ADVANCED REGRESSION

Submitted by: Smruti Ranjan Parida

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:

1. **Following are the optimal values of Alpha for below models:**
   - Ridge Regression – 10
   - Lasso Regression – 0.001
2. **Changes in the model using Double value of alpha:**
   - **Ridge Regression:**
     - The coefficients of the model reduced.
     - The predictive power of the model is reduced (Lower R2 value and Higher Mean square Value).
     - The ranking of the predictor variables changed.
   - **Lasso Regression:**
     - The coefficients of the model reduced.
     - The coefficient of few additional independent variables become 0. So No predictors in the model reduced.
     - The predictive power of the model is reduced (Lower R2 value and Higher Mean square Value).
     - The ranking of the predictor variables changed.
3. **The Most Important Predictor variables after the change:**
   - **Ridge Regression:**
     - Neighborhood_Crawfor
     - OverallQual
     - Neighborhood_StoneBr
     - GrLivArea
     - SaleCondition_Normal
   - **Lasso Regression:**
     - Neighborhood_Crawfor
     - SaleType_New
     - Neighborhood_StoneBr
     - GrLivArea
     - TotalArea

ADVANCED REGRESSION

Submitted by: Smruti Ranjan Parida

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?*

Answer:

Ridge Regression: For Ridge regression we can use the optional value as 10 as the model has Higher prediction power and by increasing the alpha the error in the model will increase.

Lasso Regression: For Lasso Regression we can use the Aplha value as 0.002 as the no of paramters for the model significantly reduced from 104 to 79 and the reduction in the R2 value or the mean square is not significant.

Question 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?*

Answer:

The top5 important predictors for the model after removing five most important predictor variables are the following ones.

- 2ndFlrSF
- 1stFlrSF
- MSSubClass_70
- OverallQual
- SaleCondition_Normal

ADVANCED REGRESSION

Submitted by: Smruti Ranjan Parida

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?*

Answer:

The Model is considered to be robust and generalizable when the model performance for the test data set/unseen dataset is similar to train dataset. So the error in case of test dataset should be close to the Train dataset

To make the model more robust and generalizable we can use larger train data. In case we have small data set, We can use Cross Validation to evaluate the models.

To check this we can compare the metrics such as R2, Mean square error for the Training and test data set to make sure that the model is not Overfitting/ Underfitting the dataset.

The model which is generalized and robust learn the pattern in the dataset rather than the noise in the data set. So the accuracy of the model is higher.