

MINERVA INVESTMENT MANAGEMENT SOCIETY
Quantitative Research Division



Forecasting 101

LASSO and Ridge Regressions

Federico Lazzarini Head of Quantitative Research
Anyi Xie Quantitative Research Analyst

Academic Year 2024/2025

Abstract

In this technical series on econometrics and statistics, we explore various econometric tools, providing the underlying theory behind these models while analyzing a dataset of house prices in California, which is featured in the book *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow* by Aurélien Géron.

DISCLAIMER

This is an academic paper related to an academic project that does not represent any investment recommendation nor solicitation to buy or sell securities. The opinions expressed are subject to change. References to specific securities, asset classes and financial markets are for illustrative purposes only and are not intended to be interpreted as recommendations. Reliance upon information in this material is at the sole risk and discretion of the reader. The material was prepared only in regard to the research activity of Minerva Investment Management Society.

1 Introduction

The objective of this analysis is to compare LASSO and Ridge regressions with standard Ordinary Least Squares (OLS) in forecasting median house prices in California. Both LASSO and Ridge impose a penalty on the coefficient estimates, which helps mitigate overfitting, especially when the dataset contains many features. While these methods introduce bias into the estimates, they typically result in lower mean squared error (MSE) due to reduced variance in the coefficient estimates. As a result, they are particularly useful in forecasting applications where generalization is key.

2 Theoretical Background

In linear regression settings, we assume a linear relationship between the input vector $\mathbf{x}$ and the output y :

$$y = \mathbf{x}^T \boldsymbol{\beta} + \epsilon$$

We also assume the following conditions hold:

1. The model is linear in the parameters.
2. $\mathbf{X}$ has full column rank (i.e., $\text{rank}(\mathbf{X}) = k$).
3. The variables in $\mathbf{X}$ are strictly exogenous ($E(\epsilon|\mathbf{X}) = 0$).
4. $\text{Var}(\epsilon|\mathbf{X}) = \sigma^2 \mathbf{I}_n$.

Under these four assumptions, the OLS estimator $\boldsymbol{\beta}_{\text{OLS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ is the best linear unbiased estimator (BLUE). In other words, the OLS estimator has the lowest variance among all unbiased linear estimators.

However, LASSO and Ridge estimators may yield lower mean squared errors (MSE) compared to the OLS estimator. This is because LASSO and Ridge regressions add penalties to the residual sum of squares (RSS) in order to prevent overfitting. Specifically, LASSO applies an L1 penalty and Ridge applies an L2 penalty.

While OLS minimizes the sum of squared errors (RSS), LASSO and Ridge minimize the following loss functions, respectively:

$$\mathcal{L}_{\text{LASSO}}(\boldsymbol{\beta}, \lambda) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1$$

$$\mathcal{L}_{\text{Ridge}}(\boldsymbol{\beta}, \lambda) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2$$

where the first term in both equations represents the usual residual sum of squares (RSS), and the second term is the regularization penalty that depends on the hyperparameter λ .

While Ridge regression has a closed-form solution to compute the estimates, given by

$$\boldsymbol{\beta}_{\text{Ridge}} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y},$$

for LASSO, we must resort to optimization techniques, such as gradient descent, where the estimates are iteratively updated to minimize the loss function.

Due to the differing penalization methods, LASSO and Ridge regression have distinct effects on the estimated coefficients. LASSO may set some coefficients exactly to zero, effectively excluding certain predictors from the model. While this enhances interpretability and facilitates feature selection, it can introduce more bias compared to Ridge regression. In contrast, Ridge regression shrinks the coefficients towards zero without eliminating any predictors, retaining all features in the model but reducing their influence proportionally.

Moreover, LASSO is particularly well-suited for high-dimensional datasets where only a subset of features are relevant to predicting the target variable. In contrast, Ridge regression is more appropriate for datasets where all variables are expected to contribute meaningfully to the forecast.

3 Methodology

Dataset Description & Preprocessing

Feature	Count	Mean	Std	Min	Median	Max
longitude	20640	-119.57	2.00	-124.35	-118.49	-114.31
latitude	20640	35.63	2.14	32.54	34.26	41.95
housing_median_age	20640	28.64	12.59	1.00	29.00	52.00
total_rooms	20640	2635.76	2181.62	2.00	2127.00	39320.00
total_bedrooms	20433	537.87	421.39	1.00	435.00	6445.00
population	20640	1425.48	1132.46	3.00	1166.00	35682.00
households	20640	499.54	382.33	1.00	409.00	6082.00
median_income	20640	3.87	1.90	0.50	3.53	15.00
median_house_value	20640	206855.82	115395.62	14999.00	179700.00	500001.00

Table 1: Summary Statistics of Dataset Features

Data Preprocessing Steps:

- **Handling Missing Values:** Rows with missing values were removed to maintain data quality, particularly in columns such as `total_bedrooms`.
- **Feature Scaling:** Numeric features were standardized using `StandardScaler` to ensure consistent scales across features, crucial for both Lasso and Ridge regression models.
- **Categorical Encoding:** The categorical feature `ocean_proximity` was transformed using one-hot encoding, making it interpretable for the models.

Model Development & Result Exhibition

Model Development

- **Train-Validation Split:**
 - Data was split into 80% for training and 20% for validation to assess model accuracy.

