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:

a)The optimal value of alpha for ridge regression is 10 and for lasso regression is 1.

b) The difference is

RMSE for ridge regression on testing set: 6.39352483552532e-10 and RMSE for lasso regression on testing set: 5.343131161993016

RMSE for ridge regression with alpha=20 on testing set: 3754.3627037847928 and RMSE for lasso regression with alpha=2 on testing set: 9.533715555818084

c) The important predictor variables will be 'OverallQual', 'Fireplaces', 'SalePrice', 'MSZoning_RM', 'Neighborhood_ClearCr', 'BldgType_Twnhs', 'ExterQual_Gd', 'BsmtQual_Ex', 'BsmtQual_Gd', 'KitchenQual_Gd'.

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:

The optimal values of lambda for both Ridge and Lasso regression is found using GridSearchCV. The optimal value of lambda is the one that gives the best performance on the validation set. The optimal value of lambda for both Ridge and

Lasso regression is 1e-15. In this case, both Ridge and Lasso regression will behave similarly to a linear regression model without regularization.

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 :

BsmtQual_Ex', 'ExterQual_Gd', 'BsmtQual_Gd', 'KitchenQual_Gd', 'BldgType_Twnhs'

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:

To make a model robust and generalizable, you can use techniques like cross-validation, regularization, feature selection, and ensemble methods. Cross-validation is a technique used to assess the performance of a model on an independent dataset. Regularization is a technique used to prevent overfitting by adding a penalty term to the loss function. Feature selection involves selecting a subset of the available features to use in the model. Ensemble methods involve combining the predictions of multiple models to produce a final prediction. These techniques can help improve the robustness and generalizability of the model.