Assignment-based 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

1. The optimal value of alpha in Ridge regression is 3.0 and Lasso regression is 0.0001.
2. After doubling the alpha in Ridge regression and Lasso regression there is a slight decrease scores which is not noticeable.
3. Top 5 variables in ridge are:
    - **GarageArea**
    - **HouseStyle_2.5Unf**
    - **LotFrontage**
    - **OverallCond**
    - **Neighborhood_StoneBr**
4. Top 5 variables in Lasso are:
    - **GarageArea**
    - **LotArea**
    - **HouseStyle_2.5Unf**
    - **LotFrontage**
    - **PoolArea**

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

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	8.787590e-01	0.864489	0.863984
1	R2 Score (Test)	-9.864883e+23	0.773792	0.778343
2	RSS (Train)	2.121881e+00	2.371622	32577.284224
3	RSS (Test)	6.865299e+24	1.574257	5640.669805
4	MSE (Train)	4.567726e-02	0.048291	0.048380
5	MSE (Test)	1.254835e+11	0.060089	0.059481

- As we can see in the above figure both ridge regression and lasso regression performed well in case of r2 score and they have similar r2 score. but if I consider RSS ridge regression had less r2 score than lasso so I will choose Ridge regression instead of Lasso regression.

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:

- These are 5 most import variables after removing the first 5 variables
  a. **Condition2_PosA**
  b. **Neighborhood_NridgHt**
  c. **Neighborhood_NoRidge**
  d. **Neighborhood_StoneBr**
  e. **Heating_GasW**

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:

- If a model doesn't overfit and it is as simple as possible then we can say our model is robust and generalisable.
- The implications are if a model overfits then we can see that the accuracy for it in train set is very high and on test it is very low.
- If we say that our model is robust and generalisable then our training and testing accuracy should be good.