

Homework 9

Juan Diego Sánchez-Torres.¹

21 April, 2021

¹ Department of Mathematics and
Physics, ITESO. dsanchez@iteso.mx

This assignment will solve several linear regression and logistic regression problems using some R software packages. This tool allows the estimation and validation of this type of model; besides, it presents methods for selecting and analyzing the explanatory variables.

Introduction

Of the models with a convex representation of their parametric estimation, generalized linear models (GLM) are a crucial case. The two most frequent examples of GLM are linear regression and logistic regression. Therefore, in this activity, several linear regression and logistic regression exercises will be solved with R software and some of its packages.

Activities

Let's start with an introduction to linear regression and logistic regression using the R packages *olsrr* and *blorr*.

Problem 1: Warm up

Note that some details are missing for all the following examples, the problems lack a complete explanation, and the code may need adequate comments. In this form, you must present a proper mathematical formulation, a brief background of the problem (and its bibliographical references) and, a much better explanation.

1. The *olsrr* Package

- (a) Introduction to *olsrr*
- (b) Variable Selection Methods
- (c) Residual Diagnostics
- (d) Heteroscedasticity
- (e) Measures of Influence
- (f) Collinearity Diagnostics, Model Fit & Variable Contribution

2. The *blorr* Package

- (a) A Short Introduction to the *blorr* Package

Upload your results to [Github](#) in the form of a [R Notebook](#).

Then, consider the following application to real problems

Problem 2: Application Problems

1. Solve a regression problem of your choice from the [UC Irvine Machine Learning Repository](#).
2. Solve a classification problem of your choice from the [UC Irvine Machine Learning Repository](#).

Upload your results to [Github](#) in the form of a [R Notebook](#).

Note: Do not rule out the use of symbolic transformers to improve the performance of your model, which should be a linear regression or logistic regression, depending on the application.

Further Lectures and Bibliography

Consider the following references:

- Dobson's book¹ (Chapter 3., pp. 49-64) offers an overview of the exponential family and generalized linear models.
- James' book on introductory statistical learning theory².

References

- [1] A.J. Dobson and A.G. Barnett. *An Introduction to Generalized Linear Models, Fourth Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2018. ISBN 9781138741515. URL <https://www.routledge.com/An-Introduction-to-Generalized-Linear-Models/Dobson-Barnett/p/book/9781138741515>.
- [2] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer New York, 2013. ISBN 9781461471387. URL <https://www.statlearning.com/>.