# Depth Functions for Partial Orders with a Descriptive Analysis of Machine Learning Algorithms

Hannah Blocher, Georg Schollmeyer, Christoph Jansen, Malte Nalenz

*Department of Statistics, Ludwig-Maximilians-Universität Munich*

hannah.blocher@stat.uni-muenchen.de

## General Situation

- A *sample of partial orders* (e.g. compare the performance of ml algorithms on a fixed data set w.r.t multiple performance measures at once).
- Each partial order is indeed a precise observation. Thus, items are allowed to be incomparable.

$\Rightarrow$ Partial orders as a special case of *non-standard data*.

## Contribution

- Introducing the ufg depth function (gives a center-outward order) to obtain a description of the distribution of partial orders ($\mathcal{P}$) based on a sample.
- Explicitly addressing incomparability in the data description.

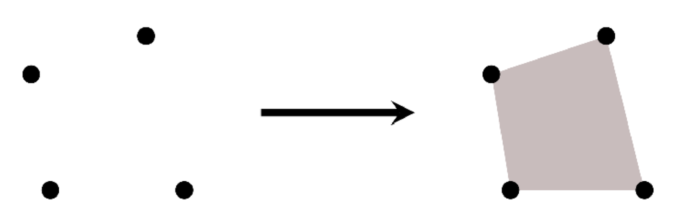
## Idea: Adaptation of the Simplicial Depth

*Depth functions* measure the centrality and outlingness of a data point with respect to a data cloud or an underlying distribution.

### Define the Closure Operator/System

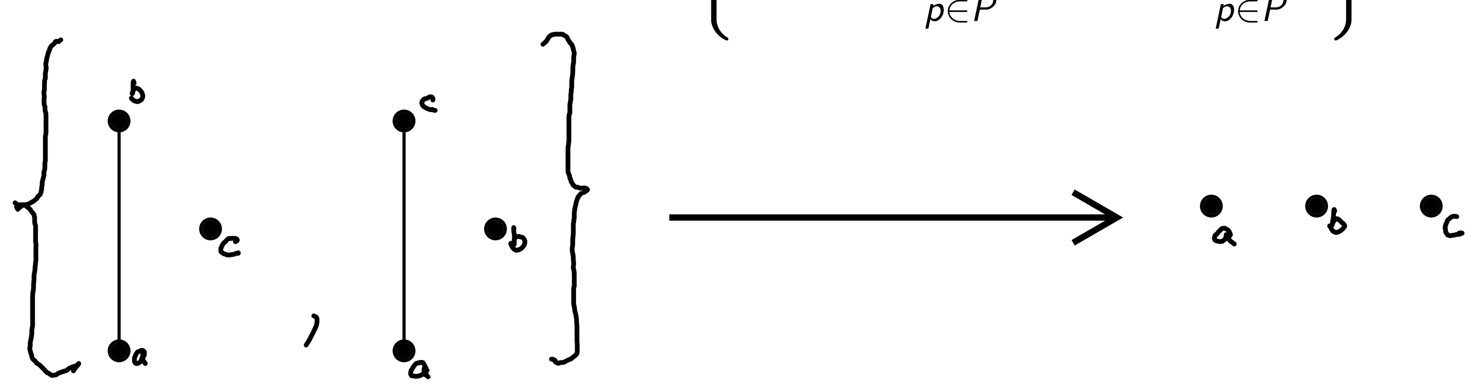
**Simplicial Depth**

$$\gamma_{\mathbb{R}^d}\colon 2^{\mathbb{R}^d} \to 2^{\mathbb{R}^d};\ A \mapsto \left\{ x\in\mathbb{R}^d \,\middle|\, \begin{array}{l} x=\sum_{i=1}^k \lambda_i a_i \text{ with } a_i\in A,\\ \lambda_i\in[0,1], \sum_{i=1}^k\lambda_i=1, k\in\mathbb{N}\end{array}\right\}$$



