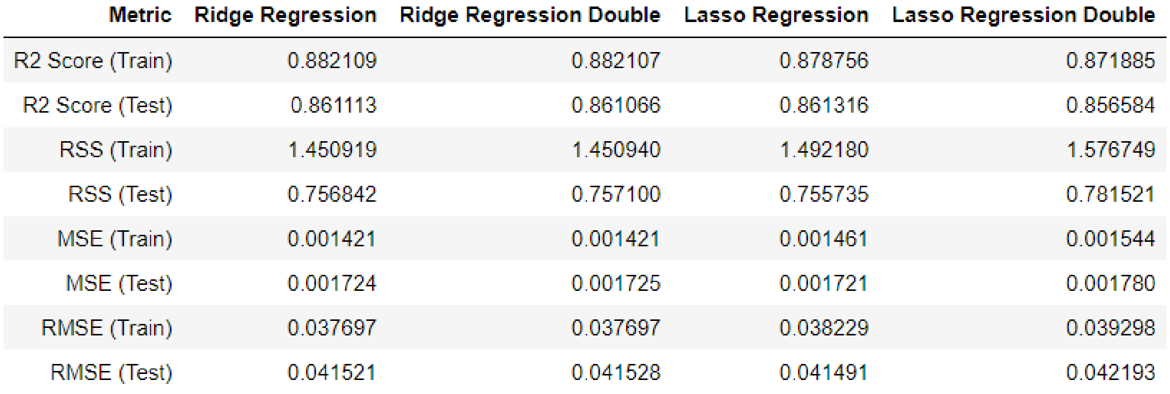
1. **What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?**

A. For the features selected post feature elimination, the alpha values are as follows
The optimal value of alpha for Ridge Regression is 0.01
The optimal value of alpha for Lasso Regression is 0.0001

If the above values of alpha are doubled, then alpha value for Ridge and Lasso will be as follows:

- Ridge Regression alpha is 0.02
- Ridge Regression alpha is 0.0002

With the above alpha values upon rebuilding the model to understand the changes caused by doubling the alpha values following observations were made:



| Metric | Ridge Regression | Ridge Regression Double | Lasso Regression | Lasso Regression Double |
|---|---|---|---|---|
| R2 Score (Train) | 0.882109 | 0.882107 | 0.878756 | 0.871885 |
| R2 Score (Test) | 0.861113 | 0.861066 | 0.861316 | 0.856584 |
| RSS (Train) | 1.450919 | 1.450940 | 1.492180 | 1.576749 |
| RSS (Test) | 0.756842 | 0.757100 | 0.755735 | 0.781521 |
| MSE (Train) | 0.001421 | 0.001421 | 0.001461 | 0.001544 |
| MSE (Test) | 0.001724 | 0.001725 | 0.001721 | 0.001780 |
| RMSE (Train) | 0.037697 | 0.037697 | 0.038229 | 0.039298 |
| RMSE (Test) | 0.041521 | 0.041528 | 0.041491 | 0.042193 |

From the above table in both Ridge and Lasso regression when alpha is doubled the following observations are obtained:

1. There is slight decrease in the value R2 score
2. There is slight increase in the value of RSS
3. There is slight increase in value of MSE
4. There is slight increase in value of RMSE

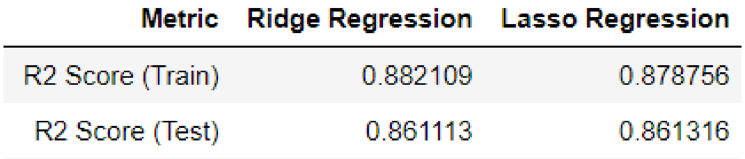
In a nutshell as the alpha value increases the Bias/Residual Error of the model increases thereby reducing the Complexity/Variance of the model. In my model the difference in very small when the alpha is doubled as the alpha values chosen by Ridge and Lasso Cross Validation algorithms is small. This happens when the features chosen for modeling are already being well explained by the model and the model's existing cost function does not need much penalization.

The most important predictor variable after the change is implemented in Ridge and Lasso Regression are:

1. GrLivArea
2. OverallQual
3. Neighborhood_NoRidge
4. YearBuilt
5. GarageCars

**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?**

A. The accuracy of the model in terms of R2 score on train and test dataset for Ridge and Lasso regression are as follows:



| Metric | Ridge Regression | Lasso Regression |
|---|---|---|
| R2 Score (Train) | 0.882109 | 0.878756 |
| R2 Score (Test) | 0.861113 | 0.861316 |

You can observe that the difference in R2 score between Train and Test dataset for Ridge Regression is 0.021 and for Lasso Regression is 0.0174 correct to 3 decimal places.

