Relational methods for statistical analysis and decision making in the context of non-standard data- and information structures

Habilitation defense, Georg Schollmeyer

19.01.2024

This (concerning §12(1) Nr.2) is joint work with ... (in alphabetical order) ...

Thomas Augustin (TA), Jean Baccelli (JB), Hannah Blocher (HB), Christoph Jansen (CJ) and Malte Nalenz (MN)

Three parts:

A: Relational data analysis for non-standard data;

B: Decision making under weakly structured information;

C: Analysis of deficient data

The additional 3 papers concerning §12(1) Nr.1 are joint work with ... (in alphabetical order) ...

Thomas Augustin (TA), Hannah Blocher (HB), Scott Ferson (SF), Christoph Jansen (CJ), Malte Nalenz (MN), Julian Rodemann (JR) and Krasymyr Tretiak (KT)

1 Statistical models for partial orders based on data depth and formal concept analysis (HB, GS, CJ) (2022, COMM COM INF SC)

2 Data depth functions for non-standard data by use of formal concept analysis (HB, GS) (2023, under review)

3 Depth functions for partial orders with a descriptive analysis of machine learning algorithms **(HB, GS, CJ, MN) (2023, Proc. of ISIPTA)**

4 Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences **(CJ, GS, TA) (2018, IJAR)**

5 Information efficient learning of complexly structured preferences: Elicitation procedures and their application to decision making under uncertainty **(CJ, HB, TA, GS) (2022, IJAR)**

6 Risk aversion over finite domains **(JB, GS, CJ) (2022, Theory Decis.)**

7 Computing simple bounds for regression estimates for linear regression with interval-valued covariates **(GS) (2021, Proc. of ISIPTA)**

8 A short note on the equivalence of the ontic and the epistemic view on data imprecision for the case of stochastic dominance for interval-valued data GS **(2019) (Proc. of ISPTA)**

9 **Statistical comparisons of classifiers by generalized stochastic dominance** (CJ, MN, GS, TA) (2023, JMLR)

10 **Robust statistical comparison of random variables with locally varying scale of measurement.** (CJ, GS, HB, JR, TA) (2023, PMLR)

11 **Neural network model for imprecise regression with interval dependent variables** (KT, GS, SF) (2023, Neural Netw.)

Relational data analysis

Numerical data analysis

A given set $\mathcal{O} = \{o_1, \ldots, o_m\}$ of objects (data points, statistical units) is analyzed by analyzing numerical assignments $u(o_1), \ldots, u(o_m)$.

E.g., person o_i has an income of 1200 Euro.

Relational data analysis

A given set $\mathcal{O} = \{o_1, \ldots, o_m\}$ of objects (data points, statistical units) is analyzed by either

- analyzing empirical relations R between the objects (e.g., person o_i has a higher income than person o_j, $\rightsquigarrow$ order theory) or here:
- analyzing empirical relations I between the objects and certain attributes $\mathcal{A} = \{a_1, \ldots, a_m\}$.

E.g., person o_i is male, $\rightsquigarrow$ Formal concept analysis (FCA).

The right math for relational data analysis: Formal concept analysis (FCA, Ganter and Wille [2012])

Given: **formal context (crosstable)** $\mathbb{K} := (G, M, I)$ where

- G is a set of objects,
- M is a set of attributes,
- $I \subseteq G \times M$ is a binary relation with the interpretation $(g, m) \in I$ iff object g has attribute m.
- Aim: Describe $\mathbb{K}$ with the help of so-called formal concepts.

	m_1	m_2	m_3	m_4	m_5	m_6
g_1						x
g_2	x					
g_3		x				
g_4				x		
g_5					x	
g_6			x			
g_7		x	x			x
g_8	x		x	x		
g_9	x	x			x	

Formal concept analysis (FCA)

Formal concept

Let $\mathbb{K} := (G, M, I)$ be a formal context. A pair (A, B) where $A \subseteq G$ is a set of objects and $B \subseteq M$ is a set of attributes is called a **formal concept** if

