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?

Ans:

Optimal Value of alpha for Ridge

- Ridge - 0.2
- Lasso - 0.0001

Double the values of alpha

- Ridge - 0.4
- Lasso - 0.0002 Let's analyse the model with these alpha values

Regularize coefficients to improve prediction accuracy and decrease variance, enhancing model interpretability and reducing the overall model's complexity.

Ridge Regression:

• Uses a tuning parameter called lambda, penalizing coefficients based on their square of magnitude.

• Residual sum or squares should be small, penalizing coefficients with greater values.

• As lambda value increases, variance in the model drops, and bias remains constant.

• Includes all variables in the final model, unlike Lasso Regression.

Lasso Regression:

• Uses a tuning parameter called lambda, penalizing absolute value of coefficients.

• As lambda value increases, the coefficient shrinks towards zero, making variables equal to 0.

• Also performs variable selection.

• Small lambda value performs simple linear regression.

• As lambda value increases, shrinkage occurs, and variables with 0 values are neglected.

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?

Ans:

Ridge Regression vs Lasso Analysis

• Ridge regression doesn't zero coefficients.

• Lasso zeros one or two coefficients in selected features.

• Lasso is preferred for feature elimination.

Question 3

After building the model, you realised 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?

Ans:

have excluded the five most important variable I have got prior.

Those are MSZoning_FV, GrLivArea, MSZoning_RL, OverallQual, Foundation_PConc.

After the Lass Regression I have got the other important predictors are Overall condition, Lot area, Lot shape, Condition1,

Question 4

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

Ans:

Robust and Generalizable Models

• Test accuracy should be equal to the training score.

• Outliers should not impact the model, allowing for correct values to be imputed.

• Predicted variables should be significant, determined by P-values, R2 and adjusted R2.

Implications of Accuracy:

• Gain more data to train the model.

• Fix missing values and outliers to improve model accuracy.

• Feature engineering or newly derived columns/standardize values.

• Scale values to ensure accuracy.

• Feature selection based on domain knowledge and data visualization.

• Apply the right machine learning algorithm with experience.

• Use cross validation to prevent overfitting.

Modelling for Simpler Predictions

• Models should not mugger up training data for accurate predictions.

• Models are often tested on unseen data in real-life situations.

• Assessing model adaptability in such scenarios is crucial.