Homework

# Rule Mining Homework – Using aRules on the Titanic dataset

Download the titanic dataset titanic.raw.rdata from <https://github.com/ethen8181/machine-learning/tree/master/association_rule/R> or <https://sites.google.com/a/rdatamining.com/www/data/titanic.raw.rdata?attredirects=1>

Note that it is not a cvs file, but rather an R dataset. So, to load the data, one would do:

load("….whatever…./Downloads/titanic.raw.rdata")

#You need to look at titanic.raw (the name of the R dataset)

t <- titanic.raw

Now that you have the datafile, do some descriptive statistics, getting some extra practice using R.

**Step 1: Descriptive stats**

1. Compute the percentage of people that survived.

2. Compute the percentage of people that were children.

3. Compute the percentage of people that were female.

4. Finally, compute the percentage of people that were in first class.

**Step 2: More descriptive stats**

1. What percentage of children survived?

2. What percentage of female survived?

3. What percentage of first-class people survived?

4. What percentage of third-class people survived?

**Step 3: Writing a function**

1. Write a function that returns a new dataframe of people that satisfies the specified criteria of sex, age, class and survived as parameters.

2. Write a function, using the previous function, that calculates the percentage (who lives, who dies) for a specified (parameters) of age, class and sex.

3. Use the function to compare age and third-class male survival rates.

4. Use the function to compare age and first-class female survival rates.

**Step 4: Use aRules**

1. Use aRules to calculate some rules (clusters) for the titanic dataset.

2. Visualize the results.

3. Pick the three most interesting and useful rules.

4. How does this compare to the descriptive analysis we did on the same dataset?