

Lasso Regression: From Explanation to Prediction

Lijin Zhang¹, Xiayan Wei^{1,2}, Jiaqi Lu^{1,2}, and Junhao Pan¹

1 Department of Psychology, Sun Yat-sen University, Guangzhou, China

2 Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China

Abstract: The purpose of psychology research is to "describe, explain and predict behavior." Exploring the relationship between variables is an important part of achieving this goal. Regression analysis, as a method of evaluating the relationship between variables, is widely used in psychological research. However, the traditional ordinary least squares regression method focuses on the interpretation of the current data, which tends to result in the over-fitting problem and weaken the generalizability of the model. Besides, when there is multicollinearity among independent variables in the regression model, the coefficients obtained by least-squares estimation are easily affected by the small fluctuations of the sample data, and the stability of the estimation is also poor. These problems have nonnegligible influence on the promotion and prediction of the model conclusion.

With the rapid development of the methodology, more and more statistical tools have emerged to better compensate for the limitations of traditional methods. By introducing a penalty term in the model and shrinking the regression coefficients to zero, the Lasso regression method can obtain higher model prediction accuracy and model generalizability at the expense of a certain estimation bias. Besides, Lasso regression can effectively deal with the problem of multicollinearity in the model, and it has been widely used in medicine, economics, neuroscience and other fields.

In the field of psychology, more and more researchers have pointed out that due to the limitations of computer computing power, researchers mainly focus on the use of hypothesis testing to verify the theory and explain the relationships between variables. However, with the vigorous development of the machine learning, the shift from the excessive attention to the interpretation of the regression coefficients to the ability of predicting has increasingly become an important development trend in the field of psychology.

Therefore, based on the fundamental theory and the real data analysis, this paper introduces the principle, implementation steps and advantages of the Lasso regression method. It is hoped that more and more applied researchers will be called upon to focus on the emerging tools of data science to promote the development of psychology with the help of data science.

Keywords: regression, regularization, Lasso, prediction