linear number of evaluations of the underlying closure operator and linkage function. We also show that for Kakutani closure systems [9], the algorithm always returns a half-space separation of the input sets with maximum margin if and only if the closures of the two training sets are disjoint.

We experimentally evaluated the predictive performance of our algorithm on various synthetic datasets. Our empirical results concerning point classification in Euclidean spaces show that our algorithm clearly outperforms the greedy separation algorithm in [14]. In addition, we carried out several experiments with vertex classification in trees and also in random graphs using the shortest path closure operator. Similarly to the other experiments, our algorithm consistently outperformed the greedy algorithm in [14]. For space limitations we omit further applications dealing, among others, with finite lattices, in particular, with formal concept and subsumption lattices in inductive logic programming.

The rest of the paper is organized as follows. In Section 2 we collect the necessary concepts and fix the notation. In Section 3 we introduce our notion of margin defined by means of monotone linkage functions for finite closure systems. In Section 4 we present our greedy algorithm solving closed set and half-space separation with maximum margin and prove some of its basic formal properties. In Section 5 we report our experimental results. Finally, in Section 6, we conclude and formulate some problems for further study.

2 Preliminaries

In this section we collect some basic notions concerning *closure systems* (see, e.g., [3,9]) and *linkage functions* (see, e.g., [6]) and fix the notation.

Closure Systems The power set of a set E is denoted by 2^E. A *set system* over a ground set E is a pair $(E, \mathcal{C})$, where $\mathcal{C} \subseteq 2^E$; $(E, \mathcal{C})$ is a *closure system* if it fulfills the axioms: (i) $E \in \mathcal{C}$ and (ii) $X \cap Y \in \mathcal{C}$ for all $X, Y \in \mathcal{C}$. Unless otherwise stated, by closure systems we always mean *finite* closure systems, i.e., $|E| < \infty$. It is a well-known fact (see, e.g., [3]) that closure systems give rise to closure operators and vice versa. More precisely, a *closure operator* over E is a function $\rho : 2^E \to 2^E$ satisfying

i) $X \subseteq \rho(X)$, (*extensivity*)
ii) $\rho(X) \subseteq \rho(Y)$ whenever $X \subseteq Y$, (*monotonicity*)
iii) $\rho(\rho(X)) = \rho(X)$ (*idempotency*)

for all $X, Y \subseteq E$. The following characterization is standard (see, e.g., [3]):

Proposition 1. *Let $(E, \mathcal{C})$ be a closure system and $\rho : 2^E \to 2^E$ be the map defined by $\rho(X) = \bigcap\{C \in \mathcal{C} : X \subseteq C\}$ for all $X \subseteq E$. Then ρ is a closure operator and $\mathcal{C} = \{C \subseteq E : \rho(C) = C\}$. Conversely, let ρ be a closure operator over E. Then $(E, \mathcal{C}_\rho)$ with $\mathcal{C}_\rho = \{C \subseteq E : \rho(C) = C\}$ is a closure system.*

The elements of $\mathcal{C}$ will be referred to as *closed* sets. We use the notation $\mathcal{C}_\rho$ to indicate that the closure system is defined by the closure operator ρ.

We will have a special interest in the following closure systems over Euclidean spaces, graphs, and lattices[1].

1. (*finite convex hulls in* $\mathbb{R}^d$) Let E be a finite subset of $\mathbb{R}^d$ for some $d > 0$. Then the function $\alpha : 2^E \to 2^E$ defined by

$$\alpha(X) = \mathrm{conv}(X) \cap E \tag{1}$$

[1] For space limitation, the applications concerning closure systems over finite lattices will be discussed in the long version of this paper.

for all $X \subseteq E$ is a closure operator over E, where $\text{conv}(\cdot)$ denotes the *convex hull* operator on $\mathbb{R}^d$.

2. (*shortest path closure in graphs [4]*) Let $G = (V, E)$ be a graph with vertex set V and edge set E. Then $(V, \mathcal{C}_\gamma)$ is a closure system if

$$V' \in \mathcal{C}_\gamma \iff V(P) \subseteq V' \tag{2}$$

for all $V' \subseteq V$, $u, v \in V'$, and $P \in \mathcal{S}_{u,v}$, where $\mathcal{S}_{u,v}$ is the set of all shortest paths connecting u and v in G and $V(P)$ denotes the set of vertices in P.

3. (*closed sets in lattices [15]*) Let $(L; \leq)$ be a finite lattice. Then the function $\lambda : 2^L \to 2^L$ defined by

$$\lambda : L' \mapsto \{x \in L \mid \inf L' \leq x \leq \sup L'\} \tag{3}$$

for all $L' \subseteq L$ is a closure operator, where $\inf L'$ (resp. $\sup L'$) is the greatest lower bound or bottom (resp. least upper bound or top) element of L'.

The primary focus of this work is on maximum margin separation in closure systems. To formulate this problem in Section 3, we recall some definitions concerning separations in finite closure systems from [14]. More precisely, let $(E, \mathcal{C})$ be a closure system and $A, B \subseteq E$. Then A and B are

(i) *separable* in $(E, \mathcal{C})$ if there are disjoint closed sets C_A, C_B in $\mathcal{C}$ such that $A \subseteq C_A$ and $B \subseteq C_B$,

(ii) *maximal closed set separable* in $(E, \mathcal{C})$ if there are disjoint closed sets C_A, C_B in $\mathcal{C}$ such that $A \subseteq C_A$ and $B \subseteq C_B$, and there are no disjoint closed sets $C'_A \supseteq C_A$ and $C'_B \supseteq C_B$ such that at least one of the containments is proper,

(iii) *half-space separable* if there are $C, C^c \in \mathcal{C}$ such that $A \subseteq C$ and $B \subseteq C^c$, where $C^c = E \setminus C$.