The error in R2 score between train and test dataset slightly lesser in Lasso when compared to Ridge. Hence, I will choose Lasso Regression over Ridge in this scenario. It is also noticeable that the R2 score of Lasso on Train dataset is smaller than that observed in Ridge Regression. This is because Lasso being L1 type of regularization eliminates variables by estimating corresponding beta coefficients to 0. In our model Lasso has estimated the beta coefficient of Exterior1st_ImStucc variable to 0. Ridge on the other hand being L2 type of regularization doesn't eliminate features.

**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?**

A. The top 5 variables impacting the target variable 'SalePrice' currently with an alpha of 0.0001 are:

i. GrLivArea
ii. OverallQual
iii. Neighborhood_NoRidge
iv. LotArea
v. YearBuilt

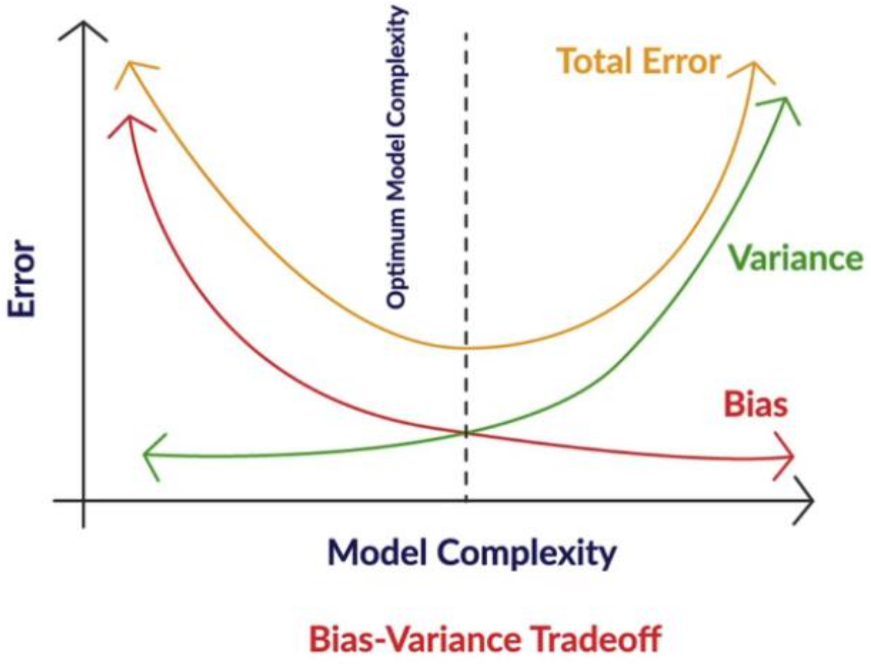
Since these are unavailable, post removing these parameters and upon rebuilding the model with an alpha of 0.0001 we obtain a model with R2 score of 0.813 on train data and 0.805 on test dataset. The top 5 important predictors now are:

i. 2ndFlrSF
ii. GarageCars
iii. BsmtFinType1_None
iv. BsmtQual_Fa
v. FullBath

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?**

A. A model is termed 'robust' when the interpretation of the model does not change wildly with the introduction to new datapoints. A model is termed 'generalizable' when the model is proficient enough to understand the underlying pattern of the data and not mug up the exact data itself. Both of these properties can be measured in terms of bias and variance of a model.

Biasness refers to the error made by model in predicting the datapoints whereas variance refers to the variation in datapoints and interpretation of the features of the model upon introduction of new datapoints. Bias and variance are inversely proportional metrics. Let's understand this better with a figure:



Picture Courtesy: UpGrad PGD AI ML Learning Material

An underfitting overly simple model would have high bias but very low variance while an overfitting complex model on the other hand would fit perfectly on the train dataset as it has memorized the entire dataset but when tested on new test dataset the model would swing wildly indicating high variance on test dataset and low bias on train dataset. Therefore, it can be generalized, as shown in the above figure that as the model complexity increases Bias decreases but Variance increases. The total error would be high on both extremes, Underfitting Model and in Overfitting Model.

A robust and generalized model is one for which the total error is low and accuracy is optimum. Often regularization helps us build such model by penalizing the cost function. Upon regularization on an overfitting model, the variance of the model is drastically reduced with some compromise on bias. Upon cross validation, with an optimum value of alpha/lambda, we obtain a robust and generic model with low error and optimum accuracy i.e., High R2 score & Low RSS, MSE and RMSE values.