Maximal Closed Set and Half-Space Separations in Finite Closure Systems

Florian Seiffarth, Tamás Horváth, Stefan Wrobel



IAIS









classical machine learning problem (Rosenblatt, 1958), well-understood



- classical machine learning problem (Rosenblatt, 1958), well-understood
- not unique

Florian Seiffarth, Tamás Horváth, Stefan Wrobel

Maximal Closed Set and Half-Space Separations in Finite Closure Systems

Definition:

- set system (E, F) : set E with F ⊆ 2^E
- *finite* set system: $|E| < \infty$

Definition:

- set system (E, F) : set E with F ⊆ 2^E
- *finite* set system: $|E| < \infty$

Motivation:

Definition:

- set system (E, F) : set E with F ⊆ 2^E
- *finite* set system: $|E| < \infty$

Motivation:

• theoretical:

Definition:

 set system (E, F) : set E with F ⊆ 2^E

• *finite* set system: $|E| < \infty$

Motivation:

Theorem (Kakutani):

Definition:

- set system (E, F) : set E with F ⊆ 2^E
- *finite* set system: $|E| < \infty$

Motivation:

Theorem (Kakutani):



Definition:

- set system (E, F) : set E with F ⊆ 2^E
- *finite* set system: $|E| < \infty$

Motivation:

Theorem (Kakutani):



Definition:

- set system (E, F) : set E with F ⊆ 2^E
- *finite* set system: $|E| < \infty$

Motivation:

Theorem (Kakutani):

• theoretical: Two sets in \mathbb{R}^d are **separable** by a hyper-plane \iff their convex hulls are **disjoint**



Florian Seiffarth, Tamás Horváth, Stefan Wrobel

Maximal Closed Set and Half-Space Separations in Finite Closure Systems

Definition:

- set system (E, F) : set E with F ⊆ 2^E
- *finite* set system: $|E| < \infty$

Motivation:

Theorem (Kakutani):



Definition:

- set system (E, F) : set E with F ⊆ 2^E
- *finite* set system: $|E| < \infty$

Motivation:

Theorem (Kakutani):

• theoretical: Two sets in \mathbb{R}^d are **separable** by a hyper-plane \iff their convex hulls are **disjoint**





Florian Seiffarth, Tamás Horváth, Stefan Wrobel

Maximal Closed Set and Half-Space Separations in Finite Closure Systems

Definition:

- set system (E, F) : set E with F ⊆ 2^E
- *finite* set system: $|E| < \infty$

Motivation:

Theorem (Kakutani):

• theoretical: Two sets in \mathbb{R}^d are **separable** by a hyper-plane \iff their convex hulls are **disjoint**



practical: structured input space (examples in a few minutes)

Florian Seiffarth, Tamás Horváth, Stefan Wrobel

Maximal Closed Set and Half-Space Separations in Finite Closure Systems

What is what:

What is what:

 $\mathbb{R}^d \longrightarrow$

What is what:

 $\mathbb{R}^d \longrightarrow E \text{ (ground set)}$

What is what:

 \mathbb{R}^{d} E (ground set) \rightarrow \rightarrow

convexity

What is what:

 $\mathbb{R}^d \longrightarrow E \text{ (ground set)}$

convexity \rightarrow closure operator

Definition:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

What is what:

 \mathbb{R}^d E (ground set) \rightarrow convexity \rightarrow

convex hull \rightarrow

closure operator

Definition:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

What is what:

 $\mathbb{R}^d \longrightarrow E \text{ (ground set)}$ convexity \rightarrow closure operator

 $\texttt{convex hull} \quad \rightarrow \quad \texttt{closed set}$

Definition:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

What is what:

 \mathbb{R}^d E (ground set) \rightarrow

- convexity closure operator \rightarrow
- convex hull closed set \rightarrow
- all convex hulls \rightarrow

Definition:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

What is what:

\mathbb{R}^{d}	\rightarrow	E (ground set)
convexity	\rightarrow	closure operator
convex hull	\rightarrow	closed set

all convex hulls \rightarrow closure system

Definition:

closure operator *c* over *E*:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

Definition:

- $\emptyset, E \in C$
- $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$

What is what:

\mathbb{R}^{d}	\rightarrow	E (ground set)
convexity	\rightarrow	closure operator
convex hull	\rightarrow	closed set
all convex hulls	\rightarrow	closure system
half-space	\rightarrow	

closure operator c over E:

Definition:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

Definition:

- $\emptyset, E \in \mathcal{C}$
- $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$

What is what:

 \mathbb{R}^{d}

- convexity
- convex hull
- all convex hulls
- half-space

- E (ground set)
- \rightarrow closure operator
- \rightarrow closed set
- \rightarrow closure system
- \rightarrow closed set with closed complement

Definition:

closure operator *c* over *E*:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

Definition:

- $\emptyset, E \in \mathcal{C}$
- $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$

