Contents

1 Data Merge .................................................. 3
  1.1 List and Data Frame .................................. 3
    1.1.1 List ............................................. 3
    1.1.2 Data Frame .................................... 4
  1.2 Merging Data Frames ................................. 4
    1.2.1 merge() ........................................ 4
    1.2.2 read.table() and read.csv() .................... 5
  1.3 Example Description ................................ 5
  1.4 Links to Datasets .................................... 5
  1.5 R Code ................................................ 5
    1.5.1 Commented Code ................................. 5
    1.5.2 Uncommented Code ................................ 5
  1.6 References ............................................. 5

2 Feature Selection .......................................... 6
  2.1 Introduction to Feature Selection ................. 7
    2.1.1 Information Theory, Entropy, CMI ............ 7
    2.1.2 Conditional Mutual Information Method ...... 7
    2.1.3 Other Feature Selection Methods ............... 7
  2.2 CMIM Implementations ............................... 7
    2.2.1 Standard Implementation ........................ 7
    2.2.2 Fast Implementation ............................ 7
  2.3 R Code ................................................. 7
    2.3.1 Commented Code ................................ 7
    2.3.2 Uncommented Code ................................ 7
  2.4 Reference ............................................... 7
3 Introduction to CVRG

4 Classification Tree Model 9
  4.1 Introduction to Classification Trees ....................... 9
  4.1.1 Comparison Indicator .................................. 9
  4.2 Instruction for the tree.revised() Function ............... 9
  4.3 R Code .................................................... 9
    4.3.1 Commented .......................................... 9
    4.3.2 Uncommented ......................................... 9
  4.4 Reference ................................................ 9

5 Painted Tree Classifier 10
  5.1 Painted Tree Classifier ................................. 10
    5.1.1 50/50 Adjustment ................................ 10
    5.1.2 Pruned Tree Classifier ............................ 10
    5.1.3 Leave One Out Cross Validation (LOOCV) ............. 10
  5.2 R Code .................................................... 10
    5.2.1 Commented Code ...................................... 10
    5.2.2 Uncommented Code .................................... 10

6 Quadrant Tree Classifier 11
  6.1 Quadrant Tree Classifier ................................. 11
    6.1.1 Sensitivity and Specificity ........................ 11
    6.1.2 Bayesian Theorem .................................. 11
    6.1.3 Survival Analysis .................................. 11
    6.1.4 Quadrant Tree Classifier ............................ 11
  6.2 R Code .................................................... 11
    6.2.1 Commented Code ...................................... 11
    6.2.2 Uncommented Code .................................... 11
Chapter 1

Data Merge

1.1 List and Data Frame

1.1.1 List

An R list is an object consisting of an ordered collection of objects known as its components. There is no particular need for the components to be of the same mode or type, and, for example, a list could consist of a numeric vector, a logical value, a matrix, a complex vector, a character array, a function, and so on.

Here is a simple example of how to make a list:

```r
Lst <- list(name="Fred", wife="Mary", no.children=3, child.ages=c(4,7,9))
```

Components of lists may be named, and referred to either by giving the component name as a character string or the number in double square brackets. For example:

```r
> Lst[[1]]
[1] "Fred"
> Lst$wife
[1] "Mary"
> Lst[[2]]
[1] "Mary"
> Lst$child.ages
[1] 4 7 9
```

This is a very useful convention as it makes it easier to get access to each individual component of lists.
1.1.2 Data Frame

A data frame is a format of dataset that we frequently use in R. It is a basically a list of data. It can be a vector, a matrix, or any multidimensional array. The definition of data frame is a list data with class "data.frame". There are restrictions on lists that may be made into data frames, namely

(1) The components must be vectors (numeric, character, or logical), factors, numeric matrices, lists, or other data frames; (2) Matrices, lists, and data frames provide as many variables to the new data frame as they have columns, elements, or variables, respectively; (3) Numeric vectors, logicals and factors are included as is, and character vectors are coerced to be factors, whose levels are the unique values appearing in the vector; (4) Vector structures appearing as variables of the data frame must all have the same length, and matrix structures must all have the same row size.

