Supplementary Notes #4

COMP 578 Data Mining and Data Warehousing

MScECT, Semester 1, 03-04

Supplementary Notes #1 Solution

1a) Find people who buy _____ are 2 times more likely to also buy _____.

Answer:

Target: find rule(s) which got lift ratio equal to 2

Let items be:

Bread	i1
Butter	i2
Coke	i3
Chips	i4
Noodles	i5

After substitution:

Trans ID	**Items Purchased**
100	i1, i2, i5
200	i2, i4
300	i2, i3
400	i1, i2, i4
500	i1, i3
600	i2, i3
700	i1, i3
800	i1, i2, i3, i5
900	i1, i2, i3

Calculate Lift Ratio:

LR of X → Y = Pr(Y|X) / Pr(Y) = Pr(X^Y) / Pr(X) Pr(Y)

Rules	**LR**	**Rules**	**LR**	**Rules**	**LR**	**Rules**	**LR**
i1 → i2	0.86	i1 → i2, i3	0.75	i1 → i2, i3, i4	0	i1, i2 → i3, i4	0
i1 → i3	1	i1 → i2, i4	0.75	i1 → i2, i3, i5	1.5	**i1, i2 → i3, i5**	**2.25**
i1 → i4	0.75	i1 → i2, i5	1.5	i1 → i2, i4, i5	0	i1, i2 → i4, i5	0

i1 → i5	1.5	i1 → i3, i4	0	i1 → i3, i4, i5	0	i1, i3 → i2, i4	0
i2 → i3	0.86	i1 → i3, i5	1.5	i2 → i1, i3, i4	0	i1, i3 → i2, i5	0
i2 → i4	1.29	i1 → i4, i5	0	i2 → i1, i3, i5	0	i1, i3 → i4, i5	0
i2 → i5	1.29	i2 → i1, i3	0.64	i2 → i1, i4, i5	0	i1, i4 → i2, i3	0
i3 → i4	0	i2 → i1, i4	1.29	i2 → i3, i4, i5	0	i1, i4 → i2, i5	0
i3 → i5	0.75	i2 → i1, i5	1.29	i3 → i1, i2, i4	0	i1, i4 → i3, i5	0
i4 → i5	0	i2 → i3, i4	0	i3 → i1, i2, i5	0	i1, i5 → i2, i3	0
		i2 → i3, i5	1.29	i3 → i1, i4, i5	0	i1, i5 → i2, i4	0
		i2 → i4, i5	0	i3 → i2, i4, i5	0	i1, i5 → i3, i4	0
		i3 → i1, i2	0.75	i4 → i1, i2, i3	0	i1 → i2, i3, i4, i5	0
		i3 → i1, i4	0	i4 → i1, i2, i5	0	i2 → i1, i3, i4, i5	0
		i3 → i1, i5	0.75	i4 → i1, i3, i5	0	i3 → i1, i2, i4, i5	0
		i3 → i2, i4	0	i4 → i2, i3, i5	0	i4 → i1, i2, i3, i5	0
		i3 → i2, i5	0.75	**i5 → i1, i2, i3**	**2.25**	i5 → i1, i2, i3, i4	0
		i3 → i4, i5	0	i5 → i1, i2, i4	0		
		i4 → i1, i2	1.125	i5 → i1, i3, i4	0		
		i4 → i1, i3	0	i5 → i2, i3, i4	0		
		i4 → i1, i5	0				
		i4 → i2, i3	0				
		i4 → i2, i5	0				
		i4 → i3, i5	0				
		i5 → i1, i2	**2.25**				
		i5 → i1, i3	1.125				
		i5 → i1, i4	0				
		i5 → i2, i3	1.125				
		i5 → i2, i4	0				
		i5 → i3, i4	0				

People who buy ***noodles*** are 2 times more likely to also buy ***bread and butter***.

People who buy ***noodles*** are 2 times more likely to also buy ***bread and butter and coke***.

People who buy ***bread and butter*** are 2 times more likely to also buy ***coke and noodles.***

1b) People who buy butter are _____ times more likely to also buy noodles.

