**Name: Kalyani Pakhale (DS41)**

**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:**

By doubling,
**Ridge:**

For Ridge, regression alpha is 1.0 and now doubling it and making it 2.0
Ridge(alpha=2.0)
Ridge Regression train r2: 0.9238505663513403
Ridge Regression test r2: 0.7526833276310245

| | Features | Coefficient | Mod |
|---|---|---|---|
| **0** | LotFrontage | 10.261566 | 10.261566 |
| **3** | OverallCond | 0.527282 | 0.527282 |
| **14** | BsmtUnfSF | 0.406426 | 0.406426 |
| **12** | BsmtFinType2 | 0.355562 | 0.355562 |
| **2** | OverallQual | 0.338037 | 0.338037 |
| **11** | BsmtFinSF1 | 0.322512 | 0.322512 |
| **9** | BsmtExposure | 0.316265 | 0.316265 |
| **33** | GarageFinish | 0.303286 | 0.303286 |
| **73** | LotConfig_CulDSac | -0.272453 | 0.272453 |
| **6** | ExterCond | 0.264079 | 0.264079 |

**Lasso:**
For Lasso regression alpha is 0.0001 and doubling it and making it 0.0002
Ridge(alpha=2.0)
Lasso Regression train r2: 0.924842196868562
Lasso Regression test r2: 0.7423837845239192

| | Feature | Coef | mod |
|---|---|---|---|
| **0** | LotFrontage | 10.174821 | 10.174821 |
| **14** | BsmtFullBath | 0.822796 | 0.822796 |
| **3** | OverallCond | 0.568885 | 0.568885 |
| **9** | CentralAir | 0.492011 | 0.492011 |
| **2** | OverallQual | 0.484268 | 0.484268 |
| **73** | Exterior1st_CBlock | -0.479688 | 0.479688 |
| **33** | MSZoning_RH | 0.379703 | 0.379703 |
| **35** | MSZoning_RM | 0.297522 | 0.297522 |
| **36** | Street_Pave | 0.246404 | 0.246404 |
| **20** | GarageQual | 0.240831 | 0.240831 |

After doubling the alpha values in the Ridge and Lasso, the prediction accuracy remains around the same.

The new model is created and demonstrated in the Jupiter notebook. Below are the changes in the coefficients.

Overall since the alpha values are small, we do not see a huge change in the model after doubling the alpha.

**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 optimum lambda value in case of Ridge and Lasso is as follows:

Ridge: {'alpha': 1.0}
Lasso: {'alpha': 0.0001}
**Prediction using ridge regression**

Ridge regression train r2:  0.9272
Ridge regression test r2:  0.7454

**Prediction using Lasso regression**

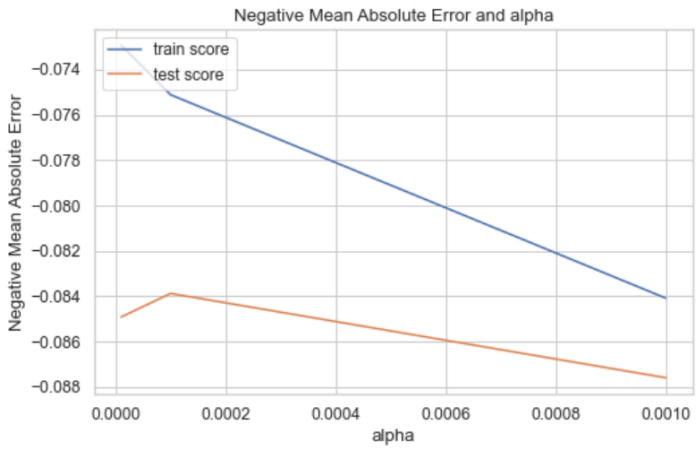
Lasso Regression train r2: 0.9280977658497441
Lasso Regression test r2: 0.7373790476938153
We can see that,

Ridge prediction is 1% more than Lasso, however, we are choosing "**Lasso Regression**" as Lasso helps with **Feature Elimination**.
Since Lasso helps in feature reduction (as the coefficient value of some of the features becomes zero), Lasso has a better edge over Ridge and should be used as the final model.

**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?**

**Answer:**

Droping the first five important predictors
['LotFrontage',
'BsmtFullBath',
'OverallCond',
'CentralAir',
'OverallQual']



A new model has been created for this in the attached Jupyter notebook.

{'alpha': 0.0001}
Lasso Regression train r2: 0.9124746336018961
Lasso Regression test r2: 0.703294795055186

The new Top 5 predictors are:-

| | Feature | Coef | mod |
|---|---|---|---|
| **0** | LotArea | 10.205140 | 10.205140 |
| **10** | FullBath | 1.029837 | 1.029837 |
| **6** | 1stFlrSF | 0.606857 | 0.606857 |
| **1** | ExterCond | 0.561906 | 0.561906 |
| **28** | MSZoning_RH | 0.512930 | 0.512930 |

---

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

As Per, Occam's Razor— given two models that show similar 'performance' in the finite training or test
data, we should pick the one that makes fewer on the test data due to the following reasons:-

Simpler models are usually more 'generic' and are more widely applicable
Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.

Simpler models are more robust.
Complex models tend to change wildly with changes in the training data set
Simple models have low variance, and high bias and complex models have low bias, high variance
Simpler models make more errors in the training set. Complex models lead to overfitting —

they work very well for the training samples, but fail miserably when applied to other tests samples.

Therefore, to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use. Regularization can be used to make the model simpler. Regularization helps to strike the delicate

balance between keeping the model simple and not making it too naive to be of any use. For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model.

Also, Making a model simple leads to Bias-Variance Trade-off:

• • A complex model will need to change for every little change in the dataset and hence is very

unstable and extremely sensitive to any changes in the training data.

• • A simpler model that abstracts out some pattern followed by the data points given is

unlikely to change wildly even if more points are added or removed.

Bias quantifies how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high.

Variance refers to the degree of changes in the model itself with respect to changes in the training data.

Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as

it minimizes the total error as shown in the below graph.