- **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?

<u>Answer: -</u> Below is the R2 score and optimum alpha value in Ridge and Lasso regression.

Optimal value of alpha in Ridge regression model
The optimum alpha for the ridge is 10.000000
Ridge Regression with 10.0
===================================
R2_score (train): 0.9099792685238111
R2_score (test): 0.8987501807731075
RMSE (train): 0.11987714672785361
RMSE (test): 0.12680116170772152
===
Optimal value of alpha in Lasso regression model
The optimum alpha for lasso is 0.001000
lasso Regression with 0.001
===================================
R2_score (train): 0.9092384262190387
R2_score (test): 0.8992035193146645
RMSE (train): 0.12036941161238388
RMSE (test): 0.1265169718517261

Now if we change the optimum alpha value for both Ridge and Lasso regression models, we get the below results.

Double the optimal value of alpha in Ridge regression model
The optimum alpha for the ridge is 20.000000
Ridge Regression with 20
===================================
R2_score (train): 0.9096835169735116
R2_score (test): 0.8992377551692294
RMSE (train): 0.120073905737566
RMSE (test): 0.12649548407518718
===
Double the optimal value of alpha in the Lasso regression model
The optimum alpha for lasso is 0.002000
lasso Regression with 0.002
===================================
R2_score (train): 0.9083209354111204
R2_score (test): 0.898450676238623

RMSE (train): 0.12097627709190853
RMSE (test): 0.1269885668782859

Ridge Regression comments: - With optimal alpha value and with the double optimal alpha value

- After we double the optimal value of alpha for ridge regression, we can see that there is a very slight change in the R^2 score and RMSE value for train and test data.
- The top 10 feature with optimal value of alpha (alpha = 10.0) are- ['1stFlrSF', '2ndFlrSF', 'OverallQual', 'BsmtFinSF1', 'SaleType_New','OverallCond', 'LotArea', 'MSZoning_RL, 'SaleCondition_Normal','Condition1_Norm']
- The top 10 feature with optimal value of alpha (alpha = 10.0) are- ['1stFlrSF', '2ndFlrSF', 'OverallQual', 'BsmtFinSF1', 'OverallCond','SaleType_New', 'LotArea', 'MSZoning_RL', 'SaleCondition_Normal','Condition1_Norm']
- We can see that the top 10 features are the same. Just the position of two features i.e., 'SaleType_New', and 'OverallCond' is changed.
- So, we can conclude that for this model after doubling of optimal alpha value there is a very minor change in the model.

Lasso Regression comments: - With optimal alpha value and with the double optimal alpha value

- After we double the optimal value of alpha for lasso regression, we can see that there is a very slight change in the R^2 score and RMSE value for train and test data.
- The top 10 feature with optimal value of alpha (alpha = 0.002) and double of optimal value of alpha(alpha=0.002) are same- ['1stFlrSF', '2ndFlrSF', 'OverallQual', 'BsmtFinSF1', 'OverallCond', 'LotArea', 'MSZoning_RL', 'SaleCondition_Normal', 'Condition1_Norm','SaleType_New']
- We can see that the top 10 features are same. But there are many features in alpha value 0.002, which having 0 coefficients. The variable is - ['Exterior1st_Plywood', 'Exterior2nd_Wd Sdng', 'MasVnrType_BrkFace','MasVnrType_None', 'MasVnrType_Stone', 'GarageType_Not_applicable','SaleCondition_Partial']
- So, for alpha 0.001 in lasso regression there are 45 features and for alpha 0.002 there are 38 features.
- There is a little shuffling of the variable when arranged in descending order.
- So, we can conclude that for the lasso model after doubling of optimal alpha value the model is changed slightly.

Overall, the features in both ridge and lasso models for optimal alpha and double optimal alpha, when they are arranged in descending order, the top 10 features are the same in all 4 models just some shuffling is there.

- ## **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 to and why?

<u>**Answer: -**</u> It is important to regularize coefficients and improve the prediction accuracy also with the decrease in variance, and make the model interpretably.

Ridge regression, uses a tuning parameter called lambda as the penalty is the square of the magnitude of coefficients which is identified by cross-validation. The residual sum of squares should be small by using the penalty. The penalty is lambda times the sum of squares of the coefficients, hence the coefficients that have greater values get penalized. As we increase the value of lambda the variance in the model is dropped and bias remains constant. Ridge regression includes all variables in the final model, unlike Lasso Regression.

