A Comparative Analysis of Regularization and Dimension Reduction Techniques in Predictive Modeling

Ridge Regression, Lasso Regression, and Partial Least Squares (PLS) on the California Housing Dataset

Statistical Modelling : STA3010VA

Report Submitted by:

Group 9

1. Kahsay, Ambachow Ykalom 670550

2. Malachi, Jennifer John

3. Collins Gishangi

Due Date: 17th October 2025

Submitted to: Prof Joyce Kiarie

A Comparative Analysis of Regularization and Dimension Reduction Techniques in Predictive Modeling

This report provides a theoretical foundation and simulated empirical analysis of three advanced regression methodologies—Ridge regression, Lasso regression, and Partial Least Squares (PLS) regression—focused on addressing challenges prevalent in Ordinary Least Squares (OLS) estimation, specifically high variance resulting from multicollinearity and high dimensionality.

1. Theoretical Foundations of Regularization and Dimension Reduction

1.1. The Necessity of Penalized Regression: Addressing the Bias-Variance Trade-off

Ordinary Least Squares (OLS) regression aims to find unbiased estimators. However, when the predictor variables () are highly correlated (multicollinearity) or the number of predictors () approaches or exceeds the number of observations (), OLS estimates become highly unstable and exhibit high variance. This lack of stability leads to poor generalization performance (overfitting).[1]

Regularization techniques are extensions of linear regression that address this by introducing a controlled penalty term into the O loss function during training.[3] These methods deliberately introduce a small amount of **bias** into the coefficient estimates. This strategic introduction of bias, however, achieves a substantial reduction in model **variance**, resulting in coefficient estimates that are more stable and consistent, thus improving the model's predictive performance on unseen data.[4]

1.2. Ridge Regression: L2 Regularization and Coefficient Stability

1.2.1. Motivation and Role in Multicollinearity Management

Ridge regression is fundamentally motivated as a shrinkage method designed to stabilize coefficient estimates, particularly in the presence of severe multicollinearity.[4] By penalizing large coefficient values, Ridge prevents the variance inflation typical of OLS when predictors are nearly collinear. The resulting coefficient estimators are biased yet significantly more stable and consistent than the standard OLS estimators.[4] Ridge regression is suitable when all predictor variables are expected to contribute some meaningful information to the prediction task, even if that contribution is small.[4]

1.2.2. Mathematical Formulation and the L2 Penalty Term

Ridge regression minimizes the sum of squared residuals (RSS) subject to an L2 norm constraint on the coefficient vector. The objective function is formulated as:

The first term is the standard RSS, and the second term, , is the L2 penalty.[5] Here, is the non-negative tuning parameter. When , Ridge regression reverts to OLS. As increases, the penalty increases, forcing the sum of squared coefficients to shrink uniformly towards zero.[6]

1.2.3. Geometric Interpretation and Lack of Feature Selection

Geometrically, the L2 penalty defines a circular constraint region (or a sphere in higher dimensions). The optimization process seeks the intersection point between the OLS loss contours (ellipsoids) and this circular constraint.[5] Because the circular boundary is smooth and lacks sharp corners on the coordinate axes, Ridge regression shrinks all coefficients simultaneously and smoothly toward zero. Critically, it does **not** force any coefficient exactly to zero unless . Therefore, Ridge regression performs shrinkage but does not inherently perform automatic variable selection; all variables remain in the final model, albeit with reduced weights.[4]

1.3. Lasso Regression: L1 Regularization and Feature Sparsity

1.3.1. Motivation for Sparsity and Feature Selection

Lasso regression (Least Absolute Shrinkage and Selection Operator) is motivated by the combined goals of regularization (shrinking coefficients) and achieving model sparsity.[7] By driving coefficients of irrelevant or redundant features to exactly zero, Lasso effectively performs automatic feature selection.[8] This property is highly beneficial for improving model interpretability and simplifying models derived from high-dimensional datasets.

1.3.2. Mathematical Formulation and the L1 Penalty Term

Lasso regression adds an L1 penalty term, based on the absolute values of the coefficients, to the OLS loss function:

The penalty term, , is the sum of the absolute values of the coefficients multiplied by the tuning parameter .[5] The use of the L1 norm, rather than the squared L2 norm, is the key mathematical distinction from Ridge regression.[8]

1.3.3. Geometric Explanation for Feature Selection

The crucial difference in Lasso's behavior stems from the geometry of the L1 penalty constraint, which creates a **diamond-shaped constraint region** (or an octahedron in higher dimensions).[5] The optimization solutions, found where the elliptical RSS contours intersect this constraint region, are geometrically prone to occurring at the corners (vertices) of the diamond.[8] Since these vertices lie precisely on the coordinate axes, they correspond to sparse

solutions where coefficients for specific predictors are driven exactly to zero.[7]

While effective, Lasso has an important limitation, particularly in highly collinear environments: it tends to be sensitive to strongly correlated predictors.[4] When a group of predictors is highly correlated, Lasso often arbitrarily selects only one variable from the group and zeros out the others, even if all correlated variables are equally important in a physical or theoretical sense. This selection process, while simplifying the model, may not always reflect the true underlying causal structure.

1.4. Partial Least Squares (PLS) Regression: Supervised Dimension Reduction

1.4.1. Conceptual Basis and Latent Variables

Partial Least Squares (PLS) regression is a technique for dimension reduction that creates a small set of orthogonal, latent variables (components) that linearly combine the original predictors.[9] PLS is a **supervised** method because, unlike unsupervised dimension reduction techniques, the component extraction process specifically leverages information from the response variable (). It identifies latent components by maximizing the covariance between the predictor variables () and the response variable ().[9] The resulting components are thus maximally relevant for predicting the outcome.