What is what:

R^d

convexity

convex hull

- all convex hulls
- half-space
- hyper-plane

- E (ground set)
- \rightarrow closure operator
- \rightarrow closed set

 \rightarrow

- \rightarrow closure system
- \rightarrow closed set with closed complement

Definition:

closure operator *c* over *E*:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

Definition:

- $\emptyset, E \in \mathcal{C}$
- $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$

What is what:

 \mathbb{R}^d

convexity

convex hull

- all convex hulls
- half-space
- hyper-plane

- E (ground set)
- \rightarrow closure operator
- \rightarrow closed set
- \rightarrow closure system
- \rightarrow closed set with closed complement
- \rightarrow no correspondance

Definition:

closure operator *c* over *E*:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

Definition:

- $\emptyset, E \in \mathcal{C}$
- $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$

What is what:

₽Rd convexity \rightarrow convex hull

- all convex hulls
- half-space hyper-plane

- E (ground set)
- closure operator
- closed set \rightarrow
- closure system \rightarrow
- closed set with closed complement \rightarrow
- no correspondance \rightarrow

Definition:

closure operator c over E:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

Definition:

closure system (E, C):

- Ø, *E* ∈ *C*
- $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$

Questions:

What is what:

₽Rd convexity \rightarrow

- convex hull
- all convex hulls
- half-space hyper-plane

- E (ground set)
- closure operator
- closed set \rightarrow
- closure system \rightarrow
- closed set with closed complement \rightarrow
- no correspondance \rightarrow

Definition:

closure operator c over E:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

Definition:

closure system (E, C):

- Ø, *E* ∈ *C*
- $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$

Questions:

separation algorithm

What is what:

₽Rd convexity \rightarrow

- convex hull
- all convex hulls
- half-space hyper-plane

- E (ground set)
- closure operator
- closed set \rightarrow
- closure system \rightarrow
- closed set with closed complement \rightarrow
- no correspondance \rightarrow

Definition:

closure operator c over E:

- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- c(c(A)) = c(A)

Definition:

closure system (E, C):

- Ø, *E* ∈ *C*
- $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$

Questions:

- separation algorithm
- classification algorithm

Florian Seiffarth, Tamás Horváth, Stefan Wrobel

Some Practical Motivations:

Some Practical Motivations: Graphs
Some Practical Motivations: Graphs



Trees

Florian Seiffarth, Tamás Horváth, Stefan Wrobel

Some Practical Motivations: Graphs





Trees

Graphs (e.g. molecule, social graph, ...)

Some Practical Motivations: Graphs





Trees

Graphs (e.g. molecule, social graph, ...)

requires some semantically meaningful definition of closure system/operator

• A node set X is **closed** if it contains all shortest paths between nodes in X.

- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.

- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.



- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.



- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.



- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.



- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.



- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.



- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.



- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.



- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.



- A node set X is **closed** if it contains all shortest paths between nodes in X.
- The **closure** c(Y) is the smallest closed set containing Y.



Some Practical Motivations: Lattices



Some Practical Motivations: Lattices



requires some semantically meaningful definition of closure system/operator

• A set X is **closed** if it contains all elements between the meet (glb) and the join (lub) of X.

- A set X is **closed** if it contains all elements between the meet (glb) and the join (lub) of X.
- The **closure** c(Y) is the smallest closed set containing Y.

- A set X is **closed** if it contains all elements between the meet (glb) and the join (lub) of X.
- The **closure** c(Y) is the smallest closed set containing Y.



- A set X is **closed** if it contains all elements between the meet (glb) and the join (lub) of X.
- The **closure** c(Y) is the smallest closed set containing Y.



Florian Seiffarth, Tamás Horváth, Stefan Wrobel

- A set X is **closed** if it contains all elements between the meet (glb) and the join (lub) of X.
- The **closure** c(Y) is the smallest closed set containing Y.



Half-Space-Separation Problem:

Given a closure system (E, C) and two sets $A, B \subseteq E$, **decide** if they are half-space separable in (E, C).

Half-Space-Separation Problem:

Given a closure system (E, C) and two sets $A, B \subseteq E$, **decide** if they are half-space separable in (E, C).

Remark: the closure system is given implicitly via a closure operator.

Florian Seiffarth, Tamás Horváth, Stefan Wrobel Maximal Closed Set and Half-Space Separations in Finite Closure Systems

Half-Space-Separation Problem:

Given a closure system (E, C) and two sets $A, B \subseteq E$, **decide** if they are half-space separable in (E, C).



Remark: the closure system is given implicitly via a closure operator.

Florian Seiffarth, Tamás Horváth, Stefan Wrobel

Half-Space-Separation Problem:

Given a closure system (E, C) and two sets $A, B \subseteq E$, **decide** if they are half-space separable in (E, C).



Remark: the closure system is given implicitly via a closure operator.

