HW 2 Problem 3

Carter Pearon

August 19, 2015

## 3

For the following problem, I tried different combinations of the parameters to see how it affected the set of association rules. I chose a support of .01, as anything much higher than that excluded all of the products. To keep things fairly simple at first I chose a maxlen of 3, as adding additional complex combinations of product mixes might make even harder to find some meaningful or insightful association rules

library(arules)

## Loading required package: Matrix  
##   
## Attaching package: 'arules'  
##   
## The following objects are masked from 'package:base':  
##   
## %in%, write

# read in the data  
groceries = read.transactions("groceries.txt", format = 'basket', sep = ',')  
  
head(groceries)

## transactions in sparse format with  
## 6 transactions (rows) and  
## 169 items (columns)

summary(groceries)

## transactions as itemMatrix in sparse format with  
## 9835 rows (elements/itemsets/transactions) and  
## 169 columns (items) and a density of 0.02609146   
##   
## most frequent items:  
## whole milk other vegetables rolls/buns soda   
## 2513 1903 1809 1715   
## yogurt (Other)   
## 1372 34055   
##   
## element (itemset/transaction) length distribution:  
## sizes  
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15   
## 2159 1643 1299 1005 855 645 545 438 350 246 182 117 78 77 55   
## 16 17 18 19 20 21 22 23 24 26 27 28 29 32   
## 46 29 14 14 9 11 4 6 1 1 1 1 3 1   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 2.000 3.000 4.409 6.000 32.000   
##   
## includes extended item information - examples:  
## labels  
## 1 abrasive cleaner  
## 2 artif. sweetener  
## 3 baby cosmetics

# create the association rules  
groceriesrules = apriori(groceries, parameter=list(support=.01, confidence=.5, maxlen=3))

##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport support minlen maxlen  
## 0.5 0.1 1 none FALSE TRUE 0.01 1 3  
## target ext  
## rules FALSE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## apriori - find association rules with the apriori algorithm  
## version 4.21 (2004.05.09) (c) 1996-2004 Christian Borgelt  
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].  
## sorting and recoding items ... [88 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 done [0.00s].  
## writing ... [15 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

#inspect all of the rules  
inspect(groceriesrules)

## lhs rhs support confidence lift  
## 1 {curd,   
## yogurt} => {whole milk} 0.01006609 0.5823529 2.279125  
## 2 {butter,   
## other vegetables} => {whole milk} 0.01148958 0.5736041 2.244885  
## 3 {domestic eggs,   
## other vegetables} => {whole milk} 0.01230300 0.5525114 2.162336  
## 4 {whipped/sour cream,   
## yogurt} => {whole milk} 0.01087951 0.5245098 2.052747  
## 5 {other vegetables,   
## whipped/sour cream} => {whole milk} 0.01464159 0.5070423 1.984385  
## 6 {other vegetables,   
## pip fruit} => {whole milk} 0.01352313 0.5175097 2.025351  
## 7 {citrus fruit,   
## root vegetables} => {other vegetables} 0.01037112 0.5862069 3.029608  
## 8 {root vegetables,   
## tropical fruit} => {other vegetables} 0.01230300 0.5845411 3.020999  
## 9 {root vegetables,   
## tropical fruit} => {whole milk} 0.01199797 0.5700483 2.230969  
## 10 {tropical fruit,   
## yogurt} => {whole milk} 0.01514997 0.5173611 2.024770  
## 11 {root vegetables,   
## yogurt} => {other vegetables} 0.01291307 0.5000000 2.584078  
## 12 {root vegetables,   
## yogurt} => {whole milk} 0.01453991 0.5629921 2.203354  
## 13 {rolls/buns,   
## root vegetables} => {other vegetables} 0.01220132 0.5020921 2.594890  
## 14 {rolls/buns,   
## root vegetables} => {whole milk} 0.01270971 0.5230126 2.046888  
## 15 {other vegetables,   
## yogurt} => {whole milk} 0.02226741 0.5128806 2.007235