Regarding (iii) above, a closed set $C \in \mathcal{C}$ is a *half-space* if its complement C^c is also closed. Finally, a closure system $(E, \mathcal{C})$ is *Kakutani* [9] if and only if all pairs of disjoint closed sets in $\mathcal{C}$ are half-space separable. It follows from the definitions that no separation is possible in $(E, \mathcal{C})$ if $\emptyset \notin \mathcal{C}$. Therefore, in the rest of the paper we always assume that the empty set is also closed, i.e., it is an element of the underlying closure system.

Monotone Linkage Functions To adapt Vapnik's idea of maximum margin separation to (abstract) finite closure systems, we need some additional formal tool to quantify the closeness between subsets of the ground set. Such an abstract measure for the proximity between elements and subsets of a ground set is provided by *monotone linkage functions* introduced by Mullat [10]. This kind of functions preserve an important elementary property of distances from points to sets in metric spaces and can therefore be regarded as a very general "distance" concept. More precisely, a *monotone linkage function* over a set E is a map $l : 2^E \times E \to \mathbb{R}$ such that

$$X \subseteq Y \implies l(X, e) \geq l(Y, e)$$

holds for all $X, Y \subseteq E$ and $e \in E$. That is, l is *anti-monotone* w.r.t. set containment, which is an essential property satisfied by distances as well. Thus, all distances give rise to monotone linkage functions; the converse is, however, not true. Note that by applying monotone linkage functions to singletons in the first argument, we obtain a pairwise proximity between the elements of the ground set. However, in contrast to metric spaces, the definition does not imply symmetry, i.e., $l(\{x\}, y)$ is not necessarily equal to $l(\{y\}, x)$. Furthermore, $l(X, e)$ is not required to be zero for $e \in X$.

There are several examples of monotone linkage functions on finite and infinite ground sets. Below we recall some of the most popular ones to illustrate the concept (c.f. [6] for further examples). The proof that the functions below are all monotone linkage is left to the reader.

(i) (*monotone linkage in* $\mathbb{R}^d$) For any distance D on $\mathbb{R}^d$, define $l : 2^{\mathbb{R}^d} \times \mathbb{R}^d \to \mathbb{R}$ by $l : (X, e) \mapsto \inf_{x \in X} \{D(x, e)\}$ for all $X \subseteq \mathbb{R}^d$ and $e \in \mathbb{R}^d$.

(ii) (*monotone linkage in (weighted) graphs*) For a (weighted) graph $G = (V, E)$ define $l : 2^V \times V \to R$ by $l : (X, e) \mapsto \min_{x \in X} \{d(x, e)\}$ for all $X \subseteq V$, where d denotes the (weighted) length of a (weighted) shortest path between vertices.

(iii) (*monotone linkage in graphs by maximum degree on induced subgraphs*) For a graph $G = (V, E)$, define $l : 2^V \times V \to \mathbb{R}$ by $l : (X, v) \mapsto \min_{x \in X} (\delta(v) - \delta_{G[X]}(x))$ for all $X \subseteq V$ and $v \in V$, where $G[X]$ is the subgraph of G induced by X, $\delta(v)$ the degree of v in G, and $\delta_{G[X]}(x)$ the degree of x in $G[X]$.

Monotone linkage functions have been studied intensively by Kempner [6,7,8] in the context of *clustering* over set systems and convex geometries. As mentioned above, we will apply them for defining margins in *arbitrary* finite closure systems. For this purpose, we will use the following notion many times in what follows. A *monotone linkage closure system* (MLCS) is a triple $(E, \mathcal{C}_\rho, l)$, where $(E, \mathcal{C}_\rho)$ is a closure system and l is a monotone linkage function on E. We will always assume that the closure operator and the linkage function are given implicitly by *oracles* under the usual complexity assumption. That is, for all $X \subseteq E$ and $e \in E$, $\rho(X)$ and $l(X, e)$ are returned in *unit time* by the oracles.

3 Maximum Margin Separations in MLCSs

Our main goal in this paper is to adapt Vapnik's idea [1] of *maximum margin* separating hyperplanes to finite closure systems. That is, given subsets A and B of some inner product (feature) space $\mathcal{F}$, in case of *support vector machines* (SVM) [1] we are interested in the hyperplane H^* having *maximum* distance to the two sets, i.e., which satisfies

$$d(A \cup B, H) \leq d(A \cup B, H^*) \tag{4}$$

for all hyperplanes H, where for all $X, Y \subseteq \mathcal{F}$, $d(X, Y) = \min_{y \in Y} d(X, y)$ with d being the distance induced by the underlying inner product. It is a well-known fact that if A and B are separable by a hyperplane, then H^* is *unique*; H^* is also referred to as the *maximum margin separating hyperplane*, where the *margin* of a separating hyperplane H is defined by

$$\mu(A, B) = d(A, H) + d(B, H) \ . \tag{5}$$

A key property of the margin is that it is *anti-monotone* w.r.t. set inclusion, i.e., $\mu(A', B') \le \mu(A, B)$ for all $A' \supseteq A$ and $B' \supseteq B$. Note that (4) implies

$$d(A, H^*) = d(B, H^*) \ .$$

Clearly, the above definitions are *not* (directly) applicable to maximum margin separation in closure systems because we do not assume E to be an inner product or a metric space and have therefore no measure in general for the distance from a point $e \in E$ to a subset $X \subseteq E$. Furthermore, while the notion of half-spaces in $\mathbb{R}^d$ has been generalized to closure systems, for hyperplanes there is no analogous definition. Hence, to be in a position to define margins, we need some suitable functions for the abstraction of "closeness" from a point to a subset of the ground set. They should *generalize* metrics, but *preserve* the anti-monotonic property above at the same time.