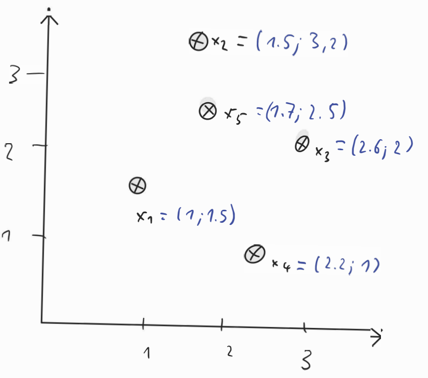
1. All objects in A have all attributes in B.
2. The set A is maximal w.r.t. the property 1.
3. The set B is maximal w.r.t. the property 1.

In such a case, we call A the **extent** and B the **intent** of the formal concept (A, B).

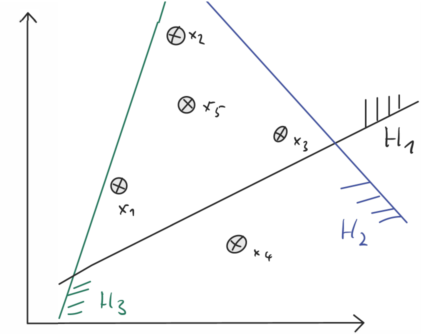
- Here, we emphasize the object sets
- The family of all concept extents, ordered by set inclusion, builds a complete lattice.

An illustrative example: Geometry

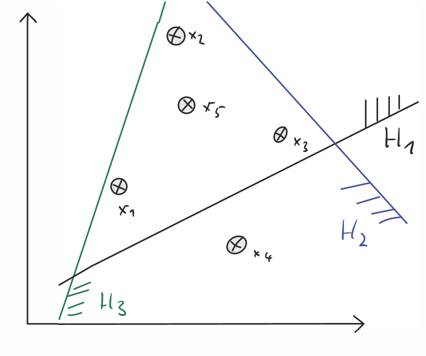
Numerical data analysis
(analytic geometry)



Relational data analysis
(synthetic geometry)



Synthetic Geometry

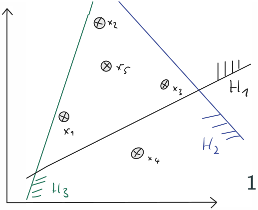


	H_1	H_2	H_3	...
x_1	×	×	×	
x_2	×	×	×	
x_3	×	×	×	
x_4		×	×	
x_5	×	×	×	
⋮				

Description languages in FCA

The structure given by a crosstable/formal context can be characterized with different tools:

1. The complete lattice of all formal concepts i.e. of all appropriate pairs of object sets and attribute sets. In pour example the object sets are the (closed) convex sets.
2. Formal object implications: A formal implication $A \longrightarrow B$ is valid in a context if the common attributes of all objects in A are also shared by the objects in B. In our example e.g., the implication $\{x_1, x_2, x_3\} \longrightarrow \{x_5\}$ is valid because x_5 lies in the convex set generated by x_1, x_2 and x_3.

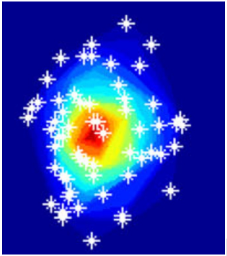


'Reminder': Data depth for $\mathbb{R}^d$

A data depth function $D(\cdot, P) : \mathbb{R}^d \longrightarrow \mathbb{R}$ measures how **deep** or **outlying** a given data point $z \in \mathbb{R}^d$ is located with respect to an observed data cloud or an assumed underlying distribution in $\mathbb{R}^d$.
It provides a **center-outward ordering** of points in $\mathbb{R}^d$.

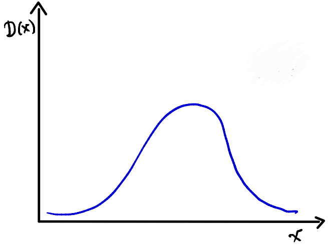
It can be used for:

- description of multivariate distributions
- outlier detection
- depth based classification and clustering
- rank and sign tests
- multivariate density estimation
- robust linear regression
- ...