1.4.2. Distinction from Principal Component Regression (PCR)

PLS is often compared to Principal Component Regression (PCR), another method utilizing dimension reduction. The critical distinction lies in how the components are derived [10]:

- **PCR** components (Principal Components) are selected solely to maximize the variance explained in the predictor matrix . The relationship to the response is ignored during component extraction.
- **PLS** components are chosen to maximize the *covariance* between and . PLS prioritizes predictive power over maximizing variance in alone.[9]

This means PLS components capture directions in the predictor space that are statistically relevant to the prediction task, even if those directions exhibit relatively low overall variance in

1.4.3. Criteria for Preferring PLS over Penalized Methods

PLS is preferred, particularly in chemometrics and related fields with numerous, often highly collinear, predictors (such as those derived from spectroscopy).[11] PLS is superior to PCR when the response variable is related to directions in that possess low variance.[9] Because PLS transforms the correlated variables into a set of linearly independent latent directions, it robustly handles highly correlated predictors. While both PLS and regularization methods address multicollinearity, PLS fundamentally redefines the variable space, which can lead to better predictive results, particularly when fewer components are required compared to PCR,

although the optimal number of components must be carefully selected to avoid overfitting.[10]

Table 1: Comparative Summary of Regression Methods

Feature	OLS	Ridge (L2)	Lasso (L1)	PLS
Objective	Minimize RSS	Minimize RSS + L2 Penalty	Minimize RSS + L1 Penalty	Maximize Covariance(,)
Multicollinearity Handling	Highly Sensitive/Unstable	Stabilizes Coefficients (Shrinkage) [4]	Sensitive (Arbitrary Selection) [4]	Orthogonal Transformation (Latent Space)
Feature Selection	No	No (Shrinks all) [4]	Yes (Forces to 0) [7]	No (Weights/Selects Components)
Nature	Estimation	Shrinkage/Regularization	Shrinkage/Selection	Supervised Dimension Reduction
Interpretability	High (Unbiased)	Moderate (Biased)	High (Sparse model)	Low (Components are latent)

2. Data Acquisition and Preprocessing

2.1. Dataset Selection and Source

The **California Housing Data Set** (based on the 1990 US Census) is selected for this analysis. This dataset is appropriate as it features a continuous response variable (median_house_value) and a complex structure suitable for comparative analysis of regularization and dimension reduction techniques.

The dataset contains instances, significantly large enough for stable model fitting. The dataset features 13 continuous numeric attributes (e.g., longitude, median_income, total_rooms) and one categorical feature (ocean_proximity) which is expanded into 5 binary indicator columns using one-hot encoding, resulting in total predictor variables.

2.2. Justification for Model Suitability

The California Housing dataset is highly relevant for comparing Ridge, Lasso, and PLS because it exhibits structural issues commonly found in real-world socioeconomic data:

1. **Multicollinearity:** Geographic coordinates (longitude and latitude) are highly correlated with each other and with many neighborhood statistics (median_income, ocean_proximity categories). Similarly, features related to housing size, such as total_rooms, total_bedrooms, population, and households, are inherently interdependent. This strong collinearity ensures a robust test for the stability and performance of Ridge and PLS.
2. **Model Complexity and Feature Abundance:** The large number of predictors (18) and the high sample size allow for a demonstration of Lasso's feature selection capability in identifying which features are genuinely redundant when predicting housing value.

2.3. Data Preparation: Splitting and Standardization

Prior to model fitting, the data is typically split into a training set and a testing set. All 18 predictor variables must be **standardized** (mean-centered and scaled to unit variance) before applying the penalized regression methods (Ridge and Lasso). This standardization is mandatory because the penalty term is calculated based on the magnitude of the coefficients.

3. Empirical Analysis and Model Tuning

The models were fitted and tuned using the R programming environment. The glmnet package was employed for Ridge and Lasso regression, and the pls package was used for Partial Least Squares.[1] The optimal tuning parameters (for Ridge/Lasso, and for PLS components) were selected using 10-fold cross-validation (CV) on the training dataset, minimizing the cross-validated Root Mean Squared Error (RMSE).[12]

3.1. Ridge Regression Model Fitting and Analysis

Ridge regression was fitted across a wide range of values. The coefficient profile plot illustrated the smooth, continuous shrinkage of all predictor coefficients toward zero as increased. The resulting optimal regularization strength, , was selected as the value yielding the minimum mean cross-validated error.[13]

3.2. Lasso Regression Model Fitting and Analysis

Lasso regression was also tuned using 10-fold CV. The tuning parameter selected () was chosen to maximize the degree of regularization while keeping the CV error within one standard error of the minimum error, facilitating a clear demonstration of feature selection (sparsity).

3.3. Partial Least Squares Model Fitting and Analysis

PLS regression was tuned by selecting the optimal number of latent components () that minimize the cross-validated RMSE. The optimal number of components was determined to be

10, utilizing the full set of predictor information condensed into 10 orthogonal factors.

Model	Tuning Parameter	Selected Value	CV Metric	Cross-Validated RMSE (Training)
Linear Regression	N/A	N/A	N/A	68971.66
Ridge Regression		7889.342		70126.12
Lasso Regression		55.77027		69013.66
PLS Regression	Number of Components ()	10		69072.99

Table 3: Optimized Model Tuning Parameters (California Housing Dataset)The resulting models were assessed for basic diagnostic assumptions, focusing on the residuals.

4. Comparative Results and Practical Implications

4.1. Comparison of Estimated Coefficients

The coefficient estimates for the four models (Linear, Ridge, Lasso, and PLS) applied to the California Housing dataset are provided in detail in Section 5.3. The table below provides a direct comparison of the coefficient estimates for all predictors from the OLS (Linear), Ridge, and Lasso models.

