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 wi
most important predictor variables
after the change is implemented?

Answer:

The optimal values found are:

1. For Lasso : 0.00049901

2. For Ridge : 0.00079901

If we need to use double of it alpha values will be as follows:

1. For Lasso : 0.00099802

2. For Ridge : 0.00159802

Below are the most important predictor variables after the changes :

	Lasso Doubled Alpha Co-Efficient
GrLivArea	0.435977
Neighborhood_NoRidge	0.429756
Neighborhood_NridgHt	0.409260
Neighborhood_StoneBr	0.380720
BsmtExposure_Gd	0.236744
SaleType_New	0.230526
Exterior1st_BrkFace	0.222502
RoofMatl_WdShngl	0.217748
Neighborhood_Somerst	0.214091

	Ridge Doubled Alpha Co-Efficient
RoofMatl_WdShngl	1.770829
RoofMatl_CompShg	1.524328
RoofMatl_Membran	1.485580
RoofMatl_WdShake	1.479181
RoofMatl_Tar&Grv	1.255095
Neighborhood_NoRidge	0.515962
Neighborhood_StoneBr	0.481206
Neighborhood_NridgHt	0.471967
GrLivArea	0.426641
LandSlope_Sev	0.396119
SaleType_New	0.358855

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?

We have selected Lasso as it has high accuracy, low MSE and Lower in comparison to Ridge.
Here are the comparison for support:

	Linear Regression	Ridge	Lasso
MSE Test	8.288314e+06	0.099043	0.072735
MSE Train	2.157934e-01	0.068053	0.060157
R2 Test	-9.458736e+13	0.863628	0.899851
R2 Train	9.399012e-01	0.912171	0.922362
RSS Test	3.008891e+16	43.380931	31.858098
RSS Train	4.754471e+01	69.482589	61.420069

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?

After removing top 5 variables , when once again Lasso model was run below are top 5 :

	Lasso Co-Efficient
RoofMatl_WdShngl	2.348365
RoofMatl_CompShg	2.152992
RoofMatl_WdShake	2.104529
RoofMatl_Membran	2.063042
RoofMatl_Tar&Grv	1.768359
RoofMatl_Metal	1.658552
MSSubClass_75	0.446919
GarageCond_Gd	0.406091
SaleCondition_Alloca	0.393761
MSSubClass_40	0.381164

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?

Robustness evaluation estimates potential failure probabilities when the model is pushed to its limits.
The classic approach towards the assessment of any machine learning model revolves around the evaluation of its generalizability i.e. its performance on unseen test scenarios.

Hence to ensure both aspects When a model accuracy is not affected much when its underlying data is changing (increasing or decreasing).
Also model should not be overfit. For which we can use Regularisations like : Ridge,Lasso.
Moreover , We can use several ways to check this robustness through Segmentation evaluation , Visual Evaluations.

A good model is one with low bias and low variance.