A data frame may for many purposes be regarded as a matrix with columns possibly of differing modes and attributes. It may be displayed in matrix form, and its rows and columns extracted using matrix indexing conventions, for example:

```R
> x<-data.frame(a=1:4,b=c(TRUE,TRUE,FALSE,TRUE),c=c("A","B","C","D"),d=13:16)
> x
   a   b   c   d
1 1 TRUE A 13
2 2 TRUE B 14
3 3 FALSE C 15
4 4 TRUE D 16
> x[,1] # the 1st column of x
[1] 1 2 3 4
> x[,2] # the 2nd column of x
[1] TRUE TRUE FALSE TRUE
> x[1,] # the 1st row of x
   a   b   c   d
1 1 TRUE A 13
> x[3,] # the 3rd row of x
   a   b   c   d
3 3 FALSE C 15
> x[3,4] # the element in 3rd row and 4th column
[1] 15
```

1.2 Merging Data Frames

1.2.1 merge()

The reason why we need merging data frames is that we often get several data files for one project. Each file contains one piece of information. In order to
have a thorough analysis, we need to have all files put together in an organized way so that the analysis can be done on the whole dataset easily.

Merging data frames (merge() ) can put two data frames together based on the common column names or row names. After the merging, the rows are by default lexicographically sorted on the common columns, unless `sort = TRUE` is specified in the merge() function. Other detailed argument explanation in merge() can be found in R help files by typing `?merge()` in R console, or simply click online help file. Merging datasets example:

In order to merge data frames, we need to have all individual data files read as data frames in R. So read.table() or read.csv() are needed here.

1.2.2 read.table() and read.csv()

1.3 Example Description

1.4 Links to Datasets

1.5 R Code

1.5.1 Commented Code

1.5.2 Uncommented Code

1.6 References
Chapter 2

Feature Selection

2.1 Introduction to Feature Selection

2.1.1 Information Theory, Entropy, CMI

2.1.2 Conditional Mutual Information Method

2.1.3 Other Feature Selection Methods

Random Sampling

Mutual Information Maximization

C4.5 Binary Trees

Fast Correlation-Based Filter

2.2 CMIM Implementations

2.2.1 Standard Implementation

2.2.2 Fast Implementation

2.3 R Code

2.3.1 Commented Code

Standard Implementation

Fast Implementation

2.3.2 Uncommented Code

Standard Implementation

Fast Implementation

2.4 Reference
Chapter 3

Introduction to CVRG
Chapter 4

Classification Tree Model

4.1 Introduction to Classification Trees
4.1.1 Comparison Indicator
4.2 Instruction for the tree.revised() Function
4.3 R Code
4.3.1 Commented
4.3.2 Uncommented
4.4 Reference
Chapter 5

Painted Tree Classifier

5.1 Painted Tree Classifier

5.1.1 50/50 Adjustment

5.1.2 Pruned Tree Classifier

5.1.3 Leave One Out Cross Validation (LOOCV)

5.2 R Code

5.2.1 Commented Code

5.2.2 Uncommented Code
Chapter 6

Quadrant Tree Classifier

6.1 Quadrant Tree Classifier

6.1.1 Sensitivity and Specificity

6.1.2 Bayesian Theorem

6.1.3 Survival Analysis

6.1.4 Quadrant Tree Classifier

6.2 R Code

6.2.1 Commented Code

6.2.2 Uncommented Code

\[ e = m \cdot c^2, \]  \hspace{2cm} (6.1)

\[ \sum_{k=1}^{n} I_k = 0. \]  \hspace{2cm} (6.2)

Kirchhoff’s voltage law can be derived . . . . which has several advantages.

\[ I_D = I_F - I_R \]  \hspace{2cm} (6.3)

is the core of a very different transistor model. . . .