[1]

'Reminder': Data depth for $\mathbb{R}^d$

- Started with Tukey's halfspace depth (Tukey [1975])
- After a bunch of further *ad hoc* depth proposals:
- Axiomatic characterization to clean up the landscape (cf., Serfling and Zuo [2000]):
- Specifying certain properties a depth function should obey, e.g., affine equivariance, monotonicity relative to deepest point, (strict) unimodality / quasiconcavity (used synonymously int the sequel)

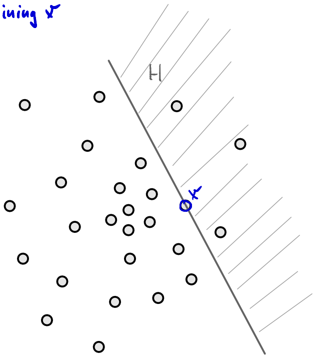


Fact:

Many data depth functions known from $\mathbb{R}^d$ can be generalized to data structures given by a formal context!

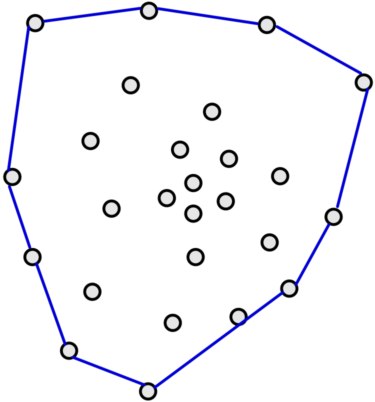
Tukey's halfspace depth

$$D(x) := \inf_{\substack{H \text{ halfspace} \\ \text{containing } x}} P(X \in H) = \dots$$

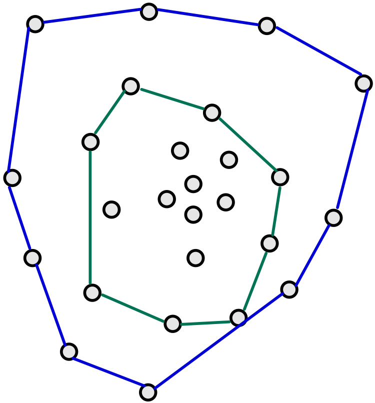


$$\dots = 1 - \sup_{\substack{C \text{ convex set} \\ \text{not contains } x}} P(X \in C)$$

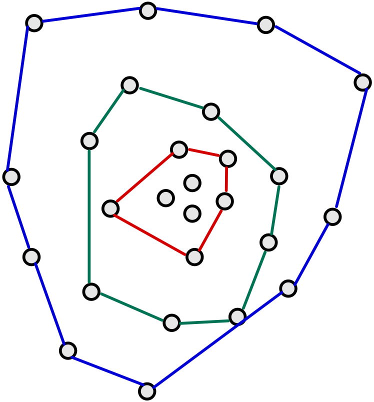
Convex hull peeling depth



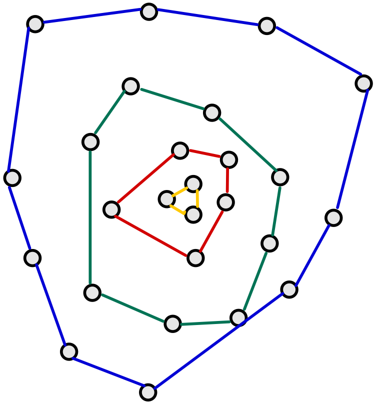
Convex hull peeling depth



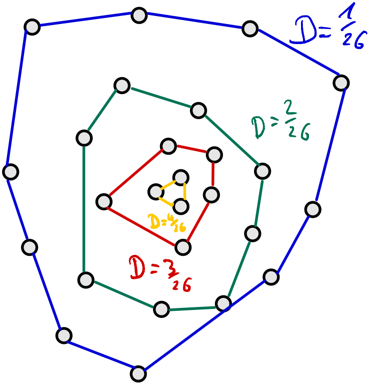
Convex hull peeling depth



Convex hull peeling depth



Convex hull peeling depth



But note:

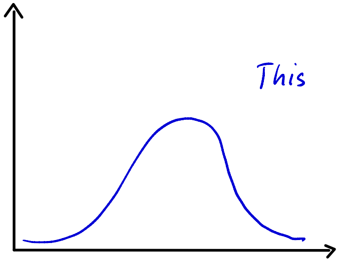
All the constructions within $\mathbb{R}^d$ become more difficult in general spaces!
(no obvious notion of a ray, (point-)symmetry, distance, infinity, etc.)