**Union-Free Generic Depth**

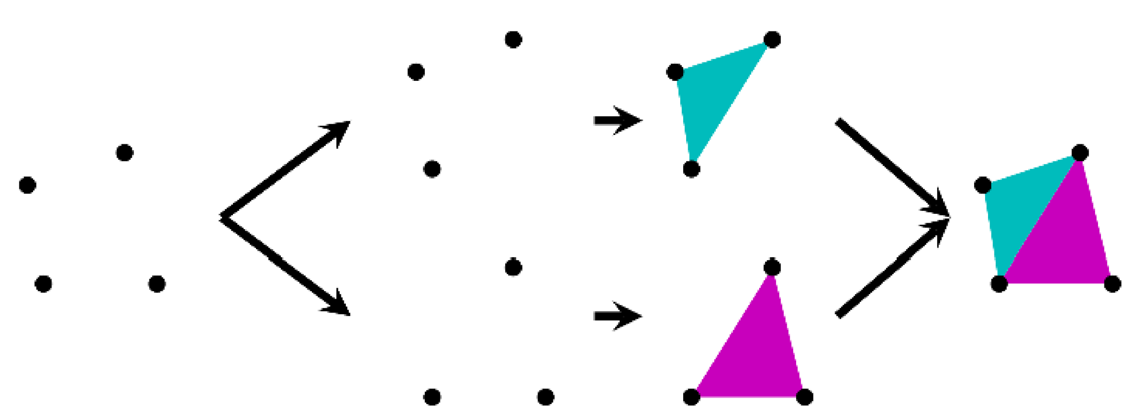
$$\gamma\colon 2^{\mathcal{P}}\to 2^{\mathcal{P}};\ P\mapsto \left\{p\in\mathcal{P} \mid \bigcap_{\tilde p\in P}\tilde p \subseteq p\subseteq \bigcup_{\tilde p\in P}\tilde p\right\}.$$




### Reduce the Input Set (It is Still Sufficient to Describe the Closure Operator)

**Simplicial Depth**

$$\mathscr{S} = \{\{a,b,c\}\subseteq\mathbb{R}\}$$

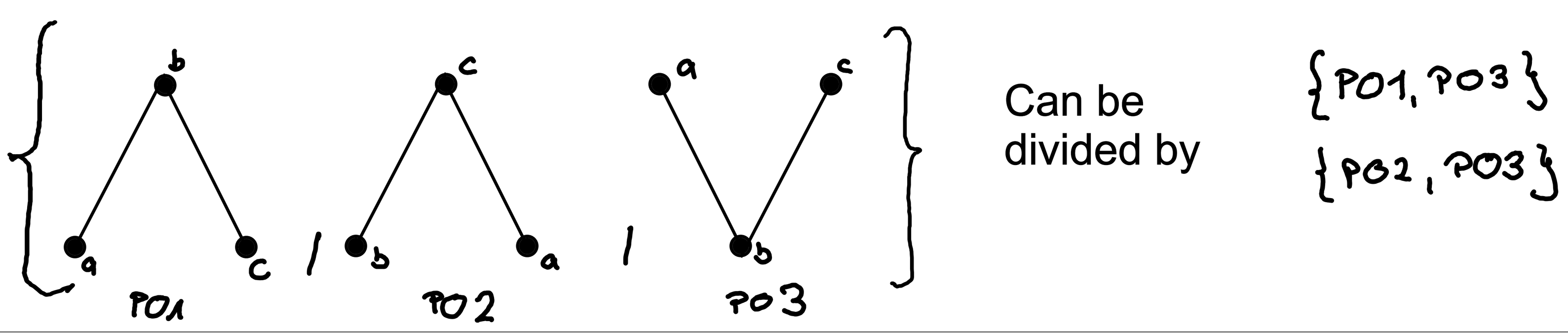


**Union-Free Generic Depth**

$$\mathscr{S} = \{P\subseteq\mathcal{P} \mid \text{Condition } (C1) \text{ and } (C2) \text{ hold }\}$$

with

- (C1) $P\subsetneq\gamma(P)$,
- (C2) There does not exist a family $(A_i)_{i\in\{1,\dots,\ell\}}$ such that for all $i\in\{1,\dots,\ell\}$, $A_i\subsetneq P$ and $\bigcup_{i\in\{1,\dots,\ell\}}\gamma(A_i)=\gamma(P)$.




