

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: Optimal value of ridge and lasso is very near to the 0. As based on model_cv params it shows that best estimator to be 10^{-6} or 0.000001. As we increase the alpha value for the regression R^2 will start decreasing. So as when doubling it. With the change as well OverallQual will be the most important predictor variable.

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: Although both of them gives a nearby result currently but with the lasso regression it is providing higher R^2 score which is why I will prefer lasso regression. Along with that the feature selection is done only in lasso which gives it an additional advantage over ridge regression.

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?

Ans: TotalBsmtSF, GrLivArea, MiscVal, PoolQC, BsmtFullBath

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: A generalisable model should not have neither a high bias nor high variance. So we have overfit the data using the attributes which are not really relatable to the overall target variable, any change in them will cause issue in predicting the target variable. Which will increase the variance value. Therefore a generalisable model should have balanced bias and variance tradeoff. Which is done by using ridge and lasso. Which helps in feature selection. The feature which has coeff near to 1 is highly related to target variable and the one near to 0 is least related to it.