- **Model Selection and Hyperparameter Tuning:**

- LASSO and Ridge regression models were implemented with regularization. Using 5-fold cross-validation, the optimal regularization parameter α was determined.
- During cross-validation, each model was trained on different subsets of the training data while being validated on the remaining folds. This provided a range of errors for different alpha values, from which the best-performing alpha was selected for each model.

Forecasts Results

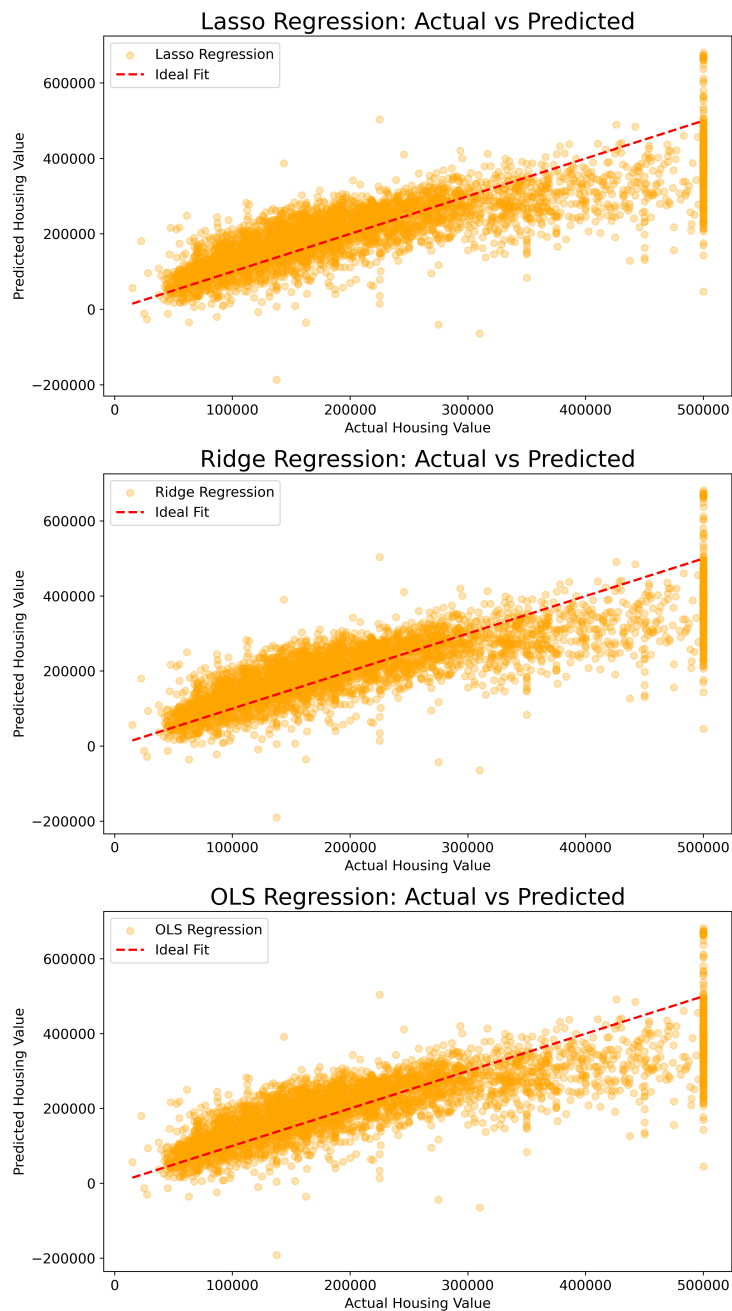


Figure 1: Forecast Results

Model Evaluation

We evaluated the models on the validation set using metrics such as MAE, MAPE, MSE, RMSE, and R-squared, allowing a detailed comparison of predictive performance.

Model	MAE	MAPE	MSE	RMSE	R_squared
LASSO	50500.43	0.29	4784523113.39	69170.25	0.6466
Ridge	50491.27	0.29	4783140549.35	69160.25	0.6467
OLS	50493.54	0.29	4782679151.64	69156.92	0.6467

Table 2: Model Evaluation Metrics

4 Conclusions

In comparing the OLS, LASSO, and Ridge regression models, we observed that while their R squared values were similar, the OLS model had a lower MSE and related error metrics (such as MAE and MAPE) than LASSO and Ridge. This suggests that the OLS model achieved a closer fit to the observed data.

The main reason for this is the lack of regularization in OLS. OLS aims to minimize error directly without penalizing large coefficients, allowing it to fit closely to the training data. As a result, OLS may capture the nuances and specific variations within the sample, leading to smaller in-sample errors. However, this flexibility also makes OLS more prone to overfitting, especially if the dataset contains noise or irrelevant features.

In contrast, LASSO and Ridge regression apply regularization (L1 for LASSO, L2 for Ridge) to reduce model complexity and avoid overfitting. This regularization controls the influence of each predictor: LASSO encourages sparsity by driving some coefficients to zero, effectively performing feature selection. This is beneficial for datasets with many predictors, where only a subset may be relevant. Ridge shrinks all coefficients proportionally, preserving each predictor’s contribution while reducing their impact. Ridge is ideal when all features contribute moderately to the target.

Therefore, LASSO and Ridge are generally better suited for data that is noisy or has many predictors, as they prevent overfitting and improve model generalization. Their regularization terms might slightly increase in-sample error metrics like MSE, but they often yield more reliable predictions on new, unseen data.