On depth functions and robustness in formal concept analysis: The (double-)peeling depth

Georg Schollmeyer and Hannah Blocher

03.08.2021

Reminder: Formal concept analysis

- Formal context $\mathbb{K} = (G, M, I)$:
  - $G \ldots$ objects,
  - $M \ldots$ attributes,
  - $I \subseteq G \times M$ binary relation, $gIm \iff$ object g has attribute m.
- For $A \subseteq G : A' := \{m \in M \mid \forall g \in A : gIm\}$
  $\ldots =$ set of all attributes common to all objects in A.
- For $B \in M : B' := \{g \in G \mid \forall m \in B : gIm\}$
  $\ldots =$ set of all objects that have all attributes in B.
- For $A \subseteq G$: A''
  $\ldots =$ set of all objects that have all attributes that are common to all objects in A. In the sequel: $\gamma :=''$.
  The images of γ are called the closed sets (w.r.t. γ).
- For $A, B \subseteq G$ the object-implication $A \longrightarrow B$ is valid in a formal context if all objects in B have all attributes that are common to all objects in A.

		a	b	c	d	e	f	g	h	i
1	Fischegel	×	×					×		
2	Brasse	×	×					×	×	
3	Frosch	×	×	×				×	×	
4	Hund	×		×				×	×	×
5	Wasserpest	×	×		×		×			
6	Schilf	×	×	×	×		×			
7	Bohne	×		×	×	×				
8	Mais	×		×	×		×			

a: benötigt Wasser zum Leben, b: lebt im Wasser, c: lebt auf dem Land, d: braucht Blattgrün zur Nahrungsaufbereitung, e: zweikeimblättrig, f: einkeimblättrig, g: fähig zum Ortswechsel, h: hat Gliedmaßen, i: säugt seine Jungen

Abbildung 1.1. Kontext zu einem Lehrfilm „Lebewesen und Wasser".

[Ganter and Wille, 1996, p.18]

A short story about meet-distributive formal contexts

Definition (meet-distributive formal context)

A formal context $\mathbb{K} = (G, M, I)$ is called **meet-distributive** if one of the following equivalent conditions is fullfilled[1]:

i) Every extent A is generated by the set $\mathrm{extr}(A)$ of all its extreme points.[2]

ii) (Anti-exchange property, c.f., [Edelman, 1980]): for every extent A and every two objects $g, h \notin A$ with $g' \neq h'$ we have $g \in (A \cup \{h\})'' \Longrightarrow h \notin (A \cup \{g\})''$.

---

[1] In the sequel, only property i) will be of importance.

[2] An object $g \in A$ of a subset $A \subseteq G$ is called an extreme point of A if $(A'' \backslash \{h \mid h' = g'\})'' \subsetneq A''$.

Example

		a	b	c	d	e	f	g	h	i
1	Fischegel	×	×					×		
2	Brasse	×	×					×	×	
3	Frosch	×	×	×				×	×	
4	Hund	×		×				×	×	×
5	Wasserpest	×	×		×		×			
6	Schilf	×	×	×	×		×			
7	Bohne	×		×	×	×				
8	Mais	×		×	×		×			

a: benötigt Wasser zum Leben, b: lebt im Wasser, c: lebt auf dem Land, d: braucht Blattgrün zur Nahrungsaufbereitung, e: zweikeimblättrig, f: einkeimblättrig, g: fähig zum Ortswechsel, h: hat Gliedmaßen, i: säugt seine Jungen

Abbildung 1.1. Kontext zu einem Lehrfilm „Lebewesen und Wasser“.

