**Effort estimation with statistical and ML techniques**

**Project Management**

**Sergio de la Mata Moratilla**

**3º GII E**

**Index:**

Bibliography: 5

Considerations: 4

Estimation Models: 3

Evaluation Measures: 3

Introduction: 3

Other Techniques: 3

Summary Practice: 3

# Introduction:

This document contains all the information regarding the performance of the practice. It will be explained briefly what it was made, the estimation models used, the evaluation measures used on them and some techniques used for the performance of the practice.

# Summary Practice:

# Estimation Models:

# Evaluation Measures:

# Other Techniques:

# Considerations:

Regarding this practice, it is needed to consider that:

* It was made using RStudio IDE in order to use R language.
* All the packages used for both R files it was needed to be the first time installed with the command ‘install.packages(“name\_package”)’
* To execute any of the operations, it is needed to do the next sequence of keys: Ctrl + Enter.
* The packages used for a practice need to be loaded each time, RStudio is initialized. The command needed is “library(name\_package)”.
* At the linear regression it is needed sometimes to execute again the command given for its representation to have a correct picture of data.
* The practice is divided in different sections which can be done at the order the user wants, but it is compulsory to have executed the operations established before the section for correlation coefficient and covariance.
* It is needed to consider that the section related to the cross validation it is recommended to be executed after or without executing the rest of sections, as the variables “data\_train” and “data\_test” will be changed from the one given initially making to obtain the same results as when it is made alone.
* The operations at the function used for the cross validation are the same as the ones used for the rest of sections, but with the difference that the results obtained each operation are printed at screen and it is not needed to execute them.
* The idea of having both representations at the R document is to show a specific study having specific set of train and test data using the approximations used for this kind of studies (70-80% of data are for training and 20-30% of data are for testing) distributing them in a random way, and using a cross validation (in this case, the set of train data could only be made of one element).
* The number of validations which can be made with cross validation can be changed at any moment on the source code depending on the specifications of the user.

# Bibliography:

* Correlations in R: <https://www.statmethods.net/stats/correlations.html>
* Read table in R: <https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/read.table>
* Data structures in R: <http://ocw.uc3m.es/estadistica/aprendizaje-del-software-estadistico-r-un-entorno-para-simulacion-y-computacion-estadistica/estructuras-de-datos-en-r>
* Introduction to R: <https://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>
* Package mlr: <https://cran.r-project.org/web/packages/mlr/mlr.pdf>
* Model Training and Turing: <https://topepo.github.io/caret/model-training-and-tuning.html>
* Confidence Interval: <https://fhernanb.github.io/Manual-de-R/ic.html>
* Measures of Evaluation in Software Engineering: <http://danrodgar.github.io/DASE/evaluationSE.html>
* Mean Squared Error in Linear Regression in R: <https://stats.stackexchange.com/questions/107643/how-to-get-the-value-of-mean-squared-error-in-a-linear-regression-in-r>
* Package ftsa, error function: <https://www.rdocumentation.org/packages/ftsa/versions/3.4/topics/error>
* Documentation package Metrics: <https://cran.r-project.org/web/packages/Metrics/Metrics.pdf>
* Forecast function in R: <https://www.rdocumentation.org/packages/forecast/versions/8.12/topics/forecast>
* Documentation package ftsa: <https://cran.r-project.org/web/packages/ftsa/ftsa.pdf>
* Forecast error measure in R: <http://finzi.psych.upenn.edu/library/ftsa/html/error.html>
* Forecast error measure in R: <https://rdrr.io/cran/ftsa/man/error.html>
* Simple and multiple logistic regression in R: <https://rpubs.com/Joaquin_AR/229736>
* Introduction Artificial Neural Networks in R: <https://rpubs.com/rdelgado/402754>
* Aggregation and Restructuring data in R: <https://www.r-statistics.com/tag/transpose/>
* Conversion H2OFrame to a data frame: <https://stackoverflow.com/questions/43189340/how-to-convert-a-column-in-h2oframe-to-a-python-list>
* Median in R need numeric data: <https://stackoverflow.com/questions/13204008/median-in-r-need-numeric-data>
* Cross validation in R: <https://www.diegocalvo.es/validacion-cruzada-en-r/>
* Conditionals in R: <https://stackoverflow.com/questions/54825426/how-to-use-conditional-statement-and-return-value-for-a-function-in-r>
* Add calculated fields to data in R: <https://www.dummies.com/programming/r/how-to-add-calculated-fields-to-data-in-r/>
* Logistic Regression in R: <http://idaejin.github.io/courses/R/2019/euskaltel/regresion-logistica.html>
* Simple and multiple logistic regression: <https://rpubs.com/Cristina_Gil/Regresion_Logistica>
* Example logistic regression with R: <https://rpubs.com/emilopezcano/logit>
* Simple and multiple logistic regression in R: <https://rpubs.com/Joaquin_AR/229736>
* Creation of random numbers from time stamp: <https://decisionstats.com/2013/09/28/using-r-for-random-number-creation-from-time-stamps-rstats/>
* Introduction to cross validation in R: <https://rpubs.com/rdelgado/405322>
* Minimum sample size to make a correlation test in R: <https://www.researchgate.net/post/What_is_the_minimum_sample_size_to_run_Pearsons_R>