

Comparison and Assessment of Shrinkage Methods in Case of Multicollinearity Problem

Şevval Kılıçoğlu and Fatma Yerlikaya-Özkurt

Department of Industrial Engineering, Atılım University, Ankara, Turkey

Abstract

Nowadays, the data analysis and interpretation are very important in many fields of science. One of the most preferred methods in data analysis is linear regression due to its simplicity to interpret and ease of application. Linear regression models are used to explain the relationship between a dependent variable and one or more independent variables. One of the assumptions accepted while obtaining linear regression models is that there is no correlation between the independent variables in the model which refers to absence of multicollinearity. As a result of multicollinearity, the variance of the parameter estimates will be high and this reduces the accuracy and reliability of the linear model.

The most powerful and preferred methods for eliminating the multicollinearity problem are shrinkage methods. Shrinkage methods aim to handle the multicollinearity problem by minimizing the variance of the estimators in the model. Ridge Regression, LASSO, and Elastic-Net are well-known and popular shrinkage methods. These methods converge the values of the coefficients of the variables in the model to zero or very close to zero values.

In this study, mentioned shrinkage methods were applied to different simulated data sets with different characteristics and also three real-world data sets. After all models created, some performance criteria were calculated for each method to determine which method gives better results in the data set in which characters. Based on performance results, the methods were compared with TOPSIS, which is one of the Multi-Criteria Decision Making Methods, and the order of preference was determined for each data set.

Keywords

Multicollinearity, Shrinkage methods, Ridge regression, LASSO, Elastic Net, Multi criteria decision making, TOPSIS.

References:

Hoerl, A. E., and R. W. Kennard, (1970). Ridge Regression: Biased Estimation for Non-orthogonal Problems, *Technometrics*, vol. 12, no. 1, pp. 55–67.

- Jato-Espino, D., E. Castillo-Lopez, J. Rodriguez-Hernandez, and J. C. Canteras-Jordana, (2014). A review of application of multi-criteria decision making methods in construction, *Autom. Constr.*, vol. 45, pp. 151–162.
- Tibshirani, R., (1996). Regression Shrinkage and Selection via the Lasso, *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288.
- Uyanık, G. K., and N. Güler, (2013). A Study on Multiple Linear Regression Analysis, *Procedia - Soc. Behav. Sci.*, vol. 106, pp. 234–240.
- Zou, H., and T. Hastie, (2005). Regularization and variable selection via the elastic net, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320.