Lasso regression, uses a tuning parameter called lambda as the penalty is the absolute value of the magnitude of coefficients which is identified by cross-validation. As the lambda value increases Lasso shrinks the coefficient towards zero and it makes the variables exactly equal to 0. Lasso also does variable selection. When the lambda value is small it performs simple linear regression and as the lambda value increases, shrinkage takes place and variables with 0 value are neglected by the model.

In our model, we get the below result in ridge and lasso regression.

```
***Optimal value of alpha in Ridge regression model***
Fitting 5 folds for each of 27 candidates, totaling 135 fits
The optimum alpha for the ridge is 10.000000
Ridge Regression with 10.0
===================================
R2_score (train): 0.9099792685238111
R2_score (test): 0.8987501807731075
RMSE (train): 0.11987714672785361
RMSE (test): 0.12680116170772152
====================================================================
***Optimal value of alpha in Lasso regression model***
Fitting 5 folds for each of 27 candidates, totaling 135 fits
The optimum alpha for lasso is 0.001000
lasso Regression with 0.001
===================================
R2_score (train): 0.9092384262190387
R2_score (test): 0.8992035193146645
RMSE (train): 0.12036941161238388
RMSE (test): 0.1265169718517261
```

- We will use lasso regression with an alpha value of 0.001. As the lasso regression is providing an R2 score of 0.8992 i.e., 89.92% of the variance in the test data can be explained by the mode whereas, in ridge regression with alpha value 10, the R2 score is 0.89875 i.e., 89.88% of the variance in the test data can be explained by the model.
- The RMSE value (test data) for the lasso model is 0.1265 which means the prediction made by the model can be off by 0.1265 units. And RMSE value (test data) for the ridge model is 0.1268 which means the prediction made by the model can be off by 0.1268 units. So, the RMSE value is better in the lasso model.
- By this, we can conclude that the Lasso model is performing slightly better than the ridge model.

Question 3

After building the model, you realized 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?

<u>Answer: -</u> After dropping the 5 most important predictor variables in the lasso model, and creating another new model, we get below five most important variables.

['FullBath', 'LotArea', 'GarageArea', 'KitchenQual', 'FireplaceQu']

	Lasso
FullBath	0.094348
LotArea	0.063393
GarageArea	0.062842
KitchenQual	0.060083
FireplaceQu	0.058542

Lasso regression model after dropping top 5 predictor variable

Fitting 5 folds for each of 27 candidates, totaling 135 fits
The optimum alpha for lasso is 0.001000
lasso Regression with 0.001
===================================
R2_score (train): 0.8250016822593496
R2_score (test): 0.815819198082042
RMSE (train): 0.16714064185839472
RMSE (test): 0.17102054415528295

➢ Question 4

How can you make sure that a model is robust and generalizable? What are the impli cations of the same for the accuracy of the model and why?

Answer: -

As Per the Occam's Razor rule, when there are two models which show similar perfor mance in the finite training or test data, then we should pick the one which is simple or less complex due to the following reasons: -

