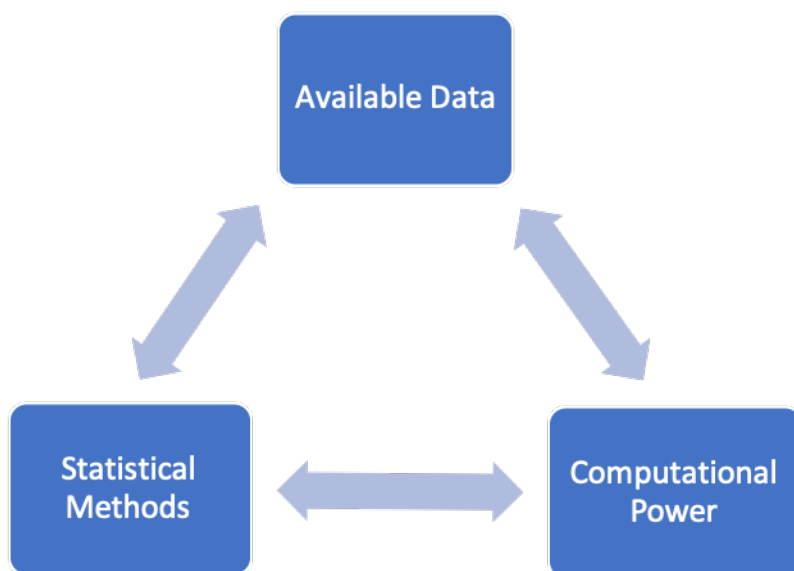


Decision Trees and Random Forests

Shaurya Jauhari, Mora Lab, GMU. (Email: shauryajauhari@gzhmu.edu.cn)

2019-05-31

Welcome to the second workshop in the series of **Machine Learning Fundamentals**. In this session, we shall explore the theory of **Decision Trees** and scale it to the broader concept of **Random Forests**. Decision trees epitomize the divide and conquer strategy to accomplish classification tasks (although they could also be implemented for regression chores as well) (Quinlan 1986, Rokach and Maimon (2005)). From the previous session on **Logistic Regression**, you may recall that if the response variable is categorical in nature (factors in R), we aim classification, or else, if the response variable is continuous, it denotes regression. Classification and regression methodologies are both categorized under supervised machine learning. The theme of machine learning was realized in 1950s and since then, coupled with the data deluge and upheaval in computational prowess, has exhibited stronghold in data analysis domain.



Decision Trees are rudimentary classification algorithms that, at a low-level, are synonymous to *if-then* conditional statements in programming languages. They follow the strategy of iterative recursion, and intuitively the leaf nodes hold the final verdict. The highest aggregate from all leaf nodes (terminals) is graded as the output of that decision tree.

In this module, we shall delve into creation of basic decision trees to have an understanding of it. For the purpose, we shall load the package **party** and make use of the function **ctree()** to calculate and analyze decision trees.

```
install.packages("party", dependencies = TRUE, repos = "https://mirrors.tuna.tsinghua.edu.cn/CRAN/")

## also installing the dependencies 'SQUAREM', 'lava', 'lmtree', 'proclim', 'measures', 'vcd', 'ipred',
##
## The downloaded binary packages are in
## /var/folders/hm/c3_fjypn62v5xh5b5ygv267m0000gn/T//Rtmp4nYh5f/downloaded_packages
library(party)

## Loading required package: grid
## Loading required package: mvtnorm
## Loading required package: modeltools
## Loading required package: stats4
## Loading required package: strucchange
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

## Loading required package: sandwich
```

All workshop study material is available at my github page (Jauhari, n.d.).

References

- Jauhari, Shaurya. n.d. “My Github Repository.” <https://github.com/shauryajauhari/Machine-Learning>.
- Quinlan, J R. 1986. “Induction of Decision Trees.” *Machine Learning* 1: 81–106.
- Rokach, Lior, and Oded Maimon. 2005. “Decision Tree.” *Data Mining and Knowledge Discovery Handbook*, pp 165–92. doi:10.1007/978-0-387-09823-4_9.