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?

The optimal alpha value for Ridge Regression is: 0.4

The optimal alpha value for Lasso Regression is: 0.0001

After doubling the optimal lambda values for both Ridge and Lasso Regression, we don't see any significant changes in both `metrics (r2 score, RSS and MSE)` and the `features`. There are few, very minor variations here and there but overall, very similar.

The below screenshot shows the most important predictor variables in their order of significance. The green highlighted portion shows the important variables along with their co-efficient values. We can see that the order of features remains the same though their coefficients might vary slightly.

	Linear	Ridge	Lasso	Ridge_Double	Lasso_Double
OverallQual	1.001830	0.977952	1.008233	0.956154	1.014596
BedroomAbvGr	0.580791	0.567406	0.574108	0.552186	0.567384
LotArea	0.539378	0.454733	0.497364	0.396282	0.455398
GarageCars	0.448252	0.450497	0.448156	0.452930	0.448066
BsmtFinSF1	0.390031	0.367451	0.369034	0.347092	0.347902
OverallCond	0.371510	0.367739	0.365301	0.363206	0.359033
Fireplaces	0.219470	0.229152	0.222487	0.237191	0.225531
Neighborhood_StoneBr	0.199406	0.197734	0.192943	0.195944	0.186504
BsmtFullBath	0.174377	0.183909	0.179172	0.190233	0.183958
SaleType_Oth	0.281333	0.198171	0.179133	0.152842	0.076955
Neighborhood_NoRidge	0.162066	0.164379	0.158980	0.166460	0.155915
Neighborhood_Crawfor	0.162664	0.160911	0.158900	0.159430	0.155164
Neighborhood_ClearCr	0.150958	0.153619	0.149765	0.154698	0.148579
Neighborhood_NridgHt	0.132871	0.135628	0.130326	0.137944	0.127795
Exterior1st_BrkFace	0.103714	0.103911	0.102347	0.103559	0.100977
BsmtUnfSF	0.105406	0.106948	0.098380	0.107619	0.091264
LandContour_HLS	0.104736	0.104293	0.095410	0.103317	0.086056
Neighborhood_Veenker	0.103455	0.099964	0.091248	0.096576	0.079050

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?

Below are the reasons why I will prefer Lasso over Ridge Regression:

- The values of R2 Score, RSS and MSE for Lasso Regression are slightly better than Ridge Regression in this model (see screenshot below)
- In Lasso Regression, we can push the model coefficients to actual zero value. This means that the features that have coefficient value of 0 can be removed from the model. This results in feature selection
- Model complexity also reduces because we can remove features with zero coefficients

Metric	Linear Regression	Ridge Regression	Lasso Regression
R2 Score (Train)	0.853738	0.853394	0.853384
R2 Score (Test)	0.834825	0.834772	0.836613
RSS (Train)	23.170141	23.224644	23.226307
RSS (Test)	12.273316	12.277272	12.140511
MSE (Train)	0.150570	0.150747	0.150753
MSE (Test)	0.167396	0.167423	0.166487

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?

Below are the initial top five most important predictor variables in the lasso model

	Linear	Ridge	Lasso
OverallQual	1.001830	0.977952	1.008233
BedroomAbvGr	0.580791	0.567406	0.574108
LotArea	0.539378	0.454733	0.497364
GarageCars	0.448252	0.450497	0.448156
BsmtFinSF1	0.390031	0.367451	0.369034

After removing the above 5 variables from model and creating another Lasso model, below are the five most important predictor variables now.

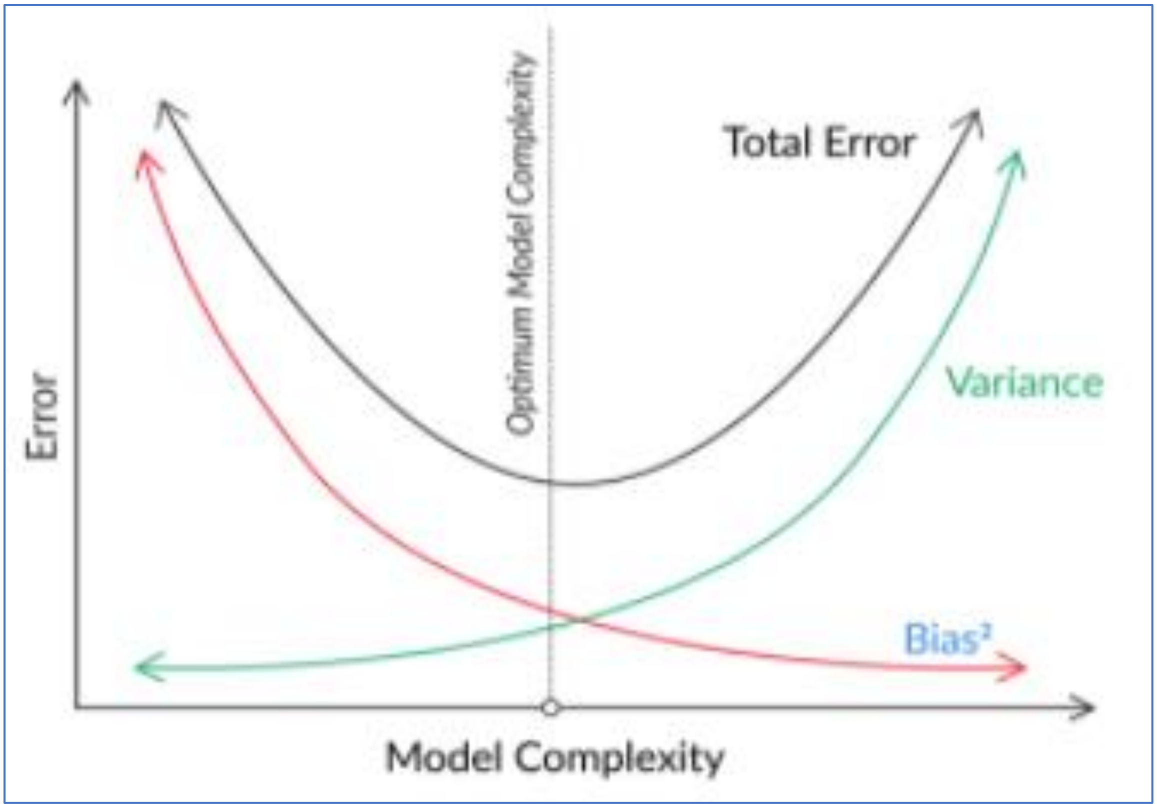
	Lasso
OverallCond	0.591532
Fireplaces	0.535998
Neighborhood_NoRidge	0.386167
BsmtUnfSF	0.359224
BsmtFullBath	0.351013

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?

A robust model has low variance. This means that an unprecedented change in one or more features does not significantly alter the value of the predicted variable. Similarly, a generalizable model has reduced model complexity. As the number of features increase in the model, it becomes more complex which usually leads to low bias but high variance. A generalizable model has just enough features that it has as much low variance as possible.

This can be observed from the Bias-Variance trade-off visual shown below.



The OLS (Ordinary least squares) regression model is very sensitive to outliers, and they induce high variance. To reduce this, we can go ahead with regularization (Ridge/Lasso) which include a penalty term in the cost function of the model. This penalty term will move the coefficients of the model towards 0 and thus it reduces model complexity (as feature addition is heavily discouraged). This reduces overfitting in the model.

So, regularization gets us high variance with a small trade-off in bias. Thus, it helps us build a model which is robust and generalizable. A robust and generalizable model will have a good, consistent train as well as test accuracy.