

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: Optimal value of alpha for ridge and lasso regression:

- Ridge Regression: 7
- Lasso Regression: 100

Double the value of alpha for both ridge and lasso: *There is a increase in coefficient values and decrease in r2 score.*

The most important predictor variables after the change is implemented

OverallQual, 1stFlrSF, Neighborhood_NoRidge, Neighborhood_NridgeHt, Neighborhood_Somerst

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 choose **Lasso** as its giving **feature selection** option also. It has removed unwanted features from model without affecting the model accuracy. Which makes are model generalized and simple and accurate.

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: After dropping the 5 important variables and then building the model. Below are the five important variables:

2ndFlrSF, MSZoning_RL, Exterior2nd_CmentBd, GarageFinish_NA, GarageFinish_Unf

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: A model is **robust** when any variation in the data does not affect its performance much. A **generalizable** model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.

Model is robust and generalisable if:

1. Model accuracy should be > 70-75%
2. it doesn't **overfit**. Test accuracy is not lesser than the training accuracy
3. P-value of all the features is < 0.05
4. VIF of all the features are < 5

If we look at it from the perspective of **Accuracy**, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.