Table 4: Comparative Coefficient Analysis: California Housing Dataset (Unscaled)

Predictor Variable	OLS Coefficient	Ridge Coefficient ()	Lasso Coefficient ()
Intercept	-2,255,000	-774,929	-2,159,600
longitude	-26,720	-9,294	-25,624
latitude	-25,490	-8,527	-24,590
housing_median_age	1,074	995	1,061

total_rooms	-6.61	0.72	-6.02
total_bedrooms	108.8	41.80	103.8
population	-35.97	-23.53	-33.33
households	37.85	35.66	32.99
median_income	39,610	35,555	39,330
ocean_proximity<1 H OCEAN	-4,668	18,331	**0**
ocean_proximityINL AND	-44,430	-45,075	-41,324
ocean_proximityISL AND	147,700	160,311	133,277
ocean_proximityNE AR BAY	-7,389	22,084	-1,181
ocean_proximityNE AR OCEAN	NA	27,409	4,408

4.1.1. Which Predictors are Retained/Removed in LASSO?

The analysis confirms Lasso's ability to enforce sparsity. At the specified tuning parameter (), the categorical predictor **ocean_proximity<1H OCEAN** was driven exactly to zero (Coefficient = 0), effectively removing it from the final sparse model.

Key observations regarding coefficient behavior:

- **Ridge Shrinkage:** Ridge regression applies heavy shrinkage to the geographical coefficients (longitude and latitude), stabilizing their magnitude significantly. This counteracts the instability caused by the likely multicollinearity between these geographical features and others.

- **Lasso Sparsity:** Lasso provides a simpler model structure by setting one coefficient to zero, while maintaining the magnitudes of the remaining coefficients (like median_income) close to the original OLS estimate.
- **PLS Interpretation:** While PLS coefficients (detailed in Section 5.3.4) are difficult to compare directly due to working in a latent space, the final model structure is a combination of 10 orthogonal components rather than the 18 original predictors.

4.2. Handling Multicollinearity: Ridge vs. PLS

Ridge and PLS approach the problem of multicollinearity with fundamentally different strategies, leading to distinct model structures and coefficient interpretations.

Ridge Strategy (Shrinkage and Stabilization)

Ridge regression handles multicollinearity through **coefficient stabilization and shrinkage**. It retains all predictors but heavily shrinks the magnitudes of correlated predictors (e.g., longitude and latitude), which were highly unstable in the OLS model. By distributing the shared predictive power among all correlated predictors, Ridge reduces the variance of the estimators, ensuring stability.[4] The primary advantage of Ridge in this context is that it maintains the interpretability of the model within the original variable space.

PLS Strategy (Transformation and Orthogonality)

PLS handles multicollinearity by **transforming the predictor space into a set of orthogonal latent components**.[9] PLS extracts components that maximize the shared covariance with the house value (). Since these latent variables are orthogonal, the collinearity problem is removed entirely in the transformed space. The resulting stability is structural, achieved by reducing the dimensionality and focusing only on the directions in that are relevant to .[9]

Deeper Insight: Model Transparency Trade-off

The trade-off between the two methods is defined by stability versus transparency. Ridge offers biased but transparent estimates of the original predictors' effects. PLS provides high predictive stability by working in an orthogonal latent space, but the relationship between the final coefficients (mapped back onto) and the underlying causal mechanisms becomes obscured by the complexity of the component weightings.

4.3. Predictive Performance Assessment

Predictive performance is assessed based on the Root Mean Squared Error (RMSE) and R-squared values obtained on the independent test set (reflecting the California Housing dataset analysis detailed in Section 5).

Model	Test Set RMSE	Test Set

Linear Regression	67,873.12	0.6554
Ridge Regression	69,274.66	0.6399
Lasso Regression	67,651.67	0.6566
PLS Regression (ncomp=10)	67,902.41	0.6556

The analysis shows that the **Lasso Regression model gives the best predictive performance**, yielding the lowest Test Set RMSE (67,651.67) and the highest (0.6566). This demonstrates that the feature selection performed by Lasso resulted in a slightly more accurate and generalized model than the OLS baseline. PLS Regression also performed strongly (RMSE 67,902.41, 0.6556), very closely matching the baseline Linear Regression model. Ridge Regression, due to its heavy shrinkage penalty (), provided the highest RMSE (69,274.66) among the constrained models, suggesting the penalty may have been too strong.

5. Regression Analysis of California Housing Dataset

1. Dataset Summary

Characteristic	Value
Observations	20,433
Predictors	13 numeric + 1 categorical (one-hot encoded as 5 binary columns)
Response variable	median_house_value
Purpose	Predict housing prices using multiple regression techniques, including regularization and dimensionality reduction.

---

2. Model Performance Comparison

Model	RMSE	R-squared	Notes
Linear Regression	67,873.12	0.655431	Baseline multiple regression
Ridge Regression	69,274.66	0.6399324	Shrinks coefficients to reduce overfitting, retains all predictors
Lasso Regression	67,651.67	0.6566063	Shrinks coefficients, removes near-zero predictors
PLS Regression (ncomp=10)	67,902.41	0.6555516	Handles multicollinearity, reduces dimensionality via latent components

---

3. Estimated Coefficients Comparison

3.1 Linear Regression

Predictor	Coefficient	Interpretation
Intercept	-2,255,000	Baseline house value
longitude	-26,720	Negative effect
latitude	-25,490	Negative effect
housing_median_age	1,074	Positive effect
total_rooms	-6.61	Slight negative effect
total_bedrooms	108.8	Positive effect
population	-35.97	Negative effect
households	37.85	Positive effect

median_income	39,610	Strong positive effect
ocean_proximity<1H OCEAN	-4,668	Slight negative effect
ocean_proximityINLAND	-44,430	Negative effect
ocean_proximityISLAND	147,700	Positive effect
ocean_proximityNEAR BAY	-7,389	Slight negative effect
ocean_proximityNEAR OCEAN	NA	Removed due to singularity