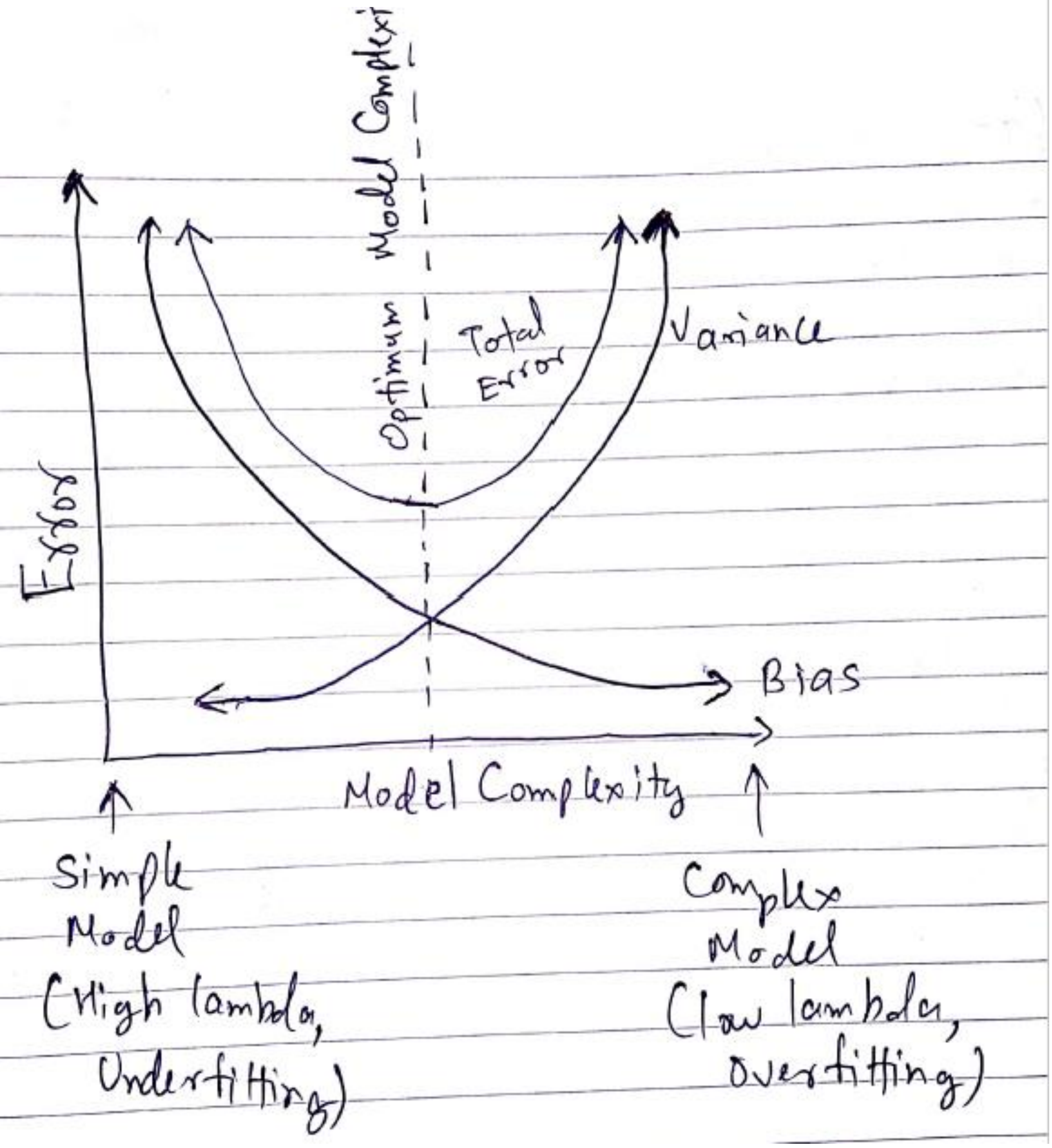
- Simpler models are usually more 'generic' and are more widely applicable.
- When the model is simple, it requires fewer training samples for effective training of the model than the more complex ones, and hence they are easier to train.
- Simpler models are more robust.
    - o Complex models tend to change wildly with changes in the training data set.
    - o Simple models have low variance, and high bias and complex models have low bi as and high variance.
- The simple model makes more errors compared to the complex ones in the trainin g data set. Complex models lead to overfitting — they work very well for the training samples but fail miserably when applied to other test samples.

So, to make the model more generalizable and robust, we need to create a simple m odel but not very simple which is not of any use. Regularization is used for making th e model simpler. Regularization helps to balance between bias and variance and kee ping the model simple and not making it too naive to be of any use. For regression, r egularization involves adding a regularization term to the cost that adds up the absol ute values or the squares of the parameters of the model

So, making a model less complex or simple brings the Bias-Variance Trade-off concep t:

- A complex model is needed to change for every little change in the dataset and so i s very unstable and very sensitive to any changes in the training data set.
- The simple model which produces some pattern followed by the data points given i s unlikely to change wildly when more points are added or removed.

Bias is how much error the model is likely to make on the test data. A complex model can perform very well when there is huge training data. Models that are too naïve, fo r e.g., one that gives the same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs is very hig h.
Variance is the degree of changes in the model itself when there is a change in the tr aining data.
So, for maintaining a balance between the bias and variance such that both are as lo w as possible, in turn, which minimizes total error, we use a bias-variance trade-off.
By bias-variance trade-off we get optimal model complexity. Below is the diagram of the bias-variance trade-off.


Optimum Model Complexity
Total Error
Variance
Error
Bias
Model Complexity
Simple Model (High lambda, Underfitting)
Complex Model (low lambda, Overfitting)