

Q) 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?

A) The optimal value of alpha for ridge is 4.0 and 0.0001 for lasso regression. More coefficients become zero when we increase the value of alpha in lasso regression which also leads to increase in bias and decrease in variance. MSZoning_FV, MSZoning_RL, MSZoning_RH, MSZoning_RM and OverallCond_3 are the 5 most important predictor variables after changing alpha to 0.0002 in Lasso regression. OverallCond_3, Neighborhood_Crawfor, OverallQual_9, OverallCond_9 and Functional_Typ are the 5 most important predictor variables after increasing alpha to 8 in Ridge regression.

Q) 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?

A) I would use Lasso regression for this assignment as the number of features are very high.

Lasso regression has the inherent property of feature selection which helps in easier understanding of the model by removing variables that are redundant and add no information to the model.

Q) 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?

A) OverallCond_9, BsmtFullBath_3, OverallCond_8, OverallQual_9 and OverallQual_10 are the 5 most important predictor variables after the five most important variables are removed.

Q) 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?

A) The model can be made more generalizable and robust by reducing the number of predictor variables further, i.e, simplifying the model by reducing the dimensionality. The trade-off would be accuracy and the model accuracy is reduced when the model is made more generalizable.