

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?

Ans. Optimum value of alpha for Lasso is 0.001 and ridge is 2. If the values of alpha is doubled then for lasso the Mean absolute error increases for test set. It is the way for penalising the model for overfitting, increasing the value of alpha will make it more prominent to overfitting. For ridge the increase in value has no effect as it is saturated after that.

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?

Ans. Ridge and lasso regression prevents the overfitting of the model.

Lasso regression shrinks the non contributing variables of the model and makes the model less complex and less collinear in between the input variables. While ridge regression focuses on making this through residual sum of squares and tries to retain the variable even when they are not contributing.

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?

Ans. 5 important variables are:

1. GrLiveArea
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. GarageArea

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?

Ans. Model has to be more robust and generalisable this can be attributed from the bias variance plot selecting the just right model which has both bias and variance as low. This is important because this also tells us as a feedback which model to select in case of less training data i.e. simple model, while complex data in case of huge amount of data. This model will also be more generalisable because less bias and less variance will be signalling the fact that model has neither overfit nor underfit and will perform with high accuracy on unseen data.