Learning Objectives:

- To understand the applications of association rule mining
- To understand the basic flow of rule induction in Apriori algorithm
- To understand the concepts of support, lift and confidence of rules
- To implement association rules using Apriori algorithm in R

Activity:

a. Association rule mining on transaction dataset, using the purchase data of stationery store across three days given below:

Trans Id	Item Purchased	Trans Id	Item Purchased
1	Pencils	5	Erasers
1	Markers	6	Envelop
1	Highlighters	7	Markers
1	Papers	7	Erasers
2	Markers	8	Pencils
2	Erasers	8	Markers
3	Stapler Pins	8	Stapler Pins
3	Papers	8	Post-It
3	Erasers	8	Highlighter
3	Card holders	8	Papers
3	Highlighters	8	Erasers
4	Papers	9	Stapler Pins
4	Erasers	9	Post-It
4	Card holders	9	Markers
5	Markers	9	Erasers
5	Post-It	10	Envelop

1. Manually compute rules using minimum support = 50%
2. Compute lift for the above rules

b. Find the support confidence and lift for the rule, {Milk, Diaper} ➔ Beer

TID	Items
1	Chips, Milk
2	Chips, Diaper, Beer, Cornflakes
3	Milk, Diaper, Beer, Pepsi
4	Chips, Milk, Diaper, Beer
5	Chips, Milk, Diaper, pepsi

c. Association Rule Mining on an Unsupervised dataset

Data: Transaction data
Steps to follow:

- Install and load 'arules' package
  install.packages("arules")
- Read 'Transactions.csv' data into R such that the arules package treats the input csv file as "transaction" data.
  trans = read.transactions(file="Transactions.csv", rm.duplicates= FALSE, format="single",sep=",",cols =c(1,2))
- Explore and understand the data and items of transaction data
  inspect(trans)
  trans
  image(trans)
- Explore the item frequency and observe frequent item sets
  itemFrequency(trans)
  itemFrequencyPlot(trans)
- Implement association mining using 'Apriori' algorithm to extract rules
  rules <- apriori(trans,parameter = list(sup = 0.1, conf = 0.6,target="rules"))
- Understanding the rules
  summary(rules)
  inspect(rules)
- Visualizing the rules
  library(arulesViz)
  plot(rules)
  plot(rules, method="graph", control=list(type="items"))
- Picking rules
  - Picking top 5 rules sorted by confidence and then support as a data.frame
    top_rules = sort(rules, by = c("confidence", "support"))
    head(as(top_rules, "data.frame"), n=5)

- Picking rules based on LHS/RHs by subsetting

```
rules.itemfilter1 <- as(subset(rules, subset = rhs %in%  "Choclates"), "data.frame")
rules.itemfilter1
rules.itemfilter2 <- as(subset(rules, subset = lhs %in% "Pencil"),  data.frame")
rules.itemfilter2
```

- Picking a subset of rules based on a given range of a quality measure

```
rules_Lift <- as(subset(rules, subset = rhs %in% "Pencil" & lift > 1.01), "data.frame")
rules_Lift
```

d. Association Rule Mining on Supervised dataset
Data: Titanic dataset

- Read in the titanic survival data set and see if its a categorical dataset

```
titanic_data <- read.csv(file = "titanic_data.csv")
head(titanic_data)
str(titanic_data)
```

- Apply Apriori on the data to find associations amongst all the attributes

```
rules <- apriori(titanic_data)
inspect(rules)
```

- Rules with rhs containing target attribute
  - From the data, to find the conditions leading to survival/death of the passengers on the board, we can filter the rules for "Survived"

```
rules <- apriori(titanic_data,parameter = list(minlen=2, supp=0.005, conf=0.8), appearance = list(rhs=c("Survived=No", "Survived=Yes"), default="lhs"))
```

- Inspect the rules

```
inspect(rules)
```

- Sort the rules based on "lift"

```
rules.sorted <- sort(rules, by="lift")
inspect(rules.sorted)
```

Additional References:

http://r-statistics.co/Association-Mining-With-R.html
https://cran.r-project.org/web/packages/arulesViz/vignettes/arulesViz.pdf
http://mhahsler.github.io/arulesViz/
http://blog.hackerearth.com/beginners-tutorial-apriori-algorithm-data-mining-r-implementation