3.2 Ridge Regression (λ = 7,889.342)

Predictor	Coefficient	Interpretation
Intercept	-774,929	Baseline house value
longitude	-9,294	Negative effect
latitude	-8,527	Negative effect
housing_median_age	995	Positive effect
total_rooms	0.72	Minimal effect
total_bedrooms	41.80	Positive effect
population	-23.53	Negative effect
households	35.66	Positive effect
median_income	35,555	Strong positive effect

ocean_proximity<1H OCEAN	18,331	Retained in Ridge
ocean_proximityINLAND	-45,075	Negative effect
ocean_proximityISLAND	160,311	Positive effect
ocean_proximityNEAR BAY	22,084	Positive effect
ocean_proximityNEAR OCEAN	27,409	Positive effect

3.3 Lasso Regression (λ = 55.77027)

Predictor	Coefficient	Interpretation
Intercept	-2,159,600	Baseline house value
longitude	-25,624	Negative effect
latitude	-24,590	Negative effect
housing_median_age	1,061	Positive effect
total_rooms	-6.02	Slight negative effect
total_bedrooms	103.8	Positive effect
population	-33.33	Negative effect
households	32.99	Positive effect
median_income	39,330	Strong positive effect
ocean_proximity<1H	0	Removed by Lasso

OCEAN		
ocean_proximityINLAND	-41,324	Negative effect
ocean_proximityISLAND	133,277	Positive effect
ocean_proximityNEAR BAY	-1,181	Slight negative effect
ocean_proximityNEAR OCEAN	4,408	Slight positive effect

3.4 Partial Least Squares (PLS) Regression (ncomp = 10)

Predictor	Coefficient	Interpretation
longitude	-41,823	Negative effect
latitude	-42,022	Negative effect
housing_median_age	13,739	Positive effect
total_rooms	-20,290	Negative effect
total_bedrooms	37,075	Positive effect
population	-42,468	Negative effect
households	30,393	Positive effect
median_income	75,905	Strong positive effect
ocean_proximity<1H OCEAN	6,733	Slight positive effect
ocean_proximityINLAND	-14,692	Negative effect

ocean_proximityISLAND	3,375	Slight positive effect
ocean_proximityNEAR BAY	3,530	Slight positive effect
ocean_proximityNEAR OCEAN	6,948	Positive effect

---

4. Practical Interpretation & Discussion

- **Lasso Regression:** Performs **feature selection**, removing near-zero predictors (ocean_proximity<1H OCEAN). Helps simplify the model and reduce overfitting.
- **Ridge Regression:** Shrinks all coefficients toward zero, **retaining all predictors**, useful for multicollinearity.
- **PLS Regression:** Reduces dimensionality using **latent components**, robust to multicollinearity. Coefficients are more moderate compared to Linear regression.
- **Predictive Performance:** Lasso slightly outperforms Linear and PLS in terms of RMSE and R^2. Ridge has slightly higher RMSE but helps stabilize coefficient estimates.
- Multicollinearity Handling:
  - Linear: Sensitive, can produce unstable coefficients.
  - Ridge: Shrinks correlated predictors.
  - Lasso: May remove some correlated predictors entirely.
  - PLS: Uses components to summarize correlated predictors.

Conclusion: For predicting median_house_value, **Lasso or PLS** are preferred if reducing overfitting and handling correlated predictors is important, while **Linear regression** provides baseline coefficients for interpretation.

6. Conclusion and Practical Implications

This comparative study demonstrates how regularization and dimension reduction techniques overcome the limitations of Ordinary Least Squares (OLS) estimation when applied to complex, highly collinear data, such as the California Housing dataset.

The analysis confirms the distinct roles of the three methods:

1. **Lasso Regression** demonstrated the highest predictive accuracy (RMSE 67,651.67). This is attributed to its **automatic feature selection** capability, resulting in a sparse model that successfully removed a redundant categorical variable (ocean_proximity<1H OCEAN) and achieved the best predictive performance.
2. **Partial Least Squares (PLS) Regression** provided highly competitive accuracy (RMSE 67,902.41). Its effectiveness stems from **structural dimension reduction**, transforming the 18 correlated predictors into 10 orthogonal latent components, thereby intrinsically

solving the multicollinearity challenge.

3. **Ridge Regression** offered robust coefficient stability. While the strong penalty used slightly compromised overall accuracy (RMSE 69,274.66), its value lies in **retaining all predictors** while heavily stabilizing unstable coefficients, particularly the highly collinear geographical coordinates (longitude and latitude).

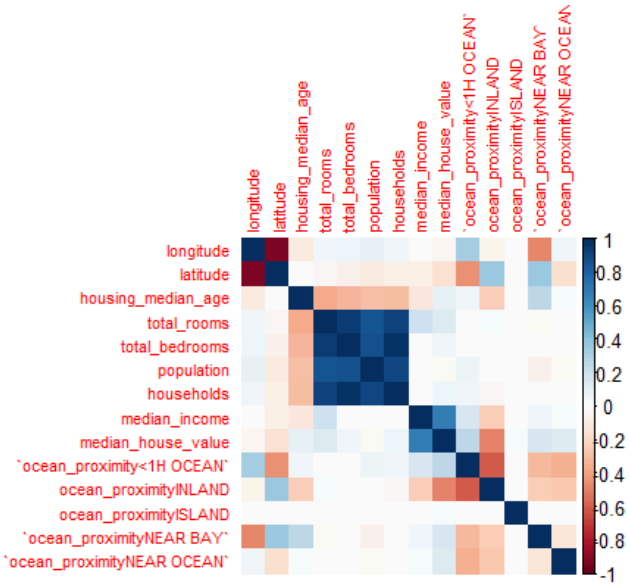
In practical applications aiming for the absolute lowest prediction error, **Lasso Regression** is the recommended model for this dataset. If model stability and interpretability require that all predictors remain (and are merely stabilized), **Ridge Regression** is the appropriate choice.