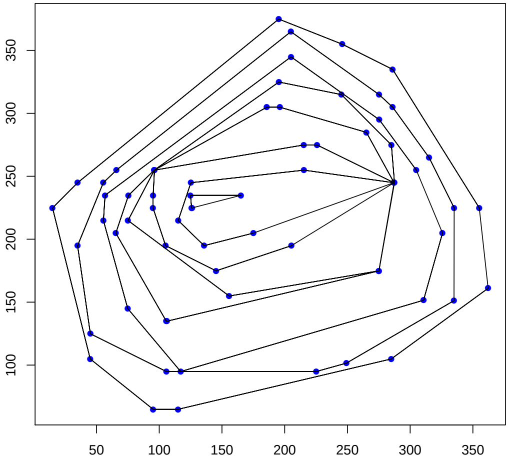
[Ganter and Wille, 1996, p.18]

- $A = \{Fischegel, Brasse, Frosch\} (= A'')$
- $\{Fischegel\} \longrightarrow \{Brasse\} \longrightarrow \{Frosch\}$
- $extr(A) = \{Fischegel\}$

Examples

- Points in $\mathbb{R}^d$ as objects, half-spaces in $\mathbb{R}^d$ as attributes with $gIm \iff$ point g lies in half-space m.
- (inter-)ordinally scaled data without ties (possibly multidimensional, with no ties in any dimension)
- Any context that is composed by meet-distributive subcontexts.

The classical convex hull peeling depth for data in $\mathbb{R}^d$

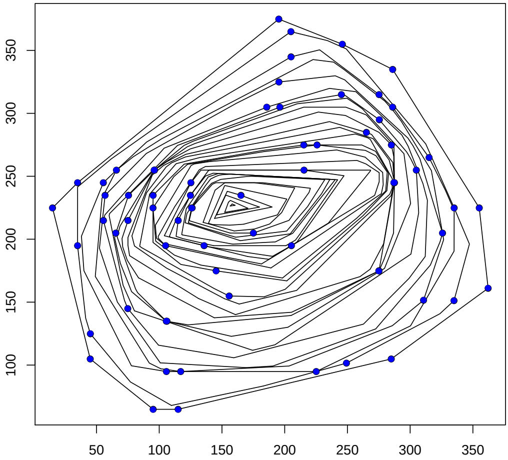


For comparison: Tukey's depth (half-space depth) for data in $\mathbb{R}^d$

$$
\begin{aligned}
D_{Tukey}(x, P_n) &= \inf\{P_n(H) \mid H \text{ half-space, } H \ni x\} \\
&= 1 - \sup\{P_n(H) \mid H \text{ half-space, } x \notin H\}
\end{aligned}
$$

with $P_n(H) =$ proportion of points in half-space H.

For comparison: Tukey's depth for data in $\mathbb{R}^d$



For comparison: Tukey's depth in formal concept analysis

$$D_{Tukey}(x, P_n) = 1 - \sup\{P_n(A) \mid A \text{ formal extent}, x \notin A\}$$

with $P_n(A) = \frac{|A|}{|G|}$

Theorem (Properties of the extreme point operator)

Let $\mathbb{K}$ be a **meet-distributive** *formal context and let $extr : 2^G \longrightarrow 2^G : A \mapsto extr(A)$ be the corresponding extreme point operator. Then, for all $A \subseteq G$ we have:*

i) **Generativity:** $extr(A)'' = A''$ *(in other words: $\gamma \circ extr = \gamma$).*

ii) **Intesiveness:** $extr(A) \subseteq A$.

iii) **Idempotence:** $extr(extr(A)) = extr(A)$.

Definition (peeling operator)

Let $\mathbb{K}$ be a formal context (not necessarily meet-distributive) and let $\gamma :='' $ be the corresponding closure operator. An operator $\nu : 2^G \longrightarrow 2^G$ is called a **peeling operator** if it is generative, intensive and idempotent. A peeling operator is called **minimal** (w.r.t γ) if for all $A \subseteq G$ and all $B \subsetneq \nu(A)$ we have $\gamma(B) \subsetneq \gamma(A)$. It is called **contour-closed** (w.r.t. γ) if for all extents A the peeled set $A'' \setminus \nu(A)$ is closed w.r.t. γ.

Remark

The extreme point operator is both minimal and contour-closed. In the general, non-meet-distributive case it seems not possible to construct a generative operator that is additionally both minimal and contour-closed.

