

Shrinkage linear regression for symbolic interval-valued variables

Oldemar Rodríguez Rojas *

Abstract

This paper proposes a new approach to fit a linear regression for symbolic interval-valued variables, which improves both the Center Method suggested by Billard and Diday in [2] and the Center and Range Method suggested by Lima-Neto, E.A. and De Carvalho, F.A.T. in [9, 10]. Just in the Centers Method and the Center and Range Method, the new methods proposed fit the linear regression model on the midpoints and in the half of the length of the intervals as an additional variable (ranges) assumed by the predictor variables in the training data set, but to make these fitments in the regression models, the methods Ridge Regression, Lasso, and Elastic Net proposed by Tibshirani, R. Hastie, T., and Zou H in [12, 8] are used. The prediction of the lower and upper of the interval response (dependent) variable is carried out from their midpoints and ranges, which are estimated from the linear regression models with shrinkage generated in the midpoints and the ranges of the interval-valued predictors. Methods presented in this document are applied to three real data sets “cardiologic interval data set”, “Prostate interval data set” and “US Murder interval data set” to then compare their performance and facility of interpretation regarding the Center Method and the Center and Range Method. For this evaluation, the root-mean-squared error and the correlation coefficient are used. Besides, the reader may use all the methods presented herein and verify the results using the RSDA package written in R language, that can be downloaded and installed directly from CRAN [14],

Keywords

Interval-valued variables, Linear Regression, Elastic Net, Lasso, Ridge Regression, Symbolic Data Analysis.

1 Introduction

Statistical and data mining methods have been developed mainly in the case in which variables take a single value. Nevertheless, in real life there are many situations in which the use of this type of

*University of Costa Rica, San José, Costa Rica; E-Mail: oldemar.rodriguez@ucr.ac.cr