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 :

The Optimal Value of alpha for ridge is 6.0
The Optimal Value of alpha for lasso is 0.001

When the lasso model is doubled the alpha becomes 0.001 and the values arrives as
Lasso Model Double the Alpha
R2 score Train : 0.9031
R2 score Test : 0.8683
RMSE Train : 0.1246
RMSE Test : 0.1443

When the ridge model is doubled the alpha becomes 12.0 and the values arrives as
Ridge Model Double the Alpha
R2 score : 0.9176
R2 score : 0.8715
RMSE : 0.1149
RMSE : 0.1425

The top 5 most important predictor variables for Lasso model after implementation

- GrLivArea
- OverallQual
- TotalBsmtSF
- Neighborhood_Crawfor
- SaleType_New

The top 5 most important predictor variables for Ridge model after implementation

- OverallQual
- TotalBsmtSF
- Neighborhood_Crawfor
- SaleType_New
- GrLivArea

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?

The R2 Score test for the Lasso is 0.8721
The R2 Score test for the Ridge is 0.8716
Both the model gave almost same value in test data but the Lasso seems slightly higher
Lasso is a simpler model compared to 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?

The following are the top 5 models that will be excluded

- GrLivArea
- OverallQual
- TotalBsmtSF
- Neighborhood_Crawfor
- SaleType_New

The Next top 5 important predictors

- OverallCond
- GarageCars
- SaleCondition_Normal
- Condition1_Norm
- Functional_Typ

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?

Robust means the model works for a broad range of inputs. If the model gets really good results at training time (it seems “more accurate”) but won’t generalize to out-of-sample data (i.e. it isn’t robust) then we call it *overfitting* and it’s considered a bad thing exactly because it’s *not* accurate

Generalizable is that the performance and adaptability of a model when applied to new conditions while maintaining the same basic set of explanatory variables

We can make sure the model is robust and generalisable by the way the model is simpler as possible though its accuracy will decrease it will be more robust and generalizable in other words the bias and variance trade off

Its implication in terms of accuracy is that the robust and generalisable model will perform equally well on both training and test data