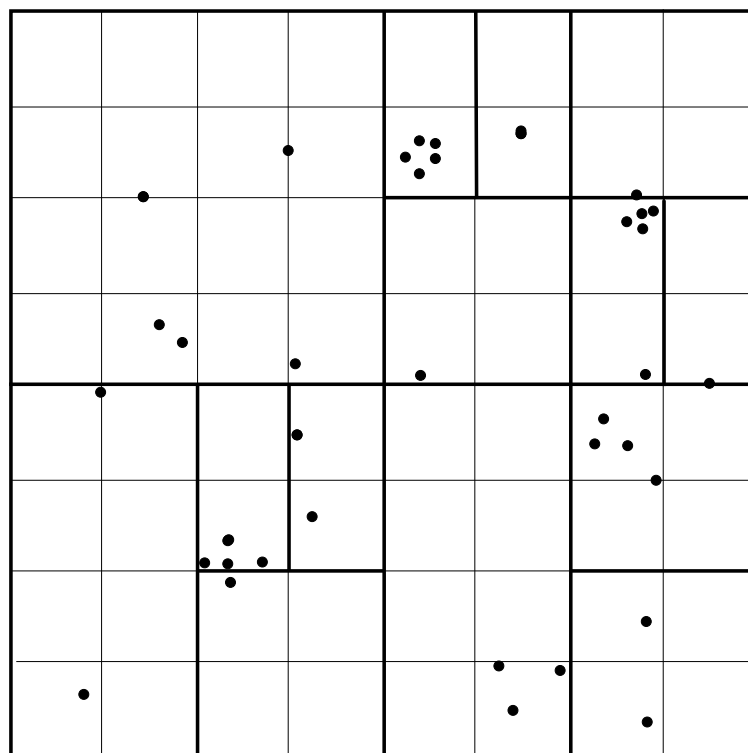


6. Übung „Knowledge Discovery“

Wintersemester 2008/2009

1 GRID-Clustering

Die folgende Abbildung zeigt Punkte der Ebene, die mittels eines *Gridfiles* mit einer Anzahl von maximal $n = 5$ Punkten pro Cluster vorgeclustert wurden (die dünnen Rasterlinien dienen nur der Orientierung).



Wenden Sie die GRID-Clustering-Methode auf den obigen Datensatz an.

2 Datenkompression zum Vor-Clustering

Erzeugen Sie einen CF-Baum aus den Datenpunkten der folgenden Tabelle. Nutzen Sie die Parameter $B = 2$, $T = 1.5$ und $L = 2$.

1	5	1.5	9	6	6.2	2
---	---	-----	---	---	-----	---

Eine detaillierte Beschreibung des Algorithmus finden Sie in [1].

3 Formale Begriffsanalyse

1. Betrachten Sie das Diagramm des Begriffsverbandes auf Folie 11 im 3. Kapitel, 3. Teil der Vorlesung. Wie lautet der Begriff der zu dem Knoten oberhalb des Knotens „Death Valley“ gehört?
2. Beschriften Sie die Knoten des Diagramms mit den zugehörigen Support-Werten und zeichnen Sie das Diagramm der Knoten mit mehr als $10/19$ Support.
3. Bestimmen Sie im folgenden Kontext alle Begriffe, indem Sie zu jeder Teilmenge $B \subseteq M$ das Paar (B', B'') bilden.

I	jung	alt	weiblich	männlich
Vater		X		X
Mutter		X	X	
Tochter	X		X	
Sohn	X			X

Literatur

- [1] T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH: an efficient data clustering method for very large databases. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data (SIGMOD'96)*, pages 103–114, 1996. <http://citeseer.ist.psu.edu/zhang96birch.html>.