Module 6 CT Option 1

Ty Cook

4/25/2020

## Read File into Dataframe

courses.df <- read.csv("Coursetopics.csv") #Read Coursetopics.csv into courses.df dataframe

## Association Rules

###Convert Dataframe into Binary Matrix

courses.mat <- as.matrix(courses.df) #Read dataframe courses.df into matrix courses.mat  
courses.mat[1:10, ] #Output top 10 Columns and Rows of the Matrix

## Intro DataMining Survey Cat.Data Regression Forecast DOE SW  
## [1,] 1 1 0 0 0 0 0 0  
## [2,] 0 0 1 0 0 0 0 0  
## [3,] 0 1 0 1 1 0 0 1  
## [4,] 1 0 0 0 0 0 0 0  
## [5,] 1 1 0 0 0 0 0 0  
## [6,] 0 1 0 0 0 0 0 0  
## [7,] 1 0 0 0 0 0 0 0  
## [8,] 0 0 0 1 0 1 1 1  
## [9,] 1 0 0 0 0 0 0 0  
## [10,] 0 0 0 1 0 0 0 0

### Convert Matrix into Transactions Database

library(arules) #Arules library to read matrix into transactions database

## Warning: package 'arules' was built under R version 3.6.3

## Loading required package: Matrix

##   
## Attaching package: 'arules'

## The following objects are masked from 'package:base':  
##   
## abbreviate, write

courses.trans <- as(courses.mat, "transactions") #Read matrix courses.mat into a transactions database courses.trans  
inspect(courses.trans[1:10, ]) #Inspect top 10 Columns and Rows of the transactions database

## items   
## [1] {Intro,DataMining}   
## [2] {Survey}   
## [3] {DataMining,Cat.Data,Regression,SW}  
## [4] {Intro}   
## [5] {Intro,DataMining}   
## [6] {DataMining}   
## [7] {Intro}   
## [8] {Cat.Data,Forecast,DOE,SW}   
## [9] {Intro}   
## [10] {Cat.Data}

###Apriori Function

rules <-apriori(courses.trans, parameter = list(supp = 0.01, conf = 0.5, target = "rules")) #Create Rules tabel using courses.trans with the parmaters of minimum support of 0.01 and confience interval of 0.5 from rules

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.5 0.1 1 none FALSE TRUE 5 0.01 1  
## maxlen target ext  
## 10 rules FALSE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 3   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[8 item(s), 365 transaction(s)] done [0.00s].  
## sorting and recoding items ... [8 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 4 done [0.00s].  
## writing ... [54 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

inspect(head(sort(rules, by = "lift"), n = 10)) #Inspect rules of first rules where lift is = 10

## lhs rhs support confidence  
## [1] {Intro,Regression,Forecast} => {DataMining} 0.01369863 0.7142857   
## [2] {Intro,Survey,DOE} => {Cat.Data} 0.01095890 0.8000000   
## [3] {Intro,DataMining,Cat.Data} => {Regression} 0.01643836 0.7500000   
## [4] {Intro,DataMining,Regression} => {Forecast} 0.01369863 0.5000000   
## [5] {Intro,Survey,Cat.Data} => {Forecast} 0.01369863 0.5000000   
## [6] {Intro,Regression,DOE} => {SW} 0.01917808 0.7777778   
## [7] {Intro,DataMining,Forecast} => {Regression} 0.01369863 0.7142857   
## [8] {Intro,Cat.Data,Forecast} => {Survey} 0.01369863 0.6250000   
## [9] {DataMining,DOE} => {Cat.Data} 0.01643836 0.6666667   
## [10] {Survey,Regression} => {Cat.Data} 0.01643836 0.6666667   
## lift count  
## [1] 4.010989 5   
## [2] 3.842105 4   
## [3] 3.601974 6   
## [4] 3.578431 5   
## [5] 3.578431 5   
## [6] 3.504801 7   
## [7] 3.430451 5   
## [8] 3.354779 5   
## [9] 3.201754 6   
## [10] 3.201754 6

**Summary**

For this assignment we are analyzing association rules in RStudio, so I choose to go with Option #1 Course Selection Association Rules. With any assignment the first step was to examine the dataset given CourseTopics.csv. From looking at the dataset I was able to see that the set was binary data of eight columns of an unknown amount of entries.

From there the next big step would be to read the dataset into a data frame called courses.df and then move on to apply my association rules. First step was to read the Courses.df into a binary matrix called courses.mat, which was pretty standard and nothing out of the ordinary. Next, I would output the top ten columns and rows of the matrix due to amount of entries in the data set would cause a larger output to present in the RMarkdown document.

The second step was to read the binary matrix into transactions database, the big hurdle I had here was that I did not know at first that you needed the library “arules” in order to perform the function of converting the binary database into a transactions database so further research had to be done on my part in order to resolve the error messages I was receiving from that function. After the errors from reading the matrix into the transactions database was resolved, all I had to do was output the transactions database using the inspect command with a limited scope of 10 entries as the output without the limiter was around 365 entries.

Lastly it was finally time to add the association rules to my transactions database using the apriori function by creating a table called rules using courses.trans, parameters assigned by the list function of minimum support if 0.01 and confidence values 0.1 and 0.5. After that I presented the output from the code using the inspect command to present the head sorted by lift and only displaying the first tern columns from the table rules.