Definition (depth function in FCA)

Let $\mathscr{K} = \{\mathbb{K} = (G, M, I) \mid \mathbb{K} \text{ is a formal context }\}$ and let $\mathscr{D} := \{(A, \mathbb{K}) \mid \mathbb{K} = (G, M, I) \text{ formal context}, A \subseteq M\}$. A depth function D is a (partial) map

$$D(\cdot, \cdot) : \mathscr{D} \longrightarrow \mathbb{R}_{\geq 0}.$$

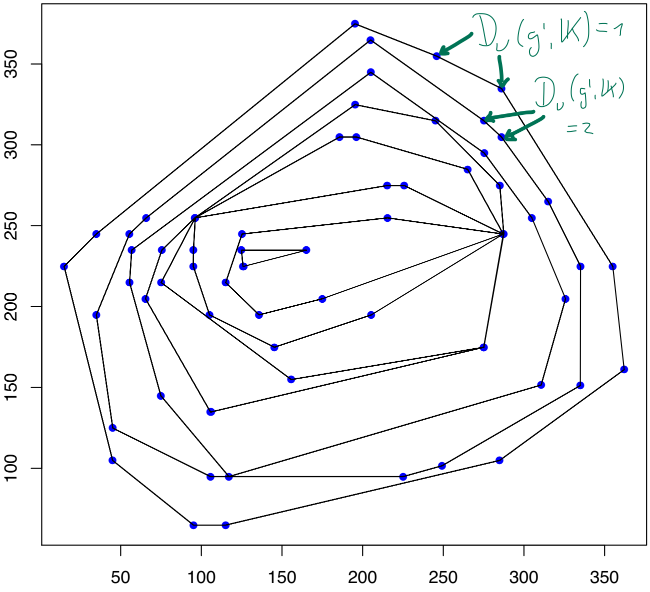
Definition (peeling depth)

Let $\mathbb{K}$ be a (finite) formal context and let ν be an arbitrary peeling operator. Then we define the peeling depth (w.r.t. ν) recursively as

$$D_\nu(g', \mathbb{K}) := 1 \qquad \text{for all } g \in \nu(G)$$

$$D_\nu(g', \mathbb{K}) := k + 1 \qquad \text{for all } g \in \nu(\{g \in G \mid D_\nu(g', \mathbb{K}) \not\leq k\})$$

Note that this defines in fact only a partial mapping, but this can be resolved (not important **here**.)



350
300
250
200
150
100
50
100
150
200
250
300
350

Some useful notions of dimension

Definition

Let $\mathbb{K}$ be a formal context, γ the corresponding closure operator, *extr* the corresponding extreme point operator and ν an arbitrary peeling operator. Then we define:

i) the extreme point dimension as $d_{extr} := \sup\{|extr(A)| \mid A \subseteq G\}$;

ii) the peeling dimension as $d_\nu : \sup\{|\nu(A)| \mid A \subseteq G\}$;

iii) and the VC dimension as $VC(\mathbb{K}) := \sup\{|A| \mid A \subseteq G, \{A \cap B \mid B \text{ extent of } \mathbb{K}\} = 2^A\}$.

VC dimension in FCA

- A set $A \subseteq G$ is called shatterable if $\{A \cap B \mid B \text{ formal extent }\} = 2^A$.
- The VC dimension is the largest cardinality of a shatterable set.
- In FCA this corresponds to the maximal cardinality of an implication-free set[3]
- Additionally, the VC dimension is the maximal cardinality of a contranominal subcontext:

	m_{j_1}	m_{j_2}	m_{j_3}	$\dots$	$m_{j_{k-1}}$	m_{j_k}
g_{i_1}	○	×	×	$\dots$	×	×
g_{i_2}	×	○	×	$\dots$	×	×
g_{i_3}	×	×	○	$\dots$	×	×
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$g_{i_{k-1}}$	×	×	×	$\dots$	○	×
g_{i_k}	×	×	×	$\dots$	×	○

[3] A set $A \subseteq G$ is called implication-free if there is no implication $C \longrightarrow D$ with $C, D \subseteq A$, $D \neq \emptyset$ and $C \cap D = \emptyset$.