The class of *monotone linkage functions* [10] defined in Section 2 fulfill both of these requirements. In addition to generality and anti-monotonicity, they have some further properties making this class an attractive candidate for our purpose. In particular, monotone linkage functions assume neither symmetry nor the triangle inequality.

To adapt the ordinary definition of margins to MLCSs, note that if a hyperplane $H \subseteq \mathbb{R}^d$ separates A and B, then (5) is equivalent to

$$\begin{aligned} \mu(A, B) &= d(A, H_2) + d(B, H_1) \\ &= d(\text{conv}(A), H_2) + d(\text{conv}(B), H_1) \ , \end{aligned} \tag{6}$$

where $H_1 \supseteq A$ and $H_2 \supseteq B$ are the closed half-spaces defined by H (i.e., $H \subseteq H_1, H_2$). That is, in case of SVM, the margin given by a hyperplane H separating A and B is defined by the sum of the distances from the *convex hull* of A to the half-space H_2 containing B and from that of B to H_1 containing A.

Analogously to distances in metric spaces, we first extend linkage functions from sets to elements to those from sets to sets. Formally, for a linkage function l on E and subsets $X, Y \subseteq E$, we define the linkage l from X to Y by $l(X, Y) = \min_{y \in Y} l(X, y)$. Note that this extended definition preserves anti-monotonicity, i.e., $l(X', Y) \le l(X, Y)$ holds whenever $X' \supseteq X$. Let H, H^c be half-spaces of an MLCS $(E, \mathcal{C}_\rho, l)$ and $A \subseteq H, B \subseteq H^c$ for some $A, B \subseteq E$. Then, by analogy with (6), our *first* definition of the *margin* of the half-space separation of A, B by H, H^c is

$$\mu^+_{H,H^c}(A, B) = l(\rho(A), H^c) + l(\rho(B), H). \tag{7}$$

While the above adaptation of the ordinary notion of margins to MLCSs is relatively natural, the generalization is less obvious for *maximum* margin half-space separations. This is because for SVM there are two *equivalent* properties characterizing maximum margin hyperplanes H^* defining the closed half-spaces $H_1 \supseteq A$ and $H_2 \supseteq B$:

(i) H^* maximizes $\mu(A, B)$ such that $d(\text{conv}(A), H_2) = d(\text{conv}(B), H_1)$.
(ii) H^* maximizes $\min\{d(\text{conv}(A), H_2), d(\text{conv}(B), H_1)\}$.

That is, the maximum margin hyperplane by (i) lies in the "middle" between the convex hulls of A and B; by (ii) it maximizes the minimum of the distances from the two convex hulls. While (i) and (ii) are equivalent in case of SVM, the situation is different for MLCSs as shown in the proposition below.

Proposition 2. *There exists an MLCS $(E, \mathcal{C}_\rho, l)$ and subsets $A, B \subseteq E$ such that $\mu^+_{H_1,H_1^c}(A, B) \neq \mu^+_{H_2,H_2^c}(A, B)$, where*

$$H_1 = \operatorname*{arg\,max}_{H,H^c \in \mathcal{C}_\rho} \mu^+_{H,H^c}(A, B) \text{ subject to } l(\rho(A), H^c) = l(\rho(B), H)$$
$$H_2 = \operatorname*{arg\,max}_{H,H^c \in \mathcal{C}_\rho} \min\{l(\rho(A), H^c), l(\rho(B), H)\} \ .$$

Proof. Consider MLCS $(E, \mathcal{C}, l)$ with $E = \{a, b, c, d\}$ and $\mathcal{C} = \{X \subseteq E : |X| \neq 3\}$. The monotone linkage function is defined by

$$l(\{a\}, b) = l(\{b\}, a) = l(\{b\}, d) = 3, \quad l(\emptyset, e) = 3 \text{ for all } e \in E, \quad l(\{a\}, c) = 2,$$
$$l(\{a\}, d) = l(\{b\}, c) = 1, \quad l(X, e) = 0 \text{ for all other } X \subseteq E \text{ and } e \in E$$

It can be easily checked that $(E, \mathcal{C})$ is a closure system and l fulfills the anti-monotonicity property. For $A = \{a\}, B = \{b\}$ there exist exactly two different separating half-spaces of size 2, i.e., $H_1 = \{a, c\}$ and $H_2 = \{a, d\}$. Using the definition of linkage on sets it follows $l(A, H_1^c) = l(B, H_1) = 1$. Moreover, $l(A, H_2^c) = 2$ and $l(B, H_2) = 3$. Thus, H_1 fulfills the first property and H_2 the second one, by noting that $2 = \min\{l(A, H_2^c), l(B, H_2)\} > \min\{l(A, H_1^c), l(B, H_1)\} = 1$. The claim then follows by $2 = \mu^+_{H_1,H_1^c}(A, B) \neq \mu^+_{H_2,H_2^c}(A, B) = 5$.

Thus, for an MLCS $(E, \mathcal{C}, l)$, maximizing the margin as defined in (7) subject to $l(\rho(A), H^c) = l(\rho(B), H)$ is *not* equivalent to maximizing

$$\mu_{H,H^c}(A, B) := \min\{l(\rho(A), H^c), l(\rho(B), H)\} \tag{8}$$

over *all* half-space separations of A and B in $(E, \mathcal{C}_\rho, l)$ (see, also, Figure 1).

Since our primary interest is in classification, we prefer the definition in (ii) above and will accordingly focus on maximizing the margin defined by (8). Note that our definition of margin differs from that in SVM, as it involves only one part of the ordinary one.

