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?

- Optimal values of alpha are Lasso(alpha=0.001) Ridge(alpha=33)
- Ridge alpha = 33

```
Train_R2:  0.9269140011785715
Test_R2:  0.9127679513418698
Train_RSS:  10.482547078227874
Test_RSS:  3.2464977762943867
Train_MSE:  0.00953825939784156
Test_MSE:  0.011805446459252316
```

- Ridge alpha = 66

```
Train_R2:  0.9257223510660966
Test_R2:  0.9121039101147994
Train_RSS:  10.653462555969579
Test_RSS:  3.271211266350096
Train_MSE:  0.009693778485868589
Test_MSE:  0.01189531369581853
```

- Lasso alpha = 0.001

```
Train_R2:  0.926177495078671
Test_R2:  0.9128757333160539
Train_RSS:  10.588182357078924
Test_RSS:  3.242486476033807
Train_MSE:  0.009634378850845246
Test_MSE:  0.011790859912850208
```

- Lasso alpha = 0.002

```
Train_R2:  0.9245276718413202
Test_R2:  0.9118918166538416
Train_RSS:  10.824812491922387
Test_RSS:  3.2791047064327046
Train_MSE:  0.009849692895288796
Test_MSE:  0.011924017114300745
```

- On doubling the alpha value for both Lasso and Ridge R2 value of train and test slightly reduced, RSS & MSE is slightly increased.

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?

- If there are too many variables and we want to perform feature selection, then we can use Lasso. If we want to reduce the coefficient magnitude then we can use Ridge Regression.

Question 3

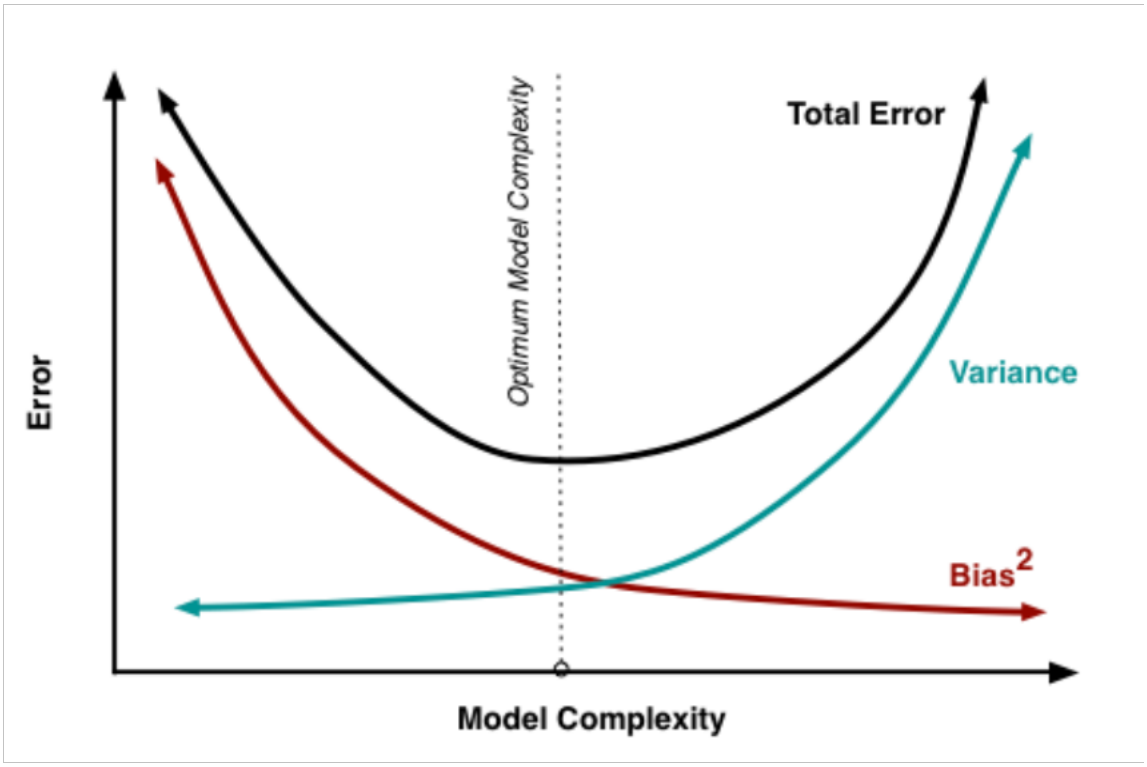
After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

- Initially, top five from Lasso regression are 'GrLivArea', 'property_age', 'OverallQual', 'OverallCond', 'TotalBsmtSF'.
- After dropping the above top 5 variables and rebuilding lasso regression gives below variable as most important predictors(new top 5).
- 'BsmtFinSF1', 'GarageArea', 'FullBath', 'BsmtUnfSF', 'KitchenQual'

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 Model will be robust and generalisable when the performance is not degraded on any variation in data and performs well new/unseen data.
- In general, model should not be too complex, If complexity increases there is a fare chance of overfitting.
- Complex model will have high accuracy on train data but not on test data, due to high variance.
- We have to reduce the variance which can introduce bias.



- We have to find the balance between vairance, bias and complexity to make our model robust and generalisable. Which can be achieved by Regularization techniques like Ridge Regression and Lasso.