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:

Optimal value of alpha for ridge and lasso regression found using GridSearchCV with scoring done with respect to negative mean absolute error:

- Ridge: 1.0
- Lasso: 0.0001

Alpha is the regularization parameter, therefore, increasing it would lead to more regularisation, that is, there would be a decrease in model complexity. Decrease in model complexity is achieved by pushing the predictor variable coefficients towards zero and reducing the number of predictor variables by making their coefficients truly zero.

Therefore, if we double the value for alpha, regularization will increase, leading to a more simpler model. Now **twice the penalty** will be charged for higher value coefficients. The model would become less complex, thereby having, more bias but less variance.

Ridge			Lasso	
	α = 1.0	α = 2.0	α = 0.0001	α = 0.0002
1.	1stFlrSF	1stFlrSF	1stFlrSF	1stFlrSF
2.	OverallQual	OverallQual	OverallQual	OverallQual
3.	2ndFlrSF	2ndFlrSF	2ndFlrSF	2ndFlrSF
4.	TotalBsmtSF	TotalBsmtSF	Neighborhood_StoneBr	Neighborhood_StoneBr
5.	Neighborhood_StoneBr	Neighborhood_StoneBr	TotalBsmtSF	TotalBsmtSF
6.	MasVnrArea	MasVnrArea	MasVnrArea	BsmtFinSF1
7.	BsmtFinSF1	BsmtFinSF1	Neighborhood_NoRidge	MasVnrArea
8.	FoundationSlab	GarageArea	BsmtFinSF1	Neighborhood_NoRidge
9.	LotArea	dummy_ExterQual	GarageArea	dummy_ExterQual
10.	GarageArea	LotArea	dummy_ExterQual	dummy_KitchenQual

For Ridge, top 7 predictor variables remain the same but changes start to happen from 8th position where GarageArea rises from 10th position to 8th position on doubling the alpha.

For Lasso, top 5 predictors remain the same but changes start to happen from the 6th position where BsmtFinSF1 (Type 1 finished square feet) moves up from 8th position to 6th position and there is a shuffle of predictor variable rank based on the absolute coefficients.

Features	Coefficient
1stFlrSF	0.1208
OverallQual	0.1090
2ndFlrSF	0.1011
KitchenAbvGr	0.0758
TotalBsmtSF	0.0735
Neighborhood_StoneBr	0.0720
MasVnrArea	0.0576
BsmtFinSF1	0.0499
Foundation_Slab	0.0465
LotArea	0.0446
GarageArea	0.0438

Table 1: Ridge (alpha=1)

Features	Coefficient
1stFlrSF	0.1081
OverallQual	0.1041
2ndFlrSF	0.0918
TotalBsmtSF	0.0718
Neighborhood_StoneBr	0.0688
KitchenAbvGr	0.0622
MasVnrArea	0.0568
BsmtFinSF1	0.0506
GarageArea	0.0444
dummy_ExterQual	0.0426
LotArea	0.0425

Table 2: Ridge (alpha=2)

Features	Coefficient
1stFlrSF	0.152779
OverallQual	0.119945
2ndFlrSF	0.112960
Neighborhood_StoneBr	0.067848
TotalBsmtSF	0.064112
KitchenAbvGr	0.061955
MasVnrArea	0.054119
Neighborhood_NoRidge	0.053425
BsmtFinSF1	0.050349
GarageArea	0.040577

Table 3: Lasso (alpha = 0.0001)

Features	Coefficient
1stFlrSF	0.162287
OverallQual	0.126212
2ndFlrSF	0.111625
Neighborhood_StoneBr	0.060656
TotalBsmtSF	0.054159
BsmtFinSF1	0.050785
MasVnrArea	0.050336
Neighborhood_NoRidge	0.050328
dummy_ExterQual	0.043688
dummy_KitchenQual	0.037091

Table 4: Lasso (alpha = 0.0002)

As we can observe from the above tables, Ridge has pushed the predictor variable coefficients towards zero on increasing alpha and Lasso apart from the top three predictor variables, has also pushed the coefficients towards zero.

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:

Our initial dataset had 1460 rows and 81 columns. After data processing and data cleaning, we had 125 columns (increase due to dummy encoding). We used Recursive Feature Elimination to reduce the number of columns to 70. Then manually tried removing columns with high p-value and high VIF score. Here we found it is manually a very tedious task to properly prune all the correct columns which violate the acceptable p-value and VIF score.

Since Lasso regularisation technique, reduces the number of columns by making the coefficients zero, I would go with Lasso regularisation. Also, the alpha for Lasso was 0.0001 using GridSearchCV with scoring equal to negative mean absolute error, but for Ridge it was 1.0. This also indicates that our columns to be important. Because if there were not, alpha for Lasso would be high.

Another advantage of Lasso is feature reduction. It eliminated five (5) columns from our dataset and gave us only 43 columns to work with.

The evaluation metric of Lasso and Ridge were very close and they have been provided in the table below:

	Lasso (alpha = 0.0001)	Ridge (alpha = 1.0)
Mean Square Error	0.0012869	0.0012576
Training R2 Score	0.88145	0.88394
Testing R2 Score	0.887433	0.88999

As we can observe, the evaluation metric shows very small difference and we get simpler model and our feature vector pruning is also taken care by Lasso. Therefore, I would choose Lasso for this dataset.

