Advanced Regression – Subjective Questions

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 for Alpha

Ridge = 500

Lasso = 0.01

Changes in Model when Alpha value is doubled

In both Ridge and Lasso, doubling the alpha value will lead to increase in bias and decrease in variance. The best fit line will reduce and will become more horizontal.

Following is the comparison between the regression metrics for Optimal Alpha Value Vs Double the optimal Alpha value

	Metric	Ridge Regression (Alpha=500.0)	Lasso Regression (Alpha=0.01)	Ridge Regression (Alpha=1000.0)	Lasso Regression (Alpha=0.02)
0	R2 Score (Train)	0.92033	0.89773	0.90097	0.85860
1	R2 Score (Test)	0.86783	0.86393	0.85316	0.82728
2	RSS (Train)	10.11823	12.98932	12.57732	17.95879
3	RSS (Test)	7.83391	8.06535	8.70358	10.23757
4	MSE (Train)	0.10266	0.11632	0.11446	0.13677
5	MSE (Test)	0.13789	0.13991	0.14535	0.15763

Comparison of Predictor Variables

	Ridge	Lasso	Ridge Double Alpha	Lasso Double Alpha
LotFrontage	0.01435	0.00916	0.01364	0.00319
LotArea	0.02216	0.02598	0.01902	0.02559
YearBuilt	0.01194	0.02119	0.00964	0.01721
YearRemodAdd	0.01742	0.03150	0.01547	0.03823
MasVnrArea	0.00719	0.00000	0.00811	0.00000
BsmtFinSF1	0.02162	0.03077	0.01784	0.02745
BsmtUnfSF	0.00691	-0.00000	0.00610	-0.00000
TotalBsmtSF	0.03510	0.07293	0.02835	0.07819
2ndFlrSF	0.01713	0.04246	0.01393	0.03591
LowQualFinSF	-0.00011	-0.00000	-0.00059	-0.00000
BsmtFullBath	0.01275	0.00166	0.01095	0.00000

Comparing the values of coefficients, we can see that doubling the alpha has pushed the coefficients closer to 0 in case of ridge regularization. In the case of Lasso, more coefficients have been made 0.

Ridge Regularization – Comparing Top Predictors (higher value coefficients)

Ridge (Optimal Value – 500.0)

	Parameters	Coefficients
0	constant	11.98300
8	TotalBsmtSF	0.02800
17	TotRmsAbvGrd	0.02500
20	GarageCars	0.02100
18	Fireplaces	0.02100
13	FullBath	0.02000
116	OverallQual_8	0.02000
2	LotArea	0.01900
6	BsmtFinSF1	0.01800
66	Neighborhood_Crawfor	0.01500
4	YearRemodAdd	0.01500

Ridge (Optimal Value – 1000.0)

	Parameters	Coefficients
0	constant	11.98300
8	TotalBsmtSF	0.02800
17	TotRmsAbvGrd	0.02500
20	GarageCars	0.02100
18	Fireplaces	0.02100
13	FullBath	0.02000
116	OverallQual_8	0.02000
2	LotArea	0.01900
6	BsmtFinSF1	0.01800
66	Neighborhood_Crawfor	0.01500
4	YearRemodAdd	0.01500

The value of coefficients when sorted based on the coefficient value in descending order seems to be the same for the top predictors

Lasso Regression – Comparing Top Predictors (higher value coefficients)

Lasso (Optimal Value – 0.01)

```
Count of Features Selected -  69
```

	Parameters	Coefficients
0	constant	11.98300
8	TotalBsmtSF	0.07300
20	GarageCars	0.05000
9	2ndFlrSF	0.04200
116	OverallQual_8	0.03900
17	TotRmsAbvGrd	0.03400
4	YearRemodAdd	0.03100
6	BsmtFinSF1	0.03100
18	Fireplaces	0.03000
2	LotArea	0.02600

Lasso (Optimal Value – 0.02)

```
Count of Features Selected -  36
```

	Parameters	Coefficients
0	constant	11.98300
8	TotalBsmtSF	0.07800
20	GarageCars	0.06000
4	YearRemodAdd	0.03800
9	2ndFlrSF	0.03600
18	Fireplaces	0.03600
17	TotRmsAbvGrd	0.03400
116	OverallQual_8	0.02800
6	BsmtFinSF1	0.02700
2	LotArea	0.02600

When Alpha value was doubled in Lasso Regularization, the coefficients of more features became 0 and the number of selected features reduced to 36. This reduces the model complexity resulting in a simpler model and will hence reduce variance.

