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 :

There is no optimal value of alpha for ridge and lasso regression, the appropriate value of alpha

depends on the specific characteristic of the dataset and the goal of the model. We need to use the cross validation to tune the alpha to build a best model.

If we double the value of alpha it means more strong regularization.

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:

I think the Lasso regression would be a better option to apply in general, usually it depends on dataset but it would help in feature elimination and the model will be more robust.

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:

“OverallCond”, “TotalBsmtSF”, “GarageArea”, “GrLivArea”, “OverallQual”

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 :

The model should always be as simple as possible, though its accuracy will/may decrease but it will be more robust and generalisable. This can also be understood using the Bias-Variance trade-off. The simpler the model the more the bias but less variance and more generalisable. Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e. the accuracy does not change much for training and test data, if the model perform well in train data set but not in the test data set, this means that the model is over fitted.

Bias : High error with simple model

Low error aith complex model

Variance : High with complex model

Low with simple model