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:

The Optimal value of alpha for ridge is 2 and for lasso it is 0.0001. This can be observed in the jupyter notebook.

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?

Lasso will be a better option as I think its makes the model robust with feature elimination.

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 5 top predictor variables.(done in jupyter notebook)

	Lasso Co-Efficient
LotFrontage	0.156221
Total_porch_sf	0.077278
HouseStyle_2.5Unf	0.055643
CentralAir_Y	0.042990
HouseStyle_2.5Fin	0.039957

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?

? What are the implications of the same for the accuracy of the model and why?

Answer:

We need to take care of the outliers for a model to be robust because outliers will change the model itself . Generalizable means the accuracy should not differ severely b/w the train and test datasets.

So by doing proper analysis of the outliers and removing them from while the training of the model will bring more accuracy to the predictions made by model.