

MACHINE LEARNING COURSE
UNIVERSITY OF PARMA - A.Y. 2020/2021

An introduction to the Caret package

AUTHOR: Francesco Vetere
E-MAIL: `francesco.vetere@studenti.unipr.it`

1 Introduction

Caret (short for **C**lassification **A**nd **R**Egression **T**raining) is a comprehensive framework for building machine learning models in R.

R is a language and environment for statistical computing and graphics, widely used in AI and ML applications.

It is open-source, provides many statistical techniques (such as statistical tests, classification, clustering, ...), and has many packages that can be used to solve different problems.

Sometimes the syntax and the way to implement ML algorithms differ across packages: **Caret**, an extremely useful package for R, provides a uniform interface for using the various modeling functions.

In particular, it offers tools for data splitting, data pre-processing, model creation and tuning, and many more.

The aim of this paper is to introduce the developer to the Caret package: starting from the installation process, it will be then shown how to use Caret to build machine learning models, focusing in particular on Multi-Layer Perceptron. Finally, some practical examples will be presented.

2 Installation and prerequisites

R is a multi-platform environment, available for various OS.
In this paper, the installation process will be shown for Ubuntu 20.04 LTS
(however, it is very similar for other Linux distros).

Precompiled binaries of R are available for various OS on **CRAN** (<https://cloud.r-project.org/>),
a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R.

However, with Ubuntu it is also possible to install it directly from terminal with `sudo`, or as a root user:

```
# update indices
apt update -qq

# install two helper packages we need
apt install --no-install-recommends software-properties-common dirmngr

# import the signing key (by Michael Rutter) for these repo
apt-key adv --keyserver keyserver.ubuntu.com --recv-keys
    E298A3A825C0D65DFD57CBB651716619E084DAB9

# add the R 4.0 repo from CRAN -- adjust 'focal' to 'groovy' or 'bionic' as
    needed
add-apt-repository "deb https://cloud.r-project.org/bin/linux/ubuntu $(
    lsb_release -cs)-cran40/"

# install R and its dependencies
apt install --no-install-recommends r-base
```

Now, R and its dependencies are installed: typing the command `R` in the terminal should launch the command interpreter.

In order to write R code, a very popular IDE is **RStudio**, available for many platforms (<https://www.rstudio.com/products/rstudio/download/>).

Once the correct version of the software has been chosen, a `.deb` package will be downloaded: it is useful to install it with `gdebi`, a command that will ensure that all additional prerequisites for RStudio are fulfilled (such as `clang` and others).

```
# install RStudio and its dependencies
gdebi rstudio-1.4.1717-amd64.deb
```

In order to install the Caret package, it's sufficient to open RStudio (or directly the R interpreter) and execute the following line of code:

```
install.packages('caret', dependencies = TRUE)
```

Caret will be downloaded from CRAN, together with its dependencies.

Once the process is ended, Caret is ready to be included in a normal R program with the following directive:

```
library('caret')
```
