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 : In Ridge regression, When we plot the curve between negative mean absolute error and alpha we can see that as the value of alpha increase from 0 ,then error term decrease and the train error is showing increasing trend when value of alpha increases .when the value of alpha is 2 the test error is minimum so we decided to go
with value of alpha equal to 2 for our ridge regression.

For lasso regression I have decided to keep a tiny value that is 0.01, when we increase the value of alpha the model tries 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.

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 :Ridge regression, uses a tuning parameter called lambda as the penalty is the 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, hence the coefficients that have greater values get penalized. As we increase the value of lambda the variance in model is dropped and bias remains constant. Ridge regression includes all variables in the final model unlike Lasso Regression.

Lasso regression, uses a 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 ,Lasso shrinks the coefficient towards zero and it makes the variables exactly equal to 0. Lasso also does variable selection. When lambda value is small it performs simple linear regression and as lambda value increases, shrinkage takes place and variables with 0 value are neglected by the model.

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?

1. OverallQual
2. GarageArea
3. FullBath
4. Foundation_PConc
5. MSZoning_RL

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?

ANS :For an ideal model ,the model should be as simple as possible, though its accuracy will decrease but it will be more robust and generalisable. It can also be understood using the Bias-Variance tradeoff.
The simpler the model the more the bias but less variance and more generalisable. Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e. the accuracy does not change much for training and test data.