Theorem

Let $\mathbb{K}$ be a formal context without duplicates and let ν be a peeling operator. Then we have:

i) $d_\nu \geq d_{extr}$. *(Reason: $\nu(A) \supseteq extr(A)$ for every $A \subseteq G$.)*

ii) *The shatterable sets are exactly the images of the operator $extr$.*

iii) *Thus, $d_{extr} = VC(\mathbb{K})$.*

iv) *If ν is minimal, then we have $d_\nu = d_{extr} = VC(\mathbb{K})$.*

Three variants of a general peeling operator

Definition (peeling operators)

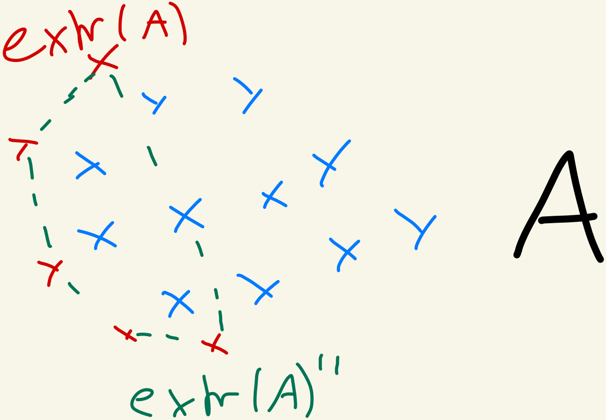
Define ν_{closed}, ν_{min} and $\nu_{minclosed}$ as mappings from 2^G to 2^G via

$$\nu_{closed}(A) := extr(A) \cup \left(A \setminus extr(A)''\right)$$

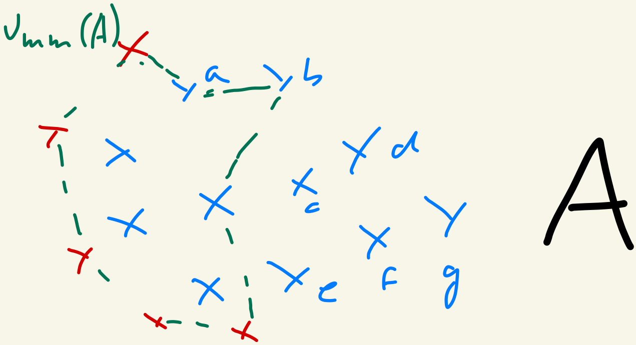
$$\nu_{min}(A) :\in \min\{B \subseteq A \mid B'' \supseteq A''\}$$

$$\nu_{minclosed}(A) :\in \min\{B \subseteq A \mid B'' \supseteq A'' \ \& \ A'' \setminus B \in im(\gamma)\}$$

V closed:



Vmm:



? B {a, b} → {c, g}

Remark

Note that ν_{min} is a minimal peeling operator. In contrast, ν_{closed} and $\nu_{minclosed}$ are generally not minimal. Note further that ν_{min} and $\nu_{minclosed}$ are generally not unique. The min operator is meant here w.r.t. $\subseteq$. (Of course, using the min operator w.r.t. cardinality of the sets would refine the choice to special sets that are also minimal w.r.t. $\subseteq$.)

A notion of robustness without metrics: An unachievable goal?

Breakdown point (according to Donoho and Gasko [1992]):

$$\varepsilon(T, X^{(n)}) := \min \left\{ \frac{m}{n+m} \;\middle|\; \sup_{Y^{(m)}} ||T(X^{(n)} \cup Y^{(m)}) - T(X^{(n)})|| = \infty \right\}$$

with $T \ldots$ a location estimator (e.g., median), $X^{(n)} \ldots$ the actually observed data set of size n, $Y^{(m)} \ldots$ potential contamination data set of size m.

