

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 :

Ridge Alpha -1

Lasso Alpha – 10

When Ridge Alpha is 1 :

--R2 Score--

Train 0.884340040460635

Test 0.8696132804688469

--RSS--

Train 584697913832.1383

Test 324449275268.702

--RMSE--

Train 654756902.3876129

Test 737384716.5197773

When Ridge Alpha is 2 :

--R2 Score--

Train 0.8820877173152852

Test 0.87108088253483

--RSS--

Train 596084124320.2516

Test 320797350989.88556

--RMSE--

Train 667507418.0517935

Test 729084888.6133763

R2 Score of training data has decreased and increased for test data

When Lasso Alpha is 10 :

--R2 Score--

Train 0.8859222400899005

Test .8646666084570094

--RSS--

Train 576699391040.1375

Test 336758382783.66766

--RMSE--

Train 645799989.9665593

Test 765359960.871972

When Lasso Alpha is 20 :

--R2 Score--

Train 0.8854019697956436

Test 0.8670105921065014

--RSS--

Train 579329522996.7144

Test 330925704432.26794

--RMSE--

Train 648745266.5136778

Test 752103873.7096999

**R2 Score of training data has decreased and increased for test data.**

**Important Predictor Variables:**

```
- LotArea----- Lot size in square feet
#- OverallQual-----Rates the overall material and finish of the house
#- OverallCond-----Rates the overall condition of the house
#- YearBuilt----- ---Original construction date
#- BsmtFinSF1-----Type 1 finished square feet
#- TotalBsmtSF-----Total square feet of basement area
#- GrLivArea-----Above grade (ground) living area square feet
#- TotRmsAbvGrd---Total rooms above grade (does not include bathrooms)
#- Street_Pave-----Pave Road access to property
#- RoofMatl_Metal--Roof material_Metal
```

### 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 R2 Score is Lasso is slightly higher than Lasso for Test Data. Hence we will choose Lasso Regression to solve this.

### 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 :**

Top 5 important variables in Lasso

1. RoofMatl\_WdShngl
2. OverallQual\_9
3. Neighborhood\_NoRidge
4. OverallQual\_8
5. GrLivArea

Top 5 variables after creating another model where the above features are not included:

1. 2ndFlrSF
2. Exterior1st\_BrkFace
3. Exterior2nd\_ImStucc
4. BsmtExposure\_Gd
5. Functional\_Typ

#### 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:**

The model should be generalized so that accuracy is not lesser than training score.

Model should not be impacted by outliers.

**1. Ridge regression helps to reduce the variance of the model and improve its generalization performance by shrinking the coefficients towards zero. This results in a smoother and more stable model, but it may also reduce the interpretability of the model.**

**2. Lasso regression helps to reduce the complexity of the model and improve its interpretability by removing some of the irrelevant or redundant predictors**