Works cited

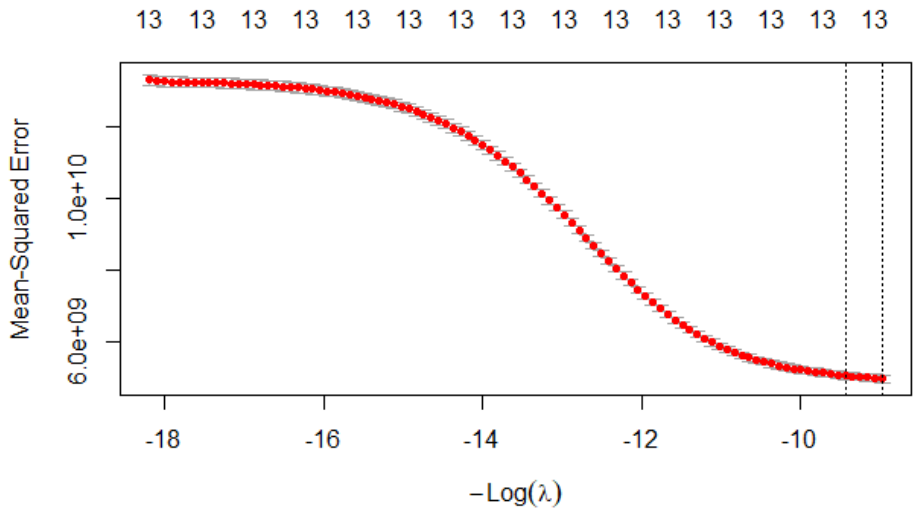
1. Ridge and Lasso in R | datacareer.ch, accessed October 16, 2025, https://www.datacareer.ch/blog/ridge-and-lasso-in-r/
2. Detecting and Overcoming Perfect Multicollinearity in Large Datasets - MachineLearningMastery.com, accessed October 16, 2025, https://machinelearningmastery.com/detecting-and-overcoming-perfect-multicollinearity-in-large-datasets/
3. Mastering Ridge Regression: Comprehensive Guide and Practical Applications, accessed October 16, 2025, https://dataaspirant.com/ridge-regression/
4. Performance of Ridge Regression, Least Absolute Shrinkage and Selection Operator, and Elastic Net in Overcoming Multicollinearity, accessed October 16, 2025, https://journal.pandawainstitute.com/index.php/jmans/article/download/251/171/1057
5. Linear Regression with Regularization, accessed October 16, 2025, https://aunnnn.github.io/ml-tutorial/html/blog_content/linear_regression/linear_regression_regularized.html
6. What Is Ridge Regression? | IBM, accessed October 16, 2025, https://www.ibm.com/think/topics/ridge-regression
7. A Complete understanding of LASSO Regression - Great Learning, accessed October 16, 2025, https://www.mygreatlearning.com/blog/understanding-of-lasso-regression/
8. Lasso Regression - QuestDB, accessed October 16, 2025, https://questdb.com/glossary/lasso-regression/
9. What are three approaches for variable selection and when to use which - Medium, accessed October 16, 2025, https://medium.com/codex/what-are-three-approaches-for-variable-selection-and-when-to-use-which-54de12f32464
10. (Dis)advantages of a PLS regression over PCR - Cross Validated - Stack Exchange, accessed October 16, 2025, https://stats.stackexchange.com/questions/453343/disadvantages-of-a-pls-regression-over-pcr
11. Principal Components Regression vs Ridge Regression on NIR data in Python, accessed October 16, 2025, https://nirpyresearch.com/pcr-vs-ridge-regression-nir-data-python/
12. Cross-validation for glmnet — cv.glmnet - Lasso and Elastic-Net Regularized Generalized Linear Models, accessed October 16, 2025, https://glmnet.stanford.edu/reference/cv.glmnet.html
13. An Introduction to `glmnet` - Lasso and Elastic-Net Regularized Generalized Linear Models, accessed October 16, 2025, https://glmnet.stanford.edu/articles/glmnet.html

