

1) $s=0.3$ και $c=0.75$

$k=1$

Item	Support(s)
A	0.5
B	0.4
C	0.7
D	0.5
E	0.5
F	0.4

Σε αυτή τη περίπτωση δε βγάζουμε κανένα στοιχείο.

$k=2$

item	Support(s)
AB	0.1
AC	0.4
AD	0.3
AE	0.3
AF	0.1
BC	0.3
BD	0.0
BE	0.0
BF	0.2
CD	0.3
CE	0.3
CF	0.2
DE	0.5
DF	0.1
EF	0.1

$k=3$

Item	Support(s)
ACD	0.2
ACE	0.2
ADE	0.3
CDE	0.3

Τερματίζει ο αλγόριθμος διότι δεν μπορούμε να δημιουργήσουμε κάποιο στοιχειοσύνολο από αυτά που είναι συχνά στο $k=3$.

Συχνά στοιχειosύνολα:

Frequent ItemSet	Support(s)
AC	0.4
AD	0.3
AE	0.3
BC	0.3
CD	0.3
CE	0.3
DE	0.5
ADE	0.3
CDE	0.3

Γνωρίζουμε ότι, ένα στοιχειosύνολο με x στοιχεία θα έχει $2^x - 2$ κανόνες άρα στη περίπτωσή μας έχουμε $7 * (2^2 - 2) + 2 * (2^3 - 2) = 26$ υποψήφιοι κανόνες.

$Lift = Confidence / P(Y)$

Κανόνες	Confidence	Lift
D → E	5/5=1	1/0.5=2
E → D	5/5=1	1/0.5=2
AE → D	3/3=1	1/0.5=2
AD → E	3/3=1	1/0.5=2
CE → D	3/3=1	1/0.5=2
CD → E	3/3=1	1/0.5=2
A → C	4/5=0.8	0.8/0.7=1.14
B → C	3/4=0.75	0.75/0.7=1.07
A → D	3/5=0.6	
D → A	3/5=0.6	
A → E	3/5=0.6	
E → A	3/5=0.6	
D → C	3/5=0.6	
E → C	3/5=0.6	
A → DE	3/5=0.6	
DE → A	3/5=0.6	
D → AE	3/5=0.6	
E → AD	3/5=0.6	
DE → C	3/5=0.6	
D → CE	3/5=0.6	
E → CD	3/5=0.6	
C → A	4/7=0.57	
C → B	3/7=0.43	
C → D	3/7=0.43	
C → E	3/7=0.43	
C → DE	3/7=0.43	

Για να πούμε ποιος-οι κανόνας-ες είναι οι πιο ισχυροί βασιζόμαστε στις 2 μετρικές που θίξαμε στο μάθημα (Confidence και Lift) διότι δεν μας αρκεί μόνο το Confidence επειδή αυτό πολλές φορές έχει μεγάλο ποσοστό χωρίς να υπάρχει στην πραγματικότητα η αντίστοιχη συσχέτιση. Οπότε όσοι κανόνες έχουν Confidence=1 και Lift=2 το οποίο είναι το μεγαλύτερο όπως φαίνεται στο πίνακα παραπάνω θα είναι οι πιο ισχυροί. Δηλαδή, βάση του παραπάνω πίνακα οι πρώτοι 6 κανόνες είναι οι πιο ισχυροί. Παρόλα αυτά οι κανόνες D->E και E->D επειδή έχουν και το max support είναι λίγο περισσότερο σημαντικοί.

2) Παρατηρούμε παρακάτω πως τα αποτελέσματα που εμφανίζει ο apriori στο Weka με s=0.3 και c=0.75 είναι τα ίδια με αυτά που έβγαλα χρησιμοποιώντας apriori με το χέρι οπότε επιβεβαιώνονται τα αποτελέσματα της 1^{ης} άσκησης.

```
Apriori
=====

Minimum support: 0.3 (3 instances)
Minimum metric <confidence>: 0.75
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Large Itemsets L(1):
A=t 5
B=t 4
C=t 7
D=t 5
E=t 5
F=t 4

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

Large Itemsets L(2):
A=t C=t 4
A=t D=t 3
A=t E=t 3
B=t C=t 3
C=t D=t 3
C=t E=t 3
D=t E=t 5

Size of set of large itemsets L(3): 2

Large Itemsets L(3):
A=t D=t E=t 3
C=t D=t E=t 3

Best rules found:

1. E=t 5 ==> D=t 5    <conf:(1)> lift:(2) lev:(0.25) [2] conv:(2.5)
2. D=t 5 ==> E=t 5    <conf:(1)> lift:(2) lev:(0.25) [2] conv:(2.5)
3. A=t E=t 3 ==> D=t 3    <conf:(1)> lift:(2) lev:(0.15) [1] conv:(1.5)
4. A=t D=t 3 ==> E=t 3    <conf:(1)> lift:(2) lev:(0.15) [1] conv:(1.5)
5. C=t E=t 3 ==> D=t 3    <conf:(1)> lift:(2) lev:(0.15) [1] conv:(1.5)
6. C=t D=t 3 ==> E=t 3    <conf:(1)> lift:(2) lev:(0.15) [1] conv:(1.5)
7. A=t 5 ==> C=t 4    <conf:(0.8)> lift:(1.14) lev:(0.05) [0] conv:(0.75)
8. B=t 4 ==> C=t 3    <conf:(0.75)> lift:(1.07) lev:(0.02) [0] conv:(0.6)
```