Until now we have concentrated on half-space separations. In case of MLCSs, two sets with disjoint closures are, however, not always half-space separable. This motivates the relaxed concept of *maximal* closed set separation [14]. Fortunately, the above definition of margin can be extended naturally to arbitrary closed

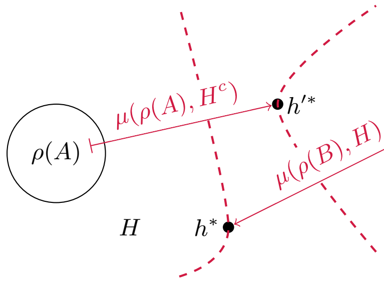


Fig. 1. Margins $\mu(\rho(A), H^c)$, $\mu(\rho(B), H)$ between closed sets $\rho(A), \rho(B)$ and half-spaces H, H^c together with the support elements h^*, h'^*.

sets. More precisely, for an MLCS $(E, \mathcal{C}_\rho, l)$, let $A, B \subseteq E$ and $C_A, C_B \in \mathcal{C}_\rho$ with $A \subseteq C_A$ and $B \subseteq C_B$. Then the *margin* for C_A and C_B is defined by

$$\mu_{C_A,C_B}(A, B) := \min\{l(\rho(A), C_A^c), l(\rho(B), C_B^c)\} \ . \tag{9}$$

Similarly to half-spaces, the definition takes only one part of the effective margin into account. Note that (8) is the special case of (9) for $C_A = H$ and $C_B = H^c$. We now show that the anti-monotonicity of monotone linkages extends to margins in MLCS. This property is essential for separations.

Lemma 3. *Let $(E, \mathcal{C}_\rho, l)$ be an MLCS, $A \subseteq A' \subseteq E$, $B \subseteq B' \subseteq E$, and $C_A \supseteq A', C_B \supseteq B'$ disjoint closed sets. Then $\mu_{C_A,C_B}(A, B) \geq \mu_{C_A,C_B}(A', B')$.*

Proof. This follows directly from the definition of margin in (9) and the anti-monotonicity of monotone linkage functions.

Moreover, maximizing the disjoint closed sets C_A and C_B in Lemma 3 maximizes the margin at the same time, as we show in the following lemma.

Lemma 4. *Let $C_A \subseteq C'_A$ and $C_B \subseteq C'_B$ be closed sets of an MLCS $(E, \mathcal{C}_\rho, l)$ with $C'_A \cap C'_B = \emptyset$ and $A \subseteq C_A, B \subseteq C_B$. Then $\mu_{C_A,C_B}(A, B) \leq \mu_{C'_A,C'_B}(A, B)$.*

Proof. From the definition of monotone linkages between sets it follows that $l(X, Y) \geq l(X, Y')$ whenever $Y \subseteq Y'$. Hence, by $C_A^c \supseteq C_A'^c$ and $C_B^c \supseteq C_B'^c$ we have

$$\begin{aligned} \mu_{C_A,C_B}(A, B) &= \min\{l(\rho(A), C_A^c), l(\rho(B), C_B^c)\} \\ &\leq \min\{l(\rho(A), C_A'^c), l(\rho(B), C_B'^c)\} \\ &= \mu_{C'_A,C'_B}(A, B) \ . \end{aligned}$$

Given a half-space separation of A, B with $A \subseteq H$ and $B \subseteq H^c$, similarly to SVM we can define the *support elements* by h^* and h'^* satisfying $l(\rho(A), H^c) = l(\rho(A), h'^*)$ and $l(\rho(B), H) = l(\rho(B), h^*)$, respectively. For example, in case of maximum margin separating half-spaces in trees, there are exactly two support elements corresponding to the two half-spaces.

4 The Maximum Margin Algorithm

Using (8) and (9) for the definition of margins for half-space and closed set separations, we are ready to formulate the separation problems in MLCS $(E, \mathcal{C}_\rho, l)$:

Maximum Margin Half-Space Separation (MMHSS) Problem: *Given* non-empty subsets A, B of E, *find* a half-space $H \in \mathcal{C}_\rho$ with $A \subseteq H, B \subseteq H^c$ that maximizes the margin, i.e., $H = \operatorname*{arg\,max}_{H_1, H_1^c \in \mathcal{C}_\rho} \mu_{H_1, H_1^c}(A, B)$, if A and B are half-space separable; o/w return "No".

Maximum Margin Closed Set Separation (MMCSS) Problem: *Given* non-empty subsets A, B of E, *find* disjoint closed sets $C_A, C_B \in \mathcal{C}_\rho$ with $A \subseteq C_A, B \subseteq C_B$ that maximize the margin, i.e., for all other disjoint closed sets $C'_A \supseteq A, C'_B \supseteq B$ it holds that $\mu_{C_A, C_B}(A, B) \geq \mu_{C'_A, C'_B}(A, B)$, if $\rho(A) \cap \rho(B) = \emptyset$; o/w return "No".

Remark 5. The MMHSS problem is a special case of the MMCSS problem for $C_A = H, C_B = H^c$. Moreover, Lemma 4 implies that for any maximum margin closed set separation there exists a maximal closed set separation of the same margin. The converse is, however, not true in general.

We solve the above problems by Algorithm 1, which is based on an adaptation of the greedy algorithm in [14]. The input to the algorithm is an MLCS $(E, \mathcal{C}_\rho, l)$ together with two sets $A, B \subseteq E$ of training examples. We assume that $\mathcal{C}_\rho$ is given by the closure operator ρ, which returns the closure for any $X \subseteq E$ in unit time. Similarly, for any $X \subseteq E$ and $e \in E$, $l(X, e)$ is returned by another oracle in unit time. Accordingly, we measure the complexity of Alg. 1 in terms of the number of closure operator calls and linkage function evaluations.

