

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?

The optimal values of Ridge & Lasso are 2 & 0.001 respectively. Doubling the alpha for Lasso do not alter the R2 score, RSS & MSE of Train & Test data, but it alters the R2 score, RSS & MSE of Train & Test data in case of Ridge Regression. Also, it change the coefficients of the variables.

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?

Comparing the Lambda value of Ridge (alpha=2) & Lasso (alpha=0.001) & MSE, it is preferred to choose the Lasso because Lasso simplifies the model by reducing the co-efficient values of some variables to zero such that the variables get eliminated and thereby reducing the complexity. Also, R2 score of the Lasso Model in this case study signifies a good fit.

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?

After excluding the first five important predictor variables, the other five variables which has become important are as follows.

- OverallQual_Excellent
- OverallQual_Very Good
- Fireplaces
- Neighborhood_Crawfor
- MSSubClass_1-STORY 1945 & OLDER

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?

The following factors need to be considered to make sure that a model is robust and generalisable.

- Removing Outliers
- The variables should be normally distributed
- Linear Relationship between Predictor variables and Target variable.
- Transformation of Data, if needed when variables doesn't satisfy the above two factors.
- Feature Scaling
- Regularization of Model Variables

The regularization reduces the accuracy of the model, thereby reducing the variance of the model such that the model is able to predict the unseen test data set with more accuracy.