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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer 1

The optimal value of alpha for Ridge Regression is **10** and for Lasso Regression it is **0.001**.

Upon doubling the current values of Ridge Regression from **10** to **20**, and Lasso Regression from **0.0001** to **0.001.** The R2 Scores of the model reduced indicating that some bias was compromised. In the case of Lasso Regression, upon doubling the value of alpha the feature selection reduced the selected feature from **79 features out of 239** to **27 features out of 239** indicating feature selection by reducing the predictors coefficient to zero.

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score Train	0.928385	0.882023	0.896843
1	R2 Score Test	0.687235	0.860226	0.873542
2	RSS Train	0.941076	1.550307	1.355566
3	RSS Test	0.913808	0.408379	0.369474
4	RMSE Train	0.029657	0.038064	0.035593
5	RMSE Test	0.058393	0.039036	0.037130

Figure 1 Before doubling the alpha values for Ridge and Lasso

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score Train	0.928385	0.868247	0.830323
1	R2 Score Test	0.687235	0.853193	0.825794
2	RSS Train	0.941076	1.731333	2.229685
3	RSS Test	0.913808	0.428928	0.508978
4	RMSE Train	0.029657	0.040225	0.045649
5	RMSE Test	0.058393	0.040006	0.043580

Figure 2 After doubling the alpha values for Ridge and Lasso

Before changing alpha (Coefficients)	After changing alpha (Coefficients)
Lasso - Condition2_positive_off_site_feature **(0.164863)** - GrLivArea **(0.314975)**	Lasso - FullBath_1 **(-0.013901)** - Neighborhood_Edwards **(-0.012640)** - GrLivArea **(0.184986)**
Ridge - FullBath_1 **(-0.030537)** - BsmtQual_Typical **(-0.027392)** - Condition2_positive_off_site_feature **(-0.025430)** - Neighborhood_Edwards **(-0.024247)** - BsmtQual_Good **(-0.022152)** - TotRmsAbvGrd_4 **(-0.021633)** - Neighborhood_Stone_Brook **(0.048034)** - Neighborhood_Northridge **(0.049798)** - FullBath_3 **(0.050796)** - KitchenQual_Excellent **(0.051184)** - GrLivArea **(0.055201)**	Ridge - Neighborhood_Stone_Brook **(0.035690)** - Fireplaces_2 **(0.035868)** - GarageCars_3 **(0.038748)** - GrLivArea **(0.039921)** - Neighborhood_Northridge **(0.040248)** - FullBath_3 **(0.043138)** - FullBath_1 **(-0.028227)** - BsmtQual_Typical **(-0.023774)** - Neighborhood_Edwards **(-0.021971)**

The above table showcases the important predictor variables after the change.

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one would you choose to apply and why?

Answer 2

The optimal value for Ridge and Lasso that was computed was 10 and 0.0001 respectively. Looking at the table below that lists the R2 Score and Residual Sum of Squares for both Ridge and Lasso Regression models, I will choose to apply the Lasso Regression here, because the R2 Score for Lasso regression for both train-test data is higher than Ridge Regression (i.e., Lasso has R2 Score of **89.7% Train**, **87.35% Test** compared to Ridge with **88.2% Train, 86.02% Test**). The value of the cost function (Root Mean Square Error) is the lowest for Lasso with RMSE of **3.5% for Train Dataset** and **3.7% for the Test Dataset**. Which means that the model generated by Lasso has a good fit on the data and is more accurate than Ridge. The data points explained by Lasso model are higher than Ridge model.