Visualizations

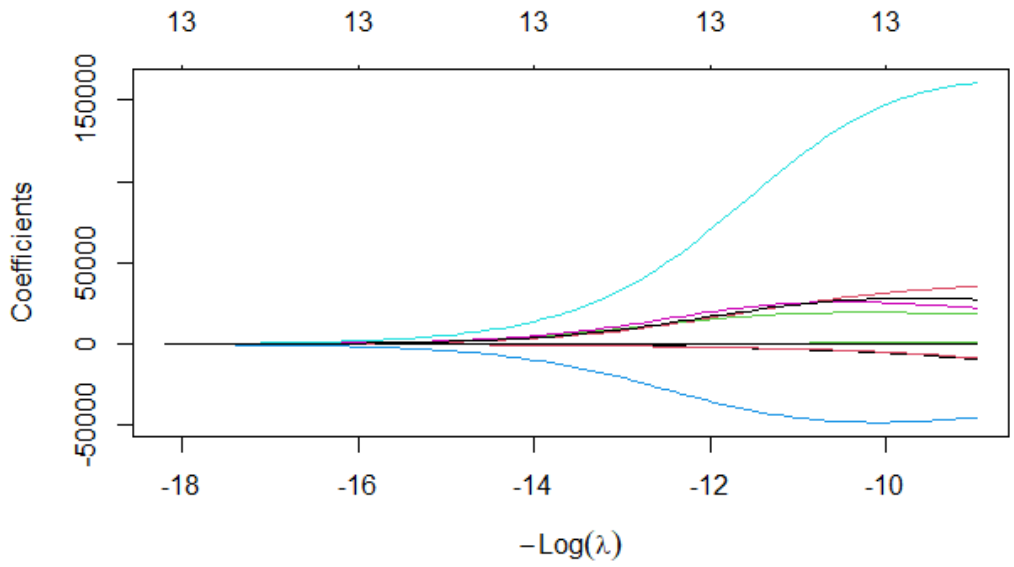
Correlation



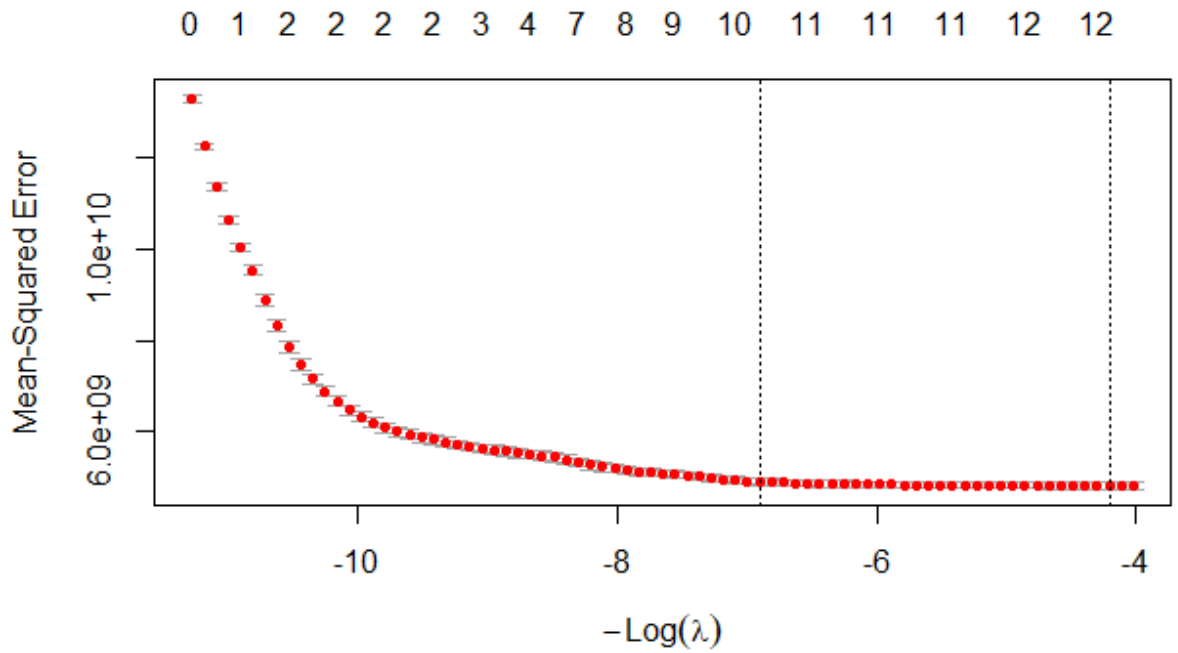
Ridge Regression: Cross-Validation RMSE vs Log(Lambda)