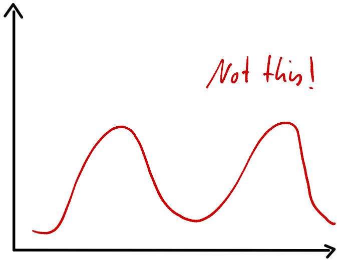
However, mathematically, things become more interesting!

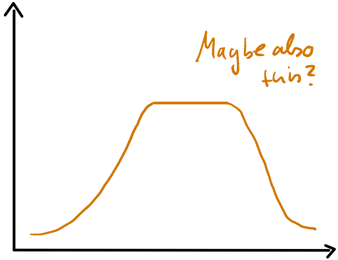
One (of many) intriguing aspects: Unimodality/Quasiconcavity

- One (possibly) desirable property of a depth function: Unimodality
- One center/mode with highest depth ...
- ... and the depth strictly decreases if one *moves away from the center*
- Two aspects:
  1. How to formalize this in higher dimensions and in more abstract spaces like in the FCA setting?
  2. How to guarantee unimodality?

Unimodality: $\mathbb{R}^1$







Unimodality: $\mathbb{R}^2$

This:

Mound grave, Cuween hill, [2]

Maybe also this:

Dike, [3]

But Not this:

Round dike, Mittelweserverband, [4]

What about this?

Star dune, [5]

Contribution 2

2 Data depth functions for non-standard data by use of formal concept analysis HB, GS (2023, under review)

Unimodality in FCA i

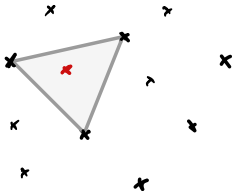
- We introduced a general FCA definition of unimodality/quasiconcavity in Contribution 2.
- This definition does already play an important role in Contribution 1.
- Contribution 2 started a systematic study in an axiomatic style inspired by the axiomatic study of data depth in $\mathbb{R}^d$: All-together 14 properties (of **representation invariance, order-preservingness, sequence behaviour and universality**) were analyzed, along with a concrete analysis of Tukey's depth w.r.t. these properties.

Unimodality in FCA ii

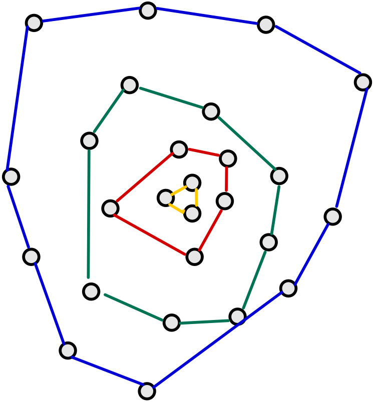
- There are certain interesting special features that only appear through the lens of FCA: One can compare data depth proposals across very different data types!
- The approach to data depth through FCA is very general. Nearly every data type can be handled (Examples: posets, mixed data structures (e.g., spatial plus ordinal plus categorical), hierarchical nominal data, etc.)

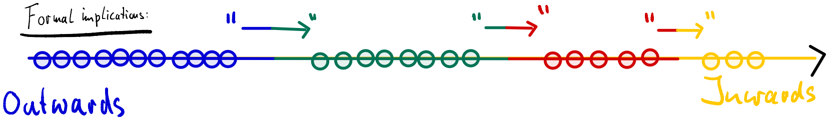
Contribution 2: Example: Unimodality in FCA: No local minima!

