

Advanced Regression Assignment

Subjective Questions

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:

Optimal alpha values are:

Ridge Regression: 100

Lasso Regression: 0.001

'GrLivArea', 'OverallQual' are the most important predictor variables.

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:

We will Lasso regression as that will remove unwanted features from model building which will be very useful for predicting most important features. Also r^2 scores for lasso are better compared to ridge regression.

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:

The top predictor variables are: 'GrLivArea', 'OverallQual', 'TotalBsmtSF', 'OverallCond', 'YearBuilt'

GrLivArea 0.128692

OverallQual 0.065629

TotalBsmtSF 0.048505

OverallCond 0.043846

YearBuilt 0.043804

After removing them, the predictor variables are:

GrLivArea 0.146928

MSZoning_RL 0.083526

MSZoning_RM 0.065606

TotalBsmtSF 0.056324

GarageCars 0.043311

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:

To make model robust and generalisable 3 features are required:

1. Model accuracy should be > 70-75%: In our case its coming 80%(Train) and 81%(Test) which is correct.
2. P-value of all the features is < 0.05
3. VIF of all the features are < 5

Thus we are sure that model is robust and generalisable