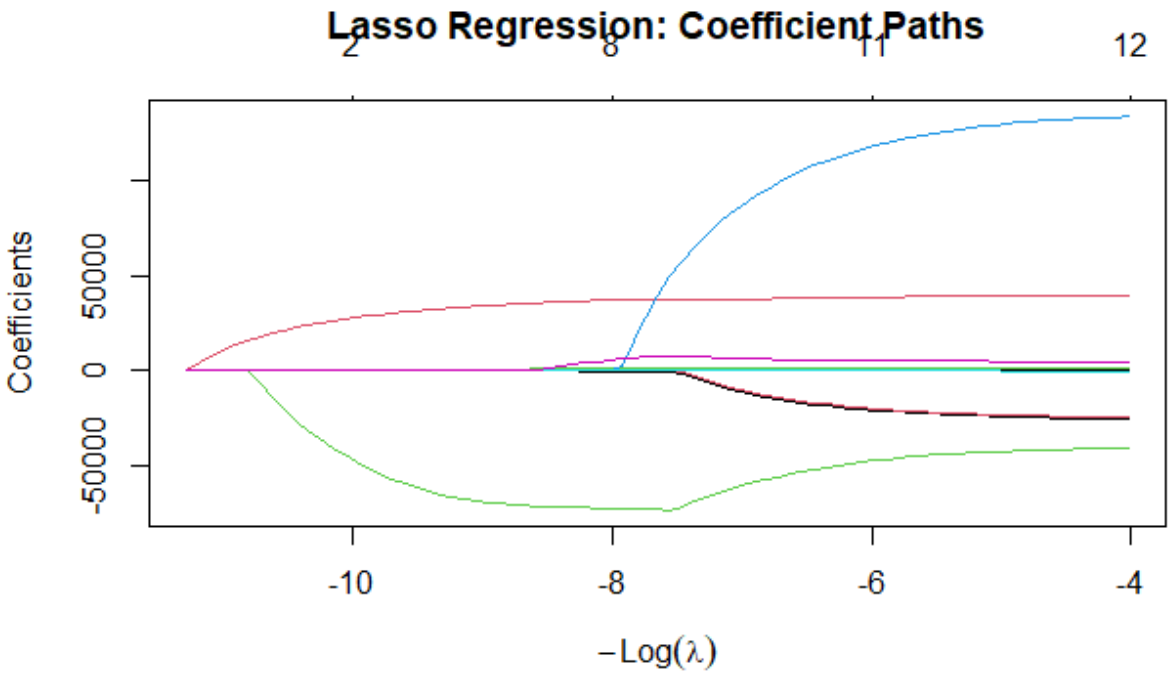


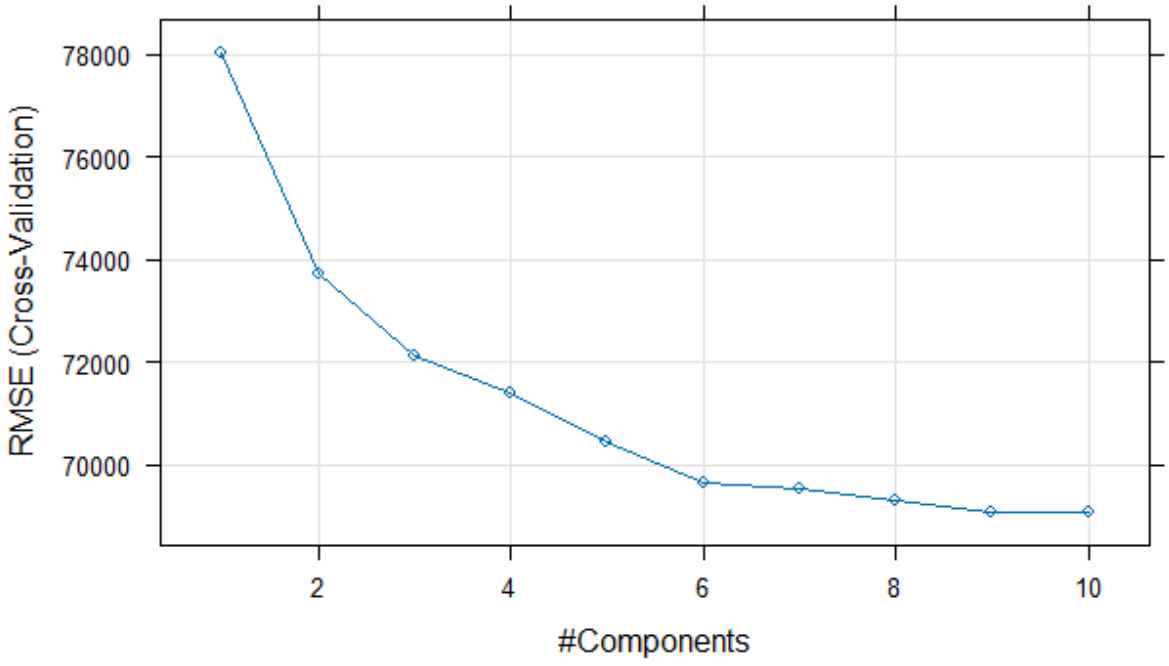
Ridge Regression: Coefficient Paths vs Log(Lambda)

