International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Study of World Country Happiness in Lasso Model

Shuwen Yang[1], Tao Jia[2], Yaoxuan Luan[3], Yuan Lu[4], Ruobai Zhao[5]

[1]School of Finance and Public Administration, Tianjin University of Finance and Economics, Zhujiang Road 25, Tianjin, 300222, China

[2]College of Electrical and Power Engineering, Taiyuan University of Technology, Ying Ze Xi Da Jie 79, Taiyuan, Shanxi, 030024, China

[3]Department of Computer Science and Software Engineering, Auburn University, 23 Samford Hall, Auburn, AL 36849, United States

[4]School of Business, University of International Business and Economics, Hui Xin Dong Jie 10, Chaoyang, Beijing 100029, China

[5]Missouri University of Science and Technology, Rolla, MO USA

Abstract: ***Study in the state of world countries' happiness begins to gain global attention and recognition. Several data sources were established to show critical factors related to the happiness of individual countries. The data that this study used is from World Happiness Report, which is a leading data source in this area. This World Happiness Report was first published in 2012. This article involved all factors that the 2019 World Happiness Report used and tried to identify the most useful information by using Lasso model to achieve the dimension reduction of the beneficial factors.***

1. Introduction

There are a total of 159 countries involved in this work. The response variable of this study is the happiness score (y). It ranges from 2.85 to 7.77. Higher score yields a better happiness level. Finland had the highest score, which is 7.769, and South Sudan got the lowest score, 2.853. The key factors that the World Happiness Report claimed to influence the happiness level are GDP per capita (x_1), Social support (x_2), Healthy life expectancy (x_3), Freedom to make life choices (x_4), Generosity (x_5), and Perceptions of corruption (x_6).

The response variable is explained by the six covariates. To check how good these factors explained the response variable, we can start with the classical method, least squares regression, or multiple regression. We fit a model, $Y = X\beta + \varepsilon$, where Y is the response vector, β is the coefficients vector, X is the design matrix, and ε is the error term. The estimated coefficients $\hat{\beta} = (x'x)^{-1}x'y$, which is to minimize $SSE = \sum(y - x\hat{\beta})^2$. We refer SSE as the sum of squared errors. The p-values of each coefficient can then be obtained to determine whether the factor is statistically significant. Furthermore, we can check whether some of the six covariates can be excluded from the study. In other words, which factors are the most important in terms of the country's happiness level.

2. Methodology

It is common to use the general model selection method, such as forward selection, backward selection, or stepwise selection. However, these methods require enough data information to fit the full model. This requirement will be violated when the dimension of covariates is large or when there are a lot of factors to be considered, or when the sample size is small. That is why we propose this least absolute shrinkage and selection operator (LASSO) method. It is a regression analysis method that performs both variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model it produces.

Other than minimizing the $SSE = \sum(y - x\hat{\beta})^2$ itself, we will minimize the $(SSE + \lambda \parallel \beta \parallel_1)$. This $\lambda \parallel \beta \parallel_1$ term is called the penalty term, which controls the magnitude in selecting the covariates (factors). λ is called the tuning parameter, which stands for how strong we want to reduce the dimension of the covariates. This $\parallel \beta \parallel_1$ term is referred to as L_1 norm, which is the sum of the absolute value of all coefficients ($\beta's$).

3. Results

The R package, *glmnet*, is introduced to run the analysis. The tuning parameter λ ranges from 0 to $+\infty$. Usually, the best λ is determined by cross-validation. However, for this study, we intentionally want to reduce the dimension; hence, we fix $\lambda = 0.1$, which is a moderate strength in dimension reduction. The estimated coefficients are reported in Table 1.

Table 1: Estimated coefficients when $\lambda = 0.1$

$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$	$\widehat{\beta_6}$
2.33	0.71	0.97	0.99	1.27	.	0.43

The Lasso model successfully reduced the dimension, which is to delete the fifth covariate. That is the same as Generosity is the least important factor among all factors under consideration. In other words, if we have to reduce one factor due to the size of the data information, we will choose Generosity to drop.

To check the influence from adjusting the tuning parameter, we fit a second model, letting $\lambda = 0.5$. The corresponding results are listed in Table 2.

$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$	$\widehat{\beta_6}$
4.06	0.44	0.51	0.46	.	.	.

When λ increased, more factors were dropped from the model. The most important factors under this setup are GDP per capita, Social support, and Healthy life expectancy.

Licensed Under Creative Commons Attribution CC BY

4. Conclusion

The key factors that the World Happiness Report claimed to influence the happiness level are proper. If, in some situations, we have to further select from this pool, we have determined the method and corresponding results.

The happiness level has become an essential area in social research. More and more related factors will be studied. There will eventually be a large pool of covariates (factors) to explain the happiness score. This Lasso method will then be more effective in such a situation.

References

[1] Kaggle.com. (2019). World Happiness Report. Sustainable Development Solutions Network. Version 2, Nov 2019.
[2] Helliwell, J., Layard R., Sachs, J. (2012). World happiness report. The London School of Economics and Political Science. Nov 2012.
[3] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 58, No. 1, pp. 267-288.
[4] Shen, G., Gao, D., Wen, Q., Magel, R. (2016). Predicting Results of March Madness Using Three Different Methods. Journal of Sports Research. Vol 3, No.1, pp.10-17.
[5] Friedman, J. (2019). R Package 'glmnet'.

Licensed Under Creative Commons Attribution CC BY