

Predictive Analytics using R

Syllabus

Professor: Luís Torgo (ltorgo@dal.ca)
Class Dates: 11,13/Mar/2019, 2 x 5h/day

1. Course Overview

The main goal of this short course is to provide hands-on experience on key predictive analytics methods using the R environment. The main focus of the short course will be on **regression methods** (i.e. numeric prediction).

R is currently one of the most used data mining tools witnessing widespread acceptance both in academia and industry. One of the reasons for this success lies on the huge amount of tools and methods that users can access freely. In this short course we will illustrate some of these methods using a hands-on approach.

After taking this course you should be able to use R for:

- 1) *Master frequently used predictive modeling techniques.* Data can be modeled in many different ways. The outcomes of these models can provide useful information for decision makers. We will address several concrete modeling tasks with frequently used techniques. We will learn how to obtain and apply these models in R.
- 2) *Correctly assess the performance of models.* Performance assessment is a key step for taking advantage of the results of data mining models. Being able to carry out this task in a reliable way is of key importance to make sure future deployment of data mining pays off.

2. Focus and Interaction

The course will illustrate the use of R for several predictive tasks. The main focus will be on how to carry out these tasks in R and not on the principles and theory behind these approaches. This means that this is a practical course that will illustrate several of the existing techniques for handling this type of problems. The course will include concrete case studies and solutions will be provided to allow you to replicate and re-use the approaches followed, in the spirit of open source software like R.

3. Readings

To follow the course you will not need to read any extra material than that provided in the course at the respective Web page:

http://web.cs.dal.ca/~ltorgo/PA_VAL

Nevertheless, we provide a series of extra sources of information that some of you may find useful in case you wish to continue your studies in this area:

- L. Torgo. *Data Mining with R: Learning with Case Studies* (2nd edition). Chapman and Hall/CRC, 2017. ISBN 9781482234893.
- D. Nolan and D. Temple Lang. *Data Science in R*. CRC Press, 2015. ISBN 9781482234817.

Other books:

- H. Wickham. *Advanced R*. CRC Press, 2014. ISBN 9781466586963 .

4. Software

This course is about using R for predictive analytics. In this context, you are required to have an up-to-date installation of R in your laptop. R can be freely obtained from <http://www.r-project.org> . RStudio is another free software tool that provides an integrated development environment to R. I strongly recommend that you use RStudio as your tool for interacting with R. RStudio can be freely downloaded from <http://www.rstudio.com> .

R comes with an extensive set of tools pre-installed. Still, it can be easily extended through the (free) installation of extra packages. We will use several of these. You can easily install any of these extra packages in RStudio.

IMPORTANT: *In order to install the above mentioned software you must have access to a computer on which you can install software. Moreover, due to the practical nature of this course you are expected to have your laptop with you during the classes, as it will be necessary for hands-on exercises.*