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:

Five most important predictor variables according to Lasso with alpha = 0.0001 were:

Features	Coefficient
1stFlrSF	0.152779
OverallQual	0.119945
2ndFlrSF	0.112960
Neighborhood_StoneBr	0.067848
TotalBsmtSF	0.064112

Removing these five columns and building the model again without these five columns gives us the following top predictors and their respective coefficient:

Features	Coefficient
TotRmsAbvGrd	0.157451
MasVnrArea	0.082690
dummy_ExterQual	0.078556
GarageArea	0.078026
Neighborhood_NoRidge	0.072516

Where,

- **TotRmsAbvGrd**: Total rooms above grade (does not include bathrooms)
- **MasVnrArea**: Masonry veneer area in square feet
- **Dummy_ExterQual**: Derived column from a ordinal column
    - **ExterQual**: Evaluates the quality of the material on the exterior
- **GarageArea**: Size of garage in square feet
- **Nerighborhood_NoRidge:** If the Neighbourhood is Northridge

	Before Deleting Top 5	After Deleting Top 5
Mean Square Error	0.0012869	0.0016901
Training R2 Score	0.88145	0.84595
Testing R2 Score	0.887433	0.852162

We can easily see that all the evaluation metric has worsen after deleting the top 5 columns. MSE value has increased, and r2 score for both training and testing data set has gone done significantly.

**Code is provided in the Jupyter Notebook for reference.

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

Occam's Razor states that a model should be as simple as possible without compromising on the quality of the model.

Simpler models are always preferred for their ease of interpretation and robustness. Advantages of simpler models:

- **Generalisability**: Simpler models have a better chances of being applicable in a wider sense and as model complexity increases, it's specificity to handle the particular dataset increases.

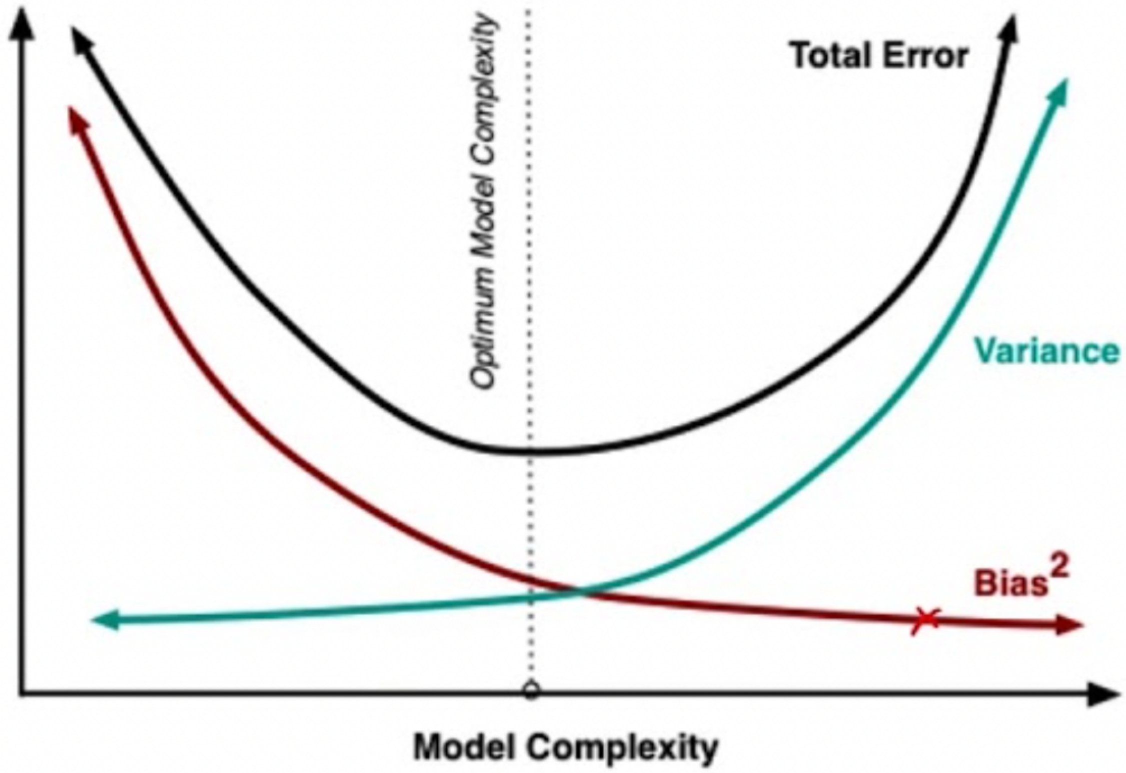
- **Robustness:** Simpler models are robust and indifferent to changes in the data set while complex model can alter their form very quickly and significantly when met with a different set of data.

- **Need for fewer datapoints:** Simpler models need fewer datapoints to train while complex models need more data. In real life, the amount of data can be limited and therefore, simpler models become the go to choice.

Therefore, to make a model robust and generalisable, it is important to make simple models but not simpler than it already needs to be. That is, make it as simple as possible without compromising with its prediction quality.

Regularisation is one of the ways to simplify the model. It does so by pushing the coefficients of the model towards zero and eliminating the number of predictor variables by making the coefficients zero. As model complexity is determined by the number of predictor variables and the magnitude of coefficients of the predictor variables, Regularisation simplifies the model and makes it robust and generalisable.

The Bias-Variance Trade-Off



As model complexity increases, bias decreases but variances increases and we have high total error. Similarly, when model becomes too simple, its variance is low but bias is very high and it also has high total error. Therefore, it is important to strike the right balance.

Bias is the measure of accuracy of model in training data. Higher the bias, more the error on training data.

Variance is the measure of consistency of the model. Higher the variance, higher the difference in performance of the model from training and testing dataset.

Therefore, if the model becomes too simple then its accuracy will be bad and if it becomes too complex then its accuracy for unseen data (test data) will be bad. Therefore, we need to strike the balance.