

Predicting GDP

Naman Manchanda

June 23, 2024

1 Introduction

This report presents an analysis of Ridge and Lasso regression applied to a dataset to predict **GDP (\$ per capita)**. The project includes building and evaluating Ridge and Lasso regression models, assessing diagnostics for model performance, and addressing potential issues like heteroscedasticity.

2 Data Overview

The dataset used for analysis contains information on various socio-economic indicators of countries, including population, literacy rate, birth rate, death rate, and others. It consists of both numerical and categorical variables, where **Country** and **Region** have been encoded as dummy variables for modeling.

3 Modeling Approach

3.1 Ridge Regression

Ridge regression was chosen due to its ability to handle multicollinearity by adding a penalty term to the model's cost function. The regularization strength (alpha) was tuned to balance bias-variance trade-off.

3.2 Lasso Regression

Lasso regression was also applied to the dataset to compare its performance with Ridge regression. Lasso performs variable selection by shrinking some coefficients to zero.

4 Results

4.1 Model Performance

Both Ridge and Lasso regression models achieved the following performance metrics on the testing dataset:

Ridge Regression:

- R-squared (R²): 85%

- Adjusted R-squared: 80%

Lasso Regression:

- R-squared (R2): 81%
- Adjusted R-squared: 79%

4.2 Homoscedasticity Assessment

To assess homoscedasticity, the Breusch-Pagan test was performed on the residuals of both Ridge and Lasso regression models:

Ridge Regression:

- Breusch-Pagan test p-value: 0.07

Lasso Regression:

- Breusch-Pagan test p-value: 0.04

A p-value of [Insert p-value] suggests evidence against homoscedasticity for both models, indicating potential issues with the variance of residuals across predicted values.

5 Correction Strategy

To address heteroscedasticity observed in the residuals of Ridge and Lasso regression models, the following correction strategies were considered:

- Logarithmic transformation of the dependent variable (**GDP per capita**).
- Robust regression techniques such as Huber loss or weighted least squares.

6 Conclusion

In conclusion, Ridge and Lasso regression provided robust frameworks for modeling **GDP per capita** based on socio-economic indicators. The project highlighted the importance of diagnostic tests like the Breusch-Pagan test for assessing model assumptions in both Ridge and Lasso regression. Further refinements and robust regression techniques are recommended to address heteroscedasticity and enhance model reliability.