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

Classification in Finite Closure Systems

Florian Seiffarth[1] — seiffarth@cs.uni-bonn.de
Tamás Horváth[1,2,3] — horvath@cs.uni-bonn.de
Stefan Wrobel[1,2,3] — wrobel@cs.uni-bonn.de

[1] Dept. of Computer Science, University of Bonn
[2] Fraunhofer IAIS
[3] Fraunhofer Center for Machine Learning

Basic Definitions

Set system

$(E, \mathcal{F})$ with $\mathcal{F} \subseteq 2^E$
- *finite* if $|E| < \infty$

Closure system

set system $(E, \mathcal{C})$ with:
- $\emptyset, E \in \mathcal{C}$
- $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$

Closure operator

function $c : E \to E$ with:
- $A \subseteq c(A)$
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$
- $c(c(A)) = c(A)$

Half-space in $(E, \mathcal{C})$

$H \in \mathcal{C}$ with $H^c \in \mathcal{C}$

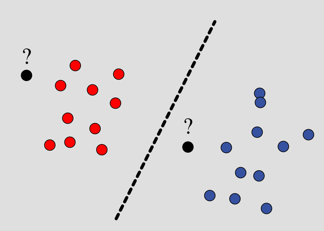
Half-space separation of $A, B \subseteq E$ in $(E, \mathcal{C})$

half-space $H \in \mathcal{C}$ with $A \subseteq H$ and $H \cap B = \emptyset$

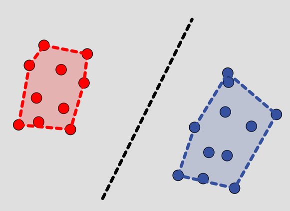
Problem setting

Classical problem: Separation in $\mathbb{R}^d$

Hyper-Plane Separation Problem: Given two sets $A, B \subseteq \mathbb{R}^d$, **find** a separating hyper-plane.



Theorem (Kakutani, 1937): Two sets in $\mathbb{R}^d$ are **separable** by a hyper-plane iff their convex hulls are **disjoint**.

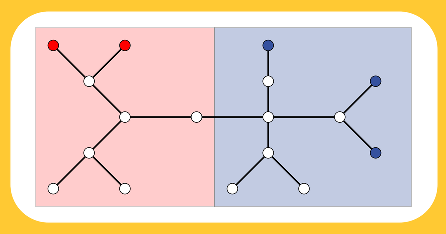


Separation in Finite closure systems

Half-Space Separation Problem: Given a closure system $(E, \mathcal{C})$ and sets $A, B \subseteq E$, **decide** if A and B are half-space separable in $(E, \mathcal{C})$.

Motivation:
- general concept
- structured input
- interesting examples

e.g. trees

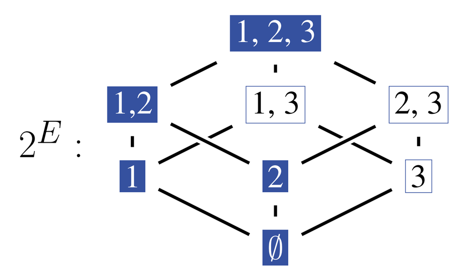


Problem: Kakutani's theorem does not hold! Example:

$E = \{1, 2, 3\}$

$\mathcal{C} = \{\emptyset, \{1\}, \{2\}, \{1, 2\}, \{1, 2, 3\}\}$

2^E :



There are **no** half-spaces separating the disjoint closed sets {1} and {2}.

Results

☹ **Theorem:** The Half-Space Separation Problem is **NP-complete**. ☹

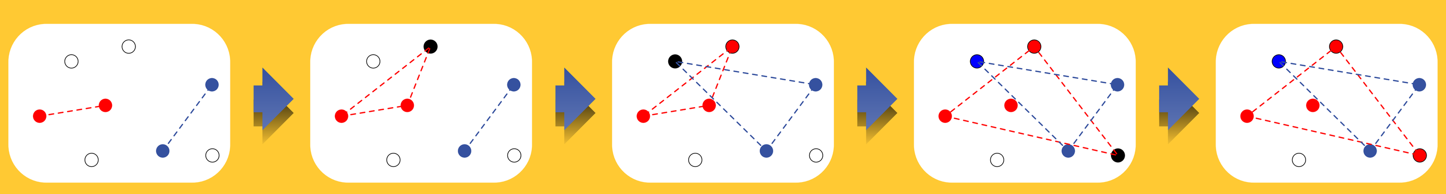
To overcome the negative result: Two approaches

simplify problem → | special closure systems →

Maximal Closed Set Separation Problem

Maximal Closed Set Separation Problem: Given a closure system $(E, \mathcal{C})$ and sets $A, B \subseteq E$, **find** maximal disjoint closed supersets of A and B.
- **caution:** maximal and not maximum

Solution: A *Simple Greedy Algorithm* processing the elements one by one.



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

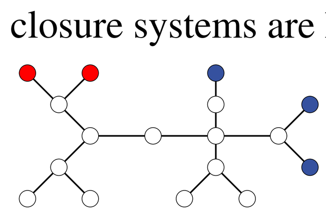
Theorem: The greedy algorithm is **optimal** w.r.t. the number of closure operator calls.

Kakutani Closure Systems

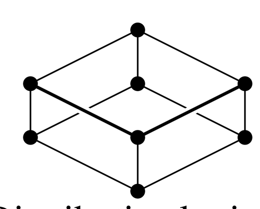
Definition: A closure system $(E, \mathcal{C})$ is Kakutani if *any* two disjoint closed sets are half-space separable.

☹ **Theorem:** Any algorithm requires in general $\Omega\left(2^{|E|/2}\right)$ closure operator calls to decide if a closure system is Kakutani.

