

Analysis of LASSO and Ridge Regression Models

Olubayode Ebenezer

April 9, 2024

1 Introduction

This report presents the findings from applying LASSO and Ridge regression models to a dataset with more features than observations. The aim was to determine the optimal regularization parameter (λ) for each model and analyze their performance in terms of the bias-variance trade-off.

2 Dataset Overview and Feature Engineering

Before proceeding into the regression analysis, it is crucial to understand the dataset's structure and the preprocessing steps undertaken. The initial dataset was subjected to extensive feature engineering, resulting in a significant expansion of the feature space.

- **Dimension of Training Data:** The training dataset encompasses 404 observations, each described by 18,565 features. This expansion in features was primarily due to the creation of polynomial features and interactions among the original variables.
- **Number of X Variables Added:** A total of 18,551 new variables were introduced as a result of feature engineering, transforming the dataset and enhancing the model's capacity to capture complex patterns in the data.

This substantial increase in dimensionality explains the challenges and considerations in model selection and regularization to prevent overfitting while striving for optimal predictive performance.

3 Estimating Linear Regression with More Columns Than Rows

For a situation in which the dataset has more features (columns) than observations (rows), a simple linear regression model cannot be directly estimated using Ordinary Least Squares (OLS) because the model would be overfit and won't generalize on new set of datasets. When this occurs, the matrix becomes singular, what this means is that it cannot be inverted, which is a necessary step for calculating the OLS solution. This situation leads to an infinite number of solutions, which will make it impossible to uniquely identify a set of regression coefficients.

Therefore, regularization methods like LASSO and Ridge regression are designed to handle such cases. They do this by adding a penalty term to the regression objective, which constrains the size of the coefficients and helps in selecting a unique solution by introducing bias into the estimator. LASSO does this by adding a penalty equal to the absolute value of the magnitude of the coefficients, potentially reducing some coefficients to zero and thus performing feature selection. Ridge regression adds a penalty equal to the square of the magnitude of the coefficients, shrinking them towards zero

4 Bias-Variance Trade-off Analysis

4.1 LASSO Regression

Optimal Alpha: $8.70519922221986 \times 10^{12}$

In-Sample RMSE: 0.2687

The extremely large optimal alpha value for LASSO indicates a strong penalization on the coefficients, with a potential of driving many of them towards zero. What this could mean is that the model is prioritizing simplicity to avoid overfitting, which is consistent with the relatively low in-sample RMSE. The low RMSE indicates the model fits the training data well and the out-of-sample RMSE for the LASSO regression model was 0.27069, closely aligning with the in-sample RMSE of 0.2687. The close proximity of in-sample and out-of-sample RMSE values indicates that the LASSO regression model generalizes well to unseen data. A model that performs similarly on both training and testing data suggests it has captured the underlying data patterns without overfitting to the noise present in the training set. However, a very high alpha might also be indicative of a model that is too simple, risking underfitting if it cannot capture all the relevant patterns in the data.

4.2 Ridge Regression

Optimal Alpha: 1×10^{-6}

Out-of-Sample RMSE: 9.1258

The optimal alpha for Ridge is extremely small, which could mean that minimal penalization on the coefficients. This implies the model is allowed more flexibility than with a higher regularization strength, potentially fitting the training data more closely. However, the high out-of-sample RMSE indicates the model does not generalize well to unseen data, suggesting overfitting. Despite the flexibility afforded by the low alpha, the model may be capturing noise in the training data, which does not translate to predictive power on new data.

5 Conclusion

The LASSO and Ridge regression models' outcomes show the critical balance in the bias-variance trade-off. While LASSO favors simplicity and robustness against overfitting, Ridge demonstrates the risks of under-regularization, where the model is overly complex and sensitive to training data noise. So by identifying the optimal regularization parameter is very essential to minimizing total error, reflecting the trade-off between model complexity and generalization capability. These observations explain the importance of

carefully tuning the regularization strength to strike a balance between bias and variance, aiming for a model that captures the essential patterns in the data without being overly complex or simplistic.