Florian Seiffarth, Tamás Horváth, Stefan Wrobel

Half-Space-Separation Problem:

Given a closure system (E, C) and two sets $A, B \subseteq E$, **decide** if they are half-space separable in (E, C).



Remark: the closure system is given implicitly via a closure operator.

Florian Seiffarth, Tamás Horváth, Stefan Wrobel

Some Negative Results

• i.e., there are disjoint closed sets which have no half-space separation

• i.e., there are disjoint closed sets which have no half-space separation

Theorem: The Half-Space-Separation problem is NP-complete

• i.e., there are disjoint closed sets which have no half-space separation

Theorem: The Half-Space-Separation problem is NP-complete

Two approaches to overcome:

• i.e., there are disjoint closed sets which have no half-space separation

Theorem: The Half-Space-Separation problem is NP-complete

Two approaches to overcome:

- problem relaxation \rightarrow find **maximal** separating closed sets
 - caution: maximal is not maximum

• i.e., there are disjoint closed sets which have no half-space separation

Theorem: The Half-Space-Separation problem is NP-complete

Two approaches to overcome:

- problem relaxation \rightarrow find **maximal** separating closed sets
 - caution: maximal is not maximum
- resort to Kakutani set systems \rightarrow (set systems where Kakutani theorem holds)
Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm

Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm

Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm

Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm

• very generic algorithm: requires only the ground set and the closure operator



Cannot added to blue because of intersection

Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm



Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm

• very generic algorithm: requires only the ground set and the closure operator



Cannot added to red because of intersection

Maximal-Closed-Set-Separation Problem:

Given a closure system (E, C) and sets $A, B \subseteq E$, find **maximal** disjoint closed sets separating A and B; print "No" if they do not exist

Our Solution: simple greedy algorithm

• very generic algorithm: requires only the ground set and the closure operator



Florian Seiffarth, Tamás Horváth, Stefan Wrobel

Maximal Closed Set and Half-Space Separations in Finite Closure Systems

Theorem: It solves the Maximal Closed Set Separation problem by calling the closure operator at most 2|E| - 2 times.

Theorem: It solves the Maximal Closed Set Separation problem by calling the closure operator at most 2|E| - 2 times.

Theorem: It is **optimal** in the number of closure operator calls.

Theorem: It solves the Maximal Closed Set Separation problem by calling the closure operator at most 2|E| - 2 times.

Theorem: It is **optimal** in the number of closure operator calls.

Theorem: It provides an **algorithmic characterization** of Kakutani closure systems.

Experimental Results on Trees



classification results of the maximal closed set separation algorithm on trees

This work: Generalization of half-space separation in ℝ^d to finite closure systems.
NP-complete

- NP-complete
- two problem relaxations:

- NP-complete
- two problem relaxations:
 - Kakutani closure systems (e.g. trees, distributive lattices with applications to ILP (long version))

- NP-complete
- two problem relaxations:
 - Kakutani closure systems (e.g. trees, distributive lattices with applications to ILP (long version))
 - maximal closed set separation

- NP-complete
- two problem relaxations:
 - Kakutani closure systems (e.g. trees, distributive lattices with applications to ILP (long version))
 - maximal closed set separation
- maximal closed set separation problem: general simple greedy algorithm

- NP-complete
- two problem relaxations:
 - Kakutani closure systems (e.g. trees, distributive lattices with applications to ILP (long version))
 - maximal closed set separation
- maximal closed set separation problem: general simple greedy algorithm
 - no specific structure is utilized

- NP-complete
- two problem relaxations:
 - Kakutani closure systems (e.g. trees, distributive lattices with applications to ILP (long version))
 - maximal closed set separation
- maximal closed set separation problem: general simple greedy algorithm
 - no specific structure is utilized
 - optimal

- NP-complete
- two problem relaxations:
 - Kakutani closure systems (e.g. trees, distributive lattices with applications to ILP (long version))
 - maximal closed set separation
- maximal closed set separation problem: general simple greedy algorithm
 - no specific structure is utilized
 - optimal
 - · excellent experimental results on trees

• further applications of the special case of (distributive) lattices

- further applications of the special case of (distributive) lattices
 - long version . . .

- further applications of the special case of (distributive) lattices
 - long version . . .
- enriching the story with distances, hyper-planes etc.

- further applications of the special case of (distributive) lattices
 - long version . . .
- enriching the story with distances, hyper-planes etc.
 - need appropriate generalizations

- further applications of the special case of (distributive) lattices
 - long version . . .
- enriching the story with distances, hyper-planes etc.
 - need appropriate generalizations
- unique half-space separation (like SVM)

- further applications of the special case of (distributive) lattices
 - Iong version ...
- enriching the story with distances, hyper-planes etc.
 - need appropriate generalizations
- unique half-space separation (like SVM)
- utilization of the VC-dimension of finite closure systems