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?

Answer

From the model, it is observed that the alpha values for Ridge and Lasso regression are – 50 and 500 respectively.

With ridge_alpha = 50 and lasso_alpha = 500

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	8.596079e-01	8.555643e-01	8.564493e-01
1	R2 Score (Test)	7.174659e-01	7.288265e-01	7.224838e-01
2	RSS (Train)	5.903061e+11	6.073082e+11	6.035870e+11
3	RSS (Test)	8.305091e+11	7.971145e+11	8.157589e+11
4	MSE (Train)	2.995199e+04	3.038027e+04	3.028705e+04
5	MSE (Test)	5.426848e+04	5.316622e+04	5.378440e+04

With ridge_alpha = 100 and lasso_alpha = 1000

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	8.596079e-01	8.504198e-01	8.503993e-01
1	R2 Score (Test)	7.174659e-01	7.327966e-01	7.233932e-01
2	RSS (Train)	5.903061e+11	6.289394e+11	6.290257e+11
3	RSS (Test)	8.305091e+11	7.854444e+11	8.130858e+11
4	MSE (Train)	2.995199e+04	3.091658e+04	3.091870e+04
5	MSE (Test)	5.426848e+04	5.277560e+04	5.369621e+04

We see the Ridge regression performs well in improving the R2 score whereas Lass regression keeps it more or less same.

	Important Predictor
ridge_alpha = 50, lasso_alpha = 500	GrLivArea
ridge_alpha = 100, lasso_alpha = 1000	

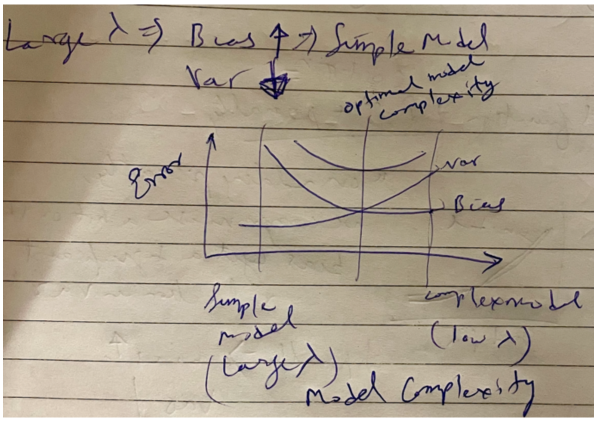
The important predictor "GrLivArea" remains the same in both the cases.

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

Larger alpha implies simpler model. Hence the choice of model should also depend on the underlying alpha value. In the assignment, lasso had a larger alpha (500) than Ridge (50).



Moreover, Lasso Regression is more favourable when it comes to furnishing predictor selection options. And this is the usual case when one wants to reduce the complexity of a model – one of the biggest factors are the number of predictor variables.

As it can be seen from the assignment analysis, with Lasso regression number of variables are seen to reduce from 62 to 44. With the reduction of predictors, the R^2 score is essentially more or less same indicating usage of redundant predictors when using either Linear or Ridge regression.

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?

Answer

The top 5 predictors in the Lasso model (before removing them) are –

```
GrLivArea              25626.537456
TotalBsmtSF            14284.829587
GarageCars              9798.824921
BsmtFinSF1              9046.563393
LotArea                 6689.373926
```

After removing these predictors, the new top 5 are –

Fireplaces	11993.277752
FullBath	10649.657550
BsmtFullBath	9587.675503
MasVnrArea	9503.014876
LotFrontage	9245.759610

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

For a model to be robust and generalizable, it must have an optimum model complexity which largely depends on the number of predictors it uses for training, and more importantly it should have optimum trade-off between variance and bias. One must be careful not to overfit a model while training as it is highly likely to fail on test data.

One technique to reduce the model complexity (and hence making it more robust) is to reduce the magnitude of the model coefficients and decreasing number of coefficients – which can be achieved by regularization. Regularization techniques include Lasso and Ridge regressions. Lasso regression plays a vital role in reducing number of predictors since it can assign zero coefficients to them while Ridge doesn't (it tries to tone them down to zero).

An implication of this could be underutilizing data. One must be careful while performing hyperparameter tuning for these regressions since if not done properly the model might not perform well on unseen data.

Hyperparameter tuning is also an important activity that one must carry out before fitting the data with predictions from a model. Lasso/Ridge regressions add penalties to the RSS cost function thereby **allowing a good fit by compromising a bit on the bias to get a significant reduction in variance**.