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 value of alpha for ridge and lasso regression is 10 and 100 respectively. If the alpha value is increased, the model tends to underfit and if the alpha value is too low, the model tends to overfit on the training data and perform poorly on the test data.

In this case, when alpha is doubled, both ridge and lasso models tend to have reduced r2 score in both train and test sets which is indicative of underfitting.

For lasso regression, as the alpha value increases, it pushes the coefficients of the columns to zero thus also performing feature selection. The model complexity reduces as the alpha value increases.

The most important predictor variables after the change is implemented are:

For Ridge:

grlivarea	22178.046728
overallqual_9	17635.668739
bsmtqual_Ex	14434.462312
neighborhood_NoRidge	13672.067196
overallqual_10	11888.287539
garagecars	10720.129424

For Lasso:

overallqual_10	75278.859108
overallqual_9	66311.727569
neighborhood_NoRidge	33139.136791
grlivarea	25995.707396
overallcond_3	-20774.653541

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:

If the lambda is zero, there will be no shrinkage of model coefficients. However when the lambda increases, the shrinkage penalty increases, which pushes the coefficients towards zero. When lambda is too small, ridge and lasso don't solve overfitting but when lambda is too high, it causes the model to underfit. Therefore a lambda value is chosen which neither causes the model to overfit not underfit which is the optimum lambda value.

For Ridge:

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda\sum_{j=1}^{m}\hat{\beta}_j^2 = ||y - X\hat{\beta}||^2 + \lambda||\hat{\beta}||^2.$$

When lambda is high, it pushes the coefficients towards zero but the coefficients don't become zero.

For Lasso:

$$L_{lasso}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda\sum_{j=1}^{m}|\hat{\beta}_j|.$$

For lasso regression, high lambda value makes the coefficients as zero thereby eliminating the feature altogether. Hence, lasso can also be used as a feature selection method.

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?

Answer:

The five most important predictor variables in the lasso model are:

overallqual_10	86335.434206
overallqual_9	69544.356056
neighborhood_NoRidge	37167.770245
neighborhood_StoneBr	26821.276143
overallcond_3	-25326.296335

After removing the above variables and running the Lasso regression again, we get a new set of most important predictor variables. They are:

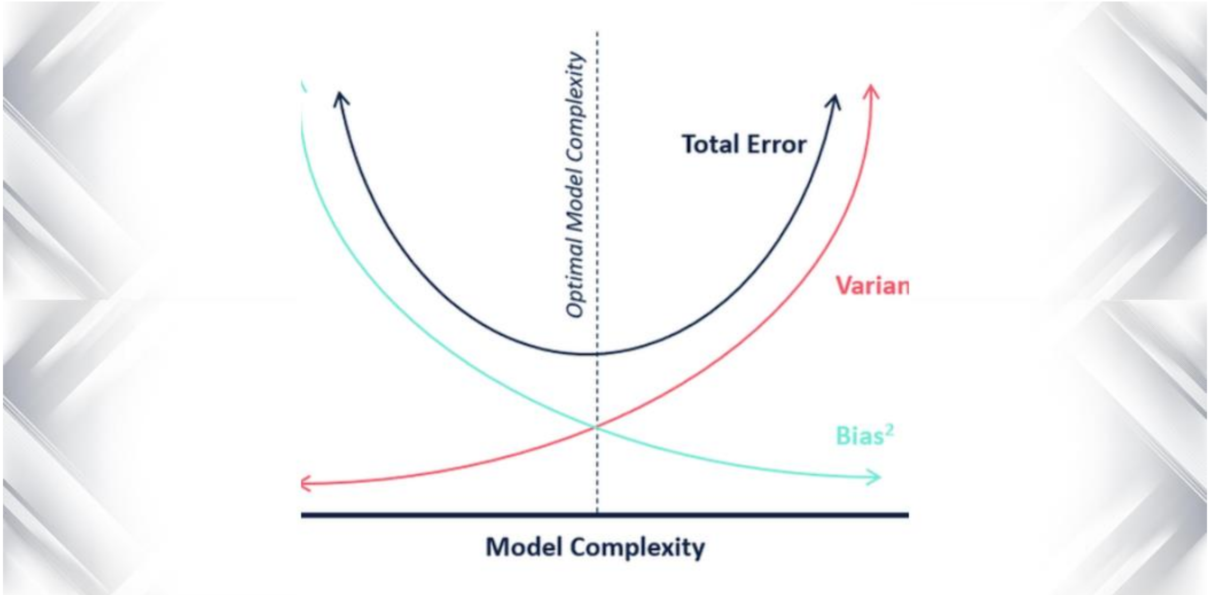
overallqual_3	-34298.311531
overallqual_6	-33576.948884
overallqual_4	-33105.491306
overallqual_5	-32888.573319
overallqual_7	-27212.296777

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?

Answer:

We can make sure that a model is robust and generalisable by understanding the bias-variance tradeoff.



We can make sure that the model is robust and more generalisable by keeping the model complexity at the optimum level. When the model complexity is low, the model tends to be excessively generalised and so leads to underfitting. We can infer that the model is underfitting if the accuracy is low on both training and test data.

When the complexity is high, the model tends to overfit on the training data. So, training accuracy will be very high but the test accuracy will be very low, hence, we can say that the model is memorising the training data and not generalized.

In lasso and ridge regression, we can create models with optimum model complexity by adjusting the lambda values.