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 of alpha for ridge is 5 and lasso is 0.0001.
- If the alpha values are doubled the error terms increase indicating reduced accuracy and variance. The coefficients reduce indicating reduction in complexity of model.
- In case of Ridge the most important predictor variable changes from **OverallQual_10** to **GrLivArea**. In case of Lasso it is unchanged at **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 prefer Lasso regression with alpha=0.0002 because number of variables are eliminated and reduces to 93. I would rather compromise a bit on the error terms but reduce the complexity of model.

```
r2_test: 0.771715259947595
rss_test: 1.346923713129009
mse_test: 0.0030892745713968096
```

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?

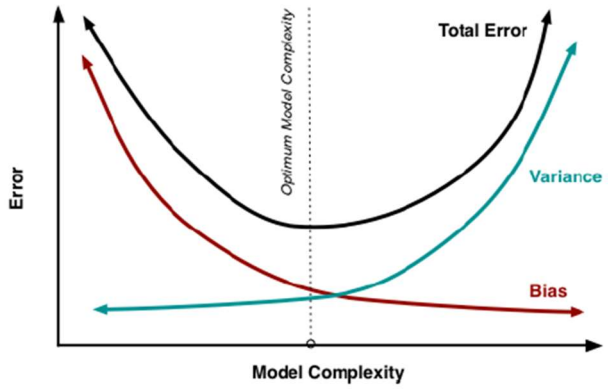
Answer:

	Lasso
1stFlrSF	0.230599
2ndFlrSF	0.194285
BsmtFinSF1	0.149897
BsmtUnfSF	0.089504
BsmtFinSF2	0.052919

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:

After we create the model we have to find out the value of evaluation metrics i.e. r2, rss, mse, etc. A model whose indicators are high with minimal differences between train and test dataset score is expected to be robust and generalizable.



As per the Bias Variance trade off we have to understand that a more complex model will have less bias but the variance on test dataset will be high. Thus we should use techniques like hyperparameter tuning to select minimum feature variables which reduce complexity and is more generalizable. In addition use cross validation technique so that model runs multiple times on the different sets of training data but sees the test data only once thus the accuracy scores obtained are more robust.

If the model is very complex and has more features it will do very well on the training set but if the test dataset indicators a very low it means that model is overfitting or memorizing the training data and is unable to predict on the test data. As per above figure we should reduce model complexity so that bias increases by a little but the variance on the test dataset is reduced significantly.