House Price Prediction Assignment Subjective Questions

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 : I got alpha(Ridge) = 8 and alpha(Lasso) = 0.0001

Metric		Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.518870e-01	0.899076	0.914119
1	R2 Score (Test)	-8.957318e+20	0.894425	0.874081
2	RSS (Train)	6.801217e-01	1.426648	1.214010
3	RSS (Test)	3.242834e+21	0.382216	0.455867
4	MSE (Train)	2.413081e-02	0.034949	0.032240
5	MSE (Test)	3.332506e+09	0.036180	0.039512

By doubling alpha for both, I got below:

Metric	Linear Regression		Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.518870e-01	0.887748	0.898094
1	R2 Score (Test)	-8.957318e+20	0.887874	0.890774
2	RSS (Train)	6.801217e-01	1.586790	1.440536
3	RSS (Test)	3.242834e+21	0.405932	0.395433
4	MSE (Train)	2.413081e-02	0.036859	0.035119
5	MSE (Test)	3.332506e+09	0.037285	0.036800

Therefore with increase in alpha, r2 value for Ridge and lasso decreases

Question 2

Q. 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. As r2 value is high in lasso with less number of features, therefore I will prefer Lasso Regression Technique

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 . New variables after removal of top 5 variables are:

1stFlrSF : Area of First Floor square feet

2ndFlrSF : Area of Second floor square feet

MasVnrArea,: Masonry veneer area in square feet

OverallQual_9 : Rates the overall material and finish of the house(rating = 9)

FullBath_3 : Full bathrooms above grade

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 : Simple Models with not overfitting are robust and generalisable. Accuracy on test and unseen data set should be not be much lesser then on training set