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?

Alpha for Ridge= 7
Alpha for Lasso= 0.0001

After we double the value of alpha, Ridge shows increase in lower value beta coefficients and decrease in higher value coefficients. Same for Lasso. In the new Lasso , we have more 0 coefficients as compared to older Lasso.

The most important predictor variable remains the same after change, which is OverallQual (overall quality) and GrLivArea (Above grade (ground) living area square feet) for Ridge and Lasso respectively.

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 will choose Lasso with alpha = .0001 because Lasso has better r2 score on train and test data than Ridge. The Lasso model is more robust than Ridge.

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?

Top 5 predictors for Ridge
1stFlrSF
TotRmsAbvGrd
MasVnrArea
OverallCond
TotalBsmtSF

Top 5 Predictors for Lasso
1stFlrSF
2ndFlrSF
OverallCond
GarageCars
MasVnrArea

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?

A model is robust when any variation in the data does not affect its performance much. Such as outliers of data. A generalizable model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model. It should provide good accuracy on test data also.

If a model is too good for prediction on training data and performs badly on testing data then our model is not robust or generalized. To make sure a model is robust and generalizable, we have to take care it doesn't overfit. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data. In other words, the model should not be too complex in order to be robust and generalizable. If we look at it from the perspective of Accuracy, a complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease. In general, we have to find some balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso.

Submitted By:-
Shravan Kumar Yadav