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:

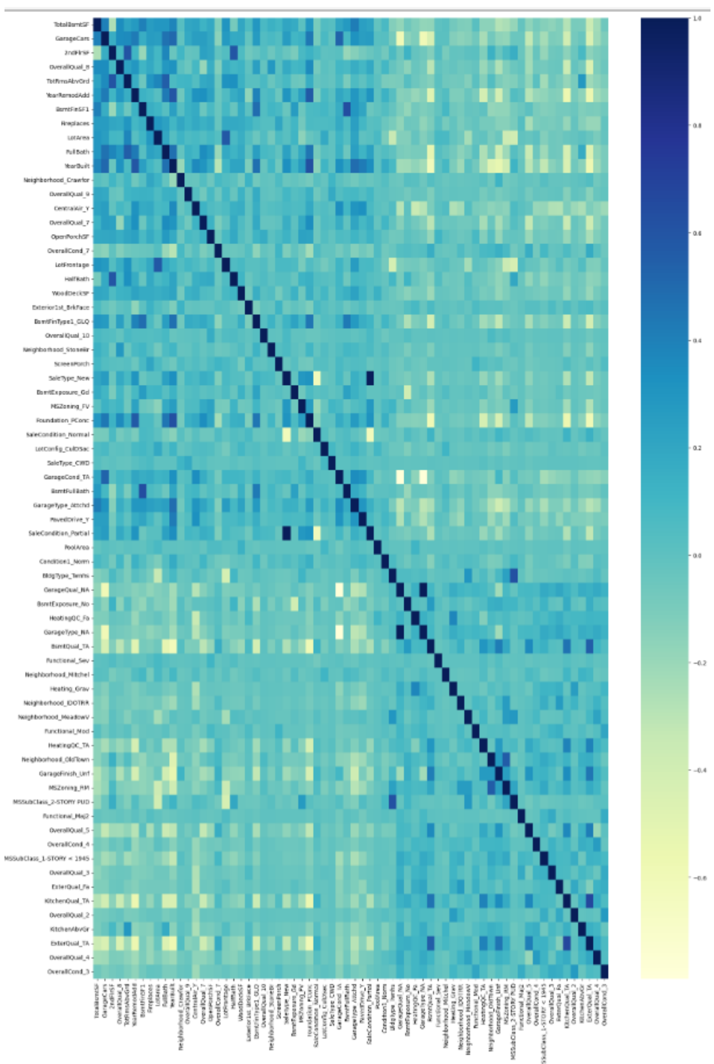
Ridge is generally used when there is significant correlation between the features. Ridge gives biased coefficients considering all the features to lower the variance.

Lasso selects a subset of features making the coefficients of less relevant features 0 simplifying the overall model.

So, Ridge and Lasso serve different purposes.

Based on the implementation, we can see that the

1) Features are correlated



2) Number of features is not very high (262 features only)

```
X = pd.concat([X, df_train_dummy], axis=1)
X = X.drop(categorical_col, axis=1)
X.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1372 entries, 0 to 1459
Columns: 262 entries, LotFrontage to SaleCondition_Partial
dtypes: float64(2), int64(26), uint8(234)
memory usage: 624.4 KB
```

3) Lower Mean Squared Error – Ridge Regularization has resulted in lower MSE value

	Metric	Ridge Regression (Alpha=500.0)	Lasso Regression (Alpha=0.01)
0	R2 Score (Train)	0.92033	0.89773
1	R2 Score (Test)	0.86783	0.86393
2	RSS (Train)	10.11823	12.98932
3	RSS (Test)	7.83391	8.06535
4	MSE (Train)	0.10266	0.11632
5	MSE (Test)	0.13789	0.13991

Considering the above, for the given problem statement, Ridge Regularization can be applied.

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:

After removing the top 5 predictors and creating a new model, below is the new set of top 5 predictors.

	Parameters	Coefficients
0	constant	11.98300
14	TotRmsAbvGrd	0.06500
5	BsmtFinSF1	0.06200
111	OverallQual_8	0.05400
10	FullBath	0.03800
2	LotArea	0.03600

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:

A robust and generalizable model is the one that performs well with unseen data. This would mean that the variance of the model will be high to be able to perform well with new data. This also means that the model need to be less complex.

In order to achieve a model that performs well with unseen data, the following can be taken care of

- A reasonable training dataset that has similar patterns to that of production data
- Thorough outlier analysis before training the model to avoid any overfitting
- Optimal bias-variance trade off to avoid overfitting or underfitting

Accuracy of the model is expected to decrease while trying to keep the model robust and generalizable. Increase in variance to keep the model robust would imply that the accuracy of the model would reduce. As mentioned above, it is important to get the bias-variance trade off and the model complexity at right levels to keep the accuracy at acceptable levels.