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:

Optimal alpha for ridge : 0.9
Optimal alpha for lasso : 0.001
If we double alpha for ridge , r2 score decreases.
If we double alpha for lasso , r2 score decreases, also a number of coeffeicients become zero.

After comparing both models after change, these variables were found to be most impiortant :YrSold_Old,KitchenQual,BsmtHalfBath.

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:

In both Ridge and Lasso we can see that the r2_scores are almost same for both of them but as lasso will penalize more on the dataset and can also help in feature elemination , I will go ahead with lasso. So lambda=alpha=0.001

Question 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?

Answer:

47	MSZoning_RL	1.517699
29	Fireplaces	0.494154
66	Neighborhood_NAmes	0.322858
24	FullBath	0.308248
12	BsmtFinSF1	0.297590

Above are the five most important varaibles.

Question 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: The model should not overfit test score and train score should not have much difference.
We tune hyperparameters in such a way that model does not overfit or underfit.