A notion of robustness without metrics: Yes indeed! i

Definition (Contamination pair)

Let $\mathbb{K} = (G, M, I)$ be a formal context. A pair (A, B) with $A, B \subseteq G$ is called a contamination pair (w.r.t. $\mathbb{K}$), if for every $C \subseteq A$ and $D \subseteq B$ the formal implication $C \longrightarrow D$ is not valid in $\mathbb{K}$.

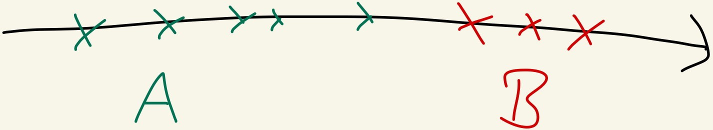
Remark

The set A can be seen as the support of the distribution of actual interest and the set B plays the role of the support of the contamination. Note further that (A, B) is a contamination pair if and only if $A \longrightarrow \{b\}$ is not valid for everey $b \in B$.

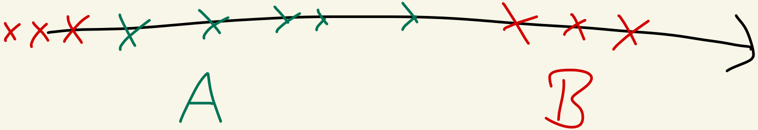
Example: $\mathbb{R}^1$ with interordinal conceptual scaling

- $G \subseteq \mathbb{R}^1$
- $M = \{\text{“}\leq x\text{”} \mid x \in G\} \cup \{\text{“}\geq x\text{”} \mid x \in G\}$.

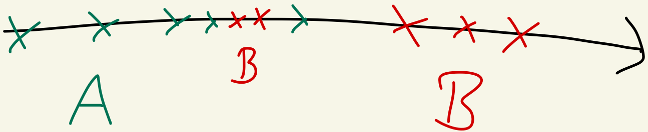
Contamination pair



Contamination pair

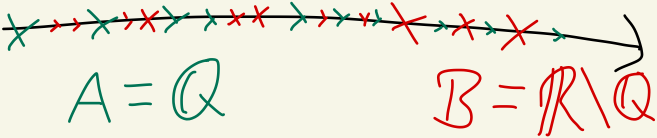


No Contamination pair



No

Contamination pair



Definition (Realized contamination breakdown point)

Let T be a mapping with domain $\mathscr{K}$ and codomain such that for all $(G, M, I) \in \mathscr{K}$ we have $T((G, M, I)) \subseteq G$. (Think of $T((G, M, I)) = \arg\max_{g \in G} D(g', (G, M, I))$ for a depth function D.)

Let furthermore $\mathbb{K} = (G, M, I)$ be a formal context. Then the realized contamination breakdown point (**RCBP**) of T (w.r.t. $\mathbb{K}$) is defined as

$$\varepsilon(T, \mathbb{K}) := \min \Big\{ \alpha \mid \exists (A, B) \text{ contamination pair w.r.t. } \mathbb{K}, \\ A \cup B = G, \frac{|B|}{|G|} = \alpha : \frac{|B \cap T(\mathbb{K})|}{|A \cap T(\mathbb{K})|} \geq \frac{|B|}{|A|} \Big\}.$$

Theorem (Realized contamination breakdown point of the peeling-median)

Let $\mathbb{K}$ *be a formal context (not necessarily meet-distributive) with* $VC(\mathbb{K}) < |G|$*. Let* ν *be a peeling operator and let* $T_\nu(\mathbb{K}) := \arg\max\{D_\nu(\cdot, \mathbb{K})\}$ *be the corresponding peeling-median. Then if* ν *is* **minimal**, *the realized contamination breakdown point of the peeling-median* T_ν *is bounded from below by*

$$\varepsilon(T_\nu, \mathbb{K}) \geq \left\lfloor \frac{1}{VC(\mathbb{K})} \right\rfloor_G, \tag{1}$$

where $VC(\mathbb{K})$ *is the VC-dimension of the closure system of all extents of the context* $\mathbb{K}$ *and where* $\lfloor \cdot \rfloor$ *is the rounding downwards to the next multiple of* $1/|G|$*. Furthermore, more concretely, if there is more than one peeling, we have*

$$\varepsilon(T_\nu, \mathbb{K}) \geq \left\lfloor \frac{\textit{number of peelings}}{|G|} \right\rfloor_G. \tag{2}$$