### Define the Depth Function ($\mathcal{M}$ is a Set of Probability Measures)

**Simplicial Depth**

$$D\colon \begin{array}{l}\mathbb{R}^d\times\mathcal{M}\to[0,1],\\ (x,\nu)\mapsto \nu(x\in\gamma_{\mathbb{R}^d}\{X_1,\dots,X_{d+1}\}),\end{array}$$

**Union-Free Generic Depth**

$$D\colon \begin{array}{l}\mathcal{P}\times\mathcal{M}\to[0,1]\\ (p,\nu)\mapsto \begin{cases} 0, & \text{if for all } S\in\mathscr{S}\colon \prod_{\tilde p\in S}\nu(\tilde p)=0\\ c\sum_{S\in\mathscr{S}\colon p\in\gamma(S)}\prod_{\tilde p\in S}\nu(\tilde p), & \text{else,}\end{cases}\end{array}$$

## Comparison of Machine Learning Algorithms

For *each data set* we *compare a set of ml algorithms* based on *multi-dimensional performance measures*. This leads to a partial order for every data set.

- Data Sets: 80 classification problems from OpenML.
- ML Algorithms: Random Forests (RF), Decision Tree (CART), Logistic regression (LR), L1-penalized logistic regression (Lasso) and k-nearest neighbours(KNN).
- Performance Measures: area under the curve, F-score, predictive accuracy and Brier score.

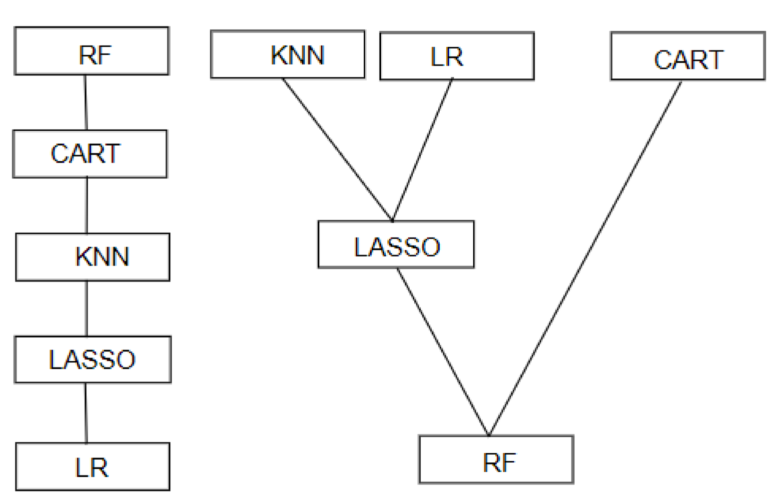



**Figure 1:** *Observed poset with maximal (left) and minimal (right) ufg depth.*

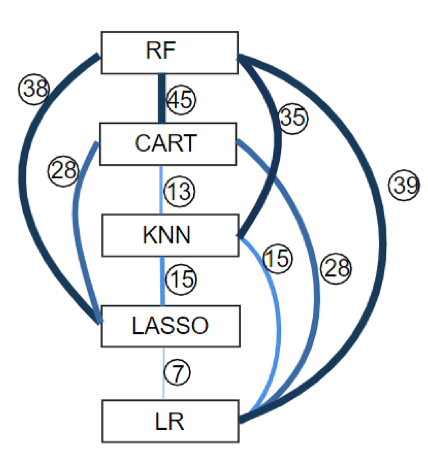



**Figure 2:** *Represents what the observed posets with the k highest depth values have in common.*

## Outlook

- Other types of non-standard data
- Statistical Inference
- Other ML problems and criteria
- ...

## References

- Blocher, Schollmeyer, Jansen, Nalenz (2023): Depth Functions for Partial Orders with a Descriptive Analysis of Machine Learning Algorithms. *Forthcoming in: ISIPTA '23.*
- Blocher, Schollmeyer, Jansen (2022): Statistical models for partial orders based on data depth and formal concept analysis. In: Ciucci, D.; Couso, I.; Medina, J.; Slezak, D.; Petturiti, D.; Bouchon-Meunier, B.; Yager, R.R. (eds): IPMU *Communications in Computer and Information Science*, vol 1602, Springer.