

### Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

The optimal alpha values for Ridge and Lasso regression depend on the data and are chosen to balance model fit and prevention of overfitting.

1. Ridge Regression: Larger alpha strengthens regularization, shrinking coefficients towards zero without making them exactly zero.
2. Lasso Regression: Higher alpha enhances regularization, driving some coefficients to exactly zero, aiding in feature selection.

Doubling alpha for both methods would:

- Ridge: Increase regularization, promoting simpler models.
- Lasso: Further, encourage feature selection, making the model sparser.

The most important predictor variables would change based on the specific dataset and the interaction of predictors. Ridge might retain more predictors with small coefficients, while Lasso could exclude more predictors, emphasizing a smaller set with significant coefficients. Cross-validation helps find optimal alphas and evaluate effects.

.....

### Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer :

Choose Ridge Regression:

If dealing with high multi collinearity among predictors.

If seeking stability in models with many correlated predictors.

If prioritizing predictive accuracy over extreme coefficient values.

Choose Lasso Regression:

If aiming for feature selection and identifying the most important predictors.

If interpretability is a top priority.

If desiring a simpler model by excluding less important predictors.

The choice hinges on the goals of analysis and dataset characteristics.

.....

### Question 3

After building the model, you realized that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Answer:

Those 5 most important predictor variables that will be excluded are :-

1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmntSF
5. GarageArea

.....

### Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer:

To ensure a model is robust and generalizable:

Cross-Validation: Assess model on different data subsets.

Train-Test Split: Test on new data not seen during training.

Out-of-Sample Testing: Validate on entirely new data.

Relevant Features: Use meaningful features, avoid noise.

Regularization: Control complexity to prevent overfitting.

Hyperparameter Tuning: Optimize parameters for best performance.

Clean Data: Ensure data quality and consistency.

Prevent Data Leakage: Avoid unintended influence from test data.

Implications for Accuracy:

Robust Model: Steady performance on new data.

Generalization: Avoids fitting to training data noise.

Real-World Performance: More likely to excel outside training set.

Balance Complexity: Prioritize reliability over extreme training accuracy.