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:

Optimal Value:

Optimal value of alpha for **ridge** is **1.0**
Optimal value of alpha for **lasso** is **0.0001**

Changes on doubling alpha:
Before doubling the alpha, we have:

Ridge
R2 of train = 0.9509648397166721
R2 of test = 0.9052772986289798

Lasso
R2 of train = 0.943994412716278
R2 of test = 0.9092244104989701

If we double the alpha for both the models, then we have:

Ridge
R2 of train = 0.9477616165424759
R2 of test = 0.9063328456144368

Lasso
R2 of train = 0.9368072103626016
R2 of test = 0.9129787556044949

We can observe that for Ridge R2 for train drops and for Lasso R2 for train improves. The changes are intuitive. Higher-than-optimal alpha penalizes the model more, forcing it to be simpler and more generic. That results in a simpler-than-optimal model that is more biased.

Most important predictors after doubling alpha:
The most important predictor for both models remain same even after doubling alpha, i.e.:

- Ridge: *OverallQual*
- Lasso: *GrLivArea*

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:

I will choose to apply **Lasso.** Since it will have better explain-ability given it has put some coefficients are zero while still having R2 values approximately same, i.e. the prediction results will be same but Lasso will can do it with less variables hence it will be easier to explain same variables to business as oppose to Ridge which will have close to 200 variables with many having negligible coefficients

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:

Existing top 5 variables in Lasso are :

```
GrLivArea       0.309101
OverallQual     0.159516
LotArea         0.112963
TotalBsmtSF     0.093961
OverallCond     0.085634
```

On excluding these and re-running the Lasso model fit, we get the following variables in top 5:

```
1stFlrSF        0.298168
2ndFlrSF        0.147612
MSZoning_FV     0.080716
BsmtFinSF1      0.070135
MSZoning_RL     0.066258
```

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:

Simpler models tend to be more generalisable and robust, here simpler means the model having less number of features or lower degree relations or simpler coefficients in comparison to other models.

Because of this explain-ability improves and the fact that there are features which have quite significant coefficients imply that general pattern of the equation is captured and change in inputs will cause too much variance unless inputs are way beyond the training bounds.

There are various ways for ensuring a model is simpler, more robust and generalisable. We can use ***regularization*** techniques where the regularization term tries to penalize the model for coefficients count and higher values there by forcing the Cost function optimization to strike a balance between learning pattern and having sane coefficients values/count.

We can also achieve increase in robustness and generalisability by using some other measures. It depends on the type of the model used. For example, in case of regression, we can limit the number of features used while, in case of decision tree, we can limit the number of nodes, the number of levels etc.

Even data preparation also contribute to more robust models. Binning of continuous variables result in lesser variance. Outliers can impact the robustness of the model therefore outliers have to be removed as they distort the model itself.

The cost of simpler model is predictive power i.e. trade-off between bias and variance where it increases bias but with huge reduction in variance which translates into lower accuracy in train set in comparison but result in consistent accuracy across unseen data with good accuracy.