

Question1:

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?

alpha of ridge and lasso model are 2 and 0.01 respectively.

if we double the value of alpha for our ridge regression the model will apply more penalty on the curve and try to make the model more generalized while making model more simpler and not thinking about to fit every data of the data set we will get more error from both test and train. Similarly when we increase the value of alpha for lasso it will try to penalize more our model and more the coefficient of the variable it will reduced to zero.

The most important variable that changes for ridge regression are:

1. MSZoning_FV
2. MSZoning_RL
3. Neighborhood_Crawfor
4. MSZoning_RH
5. MSZoning_RM

The most important variable that changes for lasso regression are:

1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmntSF
5. BsmntFinSF1

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?

It is critical to regularise coefficients and improve prediction accuracy while reducing variance and making the model interpretable.

Ridge regression employs a tuning parameter known as lambda as the penalty is the square of the magnitude of the coefficients as determined by cross validation. Using the penalty, the residual sum of squares should be small. The penalty is lambda times the sum of the squares of the coefficients, so the coefficients with higher values are penalised. As we increase the value of lambda, the variance in the model decreases while the bias remains constant. In contrast to Lasso Regression, Ridge Regression includes all variables in the final model.

we choose Lasso regression as it employs a tuning parameter known as lambda as the penalty, which is the absolute magnitude of the coefficients that are identified.

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?

the five most important predictor variables that are excluded are :-

1. GrLivArea
2. OverallQual

3. OverallCond
4. TotalBsmtSF
5. GarageArea

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?

To make model robust and generalisable it should be as simple as possible, even though its accuracy will decrease but it will be more robust and generalisable. we can also understood that using the Bias-Variance trade-off. simpler the model the more the bias but less variance and more generalizable. 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.