

1. What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose to 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 of alpha for ridge and lasso regression are 7.0 and 100. By doubling the alpha value, I see coefficients have changed but retained the same top five predictor variables. The most important predictor variables for ridge and lasso are

Ridge	lasso
Neighborhood_NoRidge	Neighborhood_NoRidge
TotRmsAbvGrd	Neighborhood_StoneBr
GrLivArea	GarageCars
2ndFlrSF	OverallQual
OverallQual	GrLivArea

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: Based on my analysis, I see Lasso regression R2 value for train and test is almost equal. There is a slight difference in R2 value for ridge regression which proves lasso performed well than ridge.

3. After building the model, you realized 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 rest of the five important predictors variables if five most important variables are missing are:

2ndFlrSF, OverallCond, Exterior2nd_ImStucc, RoofMatl_WdShngl, Neighborhood_NridgHt

4. How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer: Adopting Occam's razor principle and using cross validation and making sure that r2 values for train and test data are not showing huge variation along with will ensure that a model is robust and generalizable. (Simple model will explain the model better than complex model as per Occam's razor principle)