Problem Statement - Part II

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?

Optimal value of alpha for Ridge and Lasso regression is 6 and 0.0001 respectively

```
# Get best estimator for lambda
ridge_model_cval.best_estimator_
```

```
Ridge(alpha=6.0)
```

```
# get the best estimates for lambda

lasso_model_cvl.best_estimator_
```

```
Lasso(alpha=0.0001)
```

The below are the changes recorded when alpha values is doubled for both Ridge and Lasso.

When doubled alpha value of Ridge

Mean Squared error is 0.0136 when alpha value is 6

Mean Squared error is 0.0134 when alpha value is 12

When doubled alpha value of Lasso

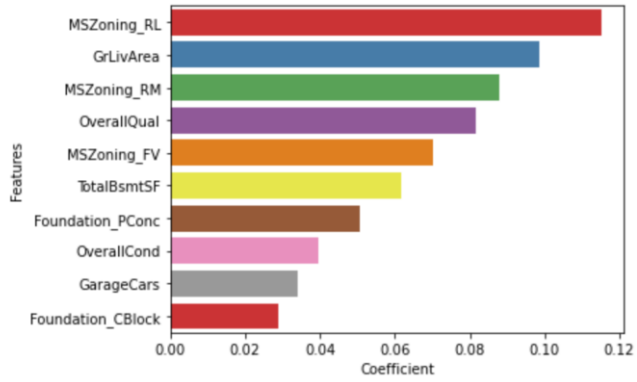
Mean Squared error is 0.0138 when alpha value is 0.0001

Mean Squared error is 0.0137 when alpha value is 0.0002

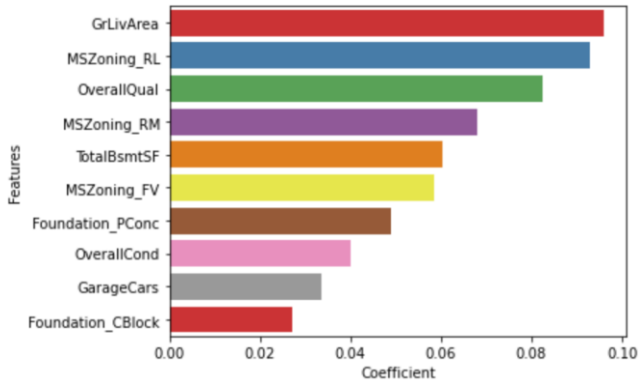
Observation – MSE is decreased when alpha value doubled but not significant change.

The below are the changes to predictable variables.

Top 10 features when ridge alpha value is at 6

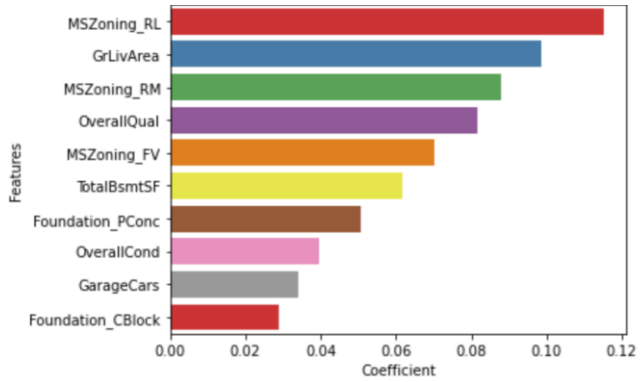


Top 10 features when Ridge alpha value is at 12

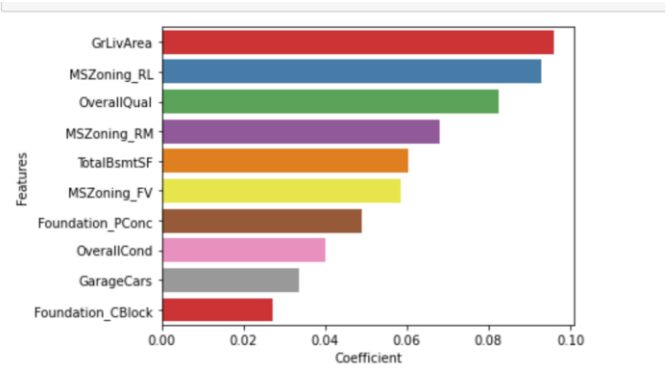


Observations: Feature ranking has been shifted for couple of variables.

Top 10 features when Lasso alpha value is at 0.0001



Top 10 features when Lasso alpha value is at 0.0002



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?

Ridge means square error little less than Lasso means squared error. As many predictors influencing response variable Ridge Regression would perform better than Lasso.

When doubled alpha value of Ridge

Mean Squared error is 0.0136 when alpha value is 6

Mean Squared error is 0.0134 when alpha value is 12

When doubled alpha value of Lasso

Mean Squared error is 0.0138 when alpha value is 0.0001

Mean Squared error is 0.0137 when alpha value is 0.0002

Ridge means square error little less than Lasso means squared error. As many predictors influencing response variable Ridge Regression would perform better than Lasso.

Lasso helps in feature reduction as the coefficient value of some features would become 0.

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?

We would have five most important predictor variables whose coefficients not equal to Zero.

The below would five most important predictors based on coefficient values

:

	Features	rfe_support	rfe_ranking	Coefficient
11	MSZoning_RL	True	1	0.146593
12	MSZoning_RM	True	1	0.115182
4	GrLivArea	True	1	0.101565
9	MSZoning_FV	True	1	0.086375
0	OverallQual	True	1	0.081252

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?

Simple models would perform well on unseen data, complex models would perform on training data set but might fail on unseen data.

We will have to ensure have right trade-off between Bias and variance to make model is robust and generalisable

Model accuracy can be efficiently managed by reducing or minimize total error.