Assignment Part-II

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: The optimal values of alpha for Ridge and Lasso regression for the model are:
Ridge: 3.0
Lasso: 0.001

The changes in the model for double the value of alpha for Ridge and Lasso?
Ridge Regression:

Alpha = 3.0

```
For Ridge Regression Model (alpha=3.0):
 ***************************************

For Train Set:
R2 score: 0.9146333125386124
MSE score: 0.08536668746138759
RMSE score: 0.29217578178450654

For Test Set:
R2 score: 0.900344689565368
MSE score: 0.08650512476172913
RMSE score: 0.2941175356243302
 ***************************************
```

Alpha = 6.0

```
For Ridge Regression Model (alpha=6.0):
 ***************************************

For Train Set:
R2 score: 0.9113880177676933
MSE score: 0.08861198223230671
RMSE score: 0.29767764819063375

For Test Set:
R2 score: 0.9017947688151875
MSE score: 0.08524639318111452
RMSE score: 0.29196984978095686
 ***************************************
```

For Ridge regression we can see with the increase or double the alpha value:

- R2 score decreases for the training set
- RMSE score increases for the training set

Lasso Regression:

Alpha = 0.001

```
For Lassp Regression Model (alpha=0.001):
 ************************************
```

```
For Train Set:
R2 score: 0.9097180649311687
MSE score: 0.09028193506883135
RMSE score: 0.30046952435951196
```

```
For Test Set:
R2 score: 0.9003168898100629
MSE score: 0.08652925615312744
RMSE score: 0.2941585561446878
 ************************************
```

Alpha = 0.002

```
For Lasso Regression Model (alpha=0.002):
 ************************************
```

```
For Train Set:
R2 score: 0.9041586149624965
MSE score: 0.09584138503750356
RMSE score: 0.3095825980857186
```

```
For Test Set:
R2 score: 0.9002818279060192
MSE score: 0.08655969140409855
RMSE score: 0.2942102843275512 5
 ************************************
```

For Lasso regression we can see with the increase or double the alpha value:

- R2 score decreases for the training set and increases for the test set
- RMSE score increases for the training set and test set

What will be the most important predictor variables after the change is implemented?

Ridge Regression:

```
For Ridge Regression for Alpha = '6.0'
The most important 10 predictor variables are:
```

```
0              GrLivArea
1           MSZoning_FV
2   Neighborhood_Crawfor
3           MSZoning_RL
4         AgeofProperty
5           OverallQual
6            CentralAir
7       Foundation_PConc
8   Neighborhood_MeadowV
9  SaleCondition_Partial
```

Lasso Regression:

```
For Lasso Regression for Alpha = '0.002'
The most important 10 predictor variables are:
```

```
                  Feature
0              GrLivArea
1   Neighborhood_Crawfor
2           MSZoning_FV
3  SaleCondition_Partial
4           OverallQual
5         AgeofProperty
6           MSZoning_RL
7       Foundation_PConc
8             CentralAir
9        GarageType_None
```

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: The optimal values of alpha for Ridge and Lasso regression for the model are:
Ridge: 3.0
Lasso: 0.001

Ridge Regression:	**Lasso Regression:**
For Ridge Regression Model (alpha=3.0): For Train Set: R2 score: 0.9146333125386124 MSE score: 0.08536668746138759 RMSE score: 0.29217578178450654 For Test Set: R2 score: 0.900344689565368 MSE score: 0.08650512476172913 RMSE score: 0.2941175356243302	For Lasso Regression Model (alpha=0.001): For Train Set: R2 score: 0.9097180649311687 MSE score: 0.09028193506883135 RMSE score: 0.30046952435951196 For Test Set: R2 score: 0.9003168898100629 MSE score: 0.08652925615312744 RMSE score: 0.2941585561446878

From the above we can infer that:

- R2 Test Score for the Lasso and Ridge regression are nearly same but the R2 Training score of Lasso regression is less than that of Ridge regression. This summarize the Lasso model is more generalized than Ridge model and would not lead to overfitting.

- The RMSE and MSE Test score for Ridge and Lasso regression are nearly the same.

- As Lasso penalize the datasets and helps in feature selection i.e., some of the predictor variables become 0. Lasso would better help in predicting variables for price prediction of the house then Ridge.

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?

Answer: The initial most import 5 predictor variables for Lasso models are:

1. MSZoning_FV
2. GrLivArea
3. Neighborhood_Crawfor
4. MSZoning_RL
5. Functional_Sev

```
The most important 5 predictor variables after removing initials 5 topmost
variables are:
```

```
1    Neighborhood_IDOTRR
2   Neighborhood_OldTown
3  SaleCondition_Partial
4   Neighborhood_MeadowV
5        Foundation_Slab
```

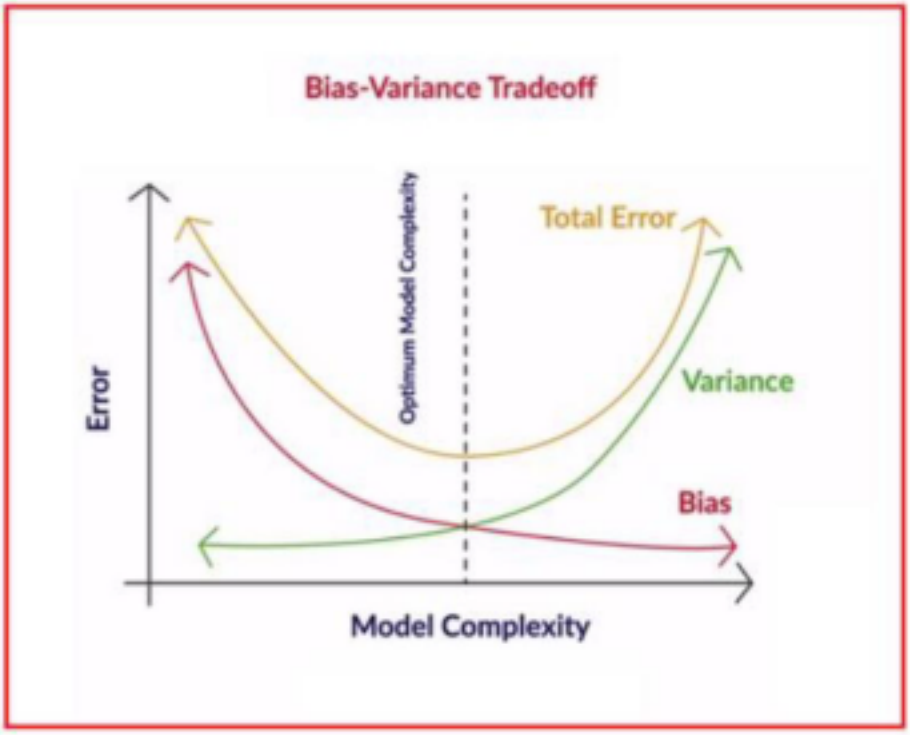
Question 4
How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer: Robustness of the model is the measure of its successfully applying to the data sets other than the one used for training. It implies the testing error of the model is equivalent to the training error.

Regularization helps in controlling between the model complexity and bias which is connected to model robustness. Regularization helps in optimal complexity of the model by penalizing the coefficients for making model complex. It helps in keeping a balance between making model optimally simple but also not to naïve. It helps in regularization between Bias and Variance Tradeoff.

A complex model would perform well in training set but would need to change with changes in dataset and hence it would have high variance and low bias.

A simple model may perform poorly in training but would need less/no change with changes in dataset and hence it would have low variance and high bias.



The accuracy of the model can be maintained by keeping balance between Bias-Variance trade off and keeping the model Optimum complex.