

Übung zur Vorlesung *Einsatz und Realisierung von Datenbanksystemen* im SoSe17

Maximilian E. Schüle (schuele@in.tum.de)
<http://db.in.tum.de/teaching/ss17/impldb/>

Blatt Nr. 08

Hausaufgabe 1

Zeigen Sie die weiteren Phasen des Apriori-Algorithmus für unser Beispiel in Abbildung 1 (hier ist lediglich bis inkl. 2. Phase dargestellt). Damit eine Menge von Produkten ein Frequentitemset ist, muss sie in mindestens $3/5$ aller Verkäufe enthalten sein, d.h. $minsupp = s_0 = 3/5$. Gehen Sie für die Assoziationsregeln von einer minimalen Konfidenz von $k_0 = 0$ aus und berechnen Sie die Konfidenz der Assoziationsregel $\{\text{Drucker}\} \Rightarrow \{\text{Papier, Toner}\}$.

VerkaufsTransaktionen	
TransID	Produkt
111	Drucker
111	Papier
111	PC
111	Toner
222	PC
222	Scanner
333	Drucker
333	Papier
333	Toner
444	Drucker
444	PC
555	Drucker
555	Papier
555	PC
555	Scanner
555	Toner

Zwischenergebnisse	
FI-Kandidat	Anzahl
{Drucker}	4
{Papier}	3
{PC}	4
{Scanner}	2
{Toner}	3
{Drucker, Papier}	3
{Drucker, PC}	3
{Drucker, Scanner}	3
{Drucker, Toner}	3
{Papier, PC}	2
{Papier, Scanner}	3
{Papier, Toner}	3
{PC, Scanner}	2
{PC, Toner}	2
{Scanner, Toner}	2

Abbildung 1: Ausgangssituation für den Apriori-Algorithmus

Hausaufgabe 2

Die in Abbildung 2 dargestellten Relationen Mietspiegel und Kindergarten dienen der Bewertung von Wohngebieten im Großraum München. Für eine junge Familie ist ausschlaggebend, wie hoch die Lebenshaltungskosten gemessen an zu zahlender Miete und zu entrichtender Gebühr für den Kindergarten im jeweiligen Wohnort ausfallen. Illustrieren

Sie die Ausführung einer Top-1-Berechnung (zur Bestimmung des günstigsten Wohnorts) für eine junge Familie mit zwei Kindern. Zeigen Sie die phasenweise Berechnung des Ergebnisses jeweils mit dem Threshold- und dem NRA-Algorithmus.

Mietspiegel		Kindergarten		WohnLage	
Ort	Miete	Ort	Beitrag	Ort	Lage
Garching	800	Grünwald	-100	Grünwald	München-Süd
Ismaning	900	Unterföhring	0	Unterföhring	München-Nord
Unterföhring	1000	Bogenhausen	100	Ismaning	München-Nord
Nymphenburg	1500	Ismaning	200	Garching	München-Nord
Bogenhausen	1600	Garching	250	Bogenhausen	München-City
Grünwald	1700	Nymphenburg	300	Nymphenburg	München-City

Abbildung 2: Münchner Wohnlagen zur Berechnung der monatlichen Kosten für eine Familie.

Hausaufgabe 3

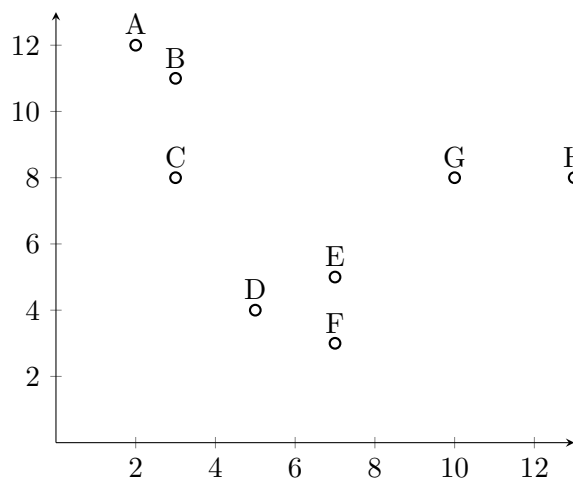
Geben die Relation Klausur:

