

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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer 1 - Optimal value of alpha for ridge regression is 0.1 and for lasso regression is 1.0. The changes in the model if choose to double the value of alpha for both ridge and lasso are coefficients change.

OverallQual is the most important predictor variables.

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 2 –

The Optimal value of lambda for ridge is 0.1

The Optimal value of lambda for lasso is 1.0

Because this will

With an increase in the value of lambda, variance reduces with a slight compromise in terms of bias. Lasso also pushes the model coefficients towards 0 to handle high variance, just like Ridge regression. But, in addition to this, Lasso also pushes some coefficients to be exactly 0 and thus performs variable selection.

This variable selection results in models that are easier to interpret.

Lasso regression would be a better option it would help in feature elimination and the model will be more robust.

Question 3

After building the model, you realized 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 3 -

Another five most important predictor variables are

'GrLivArea', 'GarageCars', 'TotalBsmtSF', '1stFlrSF' & 'BsmtCond'

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer 4 -

if its output dependent variable (label) is consistently accurate even if one or more of the input independent variables (features) or assumptions are drastically changed due to unforeseen circumstances, we can make sure that a model is robust

The model should also be generalizable so that the test accuracy is not lesser than the training score.

The model should also be generalizable so that the test accuracy is not lesser than the training score.

Outlier's heavy weightage to predict accuracy by the model is high.

One has to do outlier analysis for the relevant dataset. Then One needs to remove the outliers. If the model is unrobust, it can't be trusted for predictive analysis.

Lasso regression would be a better option it would help in feature elimination and the model will be more robust.