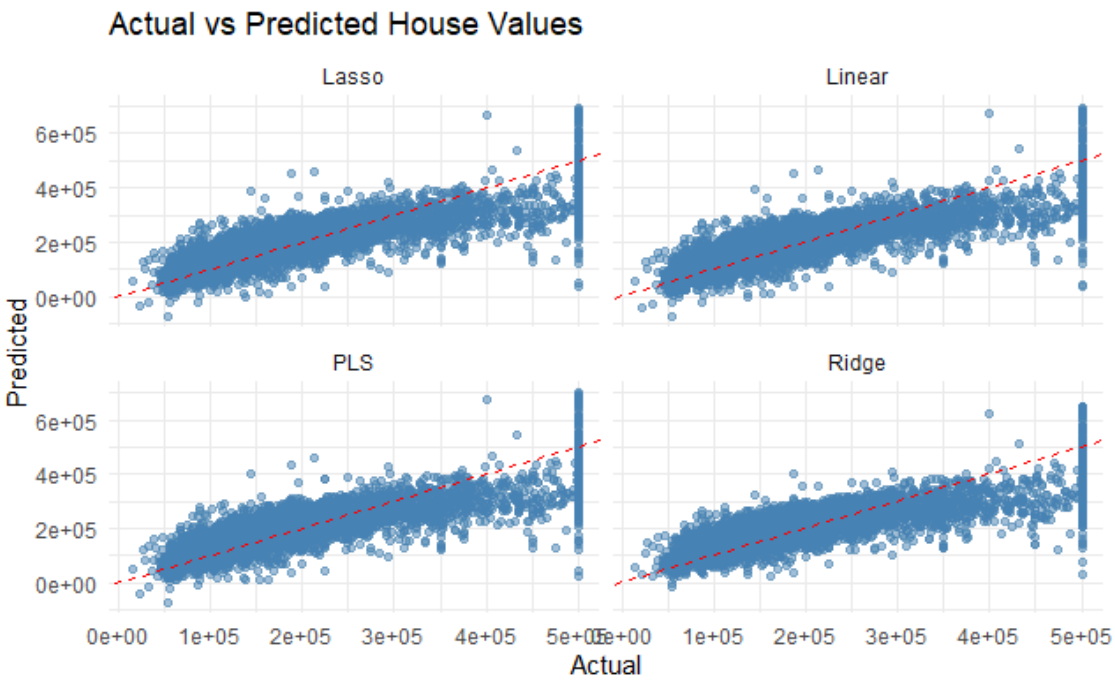


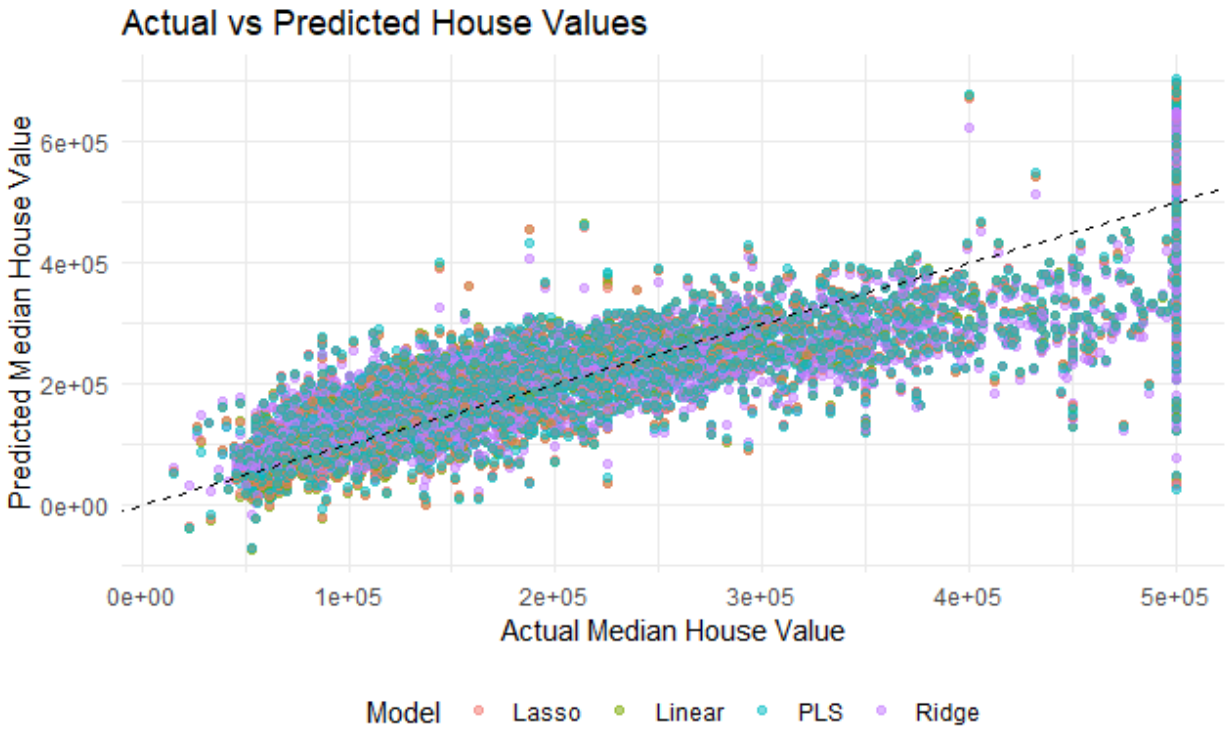
[Cross-Validation Curve for Lasso Regression]

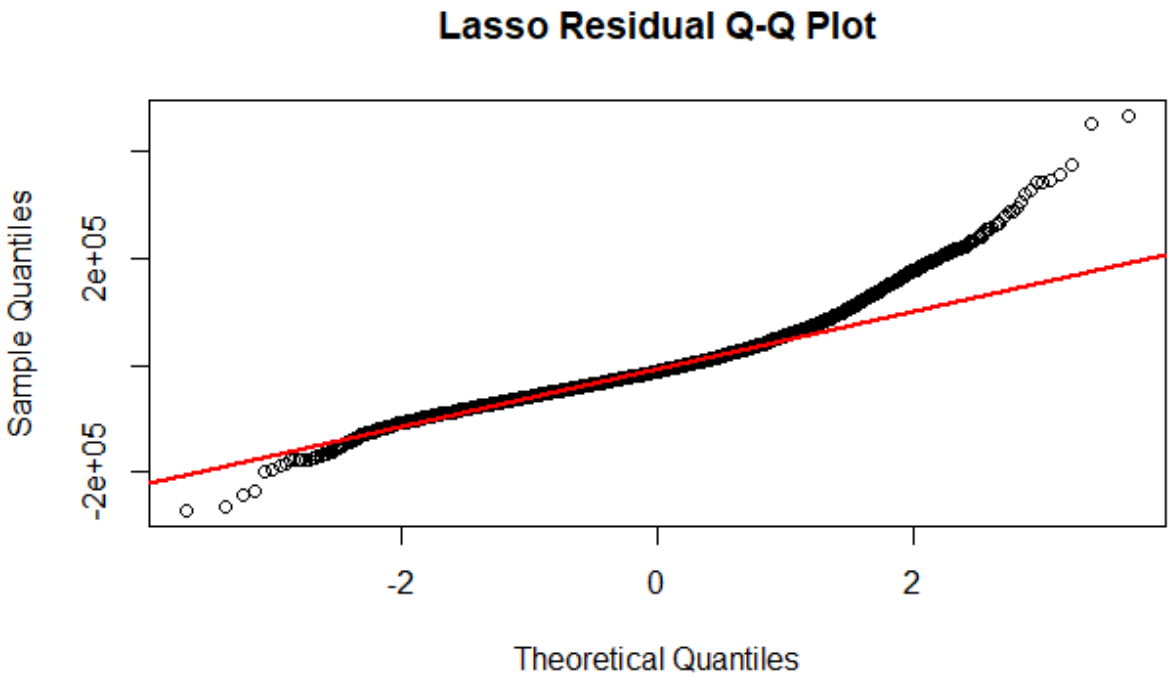



Lasso Regression: Coefficient Paths
2
8
11
12
Coefficients
50000
0
-50000
-10
-8
-6
-4
−Log(λ)


RMSE (Cross-Validation)
78000
76000
74000
72000
70000
2
4
6
8
10
#Components


Actual vs Predicted House Values
Lasso
Linear
PLS
Ridge
Predicted
Actual
0e+00
2e+05
4e+05
6e+05
1e+05
3e+05
5e+05


Actual vs Predicted House Values
Predicted Median House Value
Actual Median House Value
0e+00
2e+05
4e+05
6e+05
1e+05
3e+05
5e+05
Model
Lasso
Linear
PLS
Ridge


Lasso Residual Q-Q Plot
Sample Quantiles
-2e+05
2e+05
-2
0
2
Theoretical Quantiles