- Remember: We have no linear structure
- There is generally no notion of a ray
- But we can use formal object implications:
- Call D unimodal if $\{x_1, x_2, \ldots, x_k\} \longrightarrow \{y\}$ implies $D(y) \geq \min\{D(x_1), D(x_2), \ldots, D(x_k)\}$
- Call D strictly unimodal if $\{x_1, x_2, \ldots, x_k\} \longrightarrow \{y\}$ implies $D(y) > \min\{D(x_1), D(x_2), \ldots, D(x_k)\}$

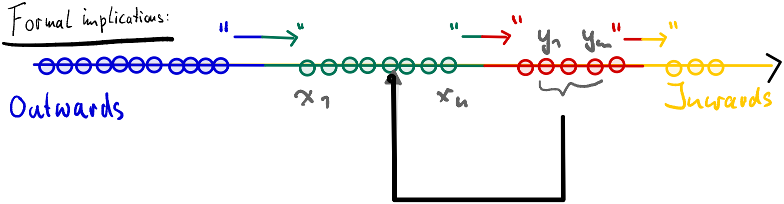


Problems with specifying a unimodal depth function





General case: This may happen



$$\{x_1,\dots,x_n\} \to \{y_1\} \Rightarrow D(y_1) > \min\{D(x_1),\dots,D(x_n)\} = D(x_k)$$

$$\{y_1,\dots,y_m\} \to \{x_k\} \Rightarrow D(x_k) > \min\{D(y_1),\dots,D(y_m)\} = D(y_1)$$

Facts:

- There are formal contexts for which there does not exist any strictly quasiconcave depth function.
- There always exist quasiconcave depth functions:
- The function $D \equiv 0$ is quasiconcave, but useless.
- Usually there are non-trivial quasiconcave depth functions.
- But they may have many ties.

'Solutions'

- Abstract solution: Rigorously define a notion of a depth function as being **as strictly quasiconcave as possible**. This is done in Contribution 2 (Universality properties, use of ideas from category theory)
- Concrete solution: Make a concrete non-trivial proposal (for a concrete data type) with presumably few ties. This is done in Contribution 1.

Contribution 1:

1 Statistical models for partial orders based on data depth and formal concept analysis (HB, GS, CJ) (2022, COMM COM INF SC)

Motivating example

- In Dittrich et al. [1998], 303 students where asked for their choices between 6 foreign universities for their semester abroad.
- Within pair comparisons, they could prefer one university over another or vice versa.
- They could also explicitly state that they have no preference between universities at all (incomparability).
- Therefore we have 303 **partial orders (posets)** as non-standard data points.

How to model poset data: The epistemic approach

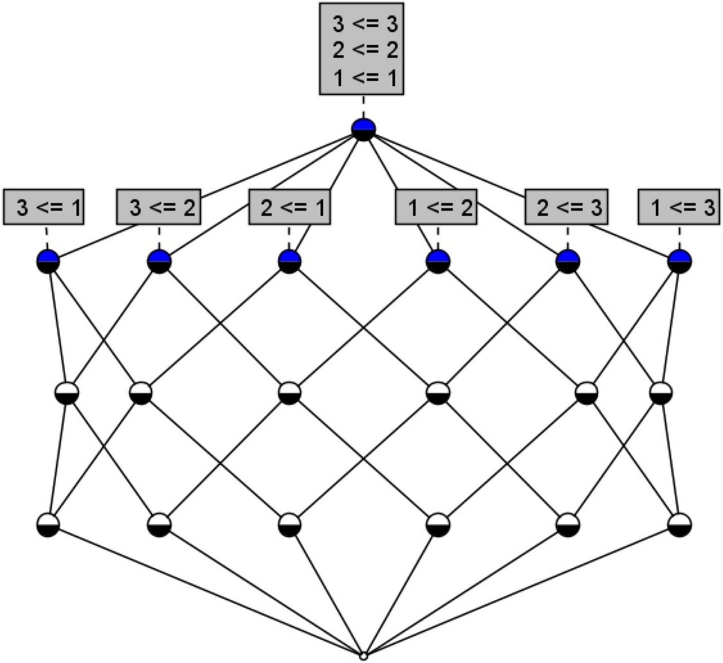
- Most approaches (e.g., Nakamura et al. [2019], Lebanon and Mao [2007]) model this by latent **total orders**, together with a coarsening process that generates the observed partial orders.
- This approach is sometimes called the *epistemic view on data imprecision* (e.g., Couso and Dubois [2014]).

