

Q1. 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-The optimal value of alpha for lasso is 0.001 and for ridge it is 5 and if we double the value of alpha there will be some reduction in r-squared value. The most important parameters are landscape, Utilities, OverallCond, HeatingQC, CentralAir

Q2. 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-We will use the ridge regression with alpha equal to 5 since it gives the best R-squared value on test data

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?

Ans-The five most important variables are
BsmtFintype1, TotalbsmtSF, Lotshape, 1stFlrSF, LotArea

Q4. 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: By selecting the important feature using P-value and VIF which will ensure that only useful features are used in model building and also checking over the difference of r-squared on the test data and train data if both are closer that mean model is more generalisable and robust. The implications on accuracy is that accuracy metrics for both the test and train are closer to each other