Remark

Inequality (2) also holds for a peeling operator ν that is not minimal. Note further that the number of peelings is dependent on the concrete peeling operator (and of course also on the context $\mathbb{K}$). In contrast, inequality (1) is only dependent on the VC demension of $\mathbb{K}$, and often, this dimension can be controlled/analyzed a priori somehow.

Remark

The bound $\left\lfloor \frac{1}{VC(\mathbb{K})} \right\rfloor_G$ for minimal ν is sharp in the sense that there exists a subset $A \subseteq G$ such that

$$\varepsilon(T_\nu, \mathbb{K}_{|A\times M}) \leq \left\lfloor \frac{1}{VC(\mathbb{K})} \right\rfloor_A \left(= \left\lfloor \frac{1}{VC(\mathbb{K}_{|A\times M})} \right\rfloor_A \right).$$

Furthermore, for arbitrary $A \subseteq G$ with $|A| > VC(\mathbb{K})$ and for arbitrary $B \subseteq M$ we have $\varepsilon(T_\nu, \mathbb{K}_{|A\times B}) \geq \left\lfloor \frac{1}{VC(\mathbb{K})} \right\rfloor_A$.

Idea of proof

- Every peeling contains at most $VC(\mathbb{K})$ data points.
- Every peeling contains at least one 'outlier' (w.r.t. every arbitrary thinkable contamination pair) as long as there are outliers at all in the remaining set of data points.

'Abstauber': Breakdown point for other depth functions

Theorem

Let $\mathbb{K}$ be a formal context, let D be an arbitrary depth function and let T_D be its corresponding median.

i) *If D is generative, then we have*

$$\varepsilon(T_D, \mathbb{K}) \geq \left\lfloor \frac{\textit{number of contours of } D}{|G|} \right\rfloor . \tag{3}$$

ii) *If D is additionally minimal, then we have*

$$\varepsilon(T_D, \mathbb{K}) \geq \left\lfloor \frac{1}{VC(\mathbb{K})} \right\rfloor . \tag{4}$$

Remark

For example, Tukey's depth is generative and under 'certain additional assumptions' also minimal.

Example: Partial ranking data

Theorem

Let $C = \{c_1, \ldots, c_q\}$ *be a set of* q *items. Let* $\mathbb{K} = (G, M, I)$ *be a formal context where every object* $g \in G$ *represents a partial ranking of the* q *items (i.e., every* $g \in G$ *is a reflexive, transitive and antisymmetric binary relation on* C*). Let further be* $M = C \times C$ *and let* $gI(c, \tilde{c}) \iff (c, \tilde{c}) \in g$*. Then the VC-dmension of* $\mathbb{K}$ *is bounded from above by*

$$VC(\mathbb{K}) \leq \left\lfloor \frac{q}{2} \right\rfloor \cdot \left\lceil \frac{q}{2} \right\rceil .$$

Furthermore, this bound is sharp in the sense that for every $q \in \mathbb{N}$ *there exists a set* $C = \{c_1, \ldots c_q\}$ *and a formal context* $\mathbb{K}$ *of the above form such that* $VC(\mathbb{K}) = \left\lfloor \frac{q}{2} \right\rfloor \cdot \left\lceil \frac{q}{2} \right\rceil$.

