

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?

The optimal value of alpha for my Ridge regression model is 2

The optimal value of alpha for my Lasso regression model is 0.001

On doubling the value of alpha for ridge regression, there is no major change in the r^2 value of train and test.

Most important predictor variables after implementing the change for Ridge regression is – GrLivArea, 1stFlrSF, 2ndFlrSF, MasVnrArea, FullBath

On doubling the value of alpha for lasso regression, there is a significant drop in the r^2 value of the train and test. The model becomes more overfit

Most important predictor variables after implementing the change for Lasso regression is – GrLivArea, BsmtQual_Ex, GarageArea, FullBath, Neighborhood_NridgHt

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?

I would choose Ridge regression because r^2 is higher for Ridge regression as compared to Lasso.

Also, the ridge regression model seems to be a better fit as the difference between r^2 of train and test data is minimal

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?

After removing the top five important predictors and creating a model again. The new top five predictors are –

MasVnrArea,

1stFlrSF

WoodDeckSF

MSZoning_RL,

LotShape_Reg

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?

We can make sure that model is robust and generalisable by checking if the model performs well not only on the given training data set , but also on new independent data.

Increase in Robustness might lead to drop in accuracy of model. Because the model can work well in wide range of datas , hence the accuracy might be reduced.

Accuracy can be high if only those outliers which doesn't seem to have any impact should be removed not all.