Tutorial Business Analytics

Tutorial 11

Exercise 11.1

Regarding a data set about taste in music with 1000 entries the association rule

R: beatles, stones → dylan, cohen

has a support of 0.4 and a confidence of 0.8.
Answer each of the following questions with an interval as small as possible.
(remark: [−∞,+∞] or a single value are valid options, too)

a) How many people like beatles and stones? 500

b) How many people like stones and Dylan?

c) What is the support of the rule "beatles, dylan, stones → cohen" ?

d) What is the lift of the above-mentioned rule (BS->DC)? Interpret your result.

a) $supp(S,S,D,C) = 0.4$

$conf(S,S,D\&C) = supp(S,S,D,C) / supp(S,S) = 0.8$

$\Rightarrow supp(S,S) = 0.5$

Exercise 11.2

Have a look the following items {Wine, Noodles, Tomato sauce, Diapers} and transactions and find all item sets that meet min. support = 0.4. Construct all possible rules that meet the min. confidence = 0.8.

Customer	Wine	Noodles	Tomato sauce	Diapers
1	1	1	1	0
2	1	0	0	1
3	0	1	1	1
4	1	1	1	1
5	0	1	1	0
6	1	1	0	1
7	0	0	0	1
8	1	1	1	1
9	0	0	1	1
10	1	1	1	0

Table 1: 10 customer transactions. 1 = bought, 0 = not bought

1-sets: supp(wine) = 0.6, supp(noodles) = 0.7,
supp(ts) = 0.7, supp(diapers) = 0.7

2-sets: supp(wine, noodles) = 0.5
wine → noodles (0.83)
supp(w, n) = 0.6
n → w (0.85), w → n (0.85)
supp(w, n) = 0.4

3-sets: supp(w, n, t) = 0.4
w, n → t (0.8)
w, t → n (1)

Exercise 11.3 Singular Value Decomposition (SVD)

Gregory registered to an online music platform, where he can stream his favorite songs.
The platform uses Singular Value Decomposition (SVD) to predict ratings for songs, based on previous ratings of a user. The following 3 tables are the result of that SVD.

U	Dim1	Dim2
User 1 (U1)	0.35	0.00
Gregory (G)	0.53	0.84
User 3 (U3)	0.84	-0.23
User 4 (U4)	0.20	0.25
…	…	…

table 1

$r_{GF} = 2 + [0.53, 0.84] \times \begin{bmatrix} 5.22 & 0 \\ 0 & 3.12 \end{bmatrix} \times [-0.13, 0.34]^T$

$= 2 + [2.77, 2.62] \times \begin{bmatrix} -0.13 \\ 0.34 \end{bmatrix} = 2.53$

S	Dim1	Dim2
Dim1	5.22	0
Dim2	0	3.12

table 2

V^T	Shape of You (S)	One Dance (O)	Closer (C)	Despacito (D)	Faded (F)	…
Dim1	0.76	0.54	0.63	0.13	-0.13	…
Dim2	0.84	0.45	0.83	-0.32	0.34	…

table 3

a) Predict the rating of Gregory (G) for the song Faded (F) considering that Gregory's average rating is $\bar{r}_G = 2$.

b) How do you interpret the values of Users 1 (U1) in table 1 and the values of Despacito (D) in table 3?

- U1 has not rated a song of (category) concept 2
- D has a positive correlation with concept 1 and negative with concept 2