Assignment-based Subjective Questions

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?

Answer1:

Lasso Regression- Optimal value of alpha is 0.0001

Ridge Regression - Optimal value of alpha is 2.0

If we choose to double the value of alpha for Lasso as 0.0002 coeff values are increasing because alpha is getting doubled. test_score from different splits are dropping. We can see more 0s as co-eff

If we choose to double the value of alpha for Ridge as 4.0: Because of double alpha coeff values are increasing. And also score dropped.

Top Features:

OverallQual	0.133
GrLivArea	0.108
1stFlrSF	0.097
2ndFlrSF	0.072
GarageCars	0.067
MasVnrArea	0.066
RoofMatl	0.061
TotRmsAbvGrd	0.058

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?

Answer2:

I will choose Ridge regression as it is giving better R2_Score and the difference of train and test R2_score are better than Lasso. Residual analysis histogram also shows better results for Ridge.

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?

Answer3:

Below are the top most important predictors:

OverallQual	0.133
GrLivArea	0.108
1stFlrSF	0.097
2ndFlrSF	0.072
GarageCars	0.067
MasVnrArea	0.066
RoofMatl	0.061
TotRmsAbvGrd	0.058
LotArea	0.049
OverallCond	0.046
BsmtFinSF1	0.038
TotalBsmtSF	0.037
BsmtFullBath	0.037
Fireplaces	0.032
LandSlope	0.032
Street	0.031
Functional	0.028
GarageArea	0.023
FullBath	0.023
WoodDeckSF	0.022
RoofStyle	0.021
ScreenPorch	0.017
Neighborhood	0.017

LandContour	0.017
SaleCondition	0.015

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?

Answer4:

To make sure model is robust and generalisable, first thing we have to make sure that model gives almost same or better accuracy on unseen test data. We also need to do the Residual Analysis to make sure model is not bias. We need to make sure model is not overfitted. For this we can use regularization parameters.

If the model is not generaized it will be biased towards training data and gived very bad accuracy on test/unseen data.