

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:

- Optimal value of lambda for Ridge Regression = **11**
- Optimal value of lambda for Lasso = **0.001**

After building the model for double value of Alpha

R2 score for lasso regression

0.9230259647800546
0.873975192135284

R2 score for ridge regression

0.8907469393193409
0.873975192135284

After change is implemented following factors are important predictors

GrLivArea, FullBath, GarageArea, Functional_Typ, Neighborhood_Crawfor

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 model we will choose to apply will depend on the use case.

- If there are too many variables and we have to do feature selection, then we will use **Lasso**.

- If we don't want to get too large coefficients and reduction of coefficient magnitude, then we will use **Ridge Regression**.

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

OverallQual, OverallCond, BsmtFullBath, FullBath, Fireplaces

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

- A model can be **robust** when the variation in the data does not affect the performance.
- To ensure a model is robust and generalizable, we have to take care it doesn't overfit.
- A generalizable model can adapt appropriately to new, previously unobserved data.
- This is so because an overfitting model has a very large variance and is highly sensitive to even modest changes in the data.
- Such a model won't be able to recognise the patterns in test data that haven't been seen, but it will recognise all the patterns in training data.
- In general, we need to achieve a balance between model complexity and accuracy. Regularisation methods like Ridge Regression and Lasso can be used to do this.