How to model poset data: The ontic approach

- Here, we hold the view that a response *incomparable* is not a non-observation of a hidden comparability, but instead a factual incomparability.
- We look at the data as without any coarsening process.
- This approach is sometimes called the **ontic view on data imprecision/non-standard data.**

Relational approach

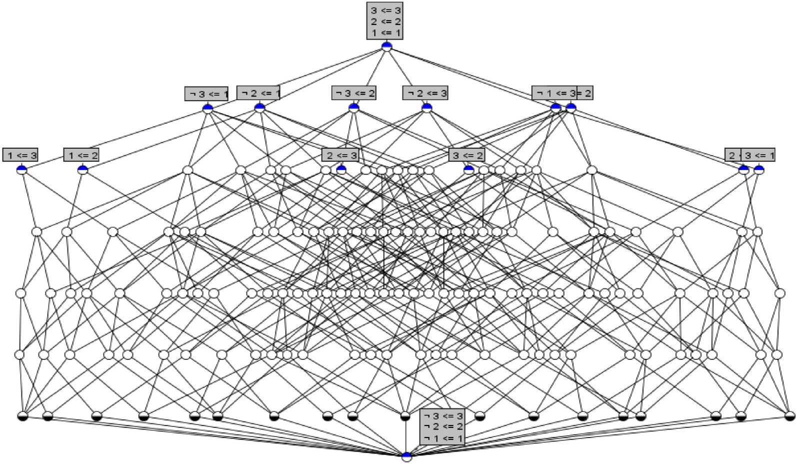
	$1 \leq 2$	$2 \leq 1$	$1 \leq 3$	$3 \leq 1$	$2 \leq 1$	$\dots$
p_1						
p_2	x					
p_3		x				
p_4				x		
p_5						
p_6			x			
p_7		x	x			
p_8	x		x	x		
$\vdots$						


3 <= 3
2 <= 2
1 <= 1
3 <= 1
3 <= 2
2 <= 1
1 <= 2
2 <= 3
1 <= 3

Relational approach: With adequate conceptual scaling* (for our purposes)

	$1 \leq 2$	$2 \leq 1$	...	$1 \not\leq 2$	$2 \not\leq 1$	...
p_1				x	x	
p_2	x				x	
p_3		x		x		
p_4				x	x	
p_5				x	x	
p_6				x	x	
p_7		x		x		
p_8	x				x	
⋮						

*:Conceptual scaling = *Dummy Coding of lattice theory*


3 <= 3
2 <= 2
1 <= 1
¬ 3 <= 1
¬ 2 <= 1
¬ 3 <= 2
¬ 2 <= 3
¬ 1 <= 3
= 2
1 <= 3
1 <= 2
2 <= 3
3 <= 2
2
3 <= 1
¬ 3 <= 3
¬ 2 <= 2
¬ 1 <= 1

Our modeling approach

$$P(X = x) = C_\lambda \cdot \Gamma(\lambda \cdot (1 - D^\mu(x))$$

2 parameters:

μ . . . modal poset (location parameter)

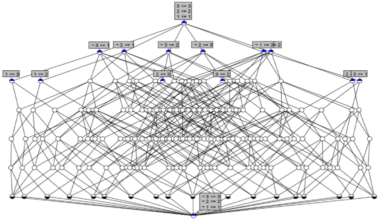
λ . . . scale parameter

and::

Γ . . . decay function (e.g., exp(-x))

D^μ . . . Data depth function with center μ

c_λ . . . normalizing constant



How to define the generic depth function D^{μ} for our model?

- Idea: start with the uniform distribution on the discrete space of all posets.
- Then, for a center/mode μ redistribute a certain amount of probability mass to μ.
- Then, apply a generic generalized depth function like Tukey's depth to get a depth function with mode μ.

