

Answer-1 optimal value of  $\alpha$  for ridge and Lasso regression

Ridge Alpha - 1

Lasso Alpha - 10

if Change The Ridge Alpha 1 from to 2

$R^2$  score - 0.87931 ( ) & 0.87102 ( )

RSS - 607995142958.1411 (train) & 320928407278.4621 (test)

RMSE - 680845624.8131 (train) & 729382743.8146 (test)

if Change The Lasso Alpha 10 from to 20

$R^2$  score - 0.88540 (train) & 0.86701 (test)

RSS - 599329522996.7144 (train) & 3309257044332.2679 (test)

RMSE - 640745266.5136 (train) & 752103873.7096 (test)

$R^2$  score of training data has decrease and it has increase on testing data for both Ridge and Lasso Alpha.

important Predictor Values:-

- (i) Lot Area - size in sqft
- (ii) OverallQual - Rates the overall material and finish of house.
- (iii) OverallCond - Rates the overall condition of house
- (iv) YearBuilt - Original construction date
- (v) BsmtFinSF1 - Type 1 finished square feet

Answer-2

The  $R^2$  score of Lasso is slightly higher than ridge for the test data set so we will choose Lasso regression to solve this.

Answer - 3 five most important predictor variables

- (i) 1stFlrSF - First floor square feet,
- (ii) GrLivArea - Above grade living area square feet
- (iii) StreetPave - Pave road access to property
- (iv) RoofMatl-Metal - Roof material-Metal
- (v) RoofStyle-Shed - Type of roof (shed)

Answer - 4

The model should be generalised so that the accuracy is not lesser than the training score. The model should be accurate for ~~dataset~~ other than the ones which were used during training. Too much importance should not be given to the outliers so that the accuracy predicted by the model is high.

To ensure that this is not the case, the outliers analysis needs to be done and only those which are relevant to the ~~dataset~~ dataset to be retained.