Problem 3

Reshms Sekar

August 19, 2015

# PROBLEM 3

# Association rule mining  
# Adapted from code by Matt Taddy  
library('arules') # has a big ecosystem of packages built around it

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

# Read in playlists from users  
groceries <- read.csv("~/Desktop/STA380-master/data/groceries.txt")  
  
  
# First create a list of baskets: vectors of items by consumer  
# Analagous to bags of words  
  
# First split data into a list of artists for each user  
groceries <- split(groceries,f=",")  
## Remove duplicates ("de-dupe")  
groceries <- lapply(groceries, unique)  
  
## Cast this variable as a special arules "transactions" class.  
groceriestrans <- read.transactions("~/Desktop/STA380-master/data/groceries.txt",format=c("basket"),sep=",",encoding="unknown",rm.duplicates=TRUE)  
# Now run the 'apriori' algorithm  
# Look at rules with support > .01 & confidence >.5 & length (# artists) <= 4  
grocrules <- apriori(groceriestrans,   
 parameter=list(support=.01, confidence=.5, maxlen=4))

##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport support minlen maxlen  
## 0.5 0.1 1 none FALSE TRUE 0.01 1 4  
## 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 4 done [0.00s].  
## writing ... [15 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

#detach(package:tm, unload)=TRUE   
# Look at the output  
inspect(grocrules)

## 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

## Choose a subset  
inspect(subset(grocrules, subset=support > .01 & 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

From the above we can infer that whole milk is brought very often in combination with curd, yoghurt and butter and vegetables, eggs and vegetables and other vegetables are brought a lot often in combination with root vegetables and fruits