Παρακάτω παραθέτω τα αποτελέσματα από το apriori με metricType=Lift αυθαίρετα έβαλα numRules=25 και τέλος έβαλα s=0.3 και minMetric=1. Στην περίπτωση αυτή έχουμε περισσότερους κανόνες διότι δε λαμβάνουμε υπόψη το Confidence αλλά το Lift και στη περίπτωση αυτήν υπάρχουν πολλοί κανόνες με Lift>1.

```
Apriori
=====

Minimum support: 0.3 (3 instances)
Minimum metric <lift>: 1
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Large Itemsets L(1):
A=t 5
B=t 4
C=t 7
D=t 5
E=t 5
F=t 4

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

Large Itemsets L(2):
A=t C=t 4
A=t D=t 3
A=t E=t 3
B=t C=t 3
C=t D=t 3
C=t E=t 3
D=t E=t 5

Size of set of large itemsets L(3): 2

Large Itemsets L(3):
A=t D=t E=t 3
C=t D=t E=t 3

Best rules found:

1. D=t 5 ==> E=t 5    conf:(1) < lift:(2)> lev:(0.25) [2] conv:(2.5)
2. E=t 5 ==> D=t 5    conf:(1) < lift:(2)> lev:(0.25) [2] conv:(2.5)
3. D=t 5 ==> A=t E=t 3    conf:(0.6) < lift:(2)> lev:(0.15) [1] conv:(1.17)
4. A=t D=t 3 ==> E=t 3    conf:(1) < lift:(2)> lev:(0.15) [1] conv:(1.5)
5. E=t 5 ==> A=t D=t 3    conf:(0.6) < lift:(2)> lev:(0.15) [1] conv:(1.17)
6. A=t E=t 3 ==> D=t 3    conf:(1) < lift:(2)> lev:(0.15) [1] conv:(1.5)
7. D=t 5 ==> C=t E=t 3    conf:(0.6) < lift:(2)> lev:(0.15) [1] conv:(1.17)
8. C=t D=t 3 ==> E=t 3    conf:(1) < lift:(2)> lev:(0.15) [1] conv:(1.5)
9. E=t 5 ==> C=t D=t 3    conf:(0.6) < lift:(2)> lev:(0.15) [1] conv:(1.17)

10. C=t E=t 3 ==> D=t 3    conf:(1) < lift:(2)> lev:(0.15) [1] conv:(1.5)
11. A=t 5 ==> D=t 3    conf:(0.6) < lift:(1.2)> lev:(0.05) [0] conv:(0.83)
12. D=t 5 ==> A=t 3    conf:(0.6) < lift:(1.2)> lev:(0.05) [0] conv:(0.83)
13. A=t 5 ==> E=t 3    conf:(0.6) < lift:(1.2)> lev:(0.05) [0] conv:(0.83)
14. E=t 5 ==> A=t 3    conf:(0.6) < lift:(1.2)> lev:(0.05) [0] conv:(0.83)
15. A=t 5 ==> D=t E=t 3    conf:(0.6) < lift:(1.2)> lev:(0.05) [0] conv:(0.83)
16. D=t E=t 5 ==> A=t 3    conf:(0.6) < lift:(1.2)> lev:(0.05) [0] conv:(0.83)
17. A=t 5 ==> C=t 4    conf:(0.8) < lift:(1.14)> lev:(0.05) [0] conv:(0.75)
18. C=t 7 ==> A=t 4    conf:(0.57) < lift:(1.14)> lev:(0.05) [0] conv:(0.88)
19. B=t 4 ==> C=t 3    conf:(0.75) < lift:(1.07)> lev:(0.02) [0] conv:(0.6)
20. C=t 7 ==> B=t 3    conf:(0.43) < lift:(1.07)> lev:(0.02) [0] conv:(0.84)
```