Answer:
Lift Ratio (i2 → i5)

= Pr(i5 | i2) / Pr(i5)
= Pr(i5^i2) / Pr(i2) Pr(i5)
= 1.29

People who buy butter are ***1.29*** times more likely to also buy noodles.

2)

min_sup = 0.2
min_conf = 0.5

a)

Item	Count	Support
i1	**6**	**0.67**
i2	**7**	**0.78**
i3	**6**	**0.67**
i4	**2**	**0.22**
i5	**2**	**0.22**

Item	Count	Support
i1, i2	**4**	**0.44**
i1, i3	**4**	**0.44**
i1, i4	1	0.11
i1, i5	**2**	**0.22**
i2, i3	**4**	**0.44**
i2, i4	**2**	**0.22**
i2, i5	**2**	**0.22**
i3, i4	0	0.00
i3, i5	**1**	**0.11**
i4, i5	0	0.00

Item	Count	Support
i1, i2, i3	**2**	**0.22**
i1, i2, i4	1	0.11
i1, i2, i5	**2**	**0.22**
i1, i3, i4	0	0.00
i1, i3, i5	1	0.11
i1, i4, i5	0	0.00
i2, i3, i4	0	0.00
i2, i3, i5	1	0.11
i2, i4, i5	0	0.00
i3, i4, i5	0	0.00

Rule	Confidence	Rule	Confidence
i1 -> i2	**0.67**	**i2 -> i1**	**0.57**
i1 -> i3	**0.67**	**i3 -> i1**	**0.67**
i1 -> i5	0.33	**i5 -> i1**	**1.00**
i2 -> i3	**0.57**	**i3 -> i2**	**0.67**
i2 -> i4	0.29	**i4 -> i2**	**1.00**
i2 -> i5	0.29	**i5 -> i2**	**1.00**
i1 -> i2, i3	0.33	**i2, i3 -> i1**	**0.50**
i2 -> i1, i3	0.29	**i1, i3 -> i2**	**0.50**
i3 -> i1, i2	0.33	**i1, i2 -> i3**	**0.50**
i1 -> i2, i5	0.33	**i2, i5 -> i1**	**1.00**
i2 -> i1, i5	0.29	**i1, i5 -> i2**	**1.00**
i5 -> i1, i2	**1.00**	**i1, i2 -> i5**	**0.50**

2c) Apriori Algorithm

Principle: Any subset of a frequent itemset must be frequent.

Hence, new candidate itemsets should be added only when all of their subsets are estimated to be frequent.

Item	Count	Support
i1	**6**	**0.67**
i2	**7**	**0.78**
i3	**6**	**0.67**
i4	**2**	**0.22**
i5	**2**	**0.22**

Item	Count	Support
i1, i2	**4**	**0.44**
i1, i3	**4**	**0.44**
i1, i4	1	0.11
i1, i5	**2**	**0.22**
i2, i3	**4**	**0.44**
i2, i4	**2**	**0.22**
i2, i5	**2**	**0.22**
i3, i4	0	0.00
i3, i5	1	0.11
i4, i5	0	0.00

Item	Count	Support
i1, i2, i3	2	0.22
i1, i2, i5	2	0.22

Rule	Confidence	Rule	Confidence
i1 -> i2	**0.67**	**i2 -> i1**	**0.57**
i1 -> i3	**0.67**	**i3 -> i1**	**0.67**
i1 -> i5	0.33	**i5 -> i1**	**1.00**
i2 -> i3	**0.57**	**i3 -> i2**	**0.67**
i2 -> i4	0.29	**i4 -> i2**	**1.00**
i2 -> i5	0.29	**i5 -> i2**	**1.00**
i1 -> i2, i3	0.33	**i2, i3 -> i1**	**0.50**
i2 -> i1, i3	0.29	**i1, i3 -> i2**	**0.50**
i3 -> i1, i2	0.33	**i1, i2 -> i3**	**0.50**
i1 -> i2, i5	0.33	**i2, i5 -> i1**	**1.00**
i2 -> i1, i5	0.29	**i1, i5 -> i2**	**1.00**
i5 -> i1, i2	**1.00**	**i1, i2 -> i5**	**0.50**