Homework 6

Jarett Smith, Max Thompson, Charlie Deaton, and Balin Allred

2023-11-20

Question 1

1. In the following, we have ten market-basket transactions:

   (a) What is the maximum size of frequent itemsets that can be extracted (assuming minsup = 0)?

Response: There are a maximum of 5 itemsets.

   (b) What is the maximum number of association rules that can be extracted from this data set (including rules that have zero support)?

```
(3^5)-(2^6)+1
```

```
## [1] 180
```

Response: There are 180 possible rules.

   (c) Compute the support, confidence, and lift for the association rule {A, D} -> {E}

```
# Support
support<- 4/10
support
```

```
## [1] 0.4
```

```
# Confidence
confidence<- 4/4
confidence
```

```
## [1] 1
```

```
# Lift
lift<- confidence/(6/10)
lift
```

```
## [1] 1.666667
```

   (d) Find all the frequent itemsets assuming minsup count = 2.

```
rules<- apriori(TRANS, parameter=list(support=0.2, conf = 0))
table<- inspect(rules)
```

```
kable(table)
```

	lhs		rhs	support	confidence	coverage	lift	count
[1]	{}	=>	{C}	0.5	0.5000000	1.0	1.0000000	5
[2]	{}	=>	{A}	0.5	0.5000000	1.0	1.0000000	5
[3]	{}	=>	{E}	0.6	0.6000000	1.0	1.0000000	6
[4]	{}	=>	{B}	0.7	0.7000000	1.0	1.0000000	7
[5]	{}	=>	{D}	0.9	0.9000000	1.0	1.0000000	9
[6]	{C}	=>	{A}	0.2	0.4000000	0.5	0.8000000	2
[7]	{A}	=>	{C}	0.2	0.4000000	0.5	0.8000000	2
[8]	{C}	=>	{E}	0.2	0.4000000	0.5	0.6666667	2
[9]	{E}	=>	{C}	0.2	0.3333333	0.6	0.6666667	2
[10]	{C}	=>	{B}	0.3	0.6000000	0.5	0.8571429	3
[11]	{B}	=>	{C}	0.3	0.4285714	0.7	0.8571429	3
[12]	{C}	=>	{D}	0.4	0.8000000	0.5	0.8888889	4
[13]	{D}	=>	{C}	0.4	0.4444444	0.9	0.8888889	4
[14]	{A}	=>	{E}	0.4	0.8000000	0.5	1.3333333	4
[15]	{E}	=>	{A}	0.4	0.6666667	0.6	1.3333333	4
[16]	{A}	=>	{B}	0.3	0.6000000	0.5	0.8571429	3
[17]	{B}	=>	{A}	0.3	0.4285714	0.7	0.8571429	3
[18]	{A}	=>	{D}	0.4	0.8000000	0.5	0.8888889	4
[19]	{D}	=>	{A}	0.4	0.4444444	0.9	0.8888889	4
[20]	{E}	=>	{B}	0.4	0.6666667	0.6	0.9523810	4
[21]	{B}	=>	{E}	0.4	0.5714286	0.7	0.9523810	4
[22]	{E}	=>	{D}	0.6	1.0000000	0.6	1.1111111	6
[23]	{D}	=>	{E}	0.6	0.6666667	0.9	1.1111111	6
[24]	{B}	=>	{D}	0.6	0.8571429	0.7	0.9523810	6
[25]	{D}	=>	{B}	0.6	0.6666667	0.9	0.9523810	6
[26]	{C, E}	=>	{D}	0.2	1.0000000	0.2	1.1111111	2
[27]	{C, D}	=>	{E}	0.2	0.5000000	0.4	0.8333333	2
[28]	{D, E}	=>	{C}	0.2	0.3333333	0.6	0.6666667	2
[29]	{B, C}	=>	{D}	0.2	0.6666667	0.3	0.7407407	2
[30]	{C, D}	=>	{B}	0.2	0.5000000	0.4	0.7142857	2
[31]	{B, D}	=>	{C}	0.2	0.3333333	0.6	0.6666667	2
[32]	{A, E}	=>	{B}	0.2	0.5000000	0.4	0.7142857	2
[33]	{A, B}	=>	{E}	0.2	0.6666667	0.3	1.1111111	2
[34]	{B, E}	=>	{A}	0.2	0.5000000	0.4	1.0000000	2
[35]	{A, E}	=>	{D}	0.4	1.0000000	0.4	1.1111111	4
[36]	{A, D}	=>	{E}	0.4	1.0000000	0.4	1.6666667	4
[37]	{D, E}	=>	{A}	0.4	0.6666667	0.6	1.3333333	4
[38]	{A, B}	=>	{D}	0.2	0.6666667	0.3	0.7407407	2
[39]	{A, D}	=>	{B}	0.2	0.5000000	0.4	0.7142857	2
[40]	{B, D}	=>	{A}	0.2	0.3333333	0.6	0.6666667	2
[41]	{B, E}	=>	{D}	0.4	1.0000000	0.4	1.1111111	4
[42]	{D, E}	=>	{B}	0.4	0.6666667	0.6	0.9523810	4
[43]	{B, D}	=>	{E}	0.4	0.6666667	0.6	1.1111111	4
[44]	{A, B, E}	=>	{D}	0.2	1.0000000	0.2	1.1111111	2
[45]	{A, D, E}	=>	{B}	0.2	0.5000000	0.4	0.7142857	2
[46]	{A, B, D}	=>	{E}	0.2	1.0000000	0.2	1.6666667	2

	lhs		rhs	support	confidence	coverage	lift	count
[47]	{B, D, E}	=>	{A}	0.2	0.5000000	0.4	1.0000000	2

(e) Find an itemset (of size 2 or larger) that has the largest support.

Response: Itemset [B, D] and [D, E] has the largest support at 6/10.

(f) Find a pair of items, say x and y, such that {x} -> {y} and {y} -> {x} have the same confidence.

Response: {A} -> {C} and {C} -> {A} have the same confidence of .4

```
rules<- apriori(TRANS, parameter=list(support=0.2, conf=0, minlen=2))
table<- inspect(subset(rules, lhs %in% c("A", "C")))
```

```
kable(head(table, 2))
```

	lhs		rhs	support	confidence	coverage	lift	count
[1]	{C}	=>	{A}	0.2	0.4	0.5	0.8	2
[2]	{A}	=>	{C}	0.2	0.4	0.5	0.8	2