KENTUCKY STATE UNIVERSITY, COMPUTER SCIENCE

Data Mining
HW7

Anil Varma B

November 17, 2019

1. Download attached dataset. Find all rules with support >= 0.2, and confidence >= 0.7. Also, "Play = Yes", or "Play = No" must appear in the rhs of rules.

```
#Reading the csv file
w<-read.csv(file.choose(),head=TRUE)
library(arules)
dim(w)
str(w)
```

```
> w<-read.csv(file.choose(),head=TRUE)
> library(arules)
> dim(w)
[1] 14  6
> str(w)
'data.frame':   14 obs. of  6 variables:
 $ Outlook    : Factor w/ 3 levels "Overcast","Rain",..: 3 3 1 2 2 2 1 3 3 2 ...
 $ Temperature: Factor w/ 3 levels "Cool","Hot","Mild": 2 2 2 3 1 1 1 3 1 3 ...
 $ Humidity   : Factor w/ 2 levels "High","Normal": 1 1 1 1 2 2 2 1 2 2 ...
 $ Windy      : logi  FALSE TRUE FALSE FALSE TRUE TRUE ...
 $ Play       : Factor w/ 2 levels "No","Yes": 1 1 2 2 1 1 2 1 2 2 ...
 $ X          : logi  NA NA NA NA NA NA ...
```

Figure 0.1: Dimensions of the data set

Applying association rules with default settings

```
w[]<- lapply(w,factor)
w_rules <-apriori(w[], parameter = list(support=0.2,confidence = 0.7, target = "rules"))
```

```
> w[]<- lapply(w,factor)
> w_rules <-apriori(w[], parameter = list(support=0.2,confidence = 0.7, target = "rules"))
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.7    0.1    1 none FALSE            TRUE       5     0.2      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 2

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[12 item(s), 14 transaction(s)] done [0.00s].
sorting and recoding items ... [12 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [17 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
> inspect(w_rules)
     lhs                                rhs                 support   confidence lift     count
[1]  {Temperature=Hot}                  => {Humidity=High}    0.2142857 0.7500000  1.500000 3
[2]  {Temperature=Hot}                  => {Windy=FALSE}      0.2142857 0.7500000  1.500000 3
[3]  {Outlook=Overcast}                 => {Play=Yes}         0.2857143 1.0000000  1.750000 4
[4]  {Temperature=Cool}                 => {Windy=TRUE}       0.2142857 0.7500000  1.500000 3
[5]  {Temperature=Cool}                 => {Humidity=Normal}  0.2857143 1.0000000  2.000000 4
[6]  {Windy=FALSE}                      => {Play=Yes}         0.3571429 0.7142857  1.250000 5
[7]  {Humidity=Normal}                  => {Play=Yes}         0.3571429 0.7142857  1.250000 5
[8]  {Temperature=Cool,Windy=TRUE}      => {Humidity=Normal}  0.2142857 1.0000000  2.000000 3
[9]  {Temperature=Cool,Humidity=Normal} => {Windy=TRUE}       0.2142857 0.7500000  1.500000 3
[10] {Humidity=Normal,Windy=TRUE}       => {Temperature=Cool} 0.2142857 0.7500000  2.625000 3
[11] {Outlook=Sunny,Play=No}            => {Humidity=High}    0.2142857 1.0000000  2.000000 3
[12] {Outlook=Sunny,Humidity=High}      => {Play=No}          0.2142857 1.0000000  2.333333 3
[13] {Humidity=High,Play=No}            => {Outlook=Sunny}    0.2142857 0.7500000  2.100000 3
[14] {Outlook=Rain,Play=No}             => {Windy=TRUE}       0.2142857 1.0000000  2.000000 3
[15] {Outlook=Rain,Windy=TRUE}          => {Play=No}          0.2142857 1.0000000  2.333333 3
[16] {Windy=TRUE,Play=No}               => {Outlook=Rain}     0.2142857 0.7500000  2.100000 3
[17] {Humidity=Normal,Windy=FALSE}      => {Play=Yes}         0.2142857 1.0000000  1.750000 3
>
```

Figure 0.2: Applying the apriori algorithm

Checking the summary of association rules.

```
summary(w_rules)
```

```
> summary(w_rules)
set of 17 rules

rule length distribution (lhs + rhs):sizes
 2  3
 7 10

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   2.000   3.000   2.588   3.000   3.000

summary of quality measures:
    support         confidence          lift            count
 Min.   :0.2143   Min.   :0.7143   Min.   :1.250   Min.   :3.000
 1st Qu.:0.2143   1st Qu.:0.7500   1st Qu.:1.500   1st Qu.:3.000
 Median :0.2143   Median :0.7500   Median :2.000   Median :3.000
 Mean   :0.2395   Mean   :0.8634   Mean   :1.852   Mean   :3.353
 3rd Qu.:0.2143   3rd Qu.:1.0000   3rd Qu.:2.100   3rd Qu.:3.000
 Max.   :0.3571   Max.   :1.0000   Max.   :2.625   Max.   :5.000

mining info:
 data ntransactions support confidence
  w[]            14     0.2        0.7
```