In Lines 1-4, the closures of A, B are calculated and checked for disjointness. In particular, if they are not disjoint, the algorithm terminates with "No", as in this case A and B are not separable by closed sets. Thus, the algorithm is correct for this case. Consider the case that $\rho(A) \cap \rho(B) = \emptyset$. For this case, all elements not contained in the union of the closures of A and B are first collected in F and sorted then by their minimum linkage from these two closed sets (Lines 5-6). The elements f in F will be processed one by one in this order and then immediately removed, potentially together with other untreated elements (Line 13). In particular, if the linkage from the closure of A to f is not greater than that of B or the current closed set C_B containing B cannot be extended by f, we expand the current closed set $C_A \supseteq A$ with f if it does not violate the disjointness with C_B (see Lines 9-10). Otherwise, we extend C_B by f, if $\rho(C_B \cup \{f\})$ remains disjoint with C_A (Lines 11-12). We then remove f and all other elements from F (Line 13) that have been added to C_A or to C_B in Line 10 or 12.

An example of the algorithm to the case that $(E, \mathcal{C}_\rho, l)$ is defined over graphs with the shortest path closure operator is given in Figure 2. We now show that Alg. 1 is correct (Thm. 6) and efficient (Thm. 8). Furthermore, in case of Kakutani closure systems, the sets C_A, C_B returned in Line 15 form complementary half-spaces with maximum margin whenever $\rho(A) \cap \rho(B) \neq \emptyset$ (Corollary 7).

Algorithm 1: Maximum Margin Separation

Input: a finite MLCS $(E, \mathcal{C}_\rho, l)$ and sets $A, B \subseteq E$
Output: *maximum* margin closed sets $C_A, C_B \in \mathcal{C}_\rho$ with $A \subseteq C_A$ and $B \subseteq C_B$ if $\rho(A) \cap \rho(B) = \emptyset$; "No" otherwise

$\overline{A}, C_A \leftarrow \rho(A); \overline{B}, C_B \leftarrow \rho(B)$;
if $C_A \cap C_B \neq \emptyset$ **then**
  return No;
end
$F \leftarrow E \setminus \{C_A \cup C_B\}$;
compute $\min\{l(\overline{A}, f), l(\overline{B}, f)\}$ for all $f \in F$ and sort F by these values;
while $F \neq \emptyset$ **do**
  take the smallest element $f \in F$;
  if $(l(\overline{A}, f) \leq l(\overline{B}, f) \vee \rho(C_B \cup \{f\}) \cap C_A \neq \emptyset) \wedge \rho(C_A \cup \{f\}) \cap C_B = \emptyset$ **then**
    $C_A \leftarrow \rho(C_A \cup \{f\})$;
  else if $\rho(C_B \cup \{f\}) \cap C_A = \emptyset$ **then**
    $C_B \leftarrow \rho(B_B \cup \{f\})$;
  $F \leftarrow F \setminus (C_A \cup C_B \cup \{f\})$;
end
return C_A, C_B

Theorem 6. *Algorithm 1 solves the MMCSS problem correctly.*

Proof. Let $(E, \mathcal{C}_\rho, l)$ be an MLCS and $A, B \subseteq E$. By construction, the algorithm returns "No" only for the case that $\rho(A) \cap \rho(B) \neq \emptyset$, i.e., when A and B are not separable in $\mathcal{C}_\rho$, implying the correctness for this case. Otherwise, the closed sets $C_A \supseteq A, C_B \supseteq B$ returned are disjoint and hence, form a *separation* of A and B. They are *maximal*, as only such elements of E are discarded that violate the disjointness condition. All such elements can be removed ultimately from F, as they do not have to be reconsidered again for the monotonicity of ρ.

Regarding optimality, suppose for contradiction that there are other disjoint closed sets $C'_A \supseteq A, C'_B \supseteq B$ such that

$$\mu_{C'_A, C'_B}(A, B) > \mu_{C_A, C_B}(A, B) \ . \tag{10}$$

For symmetry, we can assume w.l.o.g. that there is an $e^* \in C_A^c$ such that

$$\min\{l(\rho(A), C_A^c), l(\rho(B), C_B^c)\} = l(\rho(A), e^*) \ ,$$

i.e., $\mu_{C_A, C_B}(A, B) = l(\rho(A), e^*)$. Then, by (9) and (10) we have

$$l(\rho(A), e^*) < \min\{l(\rho(A), C'^c_A), l(\rho(B), C'^c_B)\} \tag{11}$$

implying $l(\rho(A), e^*) < l(\rho(A), C'^c_A)$. Thus, $e^* \notin C'^c_A$ and hence $e^* \in C'_A \subseteq C'^c_B$. But then, together with (11), we have $l(\rho(A), e^*) < l(\rho(B), e^*)$.