Problems with quasiconcavity

- Tukey's depth is quasiconcave.
- But: Analysis shows that because of symmetry, one obtains only two different depth values, i.e., there are many ties $\rightsquigarrow$ useless.
- Peeling depth is (strictly) quasiconcave, but only in the so-called meet-distributive case (e.g., the $\mathbb{R}^d$) and here we are far away from meet-distributivity.
- In particular, here the peeling procedure is not well-defined, there are many ways of peeling at every peeling step.
- A random peeling approach (with a subsequent integration over all peeling paths) would solve this non-uniqueness issue, but the result is not quasiconcave anymore.

Our solution

- We take inspiration by the analysis of the **enclosing depth** (a newly introduced depth that is based on reversing the peeling process, but that unfortunately also suffers from non-uniqueness issues)
- Given a specified center μ one can say something about how deep certain other posets are to μ in relation to each other.
- This structural insight can be used to **weight** the Tukey's depth to get a quasiconcave depth function without many ties.
- One difficult obstacle there was: How to concretely deal with the *non-edges* of μ.

Summary part A

- We combined data depth and FCA to establish a new relational methodology for the analysis and the modeling of non-standard data.
- We introduced concrete statistical location scale models for data that are posets based on FCA and data depth (including an algorithm for sampling from such models).
- In Contribution 3 we applied another FCA-based data depth (a generalization of the simplicial depth) to machine learning: We descriptively analyzed (random) multivariate performance relations between certain classifiers.

Further research that is on its way (only part A)

- **Statistical inference** (e.g., permutation based tests using depth-depth plots)
- Develop concepts of **robustness** for FCA-based depth functions (**VC** dimension seems to be a key characteristic, here)
- Statistical **regularization** in the context of FCA-data depth. (Already first results for the newly developed *double peeling depth* that regularizes both w.r.t. statistical, as well as robustness aspects.)
- Relations to **metric** approaches: There is an abstract notion of (Tukey's) depth in metric spaces (Dai and Lopez-Pintado [2022]) that can be extended to the general theory of FCA. (Possible application I have in mind: Subgroup discovery for phylogenetic data)

Further research related to the other parts

- Part B: Data depth for the special case of data points that are itself **preference systems**.
- Part B: Study notions of **dispersion** with the help of data depth (e.g., applied to the notion of polarization in social choice theory)
- Part C: Study the case of non-standard data with a **partial epistemic** and a **partial ontic** character (e.g., in the context of elicitation of preference systems)

References

I. Couso and D. Dubois. Statistical reasoning with set-valued information: Ontic vs. epistemic views. International Journal of Approximate Reasoning, 55(7):1502–1518, 2014.

X. Dai and S. Lopez-Pintado. Tukey's depth for object data. Journal of the American Statistical Association, 0(0):1–13, 2022.

R. Dittrich, R. Hatzinger, and W. Katzenbeisser. Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. Journal of the Royal Statistical Society: Series C (Applied Statistics), 47(4):511–525, 1998.

B. Ganter and R. Wille. Formal Concept Analysis: Mathematical Foundations. Springer, 2012.

G. Lebanon and Y. Mao. Non-parametric modeling of partially ranked data. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, Advances in Neural Information Processing Systems, volume 20. Curran Associates, Inc., 2007.

K. Nakamura, K. Yano, and F. Komaki. Learning partially ranked data based on graph regularization. arXiv preprint arXiv:1902.10963, 2019.

R. Serfling and Y. Zuo. General notions of statistical depth function. The Annals of Statistics, 28(2):461 – 482, 2000.

J. Tukey. Mathematics and the picturing of data. In R. James, editor, Proceedings of the International Congress of Mathematicians Vancouver, pages 523–531, Vancouver, 1975. Mathematics-Congresses.

List of figures

[1] https://link.springer.com/article/10.1007/s10994-015-5524-x

[2] https://www.yesjanecan.com/tunnels-treasure-tombs-orkney/

[3] https://cdn02.plentymarkets.com/5g7jm0xrr7e2/frontend/blog/deich.jpg

[4] https://www.mittelweserverband.de/deich/index.html

[5] https://www.rolfgemperle.ch/wp-content/uploads/2022/04/DJI_0140-Bearbeitet-Bearbeitet.jpg