A concrete example: The wisdom of the crowd phenomena for (total) ranking data i

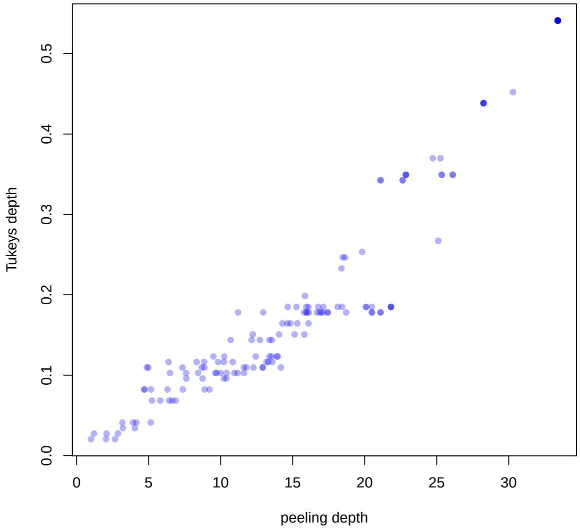
All-together, 146 undergraduates recruited from the human subjects pool at the University of California Irvine ranked 10 US holidays according to their assumed chronological order:

New Year's Day
Martin Luther King Jr. Day
President's Day
Memorial Day
Independence Day
Labor Day
Columbus Day
Halloween
Veteran's Day
Thanksgiving Day

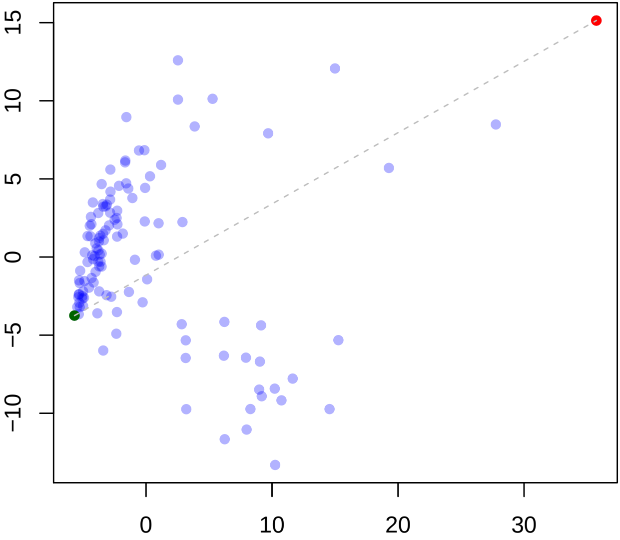
A concrete example: The wisdom of the crowd phenomena for (total) ranking data ii

- Thus, $VC(\mathbb{K}) \leq 5 \cdot 5 = 25$.
- Concretely, for this data set, $VC(\mathbb{K}) = 13$.
- Thus, $\varepsilon(T_\nu, \mathbb{K}) \geq \frac{1}{13} \approx 7.7\%$.
- There are ≈ 39 peelings.
- Thus, $\varepsilon(T_\nu, \mathbb{K}) \geq \frac{39}{146} = 26.7\%$.

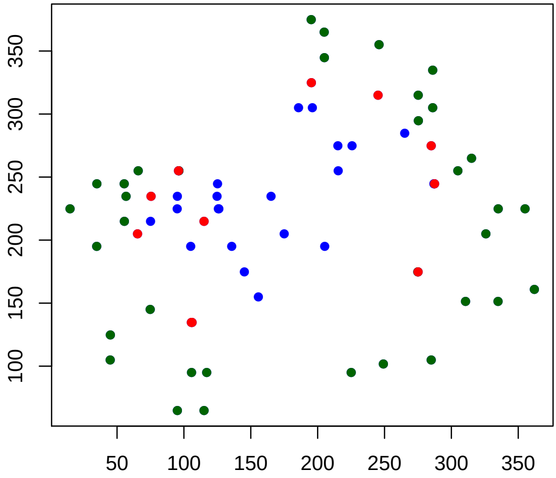
A concrete example: The wisdom of the crowd phenomena for (total) ranking data iii