We prove that $e^* \in C'_A$ and $e^* \in C^c_A$ contradicts the assumptions. Conditions $C'_A \cap C'_B = \emptyset$ and $e^* \in C'_A$ imply that $\rho(\rho(A) \cup \{e^*\}) \cap \rho(B) = \emptyset$. Since $e^* \notin C_A$, e^* has not been added to C_A, though $l(\rho(A), e^*) < l(\rho(B), e^*)$. But this can

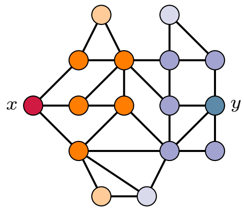


Fig. 2. Maximum margin half-space separation of x and y defined by the shortest path closure. Brighter nodes are added later to the respective class. The maximum margin between $\{x\}$ and $\{y\}$ is 2 for the linkage defined by weight 1 for all edges (see Sect. 2).

happen only if there is a non-empty set $G \subseteq F$ such that for all $g \in G$, g is before e^* in F, i.e., $\min\{l(\rho(A), g), l(\rho(B), g)\} \leq l(\rho(A), e^*)$. Assume there is a $g \in G$ such that $g \in C_A$, but $g \notin C'_A$. Then $g \in C'^c_A$ and thus,

$$\begin{aligned}\mu_{C'_A, C'_B}(A, B) &= \min\{l(\rho(A), C'^c_A), l(\rho(B), C'^c_B)\}\\ &\leq \min\{l(\rho(A), g), l(\rho(B), g)\}\\ &\leq l(\rho(A), e^*)\\ &= \mu_{C_A, C_B}(A, B)\end{aligned}$$

contradicting (10). Hence, for all $g \in G$, $g \in C_A$ implies $g \in C'_A$. In a similar way we have that $g \in C_B$ implies $g \in C'_B$ for all $g \in G$.

Since $e \in C^c_A$, $e^* \notin C_A$. There are two possible cases: (i) $e^* \in \rho(\rho(B) \cup G_B) \subseteq C'_B$, where $G_B \subseteq G$ is the set of elements added to $\rho(B)$. But this contradicts $e^* \in C'^c_B$. (ii) At the step e^* is considered for adding to C_A, there are disjoint subsets $G_A, G_B \subseteq G$ already added to $\rho(A)$ and $\rho(B)$, respectively, such that

$$\rho(\rho(\rho(A) \cup G_A) \cup \{e^*\}) \cap \rho(\rho(B) \cup G_B) \neq \emptyset.$$

But then, for $G_A \subseteq C'_A$ and $G_B \subseteq C'_B$ and for the monotonicity of ρ, we have $C'_A \cap C'_B \neq \emptyset$, as $e^* \in C'_A$; a contradiction.

Corollary 7. *For all MLCSs $(E, \mathcal{C}_\rho, l)$, Algorithm 1 solves the MMHSS-problem correctly if $(E, \mathcal{C}_\rho)$ is Kakutani.*

Proof. It is a direct implication of Thm. 6, as maximal disjoint closed sets are always half-spaces in any Kakutani closure system.

Theorem 8. *Algorithm 1 requires at most $2 \cdot |E \setminus (\rho(A) \cup \rho(B))|$ evaluations of l and $2 \cdot |E \setminus (\rho(A) \cup \rho(B))| + 2$ calls of ρ.*

Proof. To sort F, we evaluate l twice for all $f \in F$ with $|F| = |E \setminus (\rho(A) \cup \rho(B))|$. The closure is calculated twice to determine the closures of the input sets (Line 1) and twice for all $f \in F$ in the worst case (Lines 9 and 11).

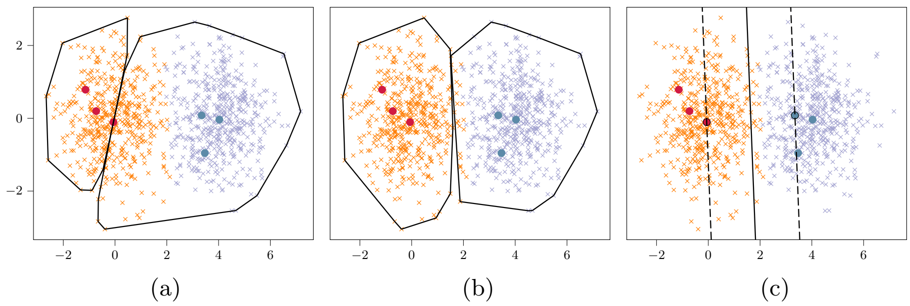


Fig. 3. Comparison of greedy separation (a), maximum margin separation (b) and ordinary support vector machines (c).

5 Empirical Evaluations

Potential applications of maximum margin closed set and half-space separation in finite closure systems include, among others, graphs, lattices (e.g., in inductive logic programming [11], formal concept analysis [5], and itemset mining [12]), and finite point sets. For space limitations, we consider only two such applications in this short version[2]. The first one, discussed in Section 5.1, is concerned with the separation of finite point sets in $\mathbb{R}^d$. For this task, we compare Alg. 1 to the greedy algorithm in [14] as well as to ordinary SVM on synthetic datasets. The other application described in Section 5.2 deals with vertex classification in random trees and graphs of different sizes and edge densities. For this task, we compare the predictive performance of our algorithm to that of the greedy algorithm in [14].

5.1 Binary Classification in Finite Point Sets

In this section we consider point separation in MLCSs over finite subsets of $\mathbb{R}^d$. The closure systems used in these experiments are given by the traces of convex hulls as defined by (1) in Section 2; the linkage function by means of the Euclidean distance. Our experimental results reported below show that the predictive performance of maximum margin separation in this kind of closure systems is comparable to that of SVM and that it outperforms the greedy separation algorithm in [14] on finite synthetic point sets in $\mathbb{R}^2, \mathbb{R}^3$ and $\mathbb{R}^4$ that are half-space separable. All datasets consist of two blobs, each with 500 points, such that the two classes are half-space separable[3]. In addition to the quantitative results below, for one of the random datasets from $\mathbb{R}^2$ we visualize the output obtained by the three algorithms (see Fig. 3). We selected three (in accordance to the VC-dimension of half-spaces in $\mathbb{R}^2$) random training examples for each

[2] The source code and the data used in the experiments reported in this section are available at https://github.com/fseiffarth/MaxMarginSeparations.

