Homework Assignment 3

IT 531, Summer 2018

**Question 1:** Association Rule Analysis

Perform Association Rule Analysis (using Apriori algorithm) using R on following dataset:

Cars Dataset: <https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

Find more details for each dataset on the UCI repository page.

Deliverables:

1. Briefly describe the dataset.
2. Find top 25 rules for each dataset. Briefly describe top five rules.
3. Do some rule pruning. Make sure right hand side of the rule is the class attributes.
4. Submit a zip file containing the data and R script file and the document describing dataset and output results

Answer:

1. This dataset has 7 attributes. The classifier of the dataset is car\_evaluation, which can held the value such as unacc,acc,good,v-good. Other 6 attributes include buying,maint(maintence) in relation to price, and doors,persons,lug\_boot,safty related to technology. The total amount of the dataset is 1728.
2. The top 25 rules by using R (support=0.1,confidence=0.6) are

|  |
| --- |
| lhs rhs support confidence  [1] {} => {classifier=unacc} 0.7002315 0.7002315  [2] {buying=med} => {classifier=unacc} 0.1550926 0.6203704  [3] {buying=high} => {classifier=unacc} 0.1875000 0.7500000  [4] {maint=low} => {classifier=unacc} 0.1550926 0.6203704  [5] {buying=vhigh} => {classifier=unacc} 0.2083333 0.8333333  [6] {maint=med} => {classifier=unacc} 0.1550926 0.6203704  [7] {maint=high} => {classifier=unacc} 0.1817130 0.7268519  [8] {doors=5more} => {classifier=unacc} 0.1689815 0.6759259  [9] {doors=4} => {classifier=unacc} 0.1689815 0.6759259  [10] {doors=3} => {classifier=unacc} 0.1736111 0.6944444  [11] {doors=2} => {classifier=unacc} 0.1886574 0.7546296  [12] {maint=vhigh} => {classifier=unacc} 0.2083333 0.8333333  [13] {lug\_boot=big} => {classifier=unacc} 0.2129630 0.6388889  [14] {lug\_boot=med} => {classifier=unacc} 0.2268519 0.6805556  [15] {safty=med} => {classifier=unacc} 0.2065972 0.6197917  [16] {safty=low} => {classifier=unacc} 0.3333333 1.0000000  [17] {lug\_boot=small} => {classifier=unacc} 0.2604167 0.7812500  [18] {persons=2} => {classifier=unacc} 0.3333333 1.0000000  [19] {lug\_boot=big,safty=low} => {classifier=unacc} 0.1111111 1.0000000  [20] {persons=2,lug\_boot=big} => {classifier=unacc} 0.1111111 1.0000000  [21] {persons=4,safty=low} => {classifier=unacc} 0.1111111 1.0000000  [22] {persons=4,classifier=unacc} => {safty=low} 0.1111111 0.6153846  [23] {persons=more,safty=low} => {classifier=unacc} 0.1111111 1.0000000  [24] {lug\_boot=med,safty=low} => {classifier=unacc} 0.1111111 1.0000000  [25] {persons=2,lug\_boot=med} => {classifier=unacc} 0.1111111 1.0000000 |

The first 5 rules indicate that classifier=unacc occurs mostly,but since the lift is 1, it doesn’t mean anything, the other 4 rules indicate that buying price is median,high,very high or the maintenance is low, the classifier will be unacc.

1. Use apperance parameter in apriori function in R, to prune the uninteresting rules.

rules <- apriori(df, parameter = list(supp=0.1, conf=0.6),appearance = list( rhs = c( 'classifier=unacc', 'classifier=acc', 'classifier=good', 'classifier=v-good' ), default = 'lhs' ))