

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 1 :

Ridge regression:

When we visualize the negative mean square error and alpha. Then we can see that when alpha is increased the train error value also increased so and at 2 the error value is less so alpha value for ridge: 2.

Lasso Regression:

The alpha value is .001.

Model	Alpha value
Lasso Regression	.001
Ridge Regression	2

If we double the value of alpha then in lasso regression that means we want to make more coefficient value zero. Given image shows that when we increase the value of alpha R2 score value decrease for training.

```
[419] lml = Lasso(alpha=0.002)
      lml.fit(X_train, y_train)

# prediction on the test set(Using R2)
y_train_pred1 = lml.predict(X_train)
print(metrics.r2_score(y_true=y_train, y_pred=y_train_pred1))
y_test_pred1 = lml.predict(X_test)
print(metrics.r2_score(y_true=y_test, y_pred=y_test_pred1))

0.9273130638676956
0.9192968049481173
```

In the case of Ridge regression if we increase the value for the alpha we want to penalize the model and R2 score is also decrease.

Most important feature after changes for lasso regression:

	Features	rfe_support	rfe_ranking	Coefficient
9	GrLivArea	True	1	0.107040
23	MSZoning_RL	True	1	0.066679
1	OverallQual	True	1	0.065221
2	OverallCond	True	1	0.051497
5	TotalBsmtSF	True	1	0.044094

	Features	rfe_support	rfe_ranking	Coefficient
24	MSZoning_RM	True	1	0.042242
101	SaleType_New	True	1	0.030474
3	BsmtFinSF1	True	1	0.027181
30	Neighborhood_Crawfor	True	1	0.026688
21	MSZoning_FV	True	1	0.026470

The most important feature in ridge is

	Features	rfe_support	rfe_ranking	Coefficient
23	MSZoning_RL	True	1	0.149449
24	MSZoning_RM	True	1	0.116829
21	MSZoning_FV	True	1	0.070772
9	GrLivArea	True	1	0.059289
1	OverallQual	True	1	0.054419
2	OverallCond	True	1	0.050085
5	TotalBsmtSF	True	1	0.048499
22	MSZoning_RH	True	1	0.040891
7	2ndFlrSF	True	1	0.039807
6	1stFlrSF	True	1	0.030116

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 2: Lasso regression would be the better option for feature elimination and RMSE of lasso is less than the ridge.

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 3:

The five most important feature are:

1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. SaleType_New

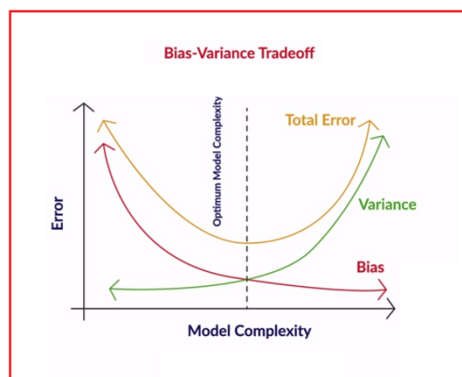
Question 4:

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 4:

For making the model robust and generalizable we can follow Occam's Razor law.

1. Model should be simple.
2. Regularization also help to make model robust
3. Bias-variance trade off



Bias: error in data. High bias means model is unable to learn the data

Variance: high variance means model perform well on training data but perform worst in test data

It is important to maintain bias-variance trade-off to overcome the under fitting and overfitting the data