

Exercise 2 :

a)

total amount of transactions is 10

support (e) $\rightarrow 8/10 = 0.8$

support(b, d) $\rightarrow 2/10 = 0.2$

support(b, d, e) $\rightarrow 2/10 = 0.2$

b)

$\{b, d\} \rightarrow \{e\}$ and $\{e\} \rightarrow \{b, d\}$

$\{b, d\} \rightarrow e$ confidence $\rightarrow 2/2 = 100\%$

$\{e\} \rightarrow \{b, d\}$ confidence $\rightarrow 2/8 = 25\%$

so the confidence is not symmetric measurement

Exercise 6

a)

six different item in sets

$$R = 3^d - 2^{d+1} + 1 = 3^6 - 2^7 + 1 = 502$$

c)

$$C(6,3) = 6! / (3! * 3!) = 6 * 5 * 4 * 3 * 2 / (6 * 6) = 20$$

d)

the item sets {bread, butter} has the most support 50%

e)

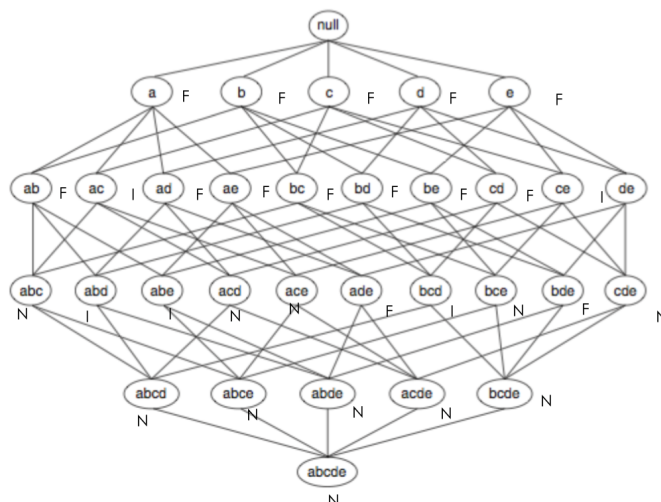
confidence of

$\{\text{bread}\} \rightarrow \{\text{butter}\} = 100\%$

$\{\text{butter}\} \rightarrow \{\text{bread}\} = 100\%$

Exercise 8

Exercise 8



b)

total itemsets = 36

frequent itemsets = $(15+1) = 16$

frequency percentage = $16/36 = 50\%$

c)

the N node amount is 11 so the pruning ratio is $11/32 = 34.3\%$

d) the items of that found to be infrequent after support counting is 5

so the false alarm rate is $5/36 = 15.6\%$

Exercise 9

a)

since in definition in books, all Candidate three itemsets was searching like this

all nodes that belong to $t = \{1,3,4,5,8\}$ should be visited and which begin at 1,2,3

so L1, L3, L5, L9 and L11 should be visited

b)

$\{1,4,5\}$ $\{4,5,8\}$

Part 2:

1) four are important

lowerBoundMinSupport (default 0.1): this define the minimum support rate lower bound, in Apriori the support rate and confident define whether the rules are 'valued'

metricType (default confidence) : this is like the above, since the confidence rate is important, this attribute is also important. we keep it in default.

minMetric (default 0.9): minimum confidence rate.

numRules (default 10) : number of rules to find. this would affect the result directly. Since we could find more than X rules.

2)

1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723 conf:(0.92)
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 conf:(0.92)
3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705 conf:(0.92)
4. biscuits=t fruit=t vegetables=t total=high 815 ==> bread and cake=t 746 conf:(0.92)

for Vote dataset:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219 conf:(1)
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198 ==> Class=democrat 198 conf:(1)
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210 conf:(1)
4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201 conf:(1)

3)

support	confidence
0.45	0.2
0.45	0.5
0.45	0.7
0.7	0.2
0.7	0.5
0.7	0.7
0.9	0.2
0.9	0.5
0.9	0.7

using attributes above.

