Problem Statement - Part II

Assignment Part-II

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?

Ans: - Optimal value of alpha for ridge is 2, and Lasso alpha value is 0.001. if we choose double value of alpha for ridge and lasso than our model accuracy will decrease and when more penalty will apply on curve than our model make more generalized.
Most imported variable for ridge is as follow

1. MSZoning_FV
2. MSZoning_RM
3. Exterior1st_AsphShn
4. SaleType_ConLI
5. MSZoning_RL
6. MSZoning_RH

7. Neighborhood_Edwards
8. Street_Pave
9. RoofMatl_Membran
10. Neighborhood_OldTown

Most Imported variable for lasso is as follow

1. GrLivArea
2. Neighborhood_Crawfor
3. Neighborhood_Somerst
4. Neighborhood_NridgHt
5. OverallQual
6. Condition1_Norm
7. MSZoning_RL
8. Exterior1st_BrkFace
9. YearBuilt

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?

Ans:- For Ridge regression uses a Hyperplane tuning parameter called lambda as the penalty is square of magnitude of coefficients which is

identified by cross validation. Residual sum or squares should be small by using the penalty. The penalty is lambda times sum of squares of the coefficients and increase the value of lambda the variance in model Ridge regression includes all variables in final model so in Ridge regression less chance to Data loss.

For Lasso regression, uses a Hyperplane tuning parameter called lambda as the penalty is absolute value of magnitude of coefficients which is identified by cross validation. As the lambda value increases curve shrinks the coefficient towards zero and it make the variables exactly equal to 0. Lasso also have feature elimination.

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?

Ans: - Lasso model top five feature is as follow

1. GrLivArea

2. Neighborhood_Crawfor
3. Neighborhood_Somerst
4. Neighborhood_NridgHt
5. OverallQual

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Ans: - Outliers not affect our training data and test data model accuracy is around or closer with train data. The simpler the model the more the bias but less variance and more generalizable.
Confidence interval we can use because it will help for standardized the prediction made by our model.
And Bias and variance it also helpful to balance in Bias and Variance to avoid overfitting and under-fitting of data.