Ranking data: Multidemsional scaling



Example: Synthetic geometry +



Enlarging the breakdown point: The double peeling depth

Observation: If ν is minimal, then every peeling $\nu(A)$ builds a contranominal substructure of the context (Here $k = |\nu(A)| \leq VC(\mathbb{K})$.):

	m_{j_1}	m_{j_2}	m_{j_3}	$\dots$	$m_{j_{k-1}}$	m_{j_k}
g_{i_1}	○	x	x	$\dots$	x	x
g_{i_2}	x	○	x	$\dots$	x	x
g_{i_3}	x	x	○	$\dots$	x	x
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$g_{i_{k-1}}$	x	x	x	$\dots$	○	x
g_{i_k}	x	x	x	$\dots$	x	○

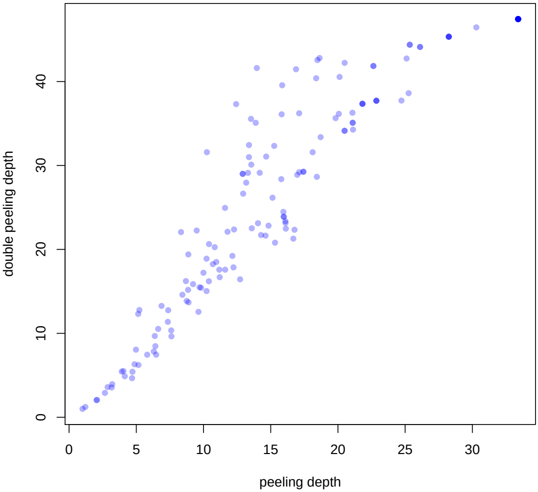
This leads to a 'perfect symmetry' between $g_{i_1}, \dots, g_{i_k}$ in the sense of: 'There exists non non-trivial implication between any of the objects of $\nu(A)$'.

Idea: break the symmetry by locally deleting the attributes $m_{j_1}, \dots, m_{j_k}$ (and possibly further attributes that are identical tome some m_{j_l} w.r.t the objects of $\nu(A)$) to uncover the hidden substructure of further attributes.

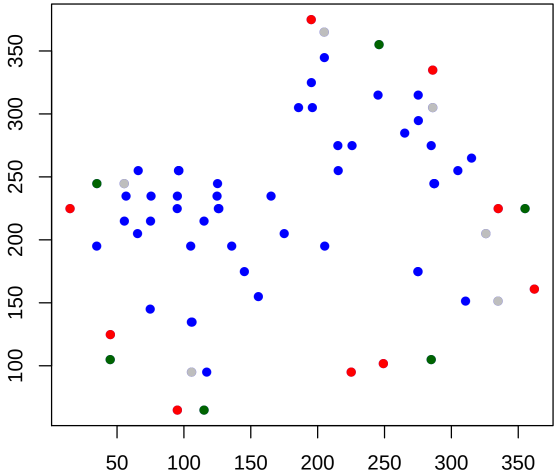
Enlarging the breakdown point: The double peeling depth

- One can repeat this local 'peeling' of attributes such that $\nu(A)$ is reduced to a subset? $\nu^*(A)$ (by removing objects from $\nu(A)$ that follow from other objects of $\nu(A)$ w.r.t. the reduced context).
- Given an envisaged $h < VC(\mathbb{K})$ one can 'always' reduce the size of the actual peeling to a size $\leq h$.
- With this one can enlarge the breakdown point of the corresponding depth function.
- **Important:** The enlarged breakdown point is of course then only valid w.r.t. a reduced class of contamination pairs (i.e. that contamination pairs that also respect the uncovered, 'stylized' implications that were introduced during the process of (locally) deleting attributes.)

Example: Ranking data ($h = 4$)



Example: Geometry



Literatur

D. L. Donoho and M. Gasko. Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness. *The Annals of Statistics*, 20(4):1803 – 1827, 1992. doi: 10.1214/aos/1176348890. URL https://doi.org/10.1214/aos/1176348890.

P. H. Edelman. Meet-distributive lattices and the anti-exchange closure. *Algebra Universalis*, 10(1):290–299, 1980.

B. Ganter and R. Wille. *Formale Begriffsanalyse: Mathematische Grundlagen*. Springer-Verlag, 1996.