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:

Ans: In the case of ridge regression:- When plotting the curve between negative mean absolute error and alpha we can see that as the alpha value increase from 0 the error term decrease and the train error shows increasing trend with the value of alpha increases .when alpha is 2 the test error is minimum so decided to continue with value of alpha as 2 for our ridge regression.

For lasso regression I decided to keep very small value of alpha that is 0.01, with increase in the value of alpha the model try to penalize more and try to make most of the coefficient value zero. Initially it came as 0.4 in negative mean absolute error and alpha. When we double the value of alpha for our ridge regression we will take the value of alpha equal to 10 the model will apply more penalty on the curve and try to make the model more generalized that is making model more simpler and no thinking to fit every data of the data set .from the graph we can see that when alpha is 10 we get more error for both test and train. Similarly when we increase the value of alpha for lasso we try to penalize more our model and more coefficient of the variable will reduced to zero, when we increase the value of our r2 square also decreases. The most important variable after the changes has been implemented for ridge regression are as follows:- 1. MSZoning_FV 2. MSZoning_RL 3. Neighborhood_Crawfor 4. MSZoning_RH 5. MSZoning_RM 6. SaleCondition_Partial 7. Neighborhood_StoneBr 8. GrLivArea 9. SaleCondition_Normal 10. Exterior1st_BrkFace The most important variable after the changes has been implemented for lasso regression are as follows:- 1. GrLivArea 2. OverallQual 3. OverallCond 4. TotalBsmtSF 5. BsmtFinSF1 6. GarageArea 7. Fireplaces 8. LotArea 9. LotArea 10. LotFrontag