[3] For a detailed description of these synthetic datasets, the reader is referred to [14].

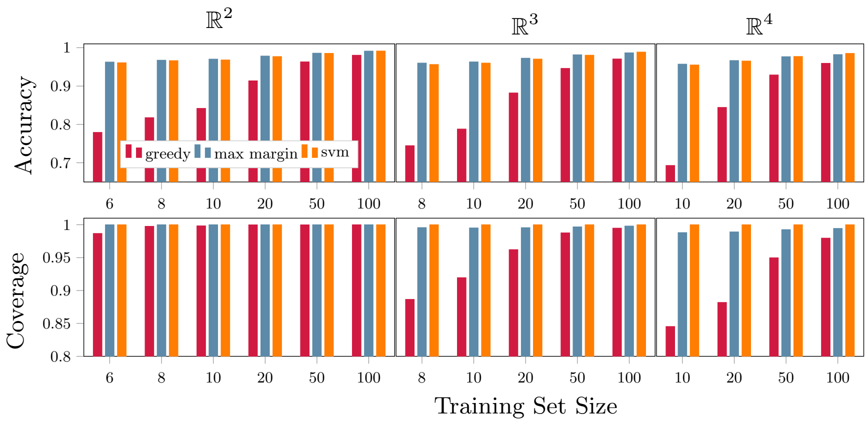


Fig. 4. Accuracy and coverage of greedy separation, maximum margin separation, and SVM for point set classification in $\mathbb{R}^2, \mathbb{R}^3, \mathbb{R}^4$.

class (denoted by dark blue resp. dark red). The class labels are indicated by light blue and light red. The predictions are given by the convex hulls for the two greedy algorithms and by the separating hyperplane for SVM.

For each of the training set sizes (see the values of the x-axes of Fig. 4), we generated $1,000$ binary labeled random sets as indicated above. Fig. 4 shows the averaged accuracy (top row) and coverage (bottom row) for the three algorithms. The results obtained clearly show that maximum margin closed set separation outperforms the greedy separation algorithm in [14] in predictive performance, especially on small training set sizes. Furthermore, at least on the random datasets we used, it is also comparable to ordinary SVM by emphasizing that our definition is *not* a generalization of SVM; it is only an adaption of the idea of maximum margin separation to finite closure systems. The accuracy of the greedy algorithm strongly depends on the training set size and the dimension of the space, while the accuracy of the maximum margin algorithm is constantly above 0.9. Regarding the coverage, for which a similar behavior can be observed, note that finite point sets in $\mathbb{R}^d$ are not half-space separable by MLCSs in general. While the average coverage for the greedy algorithm drops below 0.85 in case of $\mathbb{R}^4$ and 10 training samples, the maximum margin algorithm has an average coverage above 0.95 for all training set sizes. By definition, SVM always achieve a coverage of 1.

5.2 Vertex Classification in Random Graphs

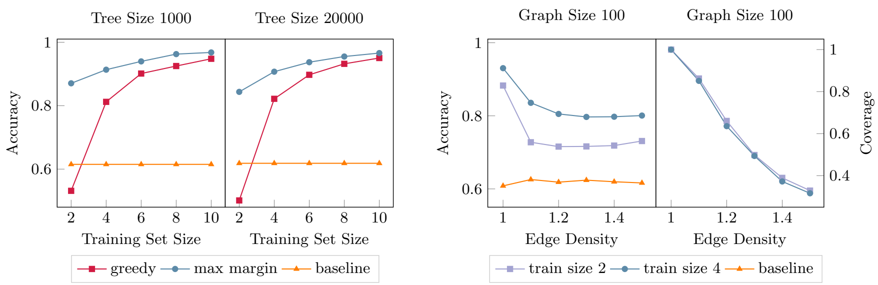
For tree and graph data we always consider the *shortest path* closure defined in (2), together with the monotone linkage function for weighted graphs as defined in Section 2. In case of graphs, we are interested in binary node predictions of random connected graphs. Of course, the distribution of the labels in the graphs plays an important role in the prediction. Clearly, in case of randomly distributed labels, it is impossible to make any acceptable prediction by MLCSs defined by the closure operator in (2). Hence, we assume the following distributions of node labels in case of trees and graphs, and analyze the predictive performance of our algorithm for different graph sizes and edge densities for the following two scenarios:

1. In case of trees, the nodes are labeled in a way that they form half-spaces, i.e., both label sets are closed and their union is the whole tree.
2. In case of graphs, we select two nodes at random and assign the labels to them. Then the labels of the other nodes are determined by their distance to these center nodes. We ensure the subgraphs induced by the same class labels to be connected and randomly flip an unbiased coin to determine the label for nodes with the same distance.

Moreover, in both cases we additionally use only graph labelings with nearly balanced class sizes, i.e., the minimum size of a class is at least 25% of the total size. In case of trees, we look at random trees of different sizes, ranging from $1,000$ to $20,000$ (see Fig. 5a). For each tree size and training sample size (see the x-axis of Fig. 5a), we generated $1,000$ binary labeled random trees in the above way. Then, for each run of the algorithm on a tree, $x/2$ training examples have been drawn at random from each of the two label sets for the input, where x is the x-axis value in Fig. 5a. For evaluation, we run the greedy algorithm from [14] and the maximum margin closed set separation algorithm on the training sets to predict the class labels of the unseen examples. The average accuracy, over all $1,000$ random trees is displayed in Fig. 5a. As a baseline, we take the percentage defined by the majority class. Note that trees induce Kakutani closure systems and hence the coverage is always 1. One can see that with increasing training set size, the accuracy increases up to more than 0.95 in case of maximum margin separation and 10 training samples. Moreover, the maximum margin separation leads to better accuracy compared to the greedy separation, especially for small training sample sizes. Somewhat surprisingly, the tree size has no significant impact on the predictive performance.