MatrNr	Vorbereitungszeit	Note
1	150	1.7
2	70	2.7
3	450	2.0
4	180	1.7
5	2500	1.3

- Formulieren Sie die Anfrage, die die MatrNr in der Skyline für die Attribute Vorbereitungszeit und Note erzeugt (kleiner ist jeweils besser) in SQL mit Hilfe des Skyline Operators.
- Formulieren Sie die Anfrage in SQL ohne Skyline Operator.
- Bestimmen Sie das Ergebnis der Anfrage.

Hausaufgabe 4

Folgende Datenpunkte im euklidischen Raum seien gegeben:



Punkt	X	Y
A	2	12
B	3	11
C	3	8
D	5	4
E	7	5
F	7	3
G	10	8
H	13	8

Clustern Sie die Punkte mithilfe des *k-means*-Verfahren in 3 Cluster. Nutzen Sie als initiale Clusterzentren die Werte *A*, *B* und *C*. Wenn ein Punkt zu mehreren Clustern die gleiche Distanz hat, wird er dem Cluster der näher am Nullpunkt liegt zugeordnet. Geben Sie für jede Iteration jeweils die Zuordnung und die Mittelpunkte der Cluster an.

Hausaufgabe 5

Entwerfen Sie einen Algorithmus, um den Klassifikationsbaum, wie er in Abbildung 3 dargestellt ist automatisch zu ermitteln.

Hausaufgabe (wird nicht in der Übung besprochen)

Die in Abbildung 4 dargestellten Relationen Autos und Unterhalt dienen der Bewertung von Autos. Eine junge Studierende sucht ein Auto mit guter Balance zwischen Sportlichkeit und Kosten. Sie überlegt sich wie die drei Werte Preis, PS und monatlicher Unterhalt in einen Score umberechnet werden können und nutzt schließlich folgende Formel:

$$\text{Preis} - (100 * PS) + 24 * \text{Unterhalt}$$

Zeigen Sie die phasenweise Berechnung der Top-3 Ergebnisse jeweils mit dem Threshold- und dem NRA-Algorithmus. Prüfen sie vor der Berechnung ob Teile der Scoringformel schon innerhalb jeder Relation vorberechnet werden können.

Schadenshöhe			
wiealt	Geschlecht	Autotyp	Schäden
45	w	Van	gering
18	w	Coupé	gering
22	w	Van	gering
38	w	Coupé	gering
19	m	Coupé	hoch
24	m	Van	hoch
40	m	Coupé	hoch
40	m	Van	gering
⋮	⋮	⋮	⋮

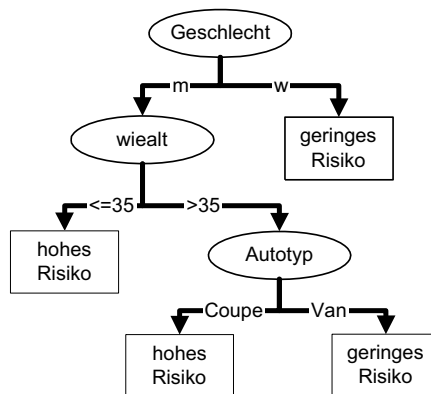


Abbildung 3: Klassifikationsschema für Haftpflicht-Risikoabschätzung.

Auto	Preis	PS	Auto	Unterhalt p. Monat
Seat Leon	25000€	200	Seat Leon	215€
Audi A1	17000€	96	Audi A1	220€
Citroen DS 4	20679€	100	Citroen DS 4	225€
Mini One	16500€	75	Mini One	262€
Mercedes C-Klasse	35000€	160	Mercedes C-Klasse	290€
Porsche Cayenne	80100€	420	Porsche Cayenne	430€

Abbildung 4: Autokauf und -Unterhaltskosten.