Now let's impose some restrictions on our association rules to see if we can hone in on some of the important relationships between product mixes for our set of transactions

In this first subset of rules I inspect rules with a lift greater than 2 to see if there are any product mixtures that are correlated with some other product mixtures. We know from probability theory that two events are independent if the product of their individual probabilities is equal to the joint probability of the two events, and thus a lift of 1 indicates no association (independence) between two sets of products. There seem to be about 14 transactions that have a lift greater than 2, however none of these rules seem to be all that interesting or telling. People who buy yogurt and curds tend to buy whole milk as well or people who buy root vegetables and tropical fruits also buy other vegetables are not all that informative from a product placement standpoint

# Look at rules with lift > 2  
inspect(subset(groceriesrules, subset=lift > 2))

## lhs rhs support confidence lift  
## 1 {curd,   
## yogurt} => {whole milk} 0.01006609 0.5823529 2.279125  
## 2 {butter,   
## other vegetables} => {whole milk} 0.01148958 0.5736041 2.244885  
## 3 {domestic eggs,   
## other vegetables} => {whole milk} 0.01230300 0.5525114 2.162336  
## 4 {whipped/sour cream,   
## yogurt} => {whole milk} 0.01087951 0.5245098 2.052747  
## 5 {other vegetables,   
## pip fruit} => {whole milk} 0.01352313 0.5175097 2.025351  
## 6 {citrus fruit,   
## root vegetables} => {other vegetables} 0.01037112 0.5862069 3.029608  
## 7 {root vegetables,   
## tropical fruit} => {other vegetables} 0.01230300 0.5845411 3.020999  
## 8 {root vegetables,   
## tropical fruit} => {whole milk} 0.01199797 0.5700483 2.230969  
## 9 {tropical fruit,   
## yogurt} => {whole milk} 0.01514997 0.5173611 2.024770  
## 10 {root vegetables,   
## yogurt} => {other vegetables} 0.01291307 0.5000000 2.584078  
## 11 {root vegetables,   
## yogurt} => {whole milk} 0.01453991 0.5629921 2.203354  
## 12 {rolls/buns,   
## root vegetables} => {other vegetables} 0.01220132 0.5020921 2.594890  
## 13 {rolls/buns,   
## root vegetables} => {whole milk} 0.01270971 0.5230126 2.046888  
## 14 {other vegetables,   
## yogurt} => {whole milk} 0.02226741 0.5128806 2.007235

Another way to find interesting rules is through a minimum confidence threshold. We already established a minimum confidence of .5 in our original model, but let's tweak this a little bit. For some association rule X --> Y, the confidence of that rule can be thought of as a conditional probability, specifically the probability of observing Y given that we've observed X. I played around with a few different balues of confidence, and .55 seemed to narrow down the rules a good amount. The results are similar to the results we saw before, there aren't any rules that stand out or are all that surpising, but we have some rules nonetheless.

inspect(subset(groceriesrules, subset=confidence > 0.55))

## lhs rhs support confidence lift  
## 1 {curd,   
## yogurt} => {whole milk} 0.01006609 0.5823529 2.279125  
## 2 {butter,   
## other vegetables} => {whole milk} 0.01148958 0.5736041 2.244885  
## 3 {domestic eggs,   
## other vegetables} => {whole milk} 0.01230300 0.5525114 2.162336  
## 4 {citrus fruit,   
## root vegetables} => {other vegetables} 0.01037112 0.5862069 3.029608  
## 5 {root vegetables,   
## tropical fruit} => {other vegetables} 0.01230300 0.5845411 3.020999  
## 6 {root vegetables,   
## tropical fruit} => {whole milk} 0.01199797 0.5700483 2.230969  
## 7 {root vegetables,   
## yogurt} => {whole milk} 0.01453991 0.5629921 2.203354