Tutorial Business Analytics

Tutorial 5

Exercise 5.1

The following table contains empirical values about your past decisions whether or not to play.

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

a) Calculate the rule set using the 1-Rule classification. Would you play if it is windy with high humidity, a sunny outlook and cool temperature? Utilize the 1-Rule classification to determine the answer.

b) Would you play if it is windy with high humidity, a sunny outlook and cool temperature? Determine the answer using Naïve Bayes classification.

c) Would you play if the evidence of outlook from b) were changed to overcast? Determine the answer using Naïve Bayes classification.

Exercise 5.2 - Bayesian Networks

Consider the following Boolean random variables related to the state of a given car:

- Battery (B): is the battery charged?
- Fuel (F): is fuel tank empty?
- Ignition (I): does the ignition system work?
- Moves (M): does the car move?
- Radio (R): can the radio be switched on?
- Starts (S): does the engine fire?

a) Represent the joint probability density function using a Bayesian network.

b) Rewrite the probabilities using the chain rule, after defining a proper set of causal relationships between the variables.

Exercise 5.3

<u>Note</u>: Use R to solve this exercise(Exercise 5.3_R-template.R).

Load the training data ("loan-train.csv") and the test data("loan-test.csv") into R. Proceed by typing *names(train)* to print the attribute names to the console.

```
library(tidyverse)
train = read_csv("admit-train.csv")
test  = read_csv("admit-test.csv")
names(train)
```

The dependent binary variable *"loan"* indicates whether an installment loan has been repaid without any issues. Except for the attribute *"age"*, which is numerical, all other independent variables are categorical. You will find further information on what each variable tells us in the file *"variables.rtf"*. We want to create a prediction model using Naïve Bayes.
<u>Note:</u> Not all functions you need are given in the exercise definition, check the provided R-template script for them.

a) Transform the independent attribute *"age"* into a categorical attribute by placing the values into buckets.
   Why are categorical variables preferable when using Naïve Bayes? What problems can occur with numerical data?

b) For using Naïve Bayes functions, import the *"e1071"* library first.

```
library(e1071)
```

   Iterate through the independent attributes to find the most suitable attribute for the 1-rule classification. What attribute would you use for a 1-rule classification?

c) Create a prediction model using the Naïve Bayes classifier and apply it on the test-dataset. Build a confusion matrix and determine the model's error rate.