Minimum support: 0.45 (2082 instances)
Minimum metric <confidence>: 0.2
Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 7

Best rules found:

1. biscuits=t 2605 ==> bread and cake=t 2083 conf:(0.8)
2. milk-cream=t 2939 ==> bread and cake=t 2337 conf:(0.8)
3. fruit=t 2962 ==> bread and cake=t 2325 conf:(0.78)
4. baking needs=t 2795 ==> bread and cake=t 2191 conf:(0.78)
5. frozen foods=t 2717 ==> bread and cake=t 2129 conf:(0.78)
6. vegetables=t 2961 ==> bread and cake=t 2298 conf:(0.78)
7. vegetables=t 2961 ==> fruit=t 2207 conf:(0.75)
8. fruit=t 2962 ==> vegetables=t 2207 conf:(0.75)
9. bread and cake=t 3330 ==> milk-cream=t 2337 conf:(0.7)
10. bread and cake=t 3330 ==> fruit=t 2325 conf:(0.7)

Minimum support: 0.45 (2082 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 7

Best rules found:

1. biscuits=t 2605 ==> bread and cake=t 2083 conf:(0.8)
2. milk-cream=t 2939 ==> bread and cake=t 2337 conf:(0.8)
3. fruit=t 2962 ==> bread and cake=t 2325 conf:(0.78)
4. baking needs=t 2795 ==> bread and cake=t 2191 conf:(0.78)
5. frozen foods=t 2717 ==> bread and cake=t 2129 conf:(0.78)
6. vegetables=t 2961 ==> bread and cake=t 2298 conf:(0.78)
7. vegetables=t 2961 ==> fruit=t 2207 conf:(0.75)
8. fruit=t 2962 ==> vegetables=t 2207 conf:(0.75)
9. bread and cake=t 3330 ==> milk-cream=t 2337 conf:(0.7)
10. bread and cake=t 3330 ==> fruit=t 2325 conf:(0.7)

```
Minimum support: 0.45 (2082 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 11
```

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 7

Best rules found:

```
1. biscuits=t 2605 ==> bread and cake=t 2083   conf:(0.8)
2. milk-cream=t 2939 ==> bread and cake=t 2337   conf:(0.8)
3. fruit=t 2962 ==> bread and cake=t 2325   conf:(0.78)
4. baking needs=t 2795 ==> bread and cake=t 2191   conf:(0.78)
5. frozen foods=t 2717 ==> bread and cake=t 2129   conf:(0.78)
6. vegetables=t 2961 ==> bread and cake=t 2298   conf:(0.78)
7. vegetables=t 2961 ==> fruit=t 2207   conf:(0.75)
8. fruit=t 2962 ==> vegetables=t 2207   conf:(0.75)
9. bread and cake=t 3330 ==> milk-cream=t 2337   conf:(0.7)
```

set support to low, when confidence increase to 0.7, the bread and cake -> fruit combine was gone from result.

when increase the support to medium and high, the result is

```
=== Run information ===
```

```
Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.1 -D 0.05 -U 1.0
Relation:     supermarket
Instances:    4627
Attributes:   217
[list of attributes omitted]
=== Associator model (full training set) ===
```

No large itemsets and rules found!

after modify the attributes, i find that 0.5 is the largest lowerBoundMinSupport. When exceed this , the results would be none will be generated.

4)
when remove bread and cake

```
Apriori
=====
```

```
Minimum support: 0.45 (2082 instances)
Minimum metric <confidence>: 0.2
Number of cycles performed: 11
```

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 1

Best rules found:

```
1. vegetables=t 2961 ==> fruit=t 2207   conf:(0.75)
2. fruit=t 2962 ==> vegetables=t 2207   conf:(0.75)
```

As the image showing,
a lot of rules was gone.

5) during this experiment,

I learn that how to using weka to analyzing the dataset to find the needed associate rules.

the effect of modifying explained with each image in above questions.

6) decision trees can miss many rules found by association rules.

tree:

physician-fee-freeze=(y)

| synfuels-corporation-cutback=(n): republican(141.7/4.0)

| synfuels-corporation-cutback!=(n)

| | mx-missile=(n)

| | | adoption-of-the-budget-resolution=(n): republican(19.28/3.31)

| | | adoption-of-the-budget-resolution!=(n)

| | | | anti-satellite-test-ban=(y): republican(2.2/0.0)

| | | | anti-satellite-test-ban!=(y): democrat(5.01/0.02)

| | mx-missile!=(n): democrat(4.99/1.02)

physician-fee-freeze!=(y): democrat(249.66/3.74)

associate rule:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219 conf:(1)

2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198 ==> Class=democrat 198 conf:(1)

3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210 conf:(1)

4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201 conf:(1)

5. physician-fee-freeze=n 247 ==> Class=democrat 245 conf:(0.99)

6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197 conf:(0.99)

7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204 conf:(0.98)

8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 203 ==> physician-fee-freeze=n 198 conf:(0.98)

9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197 conf:(0.97)

10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210 conf:(0.96)