

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: The optimal value we obtained through analysis for Lasso Regression: 0.001.

When we are making the double or increasing alpha value, the function would be penalized more. we are forcing to shrink the coefficient. Alpha is the penalty term applied to fit the data more by putting constraints on magnitude of coefficient to avoid overfitting and reducing the RMSE.

The optimal value we obtained through analysis for Ridge Regression: 0.1

Increasing alpha values mean to enhance the penalty term to reduce the magnitude of the coefficient. In Ridge regression, alpha is the square magnitude of coefficient. With increase in Alpha, RSS would and complexity of model will decrease. Small the alpha value leads to reduction in coefficient. Higher results underfitting.

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: Lambda is the tuning parameters used to controls the trade-off amide variance and bias by cross validation. Its value can be taken in between 0 and 1. Literally, its taken 0.4 based upon literature study. It's the penalty term to increase or decrease the complexity of the model without overfitting and multicollinearity .

Q3.

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 five most important predictor variables now are

MSSubClass RoofMatl _Membran,

ML2 Assignment

MSZoning_RL

MSZoning_FV

MSZoning_RH

MSZoning_RM

Condition2_PosA

Q. 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: Robust mean minimum effect of variation on the performance of the model and Generalized model means the model is able to adopt the unseen or new data. Means model behave same for unseen or new data as it is developed for the original data. Based upon analysis, we can conclude that for simple regression model the R^2 is approximate 0.74 but for Ridge and Lasso R^2 values are 0.88 and 0.85 for test set. There is no so much difference in R^2 value for train and test run. RMSE values also less for ridge and lasso test models. Therefore we can claim that models are robust and generalized.