Figure 0.3: Summary of the rules

For applying rules with rhs containing "Play" we will set rhs=c("Play=No", "Play=Yes") in appearance to make sure only that will appear in the rhs of rules.

```
rules <- apriori(w,parameter = list(minlen=2, supp=0.2, conf=0.7),appearance =
list(rhs=c("Play=No", "Play=Yes"),default="lhs"),
                 control = list(verbose=F))
inspect(rules)
```

```
> inspect(rules)
    lhs                                rhs          support   confidence lift     count
[1] {Outlook=Overcast}              => {Play=Yes} 0.2857143 1.0000000  1.750000 4
[2] {Windy=FALSE}                   => {Play=Yes} 0.3571429 0.7142857  1.250000 5
[3] {Humidity=Normal}               => {Play=Yes} 0.3571429 0.7142857  1.250000 5
[4] {Outlook=Sunny,Humidity=High}   => {Play=No}  0.2142857 1.0000000  2.333333 3
[5] {Outlook=Rain,Windy=TRUE}       => {Play=No}  0.2142857 1.0000000  2.333333 3
[6] {Humidity=Normal,Windy=FALSE}   => {Play=Yes} 0.2142857 1.0000000  1.750000 3
>
```

Figure 0.4: Association Rules with condition

Visualizing Association Rules Using arulesViz package to visualize Association Rules with baloon plot

```
library(arulesViz)
plot(rules, method="graph", control=list(type="items"))
```

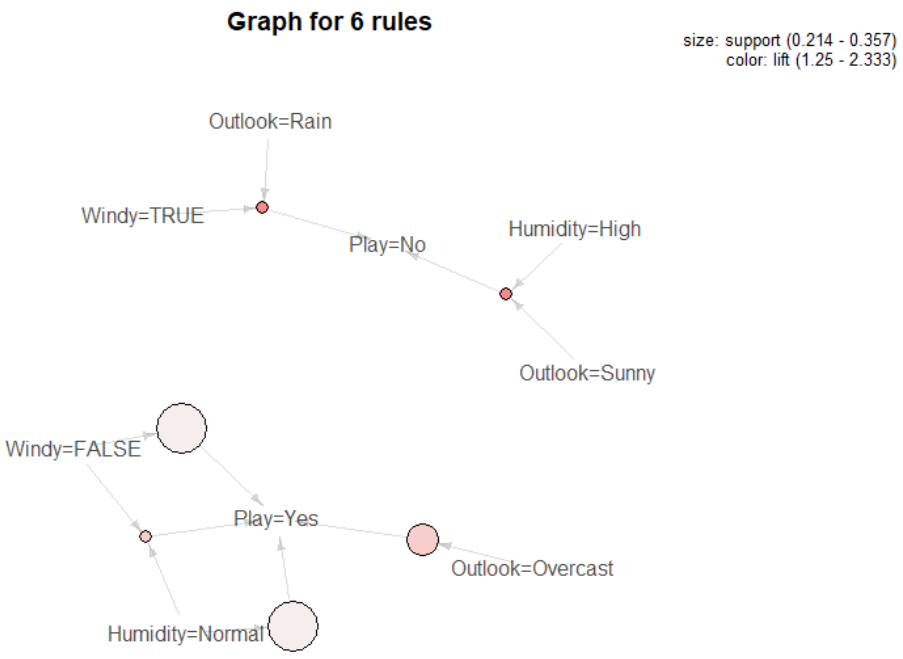

Figure 0.5: Association Rules with baloon plot

Visualizing Association Rules with parallel coordinates plot

```
plot(rules, method="paracoord", control=list(reorder=TRUE))
```

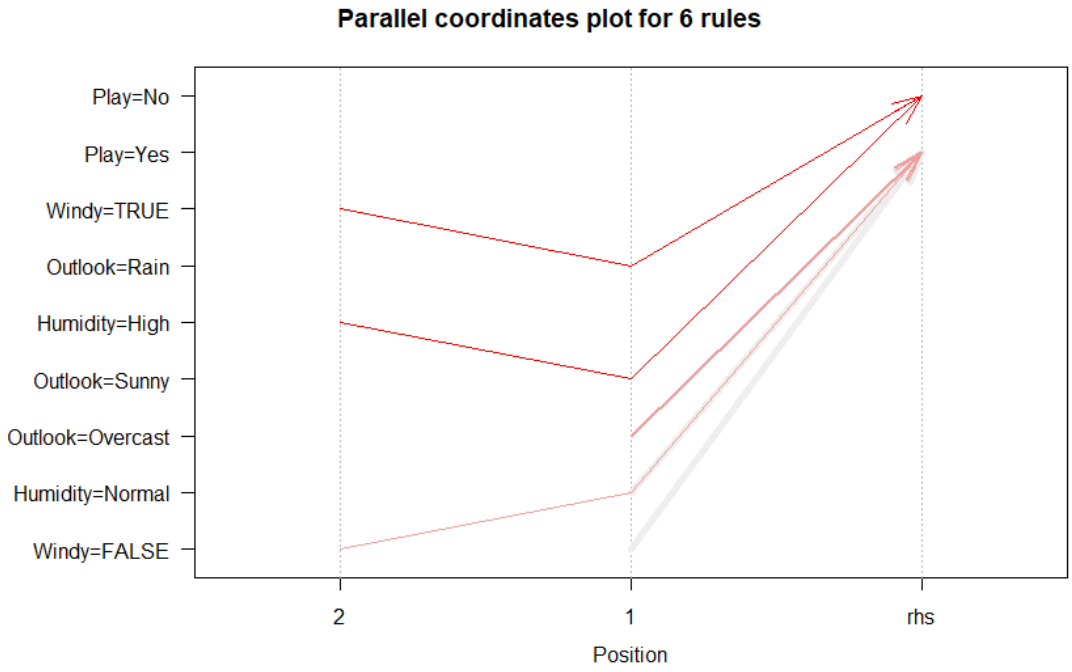

Figure 0.6: Association Rules with parallel coordinates plot