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:

After building a model Using Ridge and Lasso Regression, Optimum value for Alpha we got

For Ridge : 2

For Lasso : 0.01

If we double the value of alpha, there will be increase in the penalty on the model. The complexity of the model may decrease but there will be a increase in the error term.

Top Predictors For

Ridge:

- MSZoning_FV
- MSZoning_RH
- MSZoning_RL
- MSZoning_RM
- Street_Pave

Lasso

- GrLivArea
- OverallQual
- LotArea
- BsmtFinSF1
- TotalBsmtSF

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:

Since the value of Alpha is smaller in Lasso as compared to Ridge, So I would prefer Lasso over Ridge. Lasso model does not include all features in the model. It only includes features selection by identifying features which are significant and makes other non-significant coefficients as zero. So this helps in selecting the best predictor variables.

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:

Next five important features are
- BedroomAbvGr
- FullBath
- ScreenPorch
- OpenPorchSF
- WoodDeckSF

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 making any model robust and generalised, we should keep the model simpler. The r2 score of train and test data should be same with less difference. If the difference Is more then the model is overfitted.

For simple model, bias is more, and variance is less hence making it a generalised mode. It is very important to maintain a balance between bias and variance in order to avoid overfitting and underfitting too.

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	RMSE (Train)	0.084977	0.096752	0.153266
1	RMSE (Test)	0.154157	0.137977	0.168913
2	R2 Score (Train)	0.950725	0.936123	0.839709
3	R2 Score (Test)	0.874256	0.899265	0.849031