

Housing Investment using Ridge and Lasso Regression

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?

Ans: For Ridge optimal value for Alpha is 100 and for Lasso it is 0.01

Only change Alpha in the Ridge and Lasso Object initialization

Most important predictor variable after the change is Above grade (ground) living area square feet (GrLivArea)

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?

Ans: We will use Lasso with recommended best alpha value as using Lasso will greatly reduce dependent variable feature set to 53 as against to 188 from Ridge modelling.

As both the models are fairly performing similar.

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?

Ans: In the absence of top 5 predictor variables, next 5 predictor variable i.e;
Neighborhood_NoRidge → Neighborhood with Northridge

SaleCondition_Partial → Home was not completed when last assessed

OverallCond → Overall Condition

BsmtExposure_Gd → Basement Exposure Good

LotArea → plot Area

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?

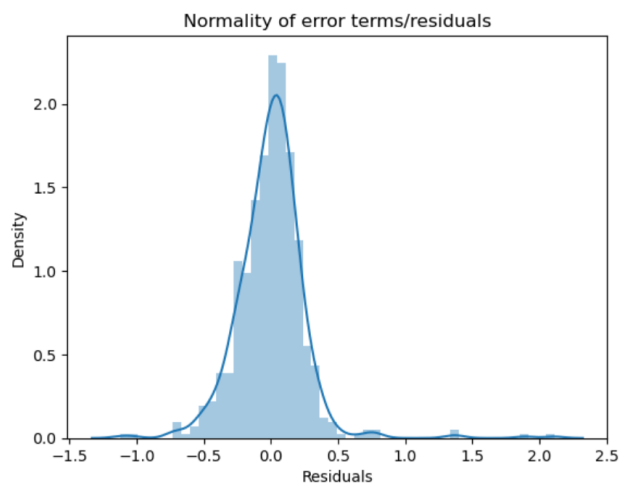
Ans: Both the Train and Test scores are good with r2-score on both ridge and lasso is 0.933 and test scoring at 0.88.

Captured summary of the metrics below;

	Metric	Linear Regression	Ridge Regression (Alpha = 100)	Lasso Regression (Alpha = 0.01)
0	R2 Score (Train)	9.498198e-01	0.934566	0.933804
1	RSS (Train)	3.216554e+01	41.943156	42.431441
2	MSE (Train)	2.240095e-01	0.255801	0.257285
3	R2 Score (Test)	-2.065633e+23	0.887964	0.896624
4	RSS (Test)	5.701146e+25	30.922023	28.531655
5	MSE (Test)	4.544923e+11	0.334718	0.321521

Residuals is normally distributed around 0 for both train and test data sets.

```
dist_plot("Target variable Error term Distribution plot", y_train, y_pred_train)
```



```
dist_plot("Target variable Error term Distribution plot", y_test, y_pred_test)
```

