

ASSIGNMENT

Ans 1:

The optimal value that we see for ridge is 9.0 and that for Lasso is 0.0004.

Range of alpha values for lasso:

1. 0 :the coefficients would be the same as linear regression.
2. Infinite :the coefficients would be 0
3. 0-Infinity :the coefficients would be same as that of simple linear regression

For ridge:

1.0 :same coefficients as simple linear regression

2.infinity :coefficients would be 0

3.0-Infinity :coefficients would be between 0 and ones for simple linear regression

If the values are doubled for both it would be 18 and 0.0008 for ridge and lasso respectively the r^2 score decreases on the train data and the top features would change accordingly.

Ans 2.

Would apply lasso regression in this case as the r^2 score is better and it would help in eliminating features and hence would make the overall model efficient and robust.

Ans 3.

The top five features before were:-

11	MSZoning_RL	True	1	0.109715
----	-------------	------	---	----------

5	GrLivArea	True	1	0.102009
12	MSZoning_RM	True	1	0.076747
1	OverallQual	True	1	0.068341
9	MSZoning_FV	True	1	0.067823

After eliminating these five and creating a new model we get the top features to be:-

OverallCond
Foundation_PConc
GarageCars
BsmtFinSF1

Ans 4:

The model should be generalized for datasets other than training. Outliers should be treated carefully and not given preference. Only the relevant data points to be taken into consideration. An unrobust model cannot be used for predictions.