Exercise 2:

a)

s({e}) = 8/10 = 0.8

s({b, d}) = 2/10 = 0.2

s({b, d, e}) = 2/10 = 0.2

b)

c(bd -> e) = s({b, d, e}) / s({b, d})

　　　　　=0.2/0.2 = 1

　　　　　=100%

c(e --> bd) = s({b, d, e}) / s({e})

　　　　　　= 0.2/0.8 = ¼

　　　　　　=0.25

　　　　　　=25%

We can see that confidence is not symmetrical

c)

s({e}) = 4/5 = 0.8

s({b, d}) = 5/5 = 1

s({b, d, e}) = 4/5 = 0.8

d)

c(bd --> e) = s({b, d, e}) / s({b, d})

　　　　　　=0.8/1

　　　　　　=0.8

　　　　　　= 80%

c(e --> bd) = s({b, d, e}) / s({e})

　　　　　　= 0.8/0.8

　　　　　　= 1

　　　　　　= 100%

e)

There is no relationship between S_1, C_1, S_2, C_2. S_1, C_1 are support and confidence values when treating each transaction ID. S_2, C_2 support and confidence values of an association rule r when treating each transaction ID. So, increase in S_1, C_1, does not mean increase in S_2, C_2.

6)

a)

items(unique)= {Milk, Beer, Diapers,Butter,Cookies,Bread}

countItems =6

$3^{countItems} - 2^{(countItems+1)} + 1$

$=3^6 - 2^{(6+1)} + 1$

=729-128+1

Number of association rules =602

b)

As the longest transaction is 4 maximum size of frequent item sets is 4.

c)
$$x = \frac{6!}{3!*3!} = 4*5 = 20$$

d)

S{Bread,Butter} is the largest

e)

c(Beer, Cookies),c(Bread, Butter) have same confidence.

8)

A)

{1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 3, 5}, {2, 3, 4}, {2, 3, 5}, {3, 4, 5}.

{1, 2, 3,4}, {1, 2, 3,5}, {1, 2, 3,6}

{1, 2, 4, 5}, {1, 2, 4, 6}, {1, 2, 5, 6}.

{1, 3, 4, 5}, {1, 3, 4, 6}, {2, 3, 4, 5}.

{2, 3, 4, 6}, {2, 3, 5, 6}

B)

{1, 2, 3,4}, {1, 2, 3,5}, {1, 2, 4,5}, (2, 3, 4, 5}, {2, 3, 4, 6}

C)

{1, 2, 3, 4} survives the pruning as it contains most frequent subsets

12)

Root:

NULL
c

Level 1:

A	B	C	D	E
C	C	C	C	C

Level 2:

AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
M	I	C	C	M	C	F	M	I	C

Level 3:

ABC	ABD	ABE	ACD	ACE	ADE	BCD	BCE	BDE	CDE
I	I	I	I	I	M	I	I	M	I

Level 4:

ABCD	ABCE	ABDE	ACDE	BCDE
I	I	I	I	I

Root:

ABCDE
I

13)

{b} → {c}

	C⁺	C-	TOTAL
B⁺	3	4	7
B⁻	2	1	3
TOTAL	5	5	10

{a}→{d}

	D⁺	D⁻	TOTAL
A⁺	4	1	5
A⁻	5	0	5
TOTAL	9	1	10

{b} −→ {d}

	D⁺	D⁻	TOTAL
B⁺	6	1	7
B⁻	3	0	3
TOTAL	9	1	10

{e} −→ {c}

	C⁺	C⁻	TOTAL
E⁺	2	4	6
E⁻	3	1	4
TOTAL	5	5	10

{c}→{a}

	A$^+$	A$^-$	TOTAL
C$^+$	2	3	5
C$^-$	3	2	5
TOTAL	5	5	10

B)

SUPPORT:

{b}→{d}	{a}→{d}	{b}→{c}	{e}→{c}	{c}→{a}
0.6	0.4	0.3	0.2	0.2
Rank:1	2	3	4	4

CONFIDENCE:

{b}→{d}	{a}→{d}	{b}→{c}	{c}→{a}	{e}→{c}
6/7	4/5	3/7	2/5	2/6
Rank:1	2	3	4	5

INTEREST:

$$Interest = \frac{SP(xy) * sp(y)}{sp(x)}$$

{b}→{d}	{a}→{d}	{b}→{c}	{c}→{a}	{e}→{c}
0.7714	0.720	0.2142	0.2	0.17
Rank:1	2	3	4	5

IS:

$$IS = \frac{SP(xy) * sp(y)}{\sqrt{suppX * suppY}}$$

{b}→{d}	{a}→{d}	{b}→{c}	{c}→{a}	{e}→{c}
0.756	0.5962	0.507	0.4	0.366
Rank:1	2	3	4	5

Klosgen:

$\sqrt{suppXY}$ * max(confidence(suppXY, suppX) - suppY, confidence(suppXY, suppY)-suppX)

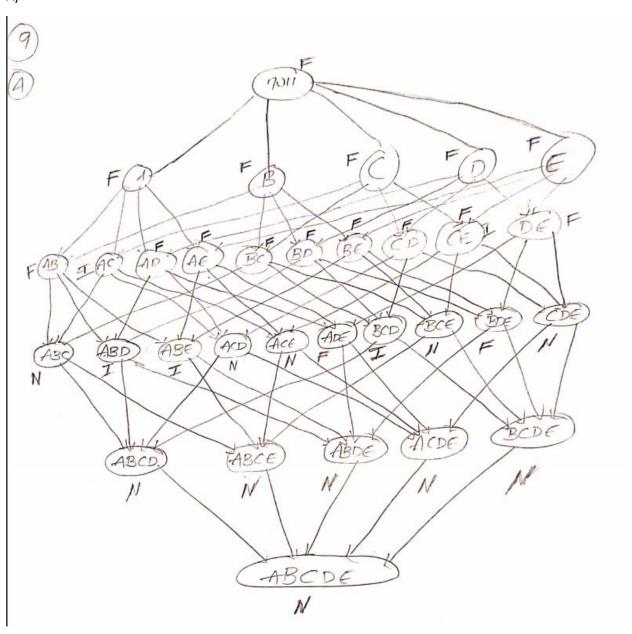| {b}→{d}	{b}→{c}	{c}→{a}	{a}→{d}	{e}→{c}
-0.033	-0.039	-0.044	-0.063	-0.074
Rank:1	2	3	4	5

Odds Ratio:

$$(XY * X^{notIn} Y^{notIn}) / (x\ Y^{notIn} *y\ X^{notIn})$$

{b}→{d}	{a}→{d}	{b}→{c}	{c}→{a}	{e}→{c}
0	0	0.3749	0.43	0.16
Rank:4	4	2	1	3

9)

A)

B)

Percentage =(16/32)*100

= 50%

C)

 Pruning Ratio=count(N)/total=11/32*100=34.4%

D)

False Alarm Rate: (count(I)/total) *100= (5/32) *100=15.625%

20)

(a) Table I.

$$B \quad \bar{B}$$
$$A \quad 9 \quad 1$$
$$\bar{A} \quad 1 \quad 89$$

	B^+	B^-	
A^+	9	1	10
A^-	1	89	90
Total	10	90	100

S(A) = 10/100=0.1

S(B) = 10/100= 0.9

S(A,B) = 9/100=0.09

Interest=0.09

φ (A,B)= (9*89) – (1*1)/√(10*10*90*90)=0.89

Confidence(A→B) = 0.9

Confidence(B→A) =0.9

B)

(b) Table II.

$$B \quad \bar{B}$$
$$A \quad 89 \quad 1$$
$$\bar{A} \quad 1 \quad 9$$

	B^+	B^-	
A^+	89	1	90
A^-	1	9	10
Total	90	10	100

S(A) =90/100 = 0.9

S(B) = 90/100 = 0.9

Interest=0.89

ϕ (A,B)= 0.89

Confidence(A→B) = 0.98

Confidence(B→A) =0.98

C)

From the result we know confidence is associative.Correlation coefficient is invariant in inverse operation. It is because coefficient considers the presence and operation of the items. The elements are positively correlated as the interest factor is close to 1.