In case of graphs with different edge densities[4], we generated $1,000$ random graphs for each edge density (see the x-axis values in Figure 5b) and assigned the nodes to one of the two classes as described above. The random graphs were generated from random trees by adding additional random edges until the required edge density has been reached. For each run of our algorithm, we selected 1 or 2 nodes from each label class at random for training such that their closures do not intersect. The accuracy results are shown in Fig. 5b. We present also the coverage values, as the underlying MLCSs are not Kakutani in general. For increasing edge density, the accuracy decreases to 0.8 in case of 4 training samples and to 0.75 in case of 2 training samples for the edge density of 1.2. For edge density 1.5, there are no obvious changes in the accuracy. This can be explained by the fact that the coverage decreases to approximately 0.38 in case of an edge density of 1.5.

[4] The edge density is the number of edges minus 1 divided by the number of nodes.



(a) Comparison of greedy algorithm and maximum margin algorithm for node prediction task in random trees.

(b) Accuracy and coverage of maximum margin separation in random graphs of different edge densities with 2 and 4 training samples.

Fig. 5. Accuracy of vertex prediction in random trees and random graphs of different sizes and edge densities.

6 Concluding Remarks

We adapted the idea of maximum margin separation in inner product feature spaces to abstract finite closure systems equipped with monotone linkage functions. Although not all properties of ordinary maximum margin separation could be preserved in this way, the anti-monotonicity property, a key feature of maximum margin separation, remains valid. Combining this concept with half-space and maximal closed set separation, we presented a simple greedy algorithm and proved that it computes a closed set separation with maximum margin correctly, using a linear number of closure operator calls and linkage function evaluations. In addition, for Kakutani closure systems the output closed sets are always complementary maximum margin half-spaces if the closures of the input sets are disjoint. Our experimental results on synthetic data clearly show that the maximum margin separation algorithm presented in this work outperforms the greedy algorithm in [14] both on point classification and vertex prediction in random trees.

We mention some interesting questions raised by this work. In contrast to ordinary SVM, the maximum margin separating half-spaces are not unique in general in Kakutani MLCSs. It would be important to *characterize* the class of Kakutani MLCSs from the point of view of *uniqueness*. In particular, how does the structure of the closure system interact with the linkage function in such a characterization, if it exists at all. Another important issue is the complexity of the maximum margin separation algorithm. Although the number of closure operator calls and linkage function evaluations is linear in the cardinality of the ground set, the algorithm is practically infeasible for MLCSs over very large ground sets (e.g., the set of vertices of the web graph). The question towards this direction is therefore to identify practically interesting classes of MLCSs for which the algorithm has sublinear complexity. For example, one can show that the greedy algorithm in [14] requires only a logarithmic number of closure operator calls for closure systems over finite lattices, as the closed sets in this kind of set systems have a very succinct representation. Last but not least, is it possible to improve the complexity of the algorithm presented in this work by relaxing the problem settings, e.g., by allowing approximate solutions.

Acknowledgments

This work was partly supported by the Ministry of Education and Research of Germany (BMBF) under project ML2R (grant number 01/S18038C) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2070 - 390732324.

References

1. Boser, B.E., Guyon, I.M., Vapnik, V.N.: Training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory. pp. 144–152. COLT '92, ACM, New York, NY, USA (1992)
2. Chepoi, V.: Separation of two convex sets in convexity structures. Journal of Geometry **50**(1-2), 30–51 (1994)
3. Davey, B.A., Priestley, H.A.: Introduction to Lattices and Order, Second Edition. Cambridge University Press (2002)
4. Farber, M., Jamison, R.E.: Convexity in Graphs and Hypergraphs. SIAM Journal on Algebraic Discrete Methods **7**(3), 433–444 (1986)
5. Ganter, B., Stumme, G., Wille, R.: Formal Concept Analysis, Foundations and Applications, vol. 3626. Springer-Verlag, Berlin, Heidelberg (2005)
6. Kempner, Y., Mirkin, B., Muchnik, I.: Monotone linkage clustering and quasi-concave set functions. Applied Mathematics Letters **10**(4), 19–24 (1997)
7. Kempner, Y., Levit, V.E.: Duality between quasi-concave functions and monotone linkage functions. Discrete Mathematics **310**(22), 3211–3218 (2010)
8. Kempner, Y., Muchnik, I.: Clustering on antimatroids and convex geometries. WSEAS Transactions on Mathematics **2**(1), 54–59 (2003)
9. Kubiś, W.: Separation properties of convexity spaces. Journal of Geometry **74**(1-2), 110–119 (2002)
10. Mullat, J.E.: Extremal Subsystems of Monotonic Systems I. Avtomatica i Telemekhanika **5**, 130–139 (1976)
11. Nienhuys-Cheng, S., de Wolf, R. (eds.): Foundations of Inductive Logic Programming, Lecture Notes in Computer Science, vol. 1228. Springer (1997)
12. Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Efficient mining of association rules using closed itemset lattices. Inf. Syst. **24**(1), 25–46 (1999)
13. Rosenblatt, F.: The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review **65**(6), 386–408 (1958)
14. Seiffarth, F., Horváth, T., Wrobel, S.: Maximal closed set and half-space separations in finite closure systems. In: Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2019. LNCS, vol. 11906, pp. 21–37. Springer (2019)
15. van de Vel, M.: Binary convexities and distributive lattices. Proceedings of the London Mathematical Society **s3-48**(1), 1–33 (1984)