Installation Instructions for Lab 1 and Lab 2 "Big data and predictive models"

Y. Pradat (MICS, CentraleSupelec),
L. Verlingue (Institut Gustave Roussy)
D. Gautheret (Institut Gustave Roussy)

Friday 24$^{\text{th}}$ January, 2020

The following document is a notice for installing the programming tools you will need in order to put into praticse the algorithms you saw in the classes. In brief, many different models exists, from the simple linear-regression to the more elaborate deep learning models that are parametrized by hundreds of thousands or even millions of parameters. In order to implement these models, we shall use code that has been developed by other people and which reduces the task to a handful of lines of code. This code is available in the form of **libraries**, some of which you may already have installed and some that you may not.

1 Git & GitHub

1.1 Git install

Git is "a distributed version-control system for tracking changes in source code during software development". In other words, it is a very popular tool that allow to efficiently keep track of different versions of files (code, text, images, etc.) so that you may revert to *any previous version* without any hassle. It is also extremely complete and resourceful for team projects that involve some sort of code collaboration (as for the preparation of the labs you will follow).

If you do not have Git installed on your computer, please install it following the instructions provided here according to your operating system. Once Git is installed, you should have access to a *Git bash* on Windows while a bash (command-line interface) is already available on Linux&MacOS systems. **Please check** that the bash command

```
git ——version
```

gives you the version of Git installed on your machine.

1.2 GitHub explained

GitHub is a company that provides hosting services (in the form of a platform www.github.com) with version control using Git. If you do not already have a GitHub account, please sign up at GitHub. Once this is done, you may download (*clone* in Git language) any *repository* available to you on GitHub, i.e public repositories or repositories to which you have been granted access.

The repository for the labs 1 and 2 of "Big data and predictive models" is available at https://github.com/gustaveroussy/IFSBM-bigdata. From the bash interface, change directory (using the command `cd`) to whichever location on your machine where you want to have the folder for the labs. Then, **execute**

```
git clone https://github.com/gustaveroussy/IFSBM-bigdata
```

2 Installation of an R environment

In order to follow the next instructions, **please open** `Rstudio` and then `File > New Project > Existing Directory > IFSM-bigdata/TP_IFSBM_module12/2020/labs > create project`. This sets the *working directory* to `labs` and creates a `labs.Rproj` in the folder with settings used by `Rstudio`. Otherwise, use a command-line interface (bash) located at the folder previously specified.

In order to facilitate the management of `R` libraries, we use **packrat**. In short, packrat makes R projects more isolated, more portable and *reproducible*. In the folder `labs` you have a subfolder named `packrat` which contains the source code (in `packrat/src`) and binaries (in `packrat/lib`, `packrat/lib-R`, `packrat/lib-ext`) of all the libraries we shall need for the labs.

You can recreate the environment with all the required libraries with *one of the following*

- (Rstudio console): packrat :: restore ()

- (Bash from the folder labs/ - Please make sure the R interpreter used here is the same as the R interpreter you use in Rstudio):

   R −e ''packrat :: restore ()''

3 Install Keras & Test

Keras is an open source *neural network library* written in Python. It is also available in R through a wrapper described here. For you to be

able to use keras, you shall execute the R comand keras :: **install** _keras().
This has already been written for you in the example Rmarkdown
`labs/lab_2/src/TP_example.Rmd`. Please do *one of the following* to run
the example file so that you can test your installation:

- (Rstudio console): Use the `Knit` button to generate a html rendering
  of the Rmarkdown.

- (Bash from the folder **labs/**): Run the associated R script.

$$\text{Rscript } \text{lab\_2/src/TP}\_\textbf{example}.\textbf{R}$$

You are more than welcome to look at the code (in the html or in the Rmd)
so as to understand how a *neural net architecture* may be implemented with
keras.