☺ Several closure systems are known to be Kakutani:



Trees



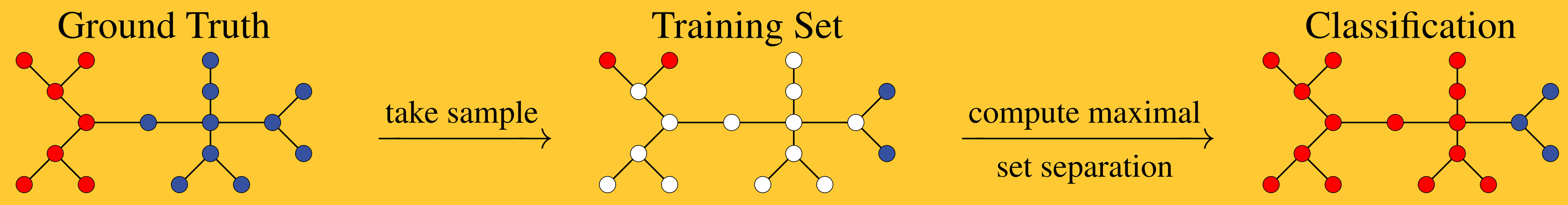
Distributive lattices

Theorem: The greedy algorithm (LHS) provides an **algorithmic characterization** of Kakutani closure systems.

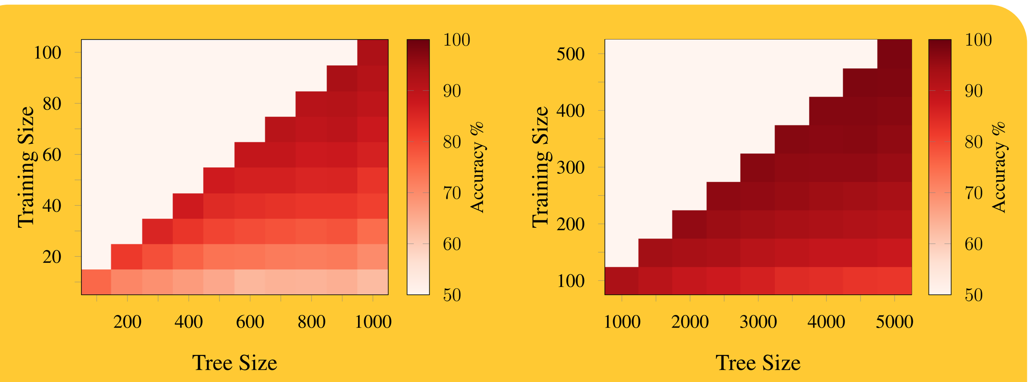
use algorithm for classification → | ← classification results

Experiments

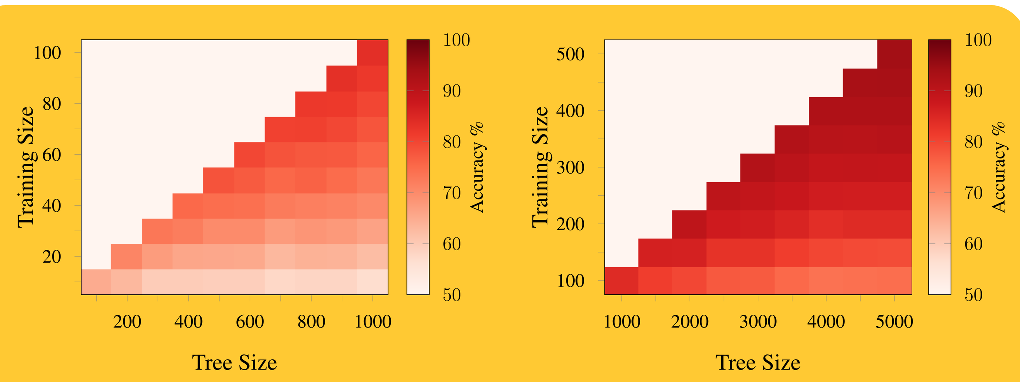
Classification Algorithm for Trees:



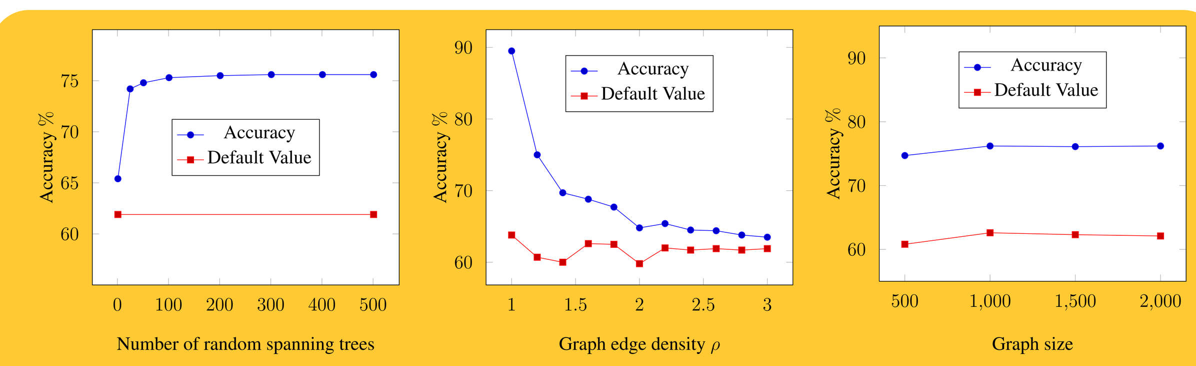
→ compare with ground truth



half-space separable vertex sets over trees



non half-space separable vertex sets over trees



vertex classification over **arbitrary** graphs