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 to 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 values of alpha for ridge and lasso regression are 10 and 200.

As the value of alpha increases, the model complexity reduces in case of ridge regression. But the high values of alpha can result in underfitting. RSS value increase and the coefficients become very very small, tends to zero.

In case of lasso regression increase in value of alpha reduces model complexity and makes more coefficients as zeros.

Ridge top features after doubling the alpha.

	Coefficients	Abs_Coefficients
OverallQual	0.069001	0.0690
GrLivArea	0.062568	0.0626
1stFlrSF	0.051077	0.0511
KitchenQual	0.048175	0.0482
TotalBsmtSF	0.047588	0.0476
TotRmsAbvGrd	0.046070	0.0461
ExterQual	0.045263	0.0453
GarageCars	0.044192	0.0442
Neighborhood_NridgHt	0.041952	0.0420
Neighborhood_NoRidge	0.041056	0.0411

Lasso top features after doubling the alpha.

	Coefficients	Abs_Coefficients
GrLivArea	0.2553	0.2553
OverallQual	0.2072	0.2072
Neighborhood_NridgHt	0.1028	0.1028
GarageCars	0.0946	0.0946
BsmtFinSF1	0.0839	0.0839
Neighborhood_NoRidge	0.0793	0.0793
Neighborhood_StoneBr	0.0674	0.0674
KitchenQual	0.0674	0.0674
ExterQual	0.0583	0.0583
BsmtExposure	0.0484	0.0484

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:

Determining the optimal value of lambda for ridge and lasso regression during the assignment Comparing each of the cross-validation scores allows me to identify the best model for predicting house prices. The ridge regression produces the highest score which implies that shrinking the coefficients for select features improves the overall explanatory power of the model. However, because the scores are so similar, using lasso eliminates coefficients and allows for a more interpretable model, therefore I will make predictions using lasso regression.

For selecting which variables are significant in predicting the price of a house we use lasso regression which results in model parameters such that the lesser important features' coefficients become zero and those who has non zero values are the features which are important for predicting price of a house.

In Lasso regression we see from the plot that the r2 score of testing and training keeps on decreasing as we increase the value of hyperparameter. So I selected the value of lambda where testing value is high for lambda value and model is simple.

In Ridge regression we see from the plot that the test r2 value first increases and then decreases forming a bell curve. But the training r2 value keeps on decreasing as we increase the value of hyperparameter, which is in accordance with the bias-variance trade-off. So I select the value of lambda where testing value is high for lambda value.

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:

The five most important variables after removing the initial five important variables.

	Coefficients	Abs_Coefficients
2ndFlrSF	0.2457	0.2457
1stFlrSF	0.1840	0.1840
BsmtQual	0.0959	0.0959
KitchenQual	0.0940	0.0940
TotalBsmtSF	0.0939	0.0939

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 say a model is robust and generalisable when it is simple. Such models works well even if there are small changes in training data. They are immune to the specifics of training data provided . It rather picks essential characteristics that is invariant across any training datasets. Generic models are bound to perform better on unseen datasets. Such models may make more errors in the training phase but it is bound to outperform complex models when it sees new data. This happens because of overfitting.

There should be a trade off between Accuracy and generalisation (robustness) of the model. It shows that with increasing levels of complexity we are obtaining diminishing returns in terms of accuracy, which never reached 100% accuracy.

The model accuracy on the training data is quite high for the complex model compared to the simple linear model. However, if we extrapolate the fitted lines for both the models, the complex model fails miserably while predicting test data.