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score Train	0.928385	0.882023	0.896843
1	R2 Score Test	0.687235	0.860226	0.873542
2	RSS Train	0.941076	1.550307	1.355566
3	RSS Test	0.913808	0.408379	0.369474
4	RMSE Train	0.029657	0.038064	0.035593
5	RMSE Test	0.058393	0.039036	0.037130

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 3

The five most important predictor variables in the lasso model while building the model were:

- **Condition2_positive_off_site_feature (-0.164863 Coef)**
- **GrLivArea (0.314975 Coef)**
- **RoofMatl_Wood_Shingles (0.103682 Coef)**
- **FullBath_3 (0.072285 Coef)**
- **Neighborhood_Stone_Brook (0.069790 Coef)**

Now to create another model excluding these five most important predictor variables, the new five most important predictor variables are:

- **FullBath_1 (-0.077370 Coef)**
- **FullBath_2 (-0.066567 Coef)**
- **TotalBsmtSF (0.164177 Coef)**
- **2ndFlrSF (0.115955 Coef)**
- **KitchenQual_Excellent (0.064166 Coef)**

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 4

To ensure that a model is robust and generalizable, it should have the optimal value of error terms such that it has optimum low bias and optimum low variance. A model that has less variance is more generalizable and performs well on unseen data, the model should bias enough to understand all the underlying patterns in the training data set. This makes the model perform better on both seen and unseen data. To achieve this, we can regularize the model or by tuning the hyperparameters to regularize the model coefficients of the model parameters such that it does not become too complex or too simple.

Regularization is a method where a penalty is added to the cost function. Such that the optimum

$$f(\theta)=\sum(y-y_p)^2+\lambda\sum(y-y_p)^2$$

value of alpha (lambda) regularizes the model by penalize when it becomes too complex. A high value of lambda will make a complex model simple and vice versa.

A model that is less biased is a complex model which understands all the patterns in the underlying training dataset, but the underperforms on unseen data (test data), a complex model has high a variance meaning if the train data is changed the model shows a high variance on the unseen data, it cannot predict new insights and need more data to re-train in order to make any predictions, hence it is not generalizable.

On the contrary a model that is simple has high bias meaning that it is unable to recognize the underlying patterns in the training data set, hence it underperforms on seen data, however a simple model has low variance, thus can perform well on unseen data making this model more generalizable. However, this model is susceptible to high errors too when introduced to complex data leading too poor accuracy due it inability to recognize complex underlying patterns.

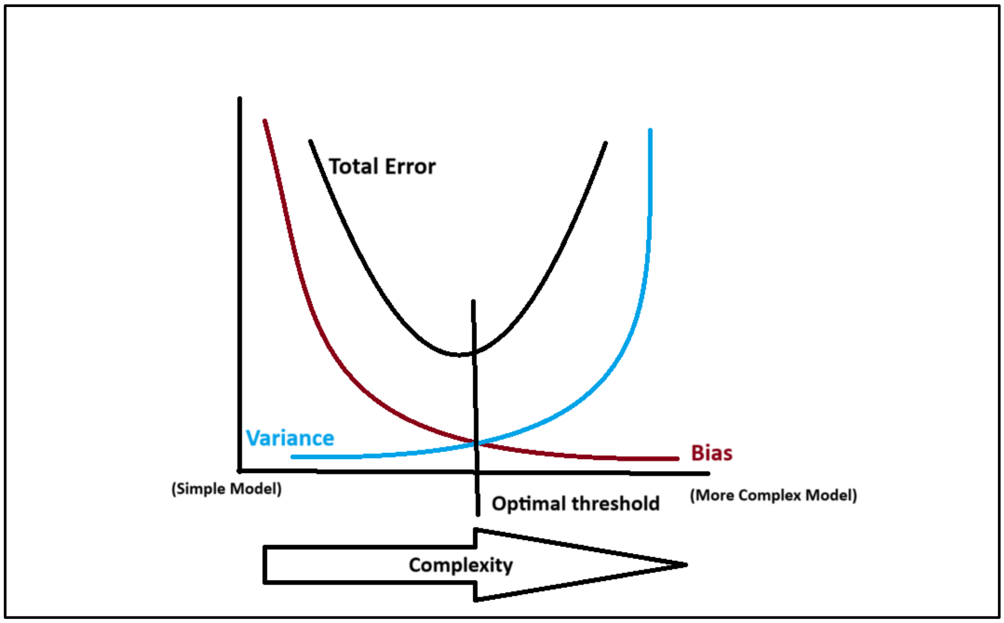

Figure 3 Bias-Variance Trade-Off